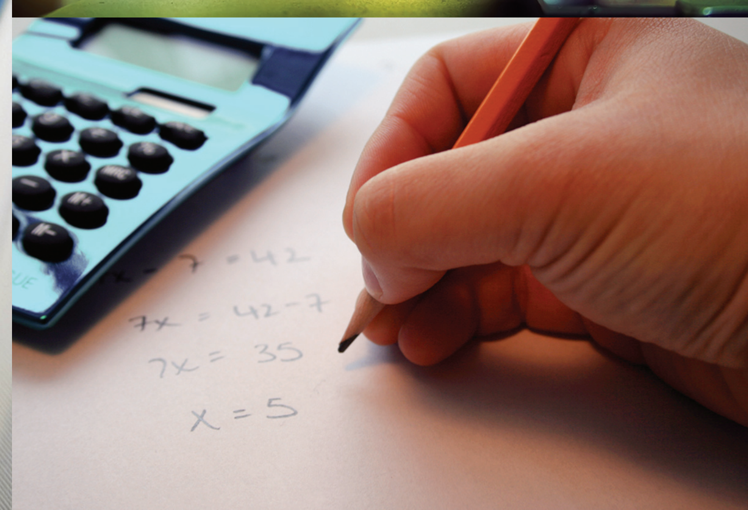
Pan-Canadian Assessment Program

PCAP-13 2007

Factors Contributing to Performance in Mathematics and Science



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

Pan-Canadian Assessment Program

PCAP-13 2007

Factors Contributing to Performance in Mathematics and Science

Developed by

Robert Crocker
Atlantic Evaluation and Research Consultants

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 987-0-88987-211-0

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 the 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 is the first 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 the mathematics/science component of PCAP 2007 and, in particular, on the factors contributing to performance in these two subjects.

The conceptual framework for this report is an educational productivity model, set within a human capital theory framework, and it takes school achievement as the main outcome of interest. Increasing average achievement and reducing disparities in achievement are treated as the immediate goals of policy-makers. Two models more directly related to teaching and learning, specifically the Carroll time model and the Wang-Haertel-Walberg proximity model, were used in developing the PCAP questionnaires and are described in this report as a means of identifying variables that might be good predictors of the two outcomes of interest in this study.

The conceptual model is operationalized through mathematical models that allow us to examine the relative contributions of various inputs (e.g., student and school characteristics) and processes (e.g., teaching and learning strategies) to achievement. Specifically, multilevel (hierarchical) models are used in this report because such models better represent data derived from complex multi-stage samples such as those in PCAP.

PCAP scores in both mathematics and science were scaled to a pan-Canadian weighted average of 500 and standard deviation of 100. The comparative results reveal statistically significant differences across jurisdictions and language groups. Although such differences are of considerable policy interest, further analysis reveals that differences among jurisdictions are small compared to differences between students and schools. This has major implications for the analysts' ability to model jurisdictional and language differences.

The results for student characteristics show that grade level and number of books in the home (the measure of SES) are the most consistently positive predictors of achievement in both subjects. Speaking French as a home language is also a positive predictor. This is a function of the high performance of Quebec francophone students; other francophone groups, in contrast, have much lower average scores. A gender effect favouring males is found for both subjects. Students born in Canada do less well in mathematics than those born outside of Canada. The opposite is true for science when taken alone, but not when other variables are controlled.

As for school characteristics, students in private schools have higher performance in the small number of jurisdictions that have sufficient private schools to permit analysis. Schools with higher proportions of Aboriginal students have lower average scores. This

effect is significantly reduced for mathematics when other factors are controlled. Schools in larger communities have higher performance in mathematics, but not in science. Students in larger classes generally show higher achievement levels. However, this effect is attenuated when other variables are controlled, which suggests that class size is related to other factors included in the full model. Specifically, class sizes are larger in larger schools and communities, both of which are positively related to achievement.

Student enjoyment of school, enjoyment of reading, and student perception of being a good reader are positive indicators of both mathematics and science performance. External attributions of success are a negative predictor of performance in the full model, but not in the bivariate model. This indicates an effect that is suppressed by other variables, in this case, most likely the variable of attitudes. Student perceptions of their performance on the PCAP test are positively related to achievement, although this is attenuated somewhat by other variables in the full model.

The ability to examine the extent to which reading, mathematics, and science performance are inter-correlated was limited by the fact that different students wrote the reading test than the mathematics/science test. Reading could therefore be correlated with mathematics/science only at the school level. Nationally, the correlation between mathematics and science is moderate. The correlations with reading at the school level are lower. At the jurisdictional/language group level, the correlations are variable, but all significantly positive. While not very definitive, these results suggest that something other than a general ability trait is being measured by the three assessments.

The influence of external resources on school programs is positively related to performance in mathematics but is only marginally so for science in the full model. Other sources of influence — specifically curriculum, teachers, and external assessment — show no significant effects.

The specific question posed for the analysis of jurisdictional effects was *"How do provincial/territorial characteristics in practices, policies, and procedures such as school size, governance model (public/private), class size, large-scale assessment procedures, and diversity of student population influence mathematics and science performance?"* Three different analytical approaches were used to examine this question. First, models were developed in which jurisdiction/language groups were included as dummy variables, first alone and then with all other variables controlled. Significant changes in the jurisdictional mathematics coefficients were found for three groups — Saskatchewan, New Brunswick French, and Nova Scotia English. Marginal changes were found for Manitoba English and New Brunswick English. These changes suggest that the differences between these jurisdictions and the reference groups are significantly related to differences among jurisdictions in the predictor variables.

The question *"Why do some jurisdictions have higher performance than others"* will be the subject of another project using the PCAP 2007 database, but we examined it briefly in this study using a 3-level model with student, school, and jurisdiction/language group as the three levels. A preliminary exploration revealed that it is technically feasible to develop and analyze such a model. However the 3-level model is limited by the small number of units at the jurisdictional level. Effective hierarchical analysis requires many more units at the highest level; also, the selected units must be a random sample of all the units in the population. Neither of these requirements can be met with the PCAP data.

In any event, the 3-level model reveals, first, that the level-3 effects, that is, the jurisdictional/language level effects account for only 3 percent of the total variance in the model. This makes it clearer that, despite their high policy interest, jurisdictional differences are much smaller than either student or school differences. The analysis also shows that adding student- and school-level variables aggregated to the jurisdiction accounts for more of the variance than using the student- and school-level variables at their "natural" or measured levels. However, none of the specific jurisdiction-level variables are statistically significant. This is almost certainly related to the small number of level-3 units, which results in large standard errors at that level. Overall, the 3-level model was judged inappropriate as a means of answering the question of differences across jurisdictions.

The issue of equity of school systems within jurisdictions and the differential impact of some predictor variables on mathematics and science achievement was investigated in an exploratory way for four selected jurisdictions. The results show some differences in the degree of variability of scores across jurisdictions at both the student and the school level. This means that some jurisdictions have managed to reduce disparities among students and schools more than others. Differences are especially large for mathematics, where Quebec English stands out as having larger disparities in student scores and across its schools than other jurisdictions. The differences in school-level variance are more striking than those for student-level variance. This is likely related to the relative diversity of school populations, although that point could not be pursued in detail in this study.

An exploratory analysis by jurisdiction/language group for four selected jurisdictions shows some comparable trends and some interaction effects. As an example, the effect of SES on achievement for the four selected jurisdictions is similar, even though average scores for the selected groups are near the extremes. The private school effect is also similar across those four jurisdictions with sufficient private schools for analysis. The school size effect is non-linear for all but Quebec French, where mathematics performance increases linearly with school size. The greatest variability in performance is found in schools in the 101–500 size range, which is the modal range for all except Quebec French, where schools are generally larger.

Most of the factors identified in this and other studies are universal in nature, in the sense that they exert much the same influence in all jurisdictions. This is actually reassuring from both a scientific and a policy perspective because it indicates that it is possible to identify some factors on which there can be universal agreement. While education systems may be local, many of the factors influencing achievement are not local. The next question, therefore, is what are these factors and what are their effects?

The results for socioeconomic status and for other variables likely related to SES (such as private versus public schools, the proportion of Aboriginal students in a school, and the number of students requiring program modification) are generally consistent with the literature. It is clear that this remains one of the most challenging aspects of policy, relating to both average achievement levels and equity.

Time is an important aspect of policy because the allocation and use of time is within the control of school systems and because learning obviously takes place within a time framework that has strong theoretical support and that is largely determined by educational policy. Only a few components of time were measured in this study. The results for variables such as homework time and days absent from school are consistent with the literature in indicating that more time is associated with higher achievement. The possible socioeconomic link to absence and homework was not pursued in this study, but it would be a useful area to investigate in studies more explicitly focused on socioeconomic effects. The question of obvious policy importance is whether ways should be found to provide more time for learning for those who are at socioeconomic disadvantage (or other forms of disadvantage, for that matter).

The grade-level results in this study are also consistent with the time model in that grade level relates to school starting age and is, thus, directly subject to policy influence. However, it is also possible that those in lower grades may have been "grade retained" and those in higher grades "grade-advanced." If this is so, then the grade-level results may represent an ability factor more than a factor of exposure to schooling. This clearly deserves further investigation and might be a source of insight into differential jurisdictional or school policies that can shed light on differences in achievement.

The results for out-of-school activities are also linked to the time model, which makes the results for "entertainment" puzzling. One would expect time spent on entertainment would detract from time on learning activities. Since the list of activities given in the questionnaire was not exhaustive, it is possible that students who spend time on the activities that were included under entertainment are doing so at the expense of even less educationally productive activities.

Class size is a key aspect of policy, both because of strong pressures to keep class sizes small and because class-size reduction is a costly initiative. The results indicate that class size seems to make little difference to achievement in either mathematics or science when other variables are controlled. Other large-scale assessments also tend to give results favouring larger classes. However, this finding conflicts with the results of the few large-scale experiments available that point to advantages for smaller classes. This conflict may be related to the lack of full control over extraneous variables in the survey research, or it may be a function of the grade levels studied because almost all the experimental studies have been in the early grades. The main point that can be made is that the research surveyed does not point to any advantage of reduced class size; hence, it does not support major and expensive class-size reduction initiatives in the intermediate grades as a means of increasing achievement.

The allocation and use of resources is one of the most obvious policy tools available to jurisdictions. Indeed, most high-level policy initiatives involve, either directly or indirectly, decisions on resource levels and on how resources are to be allocated. Except perhaps for the class-size results, this study sheds only limited light on resource issues. There is a positive effect for the influence of external resources on school programs, with schools reporting that the higher achievement is the result of the higher level of such influence. However, the question relates to principal's perceptions of influence and not to the absolute level of resources available. That issue has not been investigated in large-scale assessments in Canada, but should be a fruitful area to pursue in examining differences across jurisdictions.

This brings us, finally, to the question of differences among jurisdictions. As noted earlier, the comparative element in these assessments is typically limited to reports on differential achievement. Since such studies almost always show significant differences across jurisdictions, the next most obvious question is "Why do some jurisdictions do better than others?" Although this question is crucial from a policy perspective, the research based on large-scale assessment rarely addresses the question directly. Some effort was made to examine jurisdictional and language differences in the hope that a future, more explicit study can pursue this issue in more detail.

The results presented in this report show why it is difficult to get a clear answer to the question of jurisdictional differences — the differences are actually quite small compared to the differences between students and between schools within jurisdictions. Also, the ability to analyze data aggregated to the jurisdictional level is limited by the small number of units available and by the fact that the most important variables exert similar effects in all jurisdictions.

Our study has pointed to some possible ways of looking at this issue, but falls considerably short of a comprehensive analysis. The analysis has revealed some fundamental difficulties in conducting analysis at the jurisdictional level where data are highly aggregated. However, it is likely that important variations in jurisdictional policies and practices have been missed because the relevant data are not captured by PCAP questionnaires. Examples of missing data include curriculum content and coverage, resource allocations, and system-level elements of time. Nevertheless, this issue is too important to ignore: addressing the issue has the potential of adding considerable value to the PCAP assessments. Our final suggestion is, therefore, that the question of jurisdictional differences be placed more explicitly on the PCAP research agenda, and that the initial study (using the PCAP 2007 data) concern itself with developing clearer conceptual approaches to doing this and to more fully investigating the data requirements and the analytical techniques required.

TABLE OF CONTENTS

LIST OF TABLES AND CHARTS

1 WHAT IS THE PAN-CANADIAN ASSESSMENT PROGRAM?

Context and Aim of PCAP

The Pan-Canadian Assessment Program (PCAP) is the CMEC's most recent commitment to informing Canadians about how well their education systems are meeting the needs of students and society. School curriculum programs vary from jurisdiction to jurisdiction across the country, so comparing results from these varied programs is a complex task. However, young Canadians in the different jurisdictions learn many similar skills in reading, mathematics, and science. PCAP has been designed both to determine whether students across Canada reach similar levels of performance in these core disciplines at about the same age and to complement existing assessments in each jurisdiction.

The main goals of the program are 1) to inform educational policies as a means of improving approaches to learning; 2) to provide useful background information using complementary contextual questionnaires for students, teachers, and school administrators; and 3) to enable jurisdictions to use both national and international results to validate the results of their own assessment programs and to improve them.

PCAP Design

The first major reports (CMEC, 2008; CMEC, 2009) describe the PCAP design in detail. Briefly, the design incorporates achievement tests in reading, mathematics, and science, with one of these as the primary domain in each administration and the other two as secondary domains.[2] In 2007, the primary domain was reading and the test was administered to more students and was more comprehensive than those in the secondary domains. The assessment design also incorporates student, teacher, and school questionnaires, based on a framework developed from models of school learning and from other large-scale assessments.

In 2007, the assessment instruments were administered to stratified random samples of 13-year-old students representative of each of the Canadian jurisdictions (provinces and territories) and of the two official language groups within jurisdictions.[3] The samples were large enough to permit reporting by jurisdiction and by language group within most jurisdictions. A two-stage sampling process was used, with schools selected at the first stage and students within these schools at the second stage. Where numbers were small, all schools and/or all students meeting the criteria within the jurisdiction were selected. This method ensured that we had an adequate number of participants to allow for reporting on their achievement as if all students within the jurisdiction had participated.

The total sample for PCAP 2007 was approximately 30,000 13-year-old students. Approximately 20,000 wrote the reading segment, and about 10,000 wrote the mathematics and science segment. Approximately 15,000 wrote the reading segment

[2] This has changed for 2010, with all subjects now being taken by all students.
[3] For 2010, the samples have been drawn from the populations of Grade 8 (Secondary II in Quebec) students and will continue to be drawn from them in successive cycles.

in English, and 5,000 wrote in French. For mathematics and science, the numbers were about 7,500 in English and 2,500 in French. All students completed the student questionnaire, which was common to all test booklets. Language arts teachers of students who wrote the reading test completed the teacher questionnaire. School principals or their delegates completed the school questionnaire.

Reporting

The PCAP design incorporates a three-stage reporting process. The first, *Report on the Assessment of 13-Year-Olds in Reading, Mathematics, and Science* (CMEC, 2008), abbreviated as *PCAP-13 2007, Public Report*, gives achievement results for the primary subject (reading in 2007) and the two secondary subjects (mathematics and science in 2007) for Canada, the jurisdictions, and the two official language groups. The second stage report, referred to as the *Contextual Report* (CMEC, 2009) focuses on responses to student, teacher, and school questionnaires and, particularly, on the relationship between questionnaire response patterns and student achievement in the primary subject area.

The third reporting stage will comprise a series of Research Reports on specific topics of interest, based on more detailed analysis of the PCAP database. This report is the first of the intended third stage. Its focus is on the mathematics and science results and on relationships between these results and the patterns of response to the student and school questionnaires.

2 OVERVIEW

Research Questions

This study focuses on the mathematics/science component of PCAP 2007. The main questions addressed are:

- What does the research literature on PCAP data, particularly research based on large-scale assessments, tell us about the factors that contribute to mathematics and science achievement?
- How do student characteristics such as gender, home language, socioeconomic status, and civil status influence mathematics and science performance?
- How do school characteristics such as school size, governance model (public/private), class size, location (urban/rural), and diversity of student population influence mathematics and science performance?
- How do provincial/territorial characteristics in practices, policies, and procedures such as school size, governance model (public/private), class size, large-scale assessment procedures, and diversity of student population influence mathematics and science performance?
- How do students' attitudes toward school and their internal/external attributions of success and failure influence mathematics and science performance?
- Is there a link between students' reading behaviours and strategies and their mathematics and science performance?
- To what extent are reading, mathematics, and science performance inter-correlated?
- What effects do internal and external factors that influence school programs have on mathematics and science performance?

The first question requires a comprehensive literature review, with emphasis on models of school learning and on the results of large-scale assessments. The remaining questions can guide the literature review. However, the main focus of this report is to examine these questions directly using the PCAP database.

These questions closely parallel those addressed in the *PCAP-13 2007 Contextual Report on Student Achievement in Reading* (CMEC, 2009), which is used as a template for this report. The descriptive/comparative information given in the *Contextual Report* is not repeated here but is referenced as needed. Also, the teacher questionnaire is not examined here because that questionnaire was completed by language arts teachers — hence, it pertained mainly to the reading part of the assessment.

Other than the first, all other questions require detailed analysis of the PCAP database, using techniques to be described later in this section. The results of the database analysis will be linked to those found in the literature review, and the similarities and differences will be discussed.

Limitations of This Project

The main limitation of this project is that the PCAP 2007 questionnaires were designed mainly to capture background factors related to achievement in reading, not in mathematics and science. Thus, for the most part, only generic background factors and structural features of school systems, as identified in the research questions, can be examined. Detailed questions on teaching and learning strategies in mathematics and science were not included in the questionnaires. Such questions will be part of subsequent PCAP cycles in which mathematics (in 2010) and science (in 2013) will be the primary domains.

A second limitation, inherent in the design of this and all similar cross-sectional studies, is that the achievement results can be said to represent the cumulative learning of the students in the subjects measured. However, teaching strategies, school characteristics, and other elements relevant to educational policy are measured only at a single point for the school year in which the assessment is administered. These studies cannot measure the student's cumulative exposure to schooling. As a consequence, the effects of most direct interest for policy-making are likely to be underestimated, relative to the effects of socioeconomic status and other more stable student characteristics.

Methodology

The literature review focuses on two main areas: the first are the comprehensive syntheses of models of school achievement and the factors influencing achievement. Such a focus on the synthesis literature is necessary because it would be impossible, with the time and resources available, to review the large number of individual research studies on student achievement in science and mathematics.

The second and more comprehensive area is a review of results from the major large-scale national and international assessments of the past 15 years or so. Specifically, the main reports from the School Achievement Indicators Program (SAIP), from the Programme for International Student Assessment (PISA), and from the Trends in International Mathematics and Science Study (TIMSS)[4] will be reviewed and the relevant results for factors contributing to mathematics and science achievement will be abstracted. Results from the Organisation for Economic Co-operation and Development's (OECD) recent Teaching and Learning International Survey (TALIS) will also be included because this study addresses a model of teaching and learning strategies, although it presents no direct achievement results.

Our analysis of the PCAP database is patterned after the techniques used in the *Contextual Report*, but it is subject to the limitations noted above. The focus is mainly on the analytical models because descriptive data and comparisons across jurisdictions on the questionnaire variables already appear in the *Contextual Report*. Models are specified and run by jurisdiction to permit comparisons across jurisdictions. However, as on other reports of this nature, broad patterns of influences on achievement tend to be relatively

[4] Coverage of TIMSS in Canada is spotty because only a few provinces have participated in most cycles of this study. However, the patterns found internationally may be compared to those of other studies.

universal. In that sense, the jurisdictional analysis may be considered replications of the results. If notable differences exist across jurisdictions, they might yield implications for jurisdictional policy.

The main analytical tool used in the *Contextual Report* was hierarchical linear modelling (Raudenbush & Bryk, 2002), with students as the first level and schools as the second level in the model. This type of modelling took into account the cluster effect created by the two-stage sampling, where students within schools are likely to be more similar than the overall student population selected Canada-wide. However, it could not be applied in this study because the average number of students per school (about 6) writing the mathematics/science component of the test was insufficient to meet the accepted conditions for two-level models. The need for a two-level approach diminishes when there are fewer level-one units in each of the level-two units because the cluster effect is smaller. However, this then precludes an analysis of the relative proportions of variance accounted for by students and schools.

The modelling process is modified slightly to allow variables to be entered into the model in clusters, based on a logical (temporal) sequence from antecedent conditions to teaching and learning strategies. For example, demographic factors will be entered first, because these are taken to be relatively stable characteristics of students, teachers, and schools. We have run separate models for mathematics and science achievement and compared their results both with one another and with the results for reading.

3 CONCEPTUAL MODELS AND RESEARCH ON SCHOOL LEARNING

Conceptual Approach

The overarching conceptual framework for this report is an educational productivity model set within a human capital theory framework. A paper by Crocker (2007) gives details of this model. Briefly, the educational productivity model takes school achievement as a proximate outcome, which is a precursor to the longer-term economic and social outcomes of the human capital 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 access to higher education and ultimately on economic advantage and other longer-term outcomes that are considered within the broader human capital productivity model.

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 sometimes other economic or social outcomes) as the *dependent* variable and education and experience (usually along with other covariates) as *independent* variables. 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 —

$$\text{achievement} = f\,(\text{context, inputs, processes})$$

where f is a mathematical function, the form of which depends on the analysis being performed.

The specific definitions and measures of these variables and the specific mathematical function used in particular studies may be developed from theory, previous research, policy, or even conjecture. The typical analytical model takes the form of a regression equation, which yields a measure of the overall predictive power of the model and produces coefficients representing the relative contributions of the inputs and processes to the outcome. Some studies may focus on specific independent variables such as resource inputs, socioeconomic status, or time allocation and use. However, most large-scale assessments collect data on a large number of input and process variables, and the resulting equations are much more comprehensive.

Large-scale educational assessments, especially those at national and international levels, are usually designed to produce comparative achievement results across schools, provinces/states, or countries. Although research on factors affecting achievement is not necessarily the primary focus of such assessments, the assessments provide rich data sources for such research because they usually gather, through questionnaires, additional data on a large variety of inputs and processes. The formulation of an explicit research component within PCAP is perhaps an example of how research is attaining a higher profile within such studies, and how added value can be extracted from the databases being produced.

Models of School 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. Research on school learning is hampered by a lack of comprehensive theories that can account for the complexities of learning in a school setting. This lack makes it difficult to determine which factors are most worth investigating and, therefore, much "noise" is found in the analytical models used to account for achievement, and that noise makes the predictive power of these models much lower than desirable for policy-making.

A number of attempts have been made since about the 1960s to develop conceptual models of school learning that are more encompassing than the models derived from psychology, sociology, and other disciplines. Perhaps the best known is Carroll's time-based model (Carroll, 1963). This 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.

In his original 1963 article, Carroll set out to propose a mathematical formulation of the common-sense notion that learning takes place in a time framework. This formulation may be stated as follows:

> The degree of learning or achievement (L) is a function of the ratio of the time actually spent on learning (Tsl) to the time needed to learn (Tnl), or
>
> $L = f(Tsl/Tnl)$

Although mathematical in form, the model is essentially a conceptual one because the detailed nature of the function is unspecified in Carroll's original formulation. For example, it is not clear if the relationship is linear or if there are saturation, fatigue, or other effects that might limit the value of spending more time. Obviously, at some level, such limits exist — time is a finite resource. However, in practical terms, it is not clear whether many individuals approach these limits.

The Carroll model has been widely investigated. Carroll himself revisited the model in a 1989 retrospective, citing in particular a review by Hawley, Rosenholtz, Goodstein, and Hasselbring (1984) of some 3,000 studies based on his model. He concludes that optimizing academic learning time is one of the most important factors in improving student achievement. More recent reviews by Scheerens and Bosker (1997) and Marzano (2003) reinforce this conclusion. However, the problem that remains is how this optimization can be accomplished, especially within the overall constraints of conventional school years or days.

In an attempt to address more explicitly the issue of equity in learning, Bloom (1976) extended Carroll's model (1963) to the concept of "mastery learning." Addressing directly the equity issue, Bloom proposed that time allocated to a learning task should vary sufficiently to allow almost all students to achieve the specified learning outcomes. Those who need more time would thus get more time. Adopting this proposal is a critical policy issue, because it offers the promise of greater equity in learning through differential time allocation and use. This principle of equity can be applied to a variety of school situations such as tutoring, homework, and offering extra time on core school subjects to those needing it. More comprehensive policies — such as lengthening the school day or year for those students who need more time — might also be derived from the mastery approach. However, neither policy is easy to implement in an established school system and would likely come at high added cost. Nevertheless, the principle is embodied in activities such as summer schools, providing more time for core subjects for those who need it, or programs based on continuous progress. Repeating a grade, the traditional way of providing more time, has been shown in many studies to be ineffective because of how the student's time is used.

More recent syntheses of research on teaching have resulted in other models that help identify factors related to achievement. Perhaps the best known of these is the Wang-Haertel-Walberg model of school learning (Wang, Haertel, & Walberg, 1993). These authors use a comprehensive synthesis of research to advance a concept of "proximity" as a way of thinking about the relative effects of various factors. 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.

More specifically, the Wang-Haertel-Walberg (WHW) synthesis revealed that variables showing the strongest relationships with achievement are those in the areas of classroom management, meta-cognitive processes, cognitive processes, home environment/parental support, and student/teacher social interactions. Motivation, peer group influences, quantity of instruction, classroom climate, and other proximal variables also receive high rankings (Wang et al., 1994). In contrast, school-level variables and the variables related to broad jurisdictional and district-level education policies tend to be less influential. This point is of crucial importance because it suggests that broad policy initiatives are likely to result in improved learning *only* if translated into change at the level of the individual teacher or student.

Some more recent syntheses are consistent with this 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. The type of school climate most likely to enhance student learning is an orderly atmosphere, rules and regulations, and good student conduct and behaviour encouraged by effective classroom management strategies such as direct instruction, monitoring student progress, and instilling a positive work attitude. Marzano (2003) independently developed a list that is almost identical to that of Scheerens and Bosker.

Most of the studies covered in these syntheses have been small-scale and local in scope, and they typically cover only a few of the many variables that might be expected to influence learning. Because of the large number of variables available and the wide range of contexts used, large-scale surveys such as PCAP offer the potential to uncover more robust relationships, and to investigate the influence of context variations on the results. The extensive coverage of the PCAP database permits the analysis of particular factors that may relate to students' achievement, while also taking account of other factors that may cloud or complicate this relationship.

Socioeconomic Factors

The focus of most models of school learning has been on teaching and learning strategies because these are the factors most amenable to policy-driven interventions. However, no such model would be complete without acknowledging the importance of socioeconomic status (SES) on achievement. Socioeconomic factors are among the proximate factors in the Wang-Haertel-Walberg model and, in empirical studies, they are well established among the most influential factors. The SES factors are particularly important in discussing the equity issue because there is strong evidence that lower achievement is more prevalent among those from minority groups and lower socioeconomic backgrounds. Socioeconomic effects are of particular interest to those concerned with broad social policy and with immigration policy, where the policy issues lie outside of the education system itself.

A recent meta-analytic review (Sirin, 2005) examined the effects of socioeconomic status on achievement in the research literature from 1990 to 2000. Sirin identified the three main indicators of children's SES as parental income, parental education, and parental occupation. These measures are thought to be conceptually distinct, though inter-correlated. Household possessions such as books, computers, study space, and the availability of educational services after school are less commonly used but are somewhat more amenable to measurement through student questionnaires than are the main measures. Aggregated measures such as school-level socioeconomic status are also used in studies of school effects.

The Sirin review included 58 studies that met the selection criteria. The usual meta-analytic techniques were applied and yielded average effect sizes (measured as correlations between SES and achievement) for various measures — grade levels, student minority status, and school location (rural, suburban, urban). A mean effect size of 0.299 was found over all studies. This compared to a mean effect size of 0.343 reported in an earlier review (White, 1982), leading Sirin to conclude that the SES effect has diminished somewhat over time.

Large-scale assessments have consistently shown similar socioeconomic effects. For example, the PISA 2000 study showed statistically significant positive effects on achievement for reading, mathematics, and science for a composite index of socioeconomic status and for the specific measure of books in the home for almost all of the participating countries (Bussière et al., 2004). Similar results were found for the Canadian provinces in all three of the last SAIP cycles (CMEC, 2002; 2003; 2005) as well as in PCAP 2007 Reading (CMEC, 2009).

It is important to note that the effects of SES and of other student characteristics may be exaggerated in cross-sectional studies, relative to the effects for teaching and learning strategies, because the latter can be measured *only* for the year in which the study is conducted while measures for both student background and achievement are more likely to be stable and cumulative in nature.

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.

No recent systematic reviews of the relationship between such factors and academic achievement could be found. However, these factors are included in most large-scale assessments, which report relatively consistent findings.

Among the strongest of these findings is that for achievement correlated by gender and language. Almost all large-scale studies show a strong effect in favour of girls in reading and writing assessments (Bussière, et al., 2001; CMEC, 2002). In contrast, the same studies show smaller and less consistent results for mathematics and science. In PISA 2000, differences on mathematics and science tended to favour boys in most countries, including Canada. For PISA 2003 (OECD, 2004), where mathematics was the main focus, differences for most countries, including Canada, again tended to favour boys. The PISA 2006 report (OECD, 2007), with science as the main subject, showed a small difference internationally favouring boys, but the difference for Canadian students was not statistically significant. The final SAIP mathematics assessment (CMEC, 2003) showed only slight differences in favour of boys, while the last SAIP science assessment in 2004 (CMEC, 2005) showed no significant gender differences. This pattern continued with PCAP, the new assessment program, in 2007 when no significant gender difference was found for either mathematics or science.

Language as well as racial and ethnic effects tend to be country-specific because of different language configurations and immigration patterns. In general, the SAIP, PISA, and PCAP assessments show that, in Canada, 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 in jurisdictions outside Quebec. Immigration status has been the focus of some Canadian studies, in which the results tended to slightly favour those born in Canada over first-generation immigrants.

School Effects

Research on school effects has developed as a field distinct from other aspects studied on the topic of school learning. Research on school effects has most often emphasized the allocation and use of resources. These are commonly measured by such indicators as per-student expenditures or by such variables as teacher qualifications or class sizes, which are directly related to cost. Evidence about how resources are used is less common.

This area has proved to be highly controversial, as shown in the exchanges between Hanushek (1996) and Greenwald, Hedges, and Laine (1996). Hanushek has consistently argued that the research shows no strong or consistent evidence to support a relationship between educational resources and student achievement. Greenwald, Hedges, and Laine are equally adamant in their argument that the evidence supports that greater resources result in improved performance. One of the main differences in the two perspectives is that Hanushek typically used aggregated data such as state-level expenditures while Greenwald, Hedges, and Laine conducted meta-analyses of specific school-level factors. The latter indicates that such school-level factors as student-teacher ratio, school size, and teacher education and experience can affect achievement. Hanushek supports the results showing that some uses of resources can be effective, but has argued that much of the increase in resources goes into activities that have little or no effect. This essentially shifts the argument from resource allocation to resource use, a point consistent with the Wang–Haertel–Walberg proximity model.

Following from this, a RAND Corporation review and meta-analysis (Grissmer, Flanagan, Kawata, & Williamson, 2000) examined U.S. 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, a hypothesis that was found to be true, 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, although this seems to have had little impact on overall average performance.

Grissmer et al. also examined the results of more recent studies, concluding that there is now general support for positive effects from resources, while acknowledging the methodological challenges that plague the field. There now seems to be consensus that measures at lower levels of aggregation (e.g., student or school) yield better results than those at higher levels (e.g., state/jurisdiction or country), which is consistent with the proximity model.

Class size deserves comment because it is one of the most widely investigated resource variables, likely because of its high public profile and the strong common-sense belief that smaller classes should yield higher achievement (and other benefits). Reducing class size is an explicit policy direction in many jurisdictions, but it is a costly initiative because smaller classes require more teachers and more space and other resources. Recent 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. However, a reduction to a level of about 15 or fewer students is required to show significant effects. Such a reduction is difficult to achieve in most systems without a large increase in cost that, in turn, raises the question of cost effectiveness, a point rarely addressed in the research literature.

The results from large-scale assessments, which typically encompass students in middle or secondary grades, tend to show the opposite effect, where students in larger classes do better on the performance measures even when other school factors are controlled (CMEC, 2002; 2003; 2009). In some cases, non-linear effects are found (OECD, 2001; 2004). It is important to note that these assessments do not account for the possibility that lower-achieving students may be assigned to smaller classes. Differences in the grade levels studied as well as the time period over which class size is measured may also help account for these counter-intuitive results. 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.

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.

There is an extensive American literature on school size, much of which focuses on an argument for decreasing secondary school sizes. An ERIC digest of this research (Irmsher, 1997) concluded that larger schools do not produce higher academic performance at lower cost, as advocates have argued. However, there is evidence that larger schools work well for students of higher socioeconomic status whereas smaller schools are better for disadvantaged students. A more recent review by Leithwood and Jantzi (2009) came to essentially the same conclusion.

More recently, large-scale assessments show either very small effects for school size (OECD, 2004) or positive effects for larger schools (CMEC, 2002; 2003). It is likely that internationally and in Canada, in particular, large school size is not as clearly associated with inner-city disadvantaged populations as is the case in the United States. Also, in the large-scale assessments, the school size ranges used may not capture the effects of very large schools because of inadequate sampling at the school level. For example, the Leithwood and Jantzi recommendation that secondary schools be limited to about 1,000 students would apply to less than 10 percent of all schools in the PCAP 2007 sample.

The relative performance of public and private schools has been the subject of considerable investigation. A recent review by Coulson (2009) looked at more than 150 comparisons of public and private schools internationally, concluding that private sector schools outperform public sector schools in the overwhelming majority of cases, and that the results are more pronounced to the degree that "private" schools are driven by market forces such as competition. Although not all the studies included in the review controlled for selection factors, the author argued that the results are so strongly in one direction that selection bias is not likely to be a significant methodological flaw.

Similar public/private differences are found in most of the large-scale assessment studies (OECD, 2001; 2004). In Canada, only a few provinces have sufficient private schools for this effect to be measured, and where this is the case, the results show higher achievement in private schools (CMEC, 2009). The PCAP-13 2007 Reading results (CMEC, 2009) show that this effect is somewhat attenuated, although remaining relatively large, when socioeconomic status and other individual and school characteristics are controlled.

Similar results were found in a recent 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 been the focus of many studies of school effects. The Wöessmann et al. study cited above is an example. That study 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).

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, Sánchez, and Verdisco (2004), 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.

Instructional Climate

This review does not pursue instructional climate or teaching and learning strategies in any detail because these were not measured in the PCAP 2007 assessment for mathematics and science. However, a few generic aspects of instructional climate may be noted. These include such aspects as use of time, time allocation, homework, and absenteeism, as well as assessment practices and the accommodation of special-needs students.

Homework is one important component of time, which can vary substantially from one student to another. A recent comprehensive review of the effects of homework on academic achievement is available (Cooper, Robinson, & Patall, 2006). Of more than 900 empirical studies conducted between 1987 and 2003, about 75 met the selection criteria established by the reviewers. These studies were reviewed using established qualitative and quantitative synthesis methods. Most studies referred to homework in either language arts or mathematics.

The results showed the effects of homework to be generally positive. Effects are very small at the elementary level but increase at the higher grades. Like other factors that affect learning, the effects are not large enough to make a decisive difference for most students but they can certainly contribute to a difference between "pass" and "fail" for a marginal student. The limited information available from the research on the optimum amount of homework suggests that the upper range for secondary school students is between 90 and 150 minutes total per day. 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.

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

Research on Mathematics and Science Teaching and Learning

Mathematics and science are the foundations of the activities that distinguish the advanced from the less advanced societies. A compelling case can be made that mathematics and science are fundamental school subjects and, in this sense, mathematics and science learning is different from language learning. There are no mathematical or scientific equivalents of the bedtime story, and mathematical and scientific language and activities are not part of everyday communications, except in a rudimentary way. It is difficult to imagine that any significant level of competence in these subjects, particularly in mathematics, could be acquired outside of the school setting.

A large body of research exists on mathematics and science teaching and learning. These two disciplines each have several journals dedicated to this research, and much of the research based on large-scale assessment data is necessarily devoted to these two areas. Indeed, one such assessment, TIMSS has followed these subjects internationally for more than fifteen years, with increasing numbers of participating countries at each cycle. PISA has also now completed three full cycles, with mathematics as the primary subject in 2003 and science in 2006. Before its replacement by PCAP, SAIP also completed three assessment cycles in each of mathematics and science.

It is impossible in a brief review to do justice to this body of research. Instead, it is necessary to rely as much as possible on the emerging synthesis literature, which focuses mainly on small-scale studies reported in the major journals devoted to these areas. It is possible to delve somewhat more deeply into the large-scale assessment research because the number of studies is smaller although the scope of these studies is much larger. What follows, therefore, are brief accounts of some major syntheses of the factors related to mathematics and science learning and additional detail derived from a direct review of the SAIP and PISA[5] studies.

Mathematics

The most comprehensive recent review of factors affecting mathematics learning is that by Hiebert and Grouws (2007). These authors examine the evidence available to support the claim that "the nature of mathematics teaching significantly affects the nature of students' learning." After reviewing the early research and concluding that opportunity to learn is the key requirement, Hiebert and Grouws focus more specifically on the distinction between teaching for meaning and teaching for skill development. Citing the well-known TIMSS video study (Stigler & Hiebert, 1999), Hiebert and Grouws point out that mathematics teaching in some of the highest achieving countries is characterized by classroom interactions designed to focus on conceptual meaning. In contrast, mathematics classes in the United States and Australia are characterized by attention to lower-level skill development.[6] Many other studies cited suggest that teaching for conceptual development is associated with higher achievement and that such teaching can also enhance skill development.

[5] TIMSS is not examined in detail here because recent TIMSS reports tend to focus on comparative achievement results rather than on factors influencing achievement.

[6] No Canadian schools were included in the TIMSS video study. The PISA studies indicate that schools in English-speaking countries are more similar to each other than they are to schools in other countries. It thus seems reasonable to surmise that Canadian schools are more like those in the United States or Australia than like those in the Netherlands, Japan, or Hong Kong. Nevertheless, it is important to point out that, in recent international assessments, Canadian students do better than those in either the United States or Australia.

Hiebert and Grouws argue that we can apply the admittedly incomplete knowledge base in some useful ways immediately while other areas require further research. The following are among the points which seem to have immediate application:

- Teaching for conceptual meaning and teaching for skill development are not contradictory. Teaching that focuses on meaning is likely also to enhance skill development.
- There is a need to focus on teaching, not on teachers. General characteristics of teachers and teacher qualifications are not associated with student outcomes. The focus needs to be on what teachers and students do.
- It is important to be explicit about learning goals. (We note that this is one area in which there has been significant improvement in recent years.)

A more recent synthesis (Slavin, Lake, & Groff, 2009) looked at innovative mathematics curriculum and instructional programs in middle and secondary schools. These authors concluded that instructional practices matter more than the textbooks or technologies. In particular, the review found more positive effects for cooperative learning strategies than for other strategies. These results suggest that the strong emphasis on curriculum, as reflected in the National Council of Teachers of Mathematics (NCTM) standards and the programs that follow from them, may be misplaced and that greater emphasis is needed on strategies that enhance student engagement. This is broadly consistent with the Wang-Haertel-Walberg proximity model, because instructional practices are arguably closer to the daily lives of teachers and students than are the more abstract principles embodied in curriculum documents.

The SAIP and PISA studies warrant more detailed comment because they are among the few for which specific Canadian results are available and because they are large-scale, high-quality comparative studies capable of yielding results that are reasonably generalizable to Canadian and jurisdictional populations of students. The focus here is on the factors influencing mathematics performance.

PISA 2003 (Bussière, Cartwright, & Knighton, 2004) assessed the mathematics performance of 15-year-old students in more than 40 countries, mostly the developed countries that are members of OECD. In Canada, the sample sizes used allow provincial comparisons of achievement and of the relationship between achievement and various teaching and learning strategy indices. These indices involve measures of what was called "student engagement in mathematics." Some of the major results for Canada may be summarized as follows:

- Students using high levels of all three indicators of engagement in mathematics learning — labelled memorization, elaboration and control — had higher levels of achievement than those with low levels of engagement. Although these indicators may seem contradictory (memorization may be associated with basic skills, elaboration with conceptual understanding), all showed positive effects.
- Preference for cooperative learning situations showed a negative association with achievement while preference for competitive learning situations showed a positive association with achievement.
- Positive attitudes toward mathematics (interest in mathematics, belief in its usefulness, perceived ability, and mathematics confidence) were all positively associated with achievement, but mathematics anxiety was negatively associated with achievement.

The most recent of three SAIP mathematics assessments, conducted in 2001 (CMEC, 2003), also gave some detailed results linking mathematics achievement to student background and aspirations as well as to teaching and learning strategies. Because these are more complex than the PISA results, a summary is given in Tables 3.1 and 3.2 below.

It is important to note that these are all correlational results that do not necessarily point directly to causes of higher or lower achievement. The class size result is an obvious example of this. It is likely that this result is a consequence of larger classes being concentrated in urban areas, where other factors contribute to higher achievement. Thus, the class size effect is confounded with these other factors. Although more highly controlled studies of class size indicate that smaller classes yield improved achievement in the primary grades, it is clear from both the SAIP studies and other studies that class size is not an overriding influence on achievement, whatever the value of smaller classes for other purposes.

Table 3.1 **Factors associated with mathematics achievement: SAIP Mathematics III, 2001**

Positive effects	Negative effects
Student effects	**Student effects**
• Mother's education	• Taking mathematics tutoring
• Planning to attend university and to work in a field requiring mathematics	• Perceived difficulty of mathematics
• Time on mathematics homework	• Attribution of poor marks to bad luck
• Persistence in solving mathematics problems	• Days absent from school
• Teacher gives notes	• Work with parents on homework
• Teacher shows how to do problems	• Doing mathematics projects
• Teacher assigns homework	• Working in small groups
• Work on textbook exercises	• Off-topic discussion
• Students ask questions	• Losing 5–10 minutes because of disruptions
• Students use calculators	• Using books, magazines other than textbooks
	• Using computers
	• Using slides, videos, films
School effects	**School effects**
• Larger schools and larger communities	• Limitations on instruction due to student backgrounds, diversity, resources, community conditions
• Larger class sizes	

The negative SAIP results for tutoring are interesting in light of the seemingly widespread use of tutoring in mathematics. However, like class size, the tutoring effect is likely confounded with other factors, such as the performance of students prior to their using a tutor. The most likely explanation of this result is that most students who are tutored are likely to be performing at low levels; otherwise they would not need a tutor. For example, if many of those being tutored are high-achieving students striving for higher marks, the relationship would not be negative. While tutoring may have some positive effects on school grades, it is unlikely that it transforms low-achieving students into high-achieving ones or that it has any impact on the type of broader mathematical literacy measured by large-scale assessments. We were not able to locate a recent synthesis of research on mathematics tutoring, so its overall impact is not known. However, it is interesting to note that tutoring, like homework, is consistent with the idea of allocating additional time to students who need it. The effects of tutoring, controlling for other factors, might be expected to be at least as great as those for homework.

Table 3.2 **Summary of variables used in this report and effects found in other studies: Mathematics**

Variable	SAIP Mathematics III, 2001	PISA Mathematics 2003	Other Studies
Demographic and socioeconomic			
Gender	Small differences favouring males	Significant differences favouring males in most countries	Early studies show effects favouring males. More recent evidence suggests that the difference is diminishing
Home language	Positive if home language is the same as test language	Positive for home language the same as test language in most countries	
Grade level	Not reported	Not studied	
Immigration status	No effect	Positive for native-born students in most countries. Smaller difference in Canada than elsewhere	
Mother's education	Positive	Positive	
Books in the home		Not significant	
School size	Positive for larger schools	Positive for larger schools	Generally positive for smaller schools
School governance	Not reported	Positive for private schools	Positive for private schools
Aboriginal students	Not studied	Not studied	
ESL/FSL students			
Community size	Positive for larger communities	Positive for rural schools	
Attitudes and attributions			
Attitudes to school	No effect	Mixed effects across countries. Slightly positive for Canada	
Attitudes to reading			
External attributions of success	No effect		
External attributions of failure	Negative		
Fatalism	Negative		
Internal attributions of success and failure	Positive for persistence		
Out-of-school activities			
Reading/research			
Entertainment			
Academic/cultural activities			
Instructional climate			
Class size	Larger classes positive but inconsistent across jurisdictions	No effect	
Desired placement of special-needs students			
Length of class periods	Not studied		
Days absent from school	Negative		
Time on homework	Positive		
Frequency of use of provincial/state/national assessments	Not studied	No effect	

Science

No syntheses comparable to the Hiebert and Grouws (2007) or Slavin, Lake, and Groff (2009) studies could be found for science. Instead, we use as examples a couple of recent studies based on specific large-scale databases. The first is a report by Braun, Coley, Jia, and Trepani (2009) on factors related to grade 8 science achievement based on 2005 data from the United States NAEP. Using methods similar to those in the PCAP 2007 *Contextual Report*, these authors found several demographic and socioeconomic factors related to achievement. These factors include race, gender, home language, books in the home, absence from school, and teacher certification and specialization. All of these (with the exception of race which was not measured in PCAP) tend to be consistent with the results of the 2007 PCAP reading test.

Teaching strategies showing positive results for science achievement included reading a science textbook, doing hands-on activities, writing long answers to science tests, talking about science assessments and activities, and working with others on science activities. Giving oral science reports and using the library for science were found to have negative effects on achievement. Taking a science test, watching teachers do science demonstrations, discussing science in the news, reading books or magazines about science, and preparing written science reports had positive effects. Again, where the variables are approximately comparable, these results are consistent with those for PCAP reading and with SAIP results.

The last SAIP science assessment, conducted in 2004, gave correlational results on factors associated with science achievement (Table 3.3). These showed few significant effects at the student level and none at the school level. Positive effects were generic in nature including reading for enjoyment, interest in school work, and internal attribution of success. Negative results tended to be more science-specific, including lack of interest in science, perceived difficulty of science, and science field trips outside the school. Having quizzes or tests in science was also associated with lower achievement.

Table 3.3 **Summary of variables used in this report and effects found in other studies: Science**

Variable	SAIP Science III, 2004	PISA Science 2006	Other Studies
Demographic and socioeconomic			
Gender	No differences	No difference for most countries	
Home language	Negative for first language other than the language of the school		
Grade level	Not reported		
Immigration status		Positive for native-born students in most countries. Smaller difference in Canada than elsewhere	
Mother's education	Not reported	Positive for broader index of SES. SES gradient smaller in Canada than in many other countries.	
Books in the home	Not reported		
School size	Positive for larger schools		Generally positive for smaller schools
School governance	Positive for private schools	Positive for private schools in most countries including Canada	
Aboriginal students	Not studied		
ESL/FSL students	Not studied		
Community size	Positive for larger communities		
Attitudes and attributions			
Attitudes to school	Interest in school work positive for 16-year-olds	Not studied directly	Willson (1983) meta-analysis showed small effect size for attitudes overall with differences by gender and grade level. Achievement seems to cause attitude in elementary school but not in high school.
Attitudes to reading	Positive for time spent reading for enjoyment	Positive for high interest in science	
External attributions of success	No effect		
External attributions of failure	No effect		
Fatalism	No effect		
Internal attributions of success and failure	Positive	Positive for index of self-efficacy in science	
Perceived performance on PCAP test			
Out-of-school activities			
Reading/research			
Entertainment	Negative for 16-year-olds; no effect for 13-year-olds		
Academic/cultural activities			
Instructional climate			
Class size	Positive for larger classes		
Special-needs students	Negative for schools with more special-needs students or wide range of abilities		
Length of class periods			
Days absent from school	Negative		
Time on homework	Positive for 16-year-olds; no effect for 13-year-olds	Positive	
Existence and use of provincial/state/national assessments	Not studied	Positive for posting achievement data publicly and for allocating resources based on such data	

Summary of Research

It is important to note that our review of the research falls far short of a comprehensive examination of the many factors that affect mathematics and science achievement. In some cases, the synthesis literature is adequate to give a picture of the relevant effects; in others, it is not and it is impossible to fill the gap in a review of this nature. Indeed, each of the main areas examined could, in itself, be the subject of an extensive review.

Nevertheless, this review supports the following conclusions:

- The research consistently shows a positive effect of higher socioeconomic status on achievement. In large-scale assessments, this effect may be over-estimated, relative to effects for teaching and learning strategies. Both achievement and SES represent stable, cumulative features of the student's experience whereas teaching and learning strategies are typically measured at a single point in time, which may not reflect the student's long-term experience with schooling.
- Time is important for achievement. However, since the macro-components of time (length of school year and day) are not highly variable across schools or jurisdictions in Canada, it is difficult to give a full picture of the effects of time.
- Components of time that are fairly easily measured (such as time lost in classrooms, homework, and absenteeism) have effects in the expected direction, — more time contributes to higher achievement.
- Generally, both an immigrant background and a home language different from the language of the school contribute to lower achievement.
- Positive attitudes are associated with higher achievement because achievement is the ultimate goal. Although the direction of causality is not clear, it makes more sense to assume that better attitudes contribute to higher achievement than the other way around. Some, indeed, may consider that encouraging more positive attitudes is desirable in its own right, independent of its effect on achievement.
- The research on school size and class size yields contradictory results. Large-scale assessment results tend to show either no differences or differences that favour larger schools and larger classes. However, other studies tend to show results in the expected direction. It is possible that these contradictions are related to lack of experimental controls or lack of measurement precision in large scale assessments. However, it is also possible that the latter are more reflective of the real world than experimental studies of class size, in particular.
- Students in private schools tend to do better than those in public schools, even after controlling for SES. However, other selection factors, such as underlying student abilities, are generally not controlled in examining this effect.
- Strategies that represent an orderly, structured approach to teaching tend to have positive effects on achievement. However, the research in this area has not been adequately reviewed here, and these effects cannot be investigated in any detail in this study because the available PCAP data on teaching and learning are mostly specific to reading.
- More generally, the results lend support to the Carroll and Wang-Haertel-Walberg models of school achievement, and particularly to the idea that proximal factors are greater contributors to achievement than distal factors.

4 OVERVIEW OF PCAP 2007 MATHEMATICS AND SCIENCE ACHIEVEMENT

The details of the mathematics and science assessments are given in the PCAP 2007 report (CMEC, 2008). These are summarized briefly here to provide context for the results to follow.

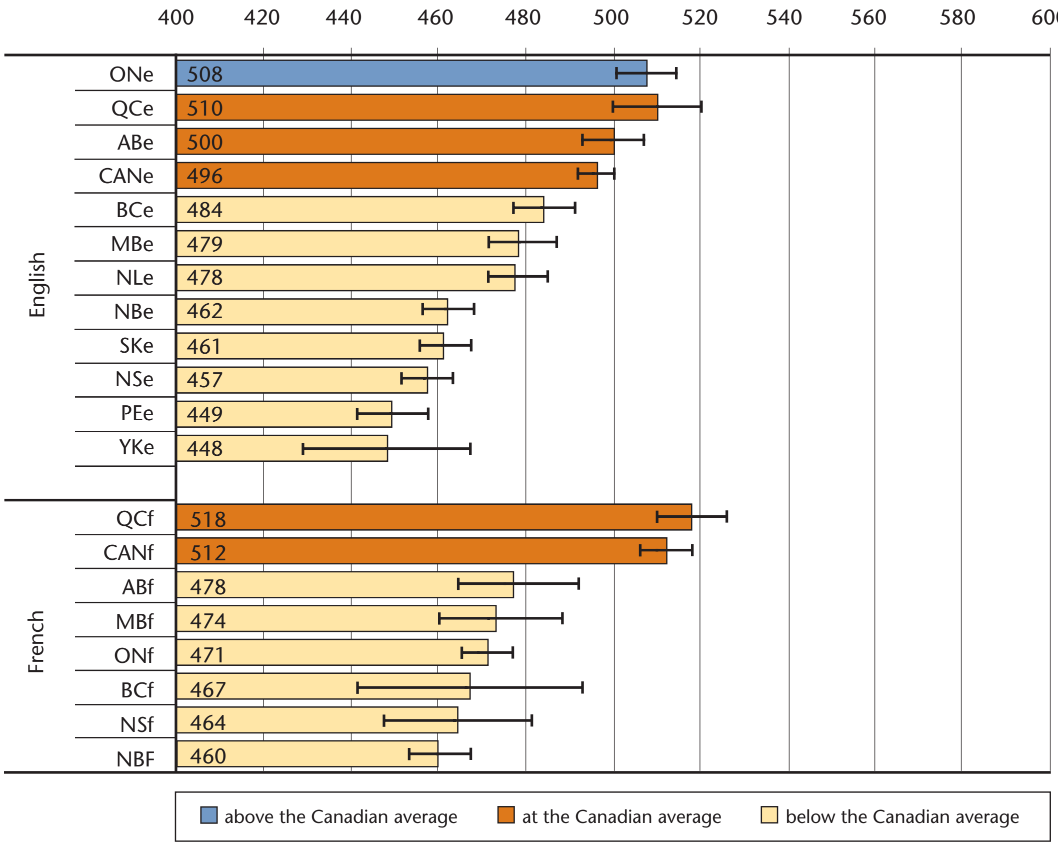
Mathematics Assessment

The mathematics assessment is organized around four domains and three processes that have been drawn from the standards of NCTM and are reflected in jurisdictional curricula. These were combined in reporting the 2007 mathematics results because, unlike the reading assessment, mathematics was a minor domain and the mathematics assessment was not comprehensive enough to give a reliable breakdown by domain.

The assessments were analyzed using item response theory and were scaled to a weighted Canada mean of 500 and a standard deviation of 100. All comparisons across jurisdictions and for other factors are expressed in terms of means and confidence intervals, allowing judgments to be made, which accounts for sampling error.

Chart 4.1 shows the mathematics results by jurisdiction and language. This chart also illustrates the format in which most of the results will be presented. The number on each bar is the mean score on the mathematics assessment for students in the category represented by the bar. The line at the end of each bar is the 95% confidence interval around the mean. The confidence interval is a measure of statistical error arising from the fact that the scores are based on samples and not on the entire populations of the entities of interest (in this case, jurisdictions and language groups within jurisdictions). The error bar indicates that we can say that the score for the population as a whole will be within the range given by the error bar 95 times out of 100. The error bars may also be used to compare groups. The difference between any two groups is said to be "statistically significant" if their error bars do not overlap.

Chart 4.1 **Mean mathematics performance by jurisdiction and language**



The results show that Ontario anglophone students performed better than the Canadian average for all anglophone students. Anglophone students in Quebec and Alberta performed at about the Canadian average. Students in all other jurisdictions performed below the Canadian average. On the francophone side, students in Quebec performed at the Canadian average while those in all other jurisdictions performed below the Canadian average.[7] It is also worth noting that the anglophone Canadian average was about at the overall Canadian average of 500, while the performance of francophone students was above the overall Canadian average.

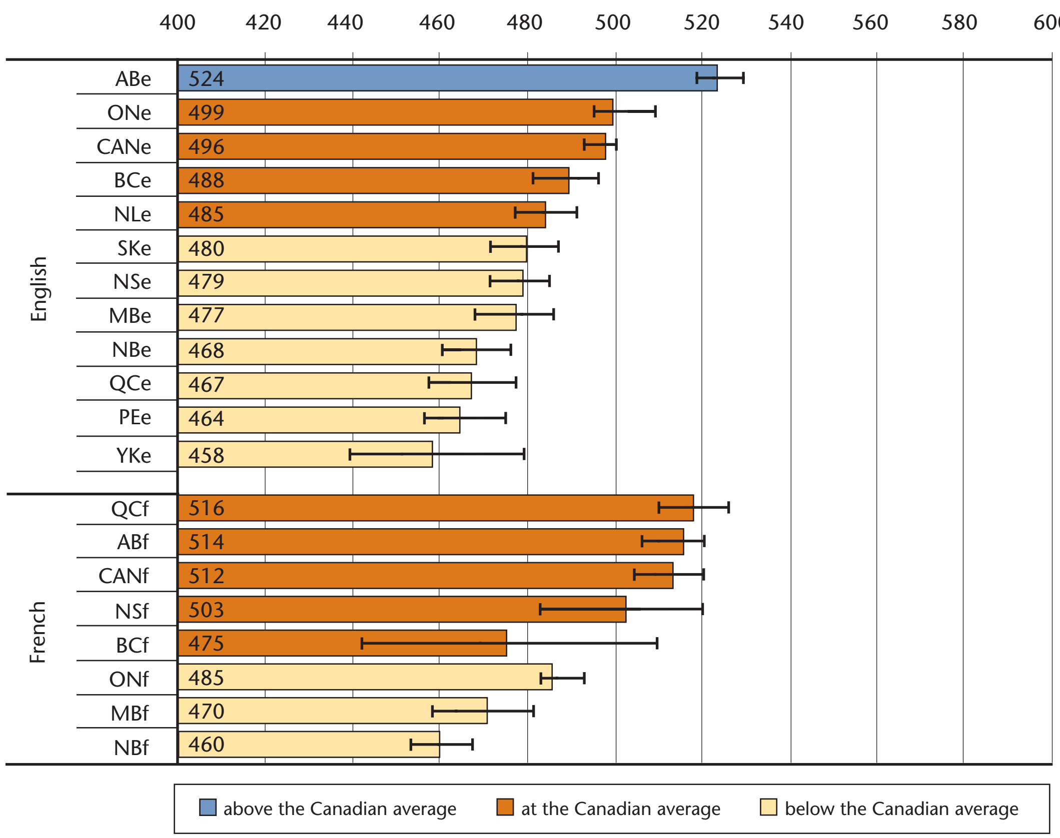
Science

The PCAP Science assessment is based on a concept of "scientific literacy." This concept encompasses three main domains — science inquiry, problem solving, and decision making — with sub-domains reflecting the nature of science and technology, scientific knowledge, skills, and attitudes. Like the mathematics assessment, these were combined in reporting the results because science was a minor domain.

[7] Only in Manitoba did all French Immersion students selected participate in French, with results included in the MBf cohort.

Chart 4.2 gives the science results by jurisdiction and language. The general pattern here is similar to that for mathematics, with some differences in detail. Among the anglophone populations, Alberta students performed better than the anglophone Canadian average. Students in Ontario, British Columbia, and Newfoundland and Labrador were at about the Canadian average. Those in all other jurisdictions were below the Canadian average. Among francophones, most jurisdictions were near the Canadian francophone average, with none above. Three francophone jurisdictions (Ontario, Manitoba,[8] and New Brunswick) were below the francophone Canadian average. More generally, anglophone students overall performed at a level close to the Canadian average while, again, the performance of francophone students was above the Canadian average.

Chart 4.2 **Mean science performance by jurisdiction and language**



Correlations Among the Subjects

A question of general interest in studies of achievement is the extent to which achievement scores represent a general ability trait or whether achievement is subject specific.[9] This question can be answered in a limited way by examining correlations among the three measures available. However, the ability to do this is limited by the fact that different students wrote the reading test from those who wrote the mathematics/science test. Fortunately, both tests were administered in most schools, so while the mathematics and science assessments can be correlated at the student level, the

[8] French Immersion students in Manitoba participated in French, with their results included in the Manitoba Francophone cohort.

[9] The question of the existence of a general ability trait as opposed to more specific traits is a major subject of debate in the literature on intelligence. A review of this literature is beyond the scope of this study.

correlation of these assessments (math and science) with reading can be done only at the school level.[10] However, for comparability in computing these correlations, all were computed at the school level, using school average scores. Student level correlations between mathematics and science followed much the same pattern but tended to be slightly smaller.

Table 4.1 shows correlations among the three subjects at a national level and for selected jurisdiction/language groups. For the most part, the pattern is one of higher correlations between mathematics and science than between these two subjects and reading. At the jurisdictional level, the correlations between mathematics and science are fairly consistent throughout. However, the correlations with reading are more variable. Because many of these correlations are based on both small samples of schools and small samples of students within schools, no strong inferences should be made from these differences.

Table 4.1 **Correlations among the subjects**

National			
	Reading	**Science**	**Mathematics**
Reading		0.357	0.381
Science			0.640
Jurisdiction/Language			
BC	Reading	0.302	0.427
	Science	1	0.497
AB	Reading	0.542	0.507
	Science	1	0.689
SK	Reading	0.347	0.399
	Science	1	0.699
MBe	Reading	0.226	0.283
	Science	1	0.732
MBf	Reading	0.620	0.641
	Science	1	0.564
ONe	Reading	0.265	0.236
	Science	1	0.588
ONf	Reading	0.135	0.126
	Science	1	0.711
QCe	Reading	0.557	0.552
	Science	1	0.827
QCf	Reading	0.565	0.597
	Science	1	0.676
NBe	Reading	0.514	0.514
	Science	1	0.714
NBf	Reading	0.273	0.065
	Science	1	0.665
NS	Reading	0.386	0.422
	Science	1	0.701
PE	Reading	-0.086	-0.252
	Science	1	0.678
NL	Reading	0.281	0.334
	Science	1	0.726
YK	Reading	0.476	0.789
	Science	1	0.806

[10] School-level correlations are less satisfactory than student-level correlations because of smaller effective sample sizes and added complexity in weighting. We understand that in PCAP 2010 all students wrote all tests, which will facilitate this type of analysis.

5 VARIABLE SELECTION AND DESCRIPTIVE STATISTICS

As indicated earlier, the research questions posed for this study are similar to those investigated in the *PCAP-13 2007 Contextual Report on Student Achievement in Reading* (CMEC, 2009). The main difference is that many of the variables examined in that report are specific to teaching and learning in reading. A more limited number of variables is available for mathematics and science. These include student, teacher, and school demographic characteristics, some aspects of time allocation and use, student attitudes and attributions of success and failure, instructional climate, and out-of-school activities. Most of these have some theoretical justification for inclusion or have been investigated in other large-scale assessments as summarized in the literature review.

The *Contextual Report* gives comparisons by jurisdiction and language for all of the variables to be used in this report. These results will not be repeated here in detail. However, in order to provide sufficient context for the models, a description of each variable and basic descriptive statistics for Canada are given in Table 5.1.

Table 5.1 **Definitions and descriptive statistics for variables in the models**

Variable	Source	Scale/Categories	Mean	SD
Dependent				
Mathematics achievement	Math test	Scaled score	500	100
Science achievement	Science test	Scaled score	500	100
Demographics (all following sources are questionnaires)				
Gender	Student	Male = 1, Female = 0	0.49	
Home language English	Student	English = 1, not English = 0	0.67	
Home language French	Student	French = 1, not French = 0	0.20	
Home language other	Student	Other = 1, not other = 0	0.13	
Grade level	Student	1–5; grades 6 to 10	3.14	0.52
Immigration status	Student	Born in Canada = 1, not born in Canada = 0	0.91	
Mother's education	Student	1–6; less than high school to university degree	4.04	1.77
Number of books in the home	Student	1–5; 0–10 books to more than 200 books	3.36	1.21
School size	School	1–4; < 100 to > 1000	2.57	0.82
Public school	School	Public = 1, not public = 0	0.92	
Private school	School	Private = 1, not private = 0	0.08	
Aboriginal students	School	1–5; 0 to more than 50%	1.99	0.83
ESL/FSL students	School	1–4; equal intervals of 25%	1.17	0.56
Community size	School	1–5; <5,000 to more than 500,000	3.19	1.53
Attitudes and attributions				
Enjoys school	Student	Derived from attitude to school scale	50	10
Sense of belonging to school	Student	Derived from attitude to school scale	50	10
Enjoys reading	Student	Derived from attitude to reading scale	50	10
Good reader	Student	Derived from attitude to reading scale		

Variable	Source	Scale/Categories	Mean	SD
External attributions of success	Student	Derived from attribution scale	50	10
External attributions of failure	Student	Derived from attribution scale	50	10
Fatalism	Student	Derived from attribution scale	50	10
Internal attributions of success and failure	Student	Derived from attribution scale		
Perceived performance on PCAP test	Student	1–3; very well to not at all well Reverse scored for analysis	2.12	0.55
Perceived fairness of PCAP test	Student	1–3; very fair to not at all fair	1.80	0.61
Out-of-school activities				
Reading/writing	Student	Derived from time scale questions	50	10
Entertainment	Student	Derived from time scale questions	50	10
Academic/cultural activities	Student	Derived from time scale questions	50	10
Instructional climate				
Class size	Teacher	1–5; fewer than 15 to 30 or more	3.72	1.00
Number of grades per class	Teacher	1 to 3 or more	1.41	0.69
Number of students requiring program modification	Teacher	1 to 5 or more	2.67	0.99
Length of class periods	School	Minutes	55	16
Days absent from school	Student	1–5; 0–2 days to more than 20 days	2.50	1.21
Time on homework	Student	1–5; less than 30 minutes to more than 3 hours	3.17	1.37
Total instructional days lost	Teacher	Sum over several categories	16	11
External influence on school programs	School	Derived from influence scale	50	10
Curriculum/teacher influence on school programs	School	Derived from influence scale	50	10
External assessment influence on school programs	School	Derived from influence scale	50	10

Multilevel (Hierarchical) Modelling

In the following sections, the mathematics and science 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. Multilevel 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, where populations represent jurisdictions and language groups (e.g., Quebec English and Quebec French are two populations). Thus, the models may be two-level (students within schools) or 3-level (students within schools within populations).

These models are developed in stages as follows:

- For a two-level model, the overall variation in student achievement is partitioned into proportions attributable to differences between students and differences between schools. This allows us to determine whether differences between schools are sufficient to justify using schools as a separate level of analysis. For a 3-level model, the variation is further partitioned into student, school, and population levels. These are referred to as "null" models because no predictor variables are entered at this stage.
- Next, the relationship between achievement and each of the predictor variables is examined separately. This yields what is referred to as the "bivariate" relationship between an individual predictors and the score. Each bivariate relationship is represented by a "coefficient," which may be interpreted as representing the change in achievement associated with a one-unit change in the predictor, taken alone, with no other variables controlled.
- Other variables are then entered into the model in a sequence determined by either logical or empirical criteria. The coefficients of these models may be expressed as the change in mathematics or science 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 particular predictor in the full model with that for the same predictor in the bivariate model allows us to examine the change in predictive power of that variable once all other variables are controlled.

The order of entry of variables into the models is an issue in analyses of this type. In this case, the order chosen was based on the conceptual model presented earlier. In its simplest form, the models may be conceptualized as based on an approximate temporal sequence, in which the student is thought of as bringing to the school certain characteristics and attributes that influence learning in ways that are independent of what happens in school. Once in school, these characteristics interact in complex (and not very well understood) ways with school and teacher characteristics, programs, and learning activities to produce school-based learning.

By this logic, stable long-term characteristics of individuals and schools are entered first into the models, followed by attitudes and out-of-school activities, all of which are, presumably influenced by factors outside as well as within the school setting. Features of the school and classroom setting, specifically instructional climate, time allocation and use, and factors influencing school programs are then entered, in that order.

It must be recognized that this approach to order-of-entry falls short of a clear causal sequence. Order of entry influences the coefficients for the various intermediate models but not the bivariate coefficients or those in the full model. The choice of order can tell us something about how the predictor variables "mediate" each other, because many of these variables are themselves inter-correlated. To take a simple example, both school size and community size may influence achievement (presumably because these features are, themselves, related to factors that influence learning). However, school size and community size are themselves correlated (i.e., larger schools tend to be found in larger communities). Taken independently, these variables will have some "bivariate" correlation with achievement. When taken together, however, the coefficient for one may change that for the other, to the point where little is added to the predictive power of the model by including both.

In this report, mediating effects are noted where they can be discerned by examining the various intermediate stages of the model. However, since multiple variables are entered at each stage, it is usually not possible to determine precisely the effects of individual mediating variables. To do so would require that many detailed variations of the model be run, which would add substantial complexity to the exercise. In some cases, correlations between specific pairs of variables are referenced as a way of looking at possible joint effects.

All of this leads to considerable complexity for the interpretation of regression models. However, in our experience, the most important underlying relationships are usually not strongly dependent on the details of how the model is built. For simplicity in presentation, only the coefficients for the bivariate and full models are presented in this section. Reference is made to the intermediate models to determine the source of any major change in these coefficients. A full table of all models is given in Appendix A. More detailed descriptive statistics on the predictors, with breakdowns by jurisdiction and language, are given in the *Contextual Report* (CMEC, 2009).

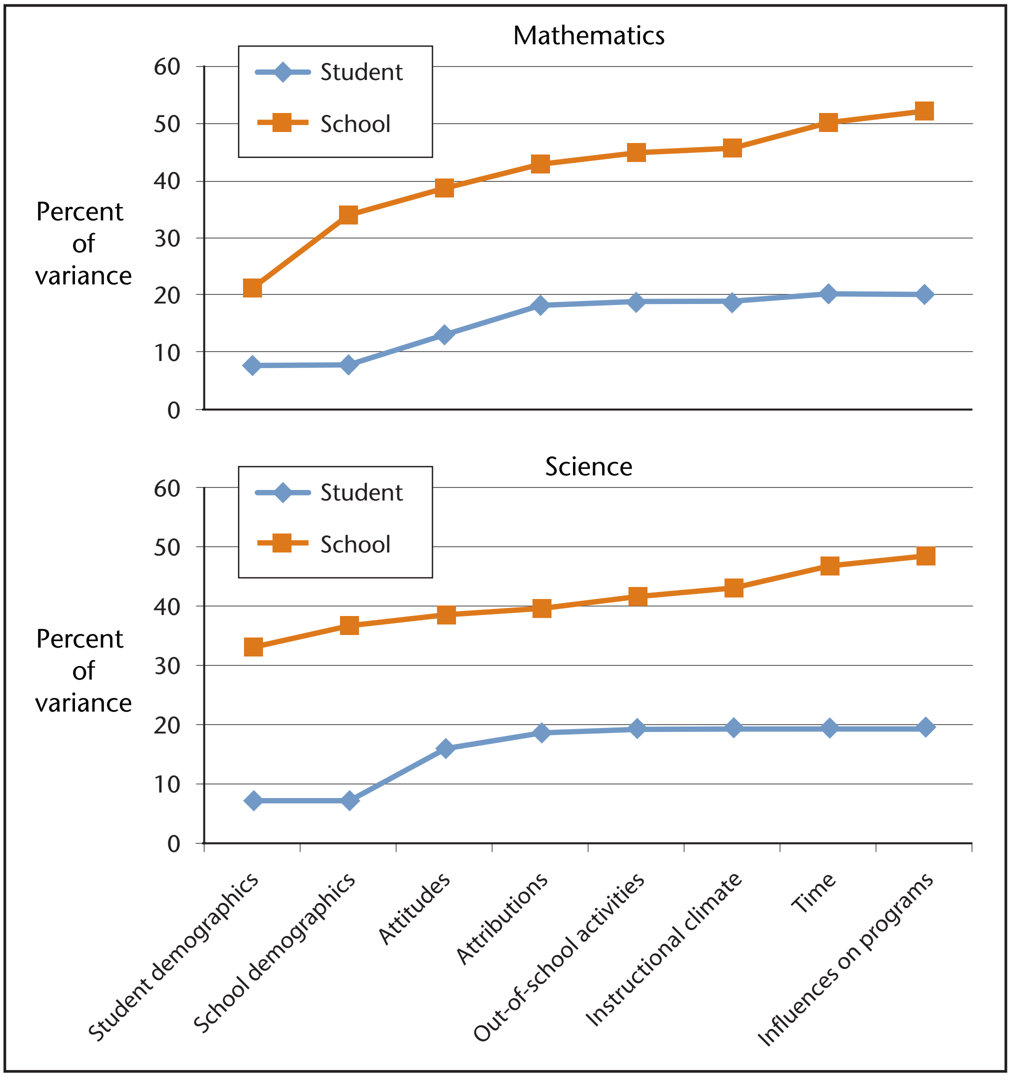
Proportions of Variance

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. If very little variation can be attributed to schools (that is, if the schools are much the same in their ability to produce the desired outcomes), the rest of the analysis can be done at the student level only. 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 81% of the variance in mathematics and 87% of the variance in science scores, leaving 19% and 13% to be accounted for by differences between schools. Taking the total student and school variance as the starting point, Chart 6.1 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 figures shown in the chart are "proportions of proportions." For example, for mathematics, entering the student demographic variables accounts for 8% of the original 81% that is student variance (that is, 8% of the original 81%) and 21% of the school variance (21% of the original 19%). Because this may seem to be somewhat complex, the important point to note is that the proportions of variance accounted for by the model increase as more variables are added and that, after entering all 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.

The detailed results are somewhat different for mathematics and science. For science, relatively more of the school variance is accounted for by student characteristics than is the case for mathematics. The proportions become similar once school characteristics are added. This indicates that school characteristics are more important for mathematics than for science. For the student variance, the data on attitudes and attributions yield relatively larger increases in the proportions than other variables. For science, the change from attitudes to attributions is non-linear, indicating that attitudes contribute more to the change than attributions. Beyond these stages, the proportion of student variance does not change much, while the proportions of school variance (explained by the added factors) increased at a fairly constant rate. This is a function of the fact that more school level variables are added to the models at later stages.

CHART 6.1 **Proportions of variance accounted for by model stages**



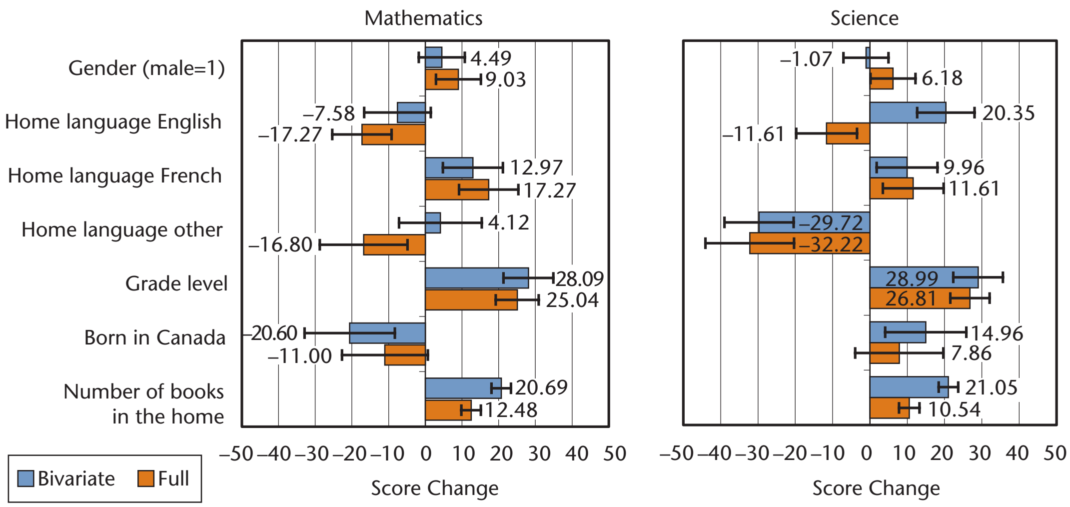
Student Characteristics

The specific variables used at this stage are student gender, grade level in school, main language spoken at home, whether or not born in Canada, and number of books in the home. The latter is considered a proxy for socioeconomic status and was used in preference to the other available SES indicator, mother's education, because these two variables were highly correlated and because there was a much larger amount of missing data for mother's education.

Chart 6.2 gives the bivariate and full model coefficients for student characteristics. For dichotomous variables, such as gender and home language, the coefficients represent the difference in mathematics scores between the two groups. For other variables, such as books in the home, the coefficient should be interpreted as the change in mathematics score for one unit change in the predictor variable, on the scale given for that variable in Table A6.1. The line at the end of each bar represents the 95% confidence interval for the coefficient. A coefficient is statistically significant if the error bar does not cross the zero point on the horizontal axis.

Looking first at the variables that yield most consistent effects across the two models, speaking French at home, grade level, and number of books in the home are all positively related to achievement in both subjects. Speaking French at home is essentially a Quebec effect, as Quebec francophone students are both high-performing and most likely to speak French at home, not to mention carrying the greatest weight among the francophone populations. In the full model, the results for those speaking English at home are essentially the mirror image of those for French, because these two languages make up by far the largest of the home language groups.[11]

Chart 6.2 **Student characteristics**[12]



The results for grade level are not surprising. Being in a higher grade makes a substantial contribution to performance, even when other variables are controlled. Under the Carroll time model, this may be interpreted as an "exposure to schooling" effect. Taken a step further, this may be seen as an argument for earlier school entry. At the same time, it is cautioned that the grade a student attains by age 13 may be related to student ability as well as to student's exposure to schooling. For example, in schools that practise grade retention, students may be in a lower grade because they have not met expected grade-level standards at some time in their school careers. Nevertheless, to the extent that this is true, the results indicate that grade retention does not enable students to catch up in achievement, a result consistent with the literature on grade retention (Shepard & Smith, 1990; Jimerson, 2001).

The results for books in the home are consistent with most other studies of socioeconomic status in that they show a positive relationship with achievement in both models. However, it is important to note that this effect is significantly smaller in the full model than in the bivariate model. The modelling sequence (see Table A6.1) indicates that this variable is mediated mainly by attitudes. More specifically, there is a reasonably

[11] English and French were entered separately into the full model, in the presence of all other variables because the two together accounted for most of the students, creating a redundancy that prevents the analysis from proceeding.

[12] It is important to note that the size of the coefficient for each variable is dependent on the scale on which the variable is measured, as well as on the size of the effect. Coefficients for the different variables are thus not directly comparable unless the variables are on the same scale. For example, in Chart 6.2, the coefficients for gender, language, and immigration status are comparable because all are measured on a dichotomous (0,1) scale. The coefficients for grade level or books in the home are comparable with each other because both of these latter variables are measured on a 5-point scale. However, they are not comparable to the coefficients for the dichotomous variables. Coefficients can be made directly comparable by using a standardized scale, but the information on the change in achievement per unit change in the scale is then lost.

high correlation (0.33) between books in the home and enjoyment of reading, suggesting that these two variables exert a joint effect on mathematics achievement.

The gender effect favours males, but is statistically significant only in the full model, for both subjects. The gender effect is most strongly mediated by the other student characteristics entered at this stage of the model (Table A6.1). More specifically, the gender effect is mediated by language, with the gender difference being larger for francophone than for anglophone students.

The relatively small number of students (16%) who speak a language other than English or French at home perform more poorly than those speaking either of the official languages in the full model (but not in the bivariate model for mathematics). Again, this is mediated for mathematics by other student characteristics. In particular, in mathematics there is a complex relationship between grade level and language, with proportionally more of those speaking other languages being in both grade 6 and grade 10 than those speaking the official languages. At both grade-level extremes, those speaking other languages do more poorly than those speaking the official languages, while the difference for grade 8/secondary II students is relatively small.

Finally, those born in Canada perform less well in mathematics but better in science than immigrant students. This difference is statistically significant in the bivariate model but not in the full model. This effect seems to be mediated mainly by school demographics and attitudes (Table A6.1).

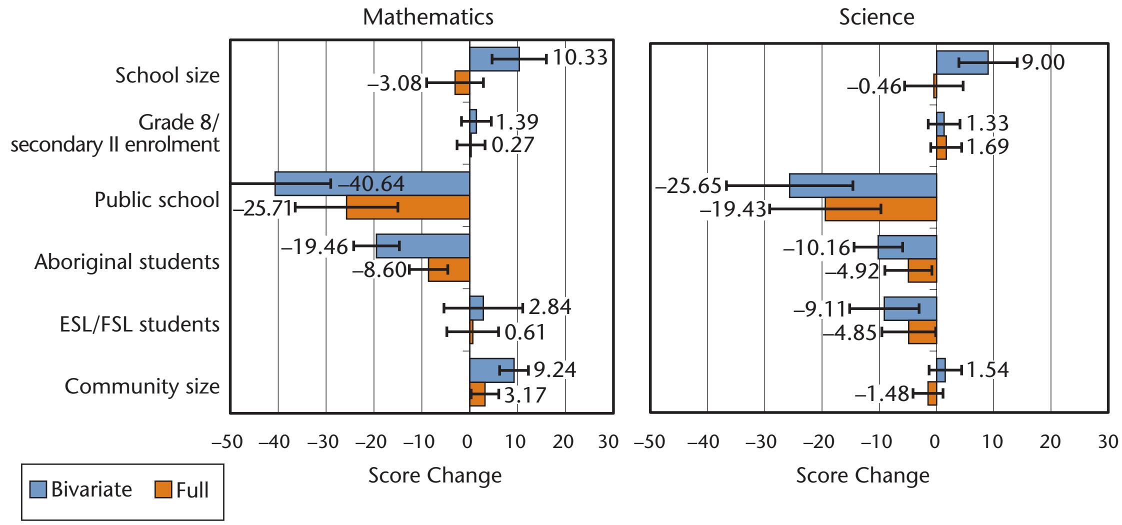
School Characteristics

Chart 6.3 shows the effects for school characteristic variables. Students in larger schools perform at a higher level than those in smaller schools. However, the effect becomes non-significant when other variables are controlled. Larger community size is associated with higher scores in mathematics but not with higher scores in science. School size is strongly correlated with community size, so this is a likely mediating factor.

Students in schools with higher percentages of Aboriginal students perform at lower levels in both subjects. Though statistically significant in both models, this effect is reduced by the effects of other variables in the full model. The largest change occurs when other student and school characteristics are entered into the model, suggesting an influence (possibly socioeconomic, language, or school or community size) on the magnitude of the Aboriginal student effect.

The model reveals that students in public schools do less well than those in private schools for both subjects. Although other variables in the full model exert some mediating effect, this does not significantly change the result. While the higher performance of private school students is sometimes attributed to selection effects, particularly on the basis of socioeconomic status, the important point here is that any such effects included in these models are insufficient to account for the private school advantage in mathematics achievement. It is quite possible that other selection factors not included in the model, such as prior academic performance or motivation, are at play in decisions to attend private schools and in student performance once in such schools. PCAP did not pursue these matters in sufficient detail to pursue this issue further.

Chart 6.3 **School characteristics**



Attitudes and Attributions

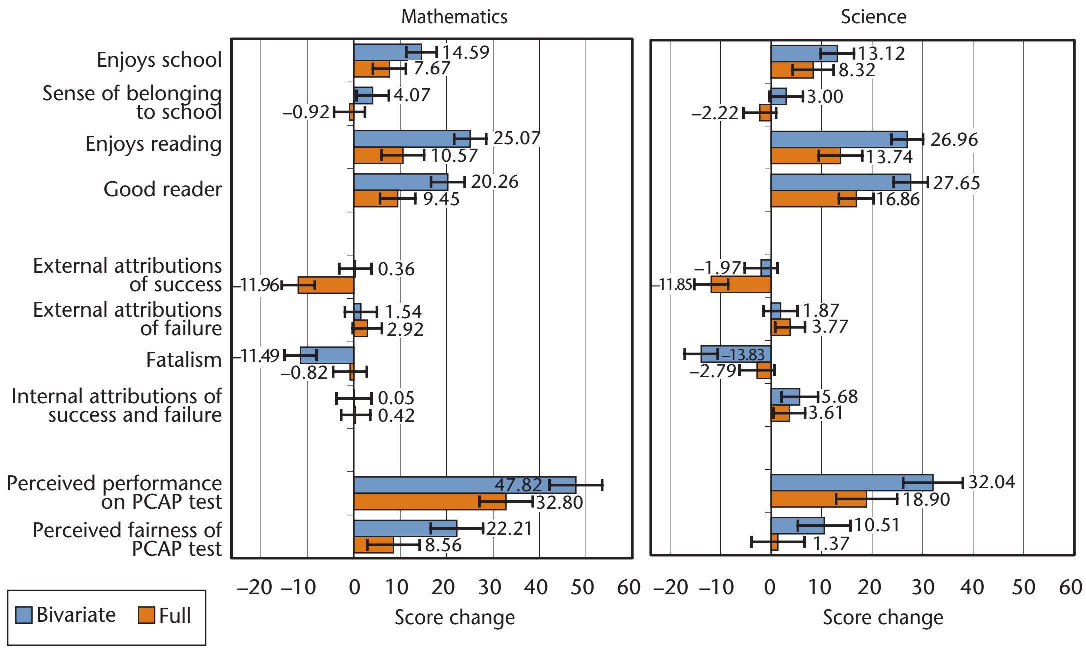
Student attitudes were measured using a series of Likert scale items on two scales, representing attitudes to school and attitudes to reading. These scales were resolved into derived variables as indicated in Chart 6.4. Likert scale items were also used to measure student attributions of success and failure to internal or external sources. This scale yielded four derived variables, also given in Chart 6.3. Finally students were asked two questions on their perceived performance on the PCAP test and their sense of the fairness of that test.

Because the attitude and attribution derived variables take the form of standard scores, the results are presented in Chart 6.4 as the change in mathematics score associated with a one standard deviation change in the derived variable. Most of the attitude results are positive for both subjects. Effects for the full model are generally smaller than those for the bivariate model, reflecting the influence of mediating variables, especially the inter-correlations among the attitude variables themselves.

Results for the attribution variables are generally non-significant. The exceptions are external attributions of success, which is significantly negative in the full model and fatalism, which is significantly negative in the bivariate model only for both subjects. The results here are different from those in reading (CMEC, 2009) in that the reading results tended to show negative effects for external attributions and positive effects for internal attributions, as predicted by attribution theory.

The results for attitudes toward the PCAP assessment are in the expected direction. Perceptions of expected performance are generally greater than those for perceived fairness of the test. The results suggest that students are reasonably good judges of their performance. However, it is also possible that students are making these judgments based on a perception of their overall school performance. In the absence of a measure of school performance, this possibility cannot be tested.

Chart 6.4 **Attitudes and attributions**

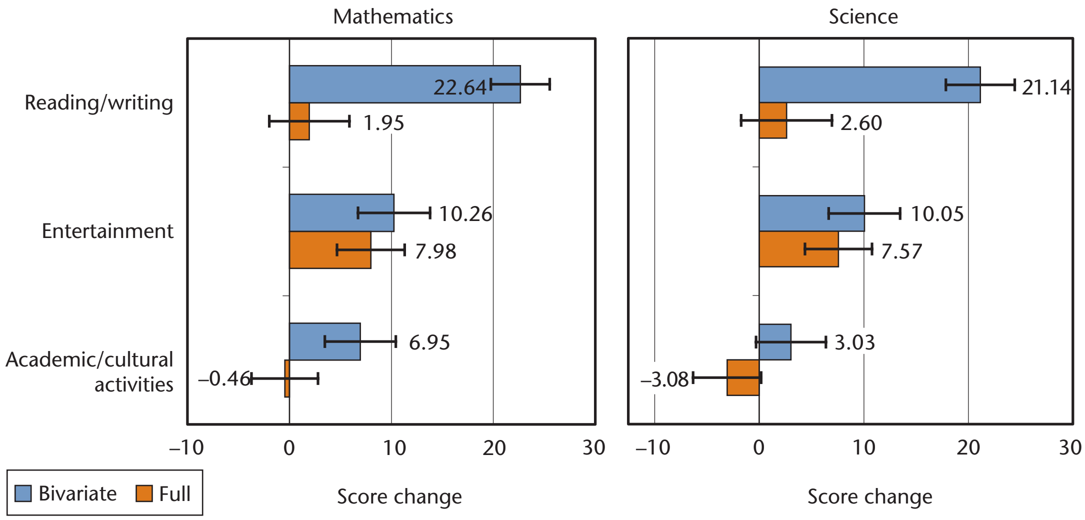


Out-of-School Activities

A series of questionnaire items about time spent on various activities outside of school yielded three derived variables labelled “reading and writing,” “entertainment,” and “academic/cultural” activities. Results for these variables are given in Chart 6.5. Again, the results should be interpreted as the change in mathematics score for one standard deviation change in the predictor variable.

The general pattern here is similar for the two subjects. All three variables show positive bivariate effects for mathematics although the effect of academic/cultural activities is non-significant for science. For the full model, only the entertainment factor remains significant. Both reading/writing and academic/cultural variables are mediated mainly by attitudes (Table A6.1). However, this is not the case for entertainment.

CHART 6.5 **Out-of-school activities**



Instructional Climate

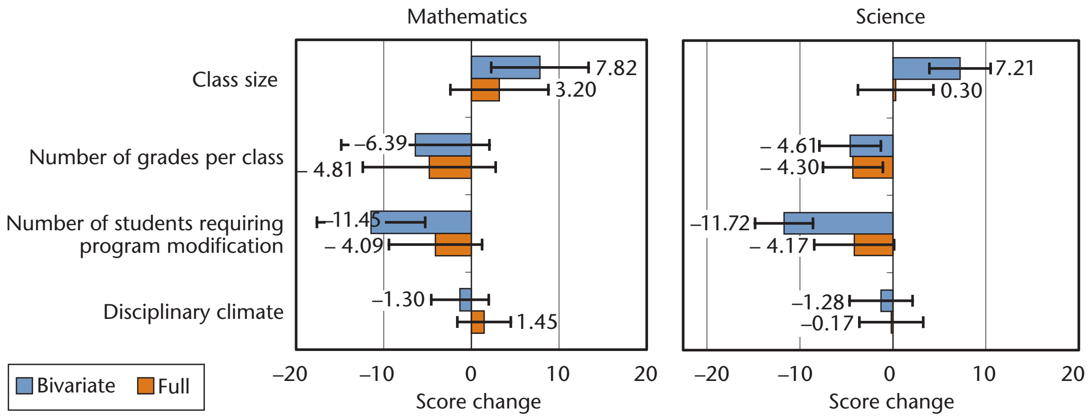
As already indicated, only a limited number of instructional climate variables are available for mathematics and science because most questions in this area were specifically related to reading. Indeed, for the variables used here, inferences are made from questions about language arts classes to classes in other subjects in the same school. For example, it is assumed that principal reports of average class size in language arts also apply to mathematics and science.

Results for the four instructional climate variables are given in Chart 6.6. Class size shows a significantly positive effect (higher scores for larger classes) in the bivariate model but no effect in the full model. Number of grades per class is not significant for mathematics but shows a negative effect for science. It is noted that no distinction is made in the teacher questionnaire between multi-grading as a deliberate structural feature of the school and multi-grading as a response to low or unbalanced grade-level enrolment. However, number of grades per class is negatively correlated with enrolment, suggesting that smaller schools are the ones more likely to have more than one grade per class.

The number of students requiring program modification is a measure of the level of individualization of programs. (This was measured for language arts classes but inferred to be a characteristic of the school.) This shows significant negative effects in the bivariate model, which become marginal negative effects in the full model.

Finally, disciplinary climate shows no effects. This is in contrast with the results of other studies in which poor disciplinary climate is a negative contributor to achievement.

Chart 6.6 **Instructional climate**

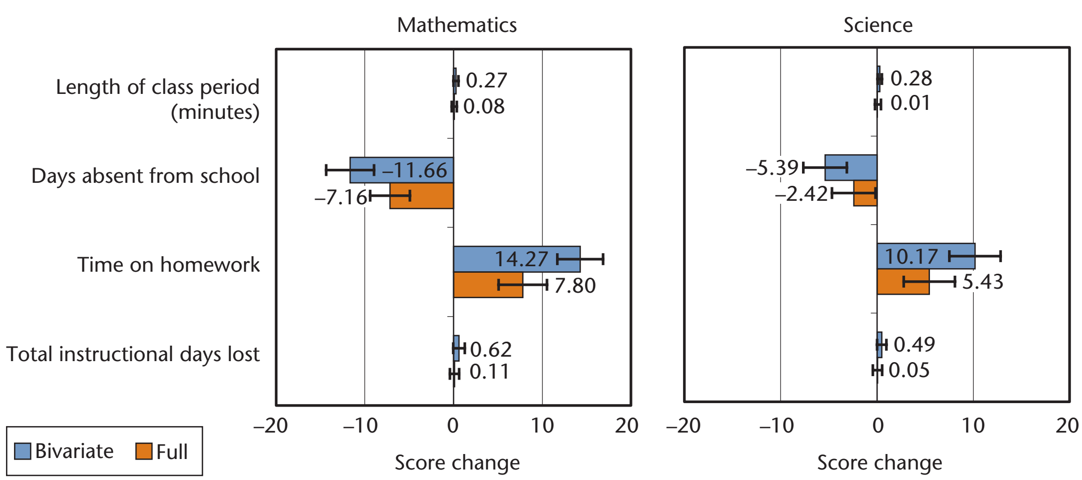


Time

Again, only a small number of time variables are available. In this case, the variables are either generic to the school (the length of class periods and the total number of instructional days lost would be expected to be a school-wide phenomenon) or to the student (days absent, total homework time) and should not be subject-specific.

Chart 6.7 gives the results for four time variables. The pattern is much the same for mathematics and science. The results for days absent from school and for total homework time are in the expected direction, negative for days absent and positive for homework time. The effects of the length of class period and instructional days lost are not statistically significant.

Chart 6.7 **Time**

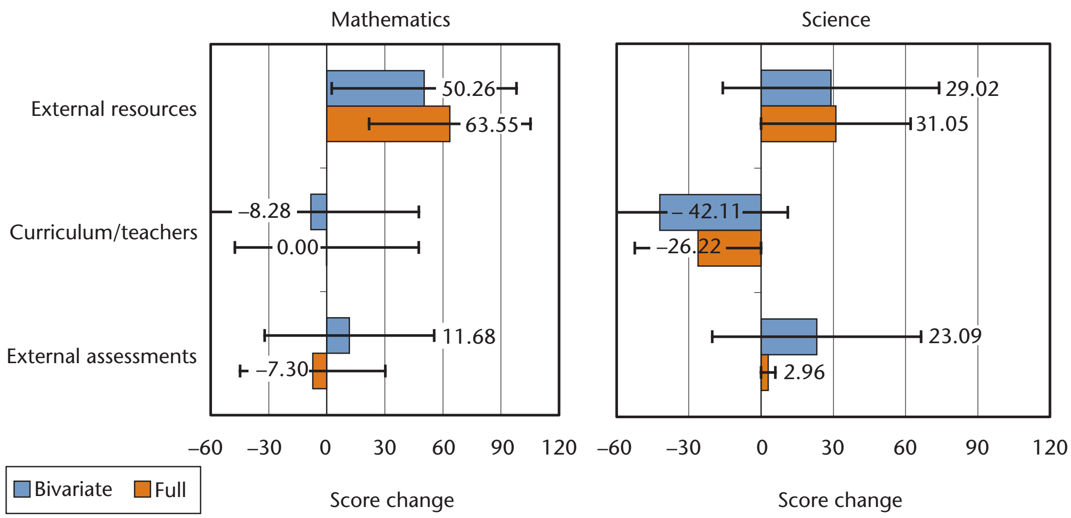


Sources of Influence on School Programs

The school questionnaire included a series of questions on factors that might influence decisions on school programs. These were conceptualized on an internal (teachers, students) to external (available resources, external assessment, publishers) continuum. The scale was a four-point one from "no influence" to a "lot of influence." This question set was resolved into three derived variables labelled "external resource influence," "curriculum/teacher influence," and "external assessment influence."

The results appear in Chart 6.8. The only factor showing a significant effect is external resources for mathematics, with a marginal effect for science. This variable derives from questions about textbooks and textbook publishers, availability of instructional resources and teacher groups external to the school. These results are not in accord with the proximity model in that they suggest that influences external to the school are more important than influences within the school. However, the external resources effect is fairly marginal relative to many of the other effects in these models and should be treated with caution in the absence of further evidence on this issue.

CHART 6.8 **Influences on school programs**



7 RESULTS BY JURISDICTION AND LANGUAGE

Although differences in achievement across jurisdictions are relatively small compared to the differences across students and schools, a difference of more than half a standard deviation is usually found between the highest and lowest performing jurisdictions in the SAIP assessments and in PCAP 2007 (see Chart 4.1). These differences have also been relatively persistent, with the same jurisdictions at the higher and lower ends of the scale in most cases.

Large-scale assessments almost always have a comparative element. To the extent that the results of these assessments attract public attention, it is usually this element that is highlighted. Up to now, however, little effort has been made to account for differences among jurisdictions in terms of the predictor variables. This issue is more explicitly identified for this study by the following research question:

> *"How do provincial/territorial characteristics in practices, policies, and procedures such as school size, governance model (public/private), class size, large-scale assessment procedures, and diversity of student population influence mathematics and science performance?"*

A recent study (Ma & Crocker, 2007) attempted to examine this issue using PISA data, but the analytical approach used was difficult to implement — not to mention for readers to comprehend. In many PISA reports, this issue is captured to some extent by conducting separate analyses for each country and comparing the results for a given variable across countries and, less explicitly, for a given country across all the variables in the model. A similar approach was taken in the *PCAP-13 2007 report on Differences in Reading Performance of 13-Year-Olds Based on Language and Minority/Majority Status*. This approach is complex because it requires that the models already described be run separately for each jurisdiction and language group, an approach that allows only for qualitative comparisons across the groups.

A research project specifically designed to address the issue of jurisdictional differences is planned as part of the PCAP research series. For that reason, the results reported here should be treated as preliminary, particularly since reading results are not examined. The intent here is to explore some possible analytical techniques, present a selected set of results, and compute a full set of jurisdictional results to be included in the Appendices for reference by those interested in more detail.

Three Analytical Approaches

Three alternative analytical approaches are explored here. First, jurisdictional codes are added to the two-level hierarchical models. This allows the effects of jurisdiction to be separated from those of other variables and the effects on jurisdictional differences of including other variables to be assessed. By extension, this also addresses the question of whether jurisdictional characteristics, other than those measured by the questionnaires and included in the model, have an influence on achievement. While this approach cannot explicitly identify these characteristics, it can help examine the question of whether other explanations should be sought outside the realm of the measured factors.

The second approach is to use a 3-level hierarchical model with population (jurisdiction/language) as the third level. Groupings were used that allow analysis for the two language groups in jurisdictions where the sample size is reasonably large. This gives a measure of the proportions of variance accounted for at each level, but is limited in its ability to model the effects because of the small number of units available at the jurisdictional level.

Finally, separate analyses by province and language are conducted for selected groups, accompanied by exploratory analysis of some specific predictor variables. This allows us to examine a variety of hypotheses that can shed some light on differences in jurisdictional characteristics that may be related to differences in achievement.

Two-Level Analysis with Jurisdiction/Language as Variables

Under this approach, each jurisdiction and the two official language groups within jurisdictions with reasonably large minority populations (Manitoba French, Ontario French, Quebec English, and New Brunswick French) was dummy-coded (1 if a member of the group, 0 if not) into the data file. These codes could then be used as variables in the model, with the coefficient for each representing the contribution of that jurisdiction/language group to the prediction of achievement. Changes to these coefficients when other predictor variables are added to the model give some information on the effects of these predictors on the jurisdictional/language groups.

The method used to build the models is essentially the same as in the previous sections. The main difference is that the initial model includes all the provincial/language groups, with one “reference group” (Ontario English) omitted to avoid the redundancy inherent in including a complete set of dummy variables in a model. The main interest here is in how much the initial coefficients for jurisdictions shift as other variables are added to the model. As before, these variables are added in stages, thereby yielding several intermediate models and, ultimately, a new full model that includes all the variables in the previous models plus the dummy-coded variables representing jurisdiction and language.

Chart 7.1 gives the proportions of variance accounted for by the model with jurisdiction/language groups only and the full model. In each case, the blue bars represent the proportion of student or school variance accounted for by the jurisdiction/language groups, with no other variables controlled. Thus, adding the jurisdiction/language groups to the model accounts for none of the student variance for either subject. However, these groups account for 17% of the original 19% school variance for mathematics and 20% of the original 13% school variance for science (see Chart 6.1).

The orange bars represent the proportions of student and school variance accounted for by a new full model, with jurisdiction/language added to all the variables already in the model. The full model accounts for close to 20% of student variance (20% of 81% for mathematics and 20% of 87% for science) in each case, and for 55% (of 19%) and 65% (of 13%) of the school variance for mathematics and science respectively. This is slightly more of the school variance, but not of the student variance, than is accounted for by the full models in the absence of the jurisdiction/language groups (see Chart 6.1). The implication is that, although differences among jurisdictions seem relatively large when examined comparatively, these differences are quite small relative to differences among students and schools.

Chart 7.1 **Proportions of variance for jurisdiction/language groups**

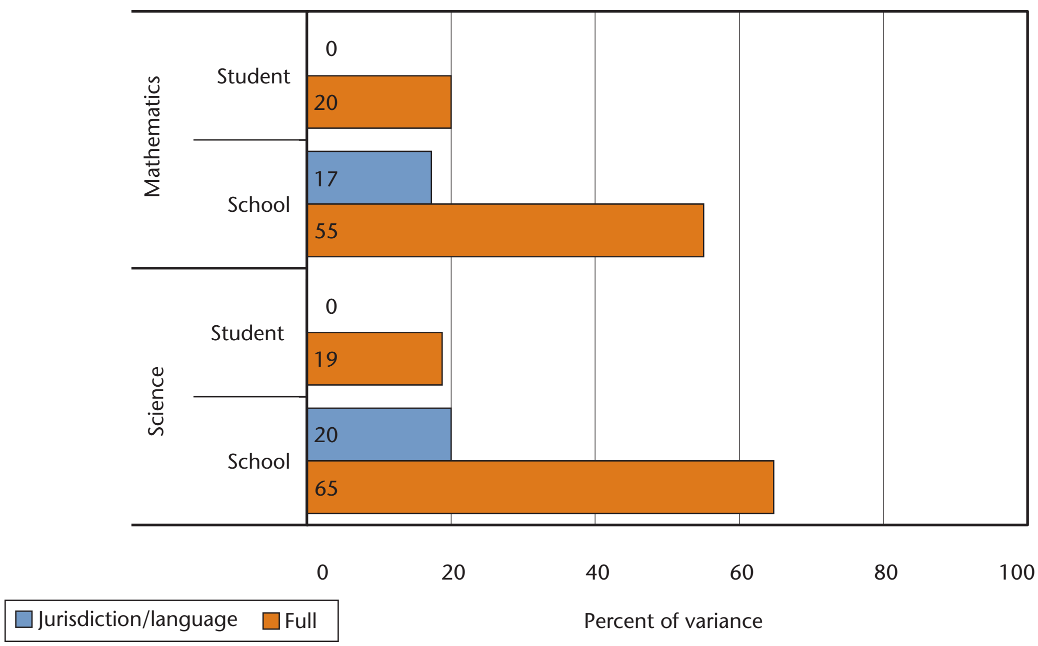
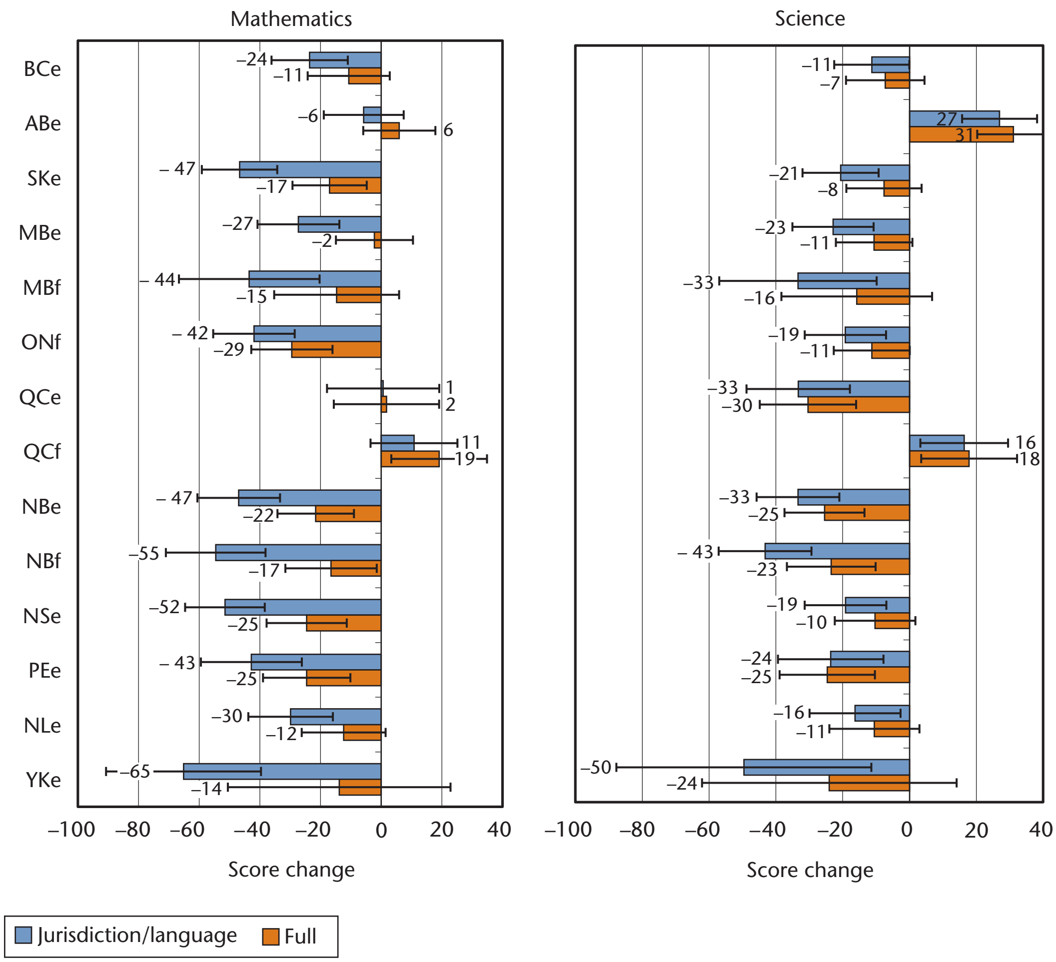


Chart 7.2 gives the coefficients for the initial and full models. The coefficients in the initial model represent the difference between the mean score for the jurisdiction and that for the reference group, Ontario English (508 for mathematics and 499 for science). These are essentially the same coefficients found by running a separate (bivariate) model for each jurisdiction except that, in the latter case, the reference point would be the grand mean (500) rather than the Ontario English mean.

For mathematics, the coefficients are smaller for the full model relative to the initial model in all cases. The shift is statistically significant for Saskatchewan English, New Brunswick French, and Nova Scotia English and marginally so for Manitoba English and New Brunswick English. The next question, of course, is which of the predictor variables contribute most to this shift. This is difficult to determine because the general pattern is a gradual shift as variables are added. However, an examination of the intermediate models (Table A7.1) indicates the following:

- For Saskatchewan English, New Brunswick English, and Nova Scotia English, student characteristics and school characteristics contribute about equally to the change.
- For New Brunswick French, almost all of the change comes from student characteristics
- For Manitoba English, most of the change comes from school characteristics.

CHART 7.2 **Regression coefficients for jurisdiction/language groups**



Although the overall change for Quebec French is not statistically significant, that group shows a different pattern of change from any of the others: the addition of student characteristic variables increases the positive coefficient for that group, suggesting that the observed difference between Quebec French and others is suppressed, relative to its true value, by student characteristics. This is consistent with the fact that Quebec francophone students do very well despite having lower average socioeconomic status than students in most other jurisdictions. Because the overall shift is not statistically significant, it is not possible to make a definitive statement about this point. However, this does warrant further research, because it suggests that Quebec francophone schools may be doing more to reduce the socioeconomic gap than is the case elsewhere.

The results in science are different from those in mathematics — the shift in coefficients for most groups is quite small and none is statistically significant.

3-Level Models

The core 3-level model treats schools as "nested" within the jurisdiction/language-group schools as in the 2-level model. This allows us to compute the proportion of variance accounted for by these groups, relative to student and school variance, and the contributions of variables aggregated to that level.

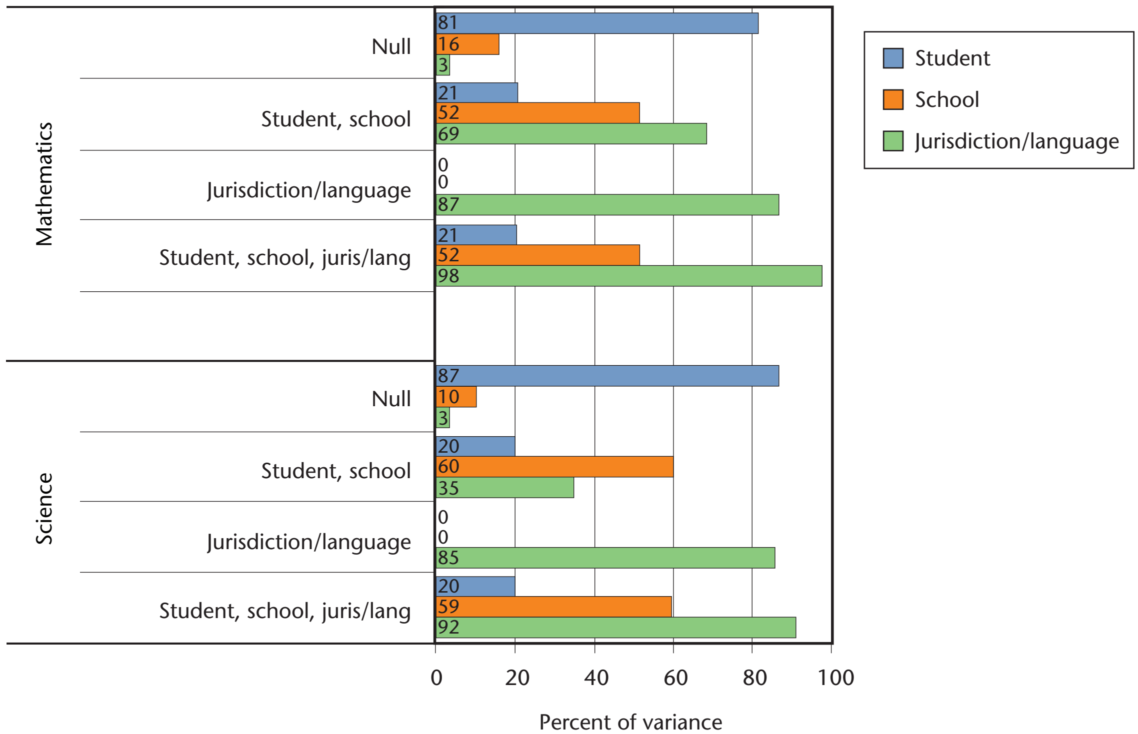
Chart 7.3 gives the proportions of variance accounted for by the null model and models with student and school variables, jurisdiction/language variables, and student, school, and jurisdiction/language variables. This chart should be interpreted as follows:

- For each of mathematics and science, the null model represents the proportions of variance accounted for by differences between students, schools, and jurisdiction/language groups respectively.
- The student/school model represents the proportions of each of the above proportions (the 81%, 16%, and 3% respectively) accounted for when all the student and school variables are added to the model.
- The student, school, jurisdiction/language model represents the proportions of each of the original variances accounted for when all the student, school, and jurisdiction/language variables are added to the model.

This shows, first, that the jurisdiction/language groups account for only about 3 percent of the total variance in the model. Adding the student and school level variables accounts for 69 percent of that variance (69% of 3%) for mathematics and 35 percent for science. Adding a selection of these same variables aggregated to the jurisdiction/language level (level 3 of the model) accounts for 87 percent and 86 percent of the variance for mathematics and science respectively. This changes only slightly when the student- and school-level variables are added to the model, along with the jurisdiction/language-level variables.

All of this seems to indicate that characteristics of the jurisdiction/language groups can account for a large proportion of the difference among these groups, either taken alone or after controlling for student and school characteristics. However, a closer look at the model coefficients revealed that the small number of Level 3 units (15 jurisdiction/language groups) yields large standard errors at the group level. Thus, none of the jurisdictional effects are statistically significant. The general conclusion, therefore, is that the 3-level model is not an appropriate one for examining jurisdictional differences. Nevertheless, this model does show that jurisdictional/language differences are actually quite small compared to differences between students and schools. While that does not diminish their interest for policy-makers, the model illustrates how difficult it is to examine the sources of these differences.

Chart 7.3 **Proportions of variance for 3-level model**



Total Variance and Variability across Schools: The Equity Issue

An argument is commonly made that jurisdictions should strive not only for high average achievement but also for greater equity in achievement. In particular, Willms (2003) and others have argued that a desired goal of schooling is a reduction in the degree of inequality across socioeconomic groups.

Following the equity argument, the goal is to ensure that the difference in achievement between the highest and lowest performing students and, particularly, between those of higher and lower socioeconomic status 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.

Equity may be examined by looking at the total variance for each of the jurisdiction/language groups. On average for Canada, the total variance is approximately 10,000, which 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 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 7.4 shows this variance. The results show the Quebec English population to be substantially more variable than other populations in mathematics. Beyond this, there is a relatively constant gradation, with Saskatchewan English and Prince Edward

Island English being the two most homogeneous groups in mathematics. The picture is somewhat different in science, with Newfoundland and Labrador having the greatest variability and Ontario English the least. Nevertheless, the two sets of variances are positively correlated (r=0.37).

CHART 7.4 **Total score variance by jurisdiction/language group**

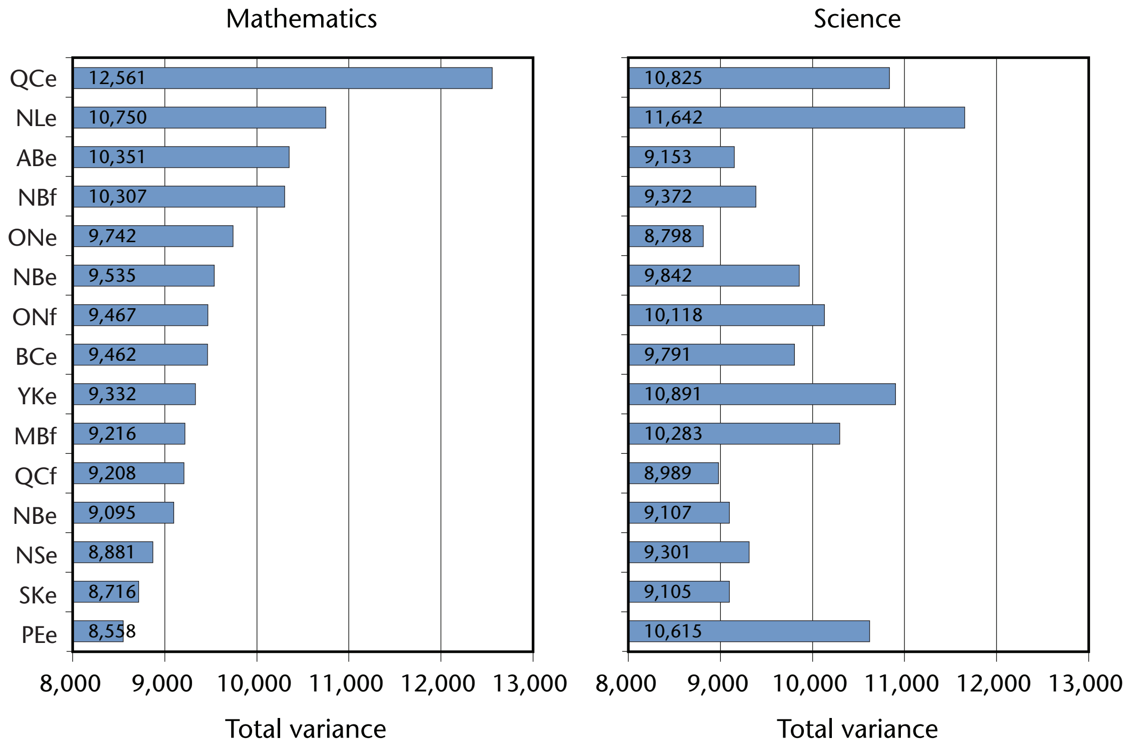
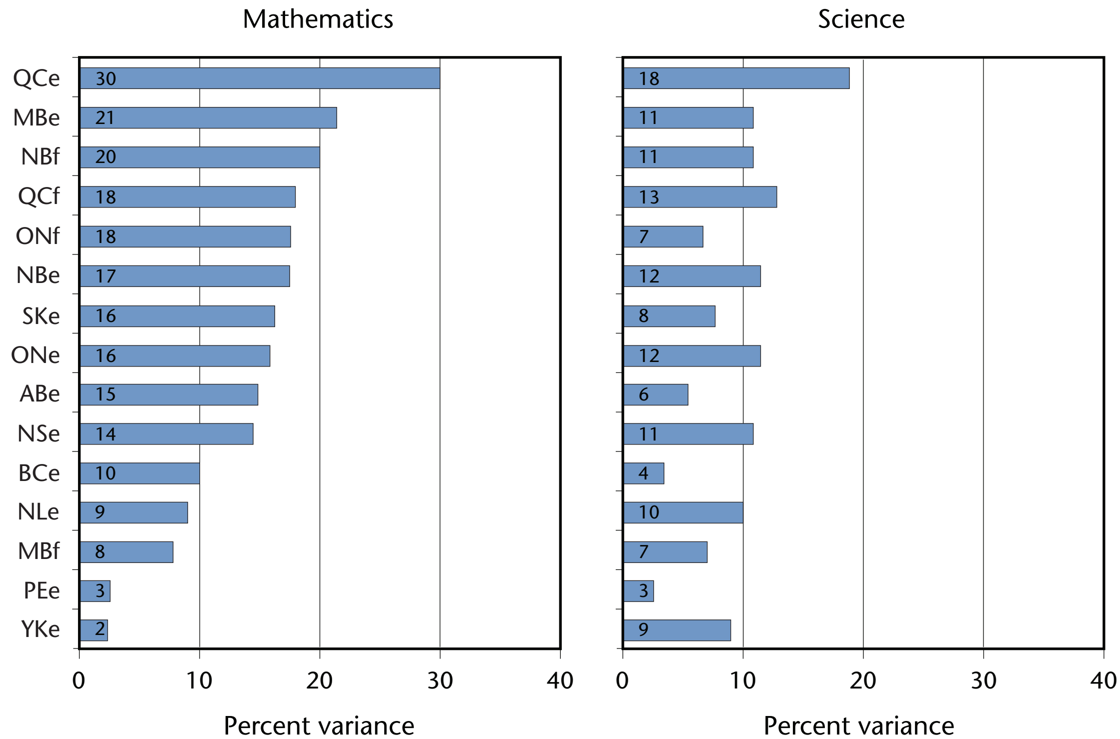


Chart 7.5 looks at the equality issue at the school level, showing the proportion of total variance that exists between schools. Again, Quebec English stands out as having the greatest variability, this time for both subjects. At the opposite extreme are Prince Edward Island English and Yukon English. In this case, although the science picture looks somewhat different from that for mathematics, the two sets of variances are highly correlated (r=0.71). Overall, however, there is less variability across schools for science than for mathematics.

Chart 7.5 **Percent of variance at the school level by jurisdiction/language group**



Exploratory Analysis of Selected Jurisdiction/Language Groups

In this section, direct comparisons are given for four jurisdiction/language groups selected for their diversity in means and variances on their mathematics scores. This analysis is intended to be illustrative rather than comprehensive. Table B1 gives a more detailed breakdown of bivariate relationships between performance and a selection of predictor variables.

The main interest here is in explicitly examining the differences across these groups. Predictor variables were therefore selected on the basis of their variability across groups, using the comparative statistics in the Contextual Report (2009), their predictive value as given in the full models in the previous section, and more qualitatively on their judged policy value as a means of helping understand the differences among the groups. The latter criterion allowed a few variables that showed low predictive power overall to be tested for their predictive value within jurisdiction/language groups.

The selected groups are identified below:

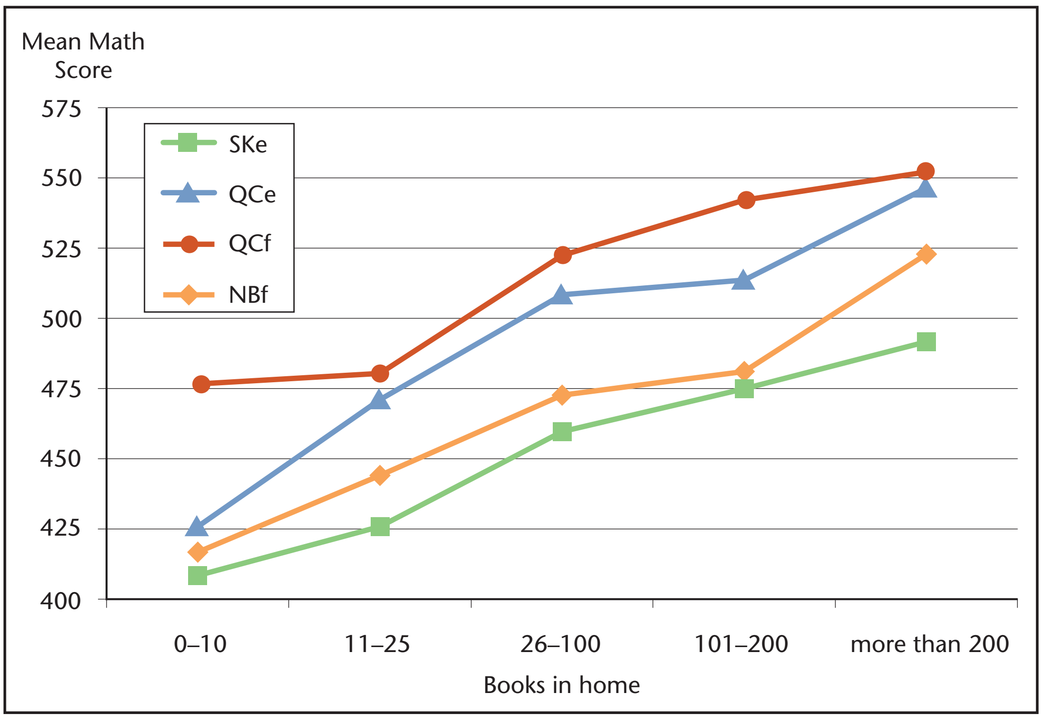
	Mean math score	Total variance
Quebec English	510	12,561
Quebec French	518	9,208
New Brunswick French	460	10,307
Saskatchewan English	461	8,716

For each of these jurisdictions, a number of variables were selected from the set used in previous analyses, based on policy relevance and differences across jurisdictions. In one instance (that for public versus private schools), the two last groups above were replaced by British Columbia English and Manitoba English, because these are the only groups outside of Quebec with sufficient private schools to permit analysis.

Socioeconomic Gradient

The four selected jurisdictions are quite similar in socioeconomic status as measured by the number of books in the home. A relatively simple picture of the socioeconomic gradient can be given by plotting the mathematics scores for each SES level, as measured by the variable "books in the home." This plot is shown for the selected groups in Chart 7.6. A two-way factorial analysis–of-variance (ANOVA)(without controlling for other variables) shows both the group effect and the SES effect to be statistically significant, but the interaction effect is not significant. This indicates that the general SES pattern is similar for all four groups. Thus, although the chart suggests that the SES trend for Quebec English, in particular, is steeper than for the other groups, this is not sufficient to mediate the general pattern of higher scores for higher SES groups.[13]

CHART 7.6 **Mean mathematics scores by books in home: Selected jurisdiction/language groups**



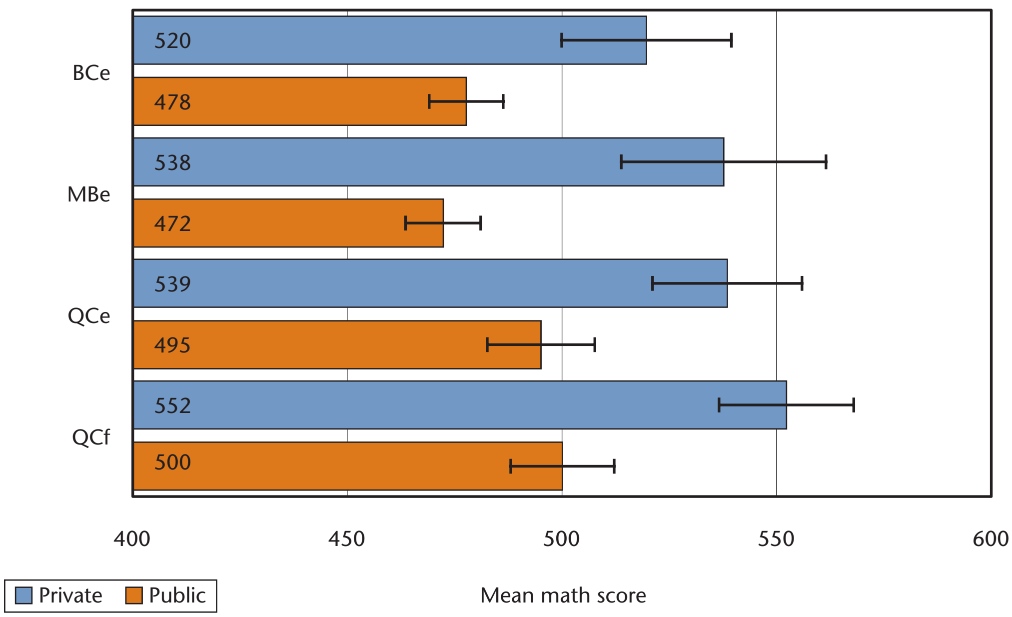
Public and Private Schools

One feature generally thought to be related to SES is public versus private schools. As it happens, the proportion of private schools in the PCAP sample is higher in Quebec than in other jurisdictions, with 37% of Quebec French and 34% or Quebec English students attending private schools. The closest other groups to this are British Columbia English with 15% and Manitoba English with 12% of students attending private schools. No other jurisdictions have sufficient private schools to permit analysis. For that reason, we compare public and private mean scores for these four jurisdictions.

[13] It is important to note that this analysis is at the student level only and does not take account of differences between schools as in other models.

These results are given in Chart 7.7. This shows a substantial private school advantage for all these jurisdictions, as was shown in the earlier models. This, of course, is consistent with the more general private/public school effect shown in the earlier models. Indeed, given the very small number of private schools in other jurisdictions, it can be said that the overall effect is almost entirely a function of the effects for these four groups. The earlier models indicate that the private school advantage is mediated somewhat by other variables, but remains significant, even in the full models. This point is examined more closely for selected jurisdictions in the analysis that follows.

Chart 7.7 **Mean mathematics score by public/private schools: Selected jurisdiction/language groups**



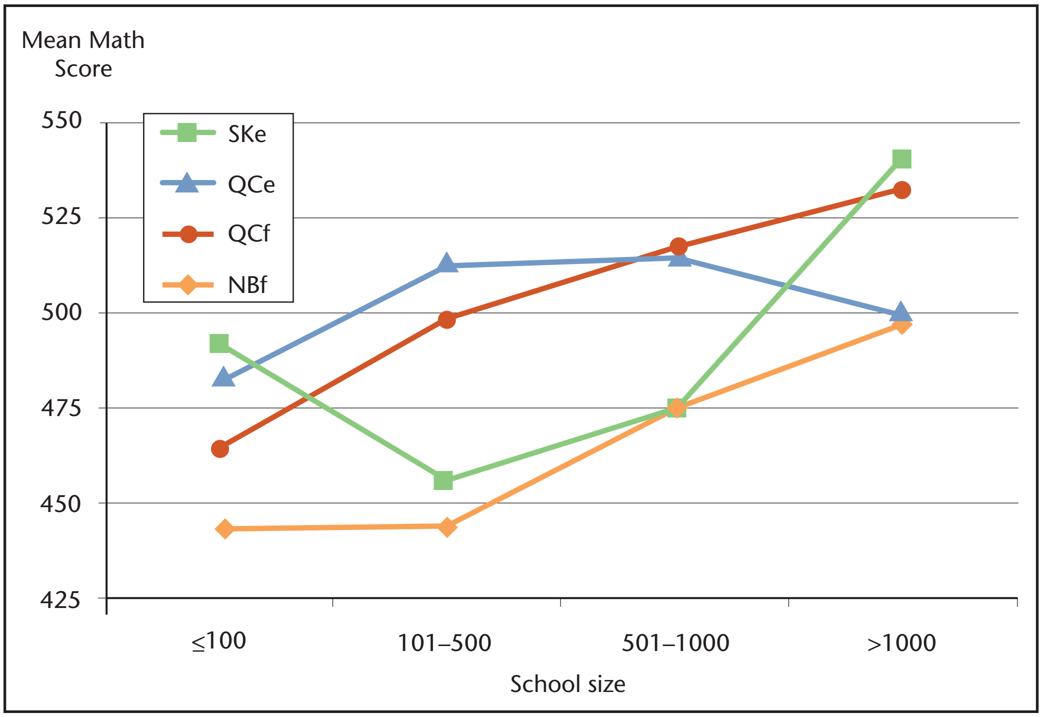
School Size

School sizes differ substantially across the four selected jurisdiction/language groups. The median school size for Saskatchewan English and New Brunswick French is in the 101–500 range. By contrast, Quebec is a province of relatively large schools, with the median for Quebec French being 1,000 or more and that for Quebec English 501–1,000.

Chart 7.8 shows the relationship between mean mathematics score and school size. In this case, the interaction effect is statistically significant, and the trend is clearly non-linear for all but Quebec French, where achievement tends to increase fairly constantly with the school size range. Given that Quebec French has the highest proportion of students in the largest schools, this indicates that being in larger schools is a substantial contributor to high scores for Quebec French students. For Saskatchewan English, those in the median size schools (101–500) perform at the lowest level. For New Brunswick French, those in the smallest two school-size ranges have the lowest scores. The trend for

Quebec English is quite different from the others, with those in the middle two school-size ranges performing at the highest levels. More generally, the greatest variability in performance is found in schools with enrolments in the 101–500 range.

CHART 7.8 **Mean mathematics score by school size: Selected jurisdiction/language groups**



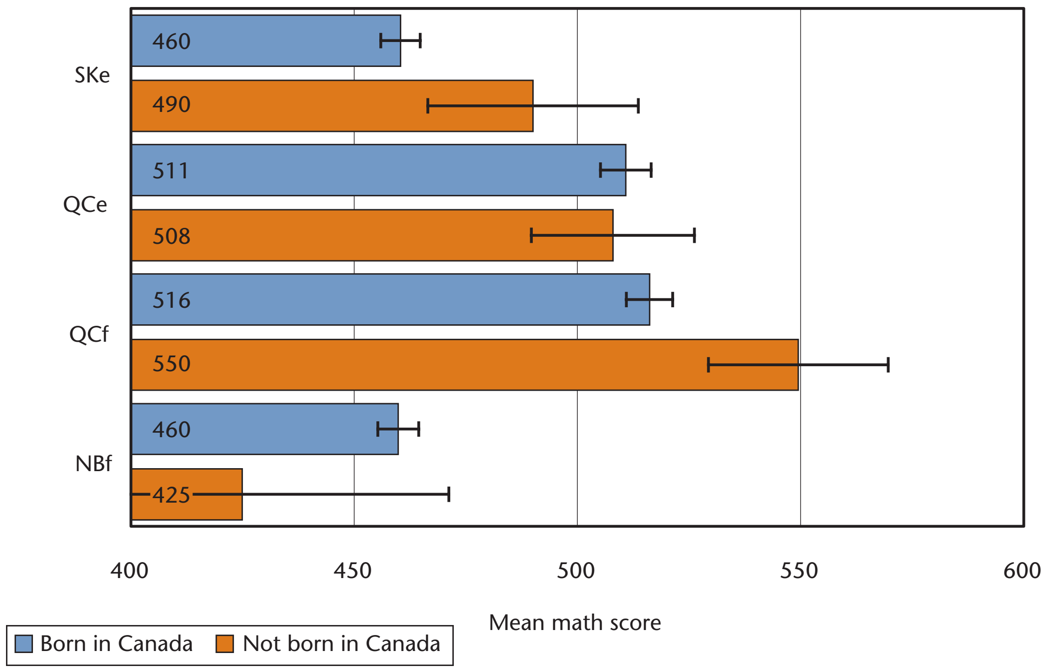
Student Diversity

Student diversity was measured in several ways. In this section, we focus on the number of immigrant students and the number of students requiring program modification.

The proportion of students not born in Canada is relatively small for the selected jurisdiction/language groups, although variable across groups, at 3 percent for Saskatchewan, 9 percent for Quebec English, 6 percent for Quebec French, and 1 percent for New Brunswick French. While not large enough to have a major effect on overall performance, the question of how immigrant students perform relative to their Canadian-born counterparts is of substantial policy interest.

Chart 7.9 shows the difference in mean mathematics scores for the two immigration status categories. Those born outside Canada have significantly higher scores than Canadian-born students in two jurisdictions, Saskatchewan English and Quebec French. It is worth noting that the two jurisdictions with the highest proportion of immigrant students — Ontario English and British Columbia English — both also show significant differences in favour of immigrant students.

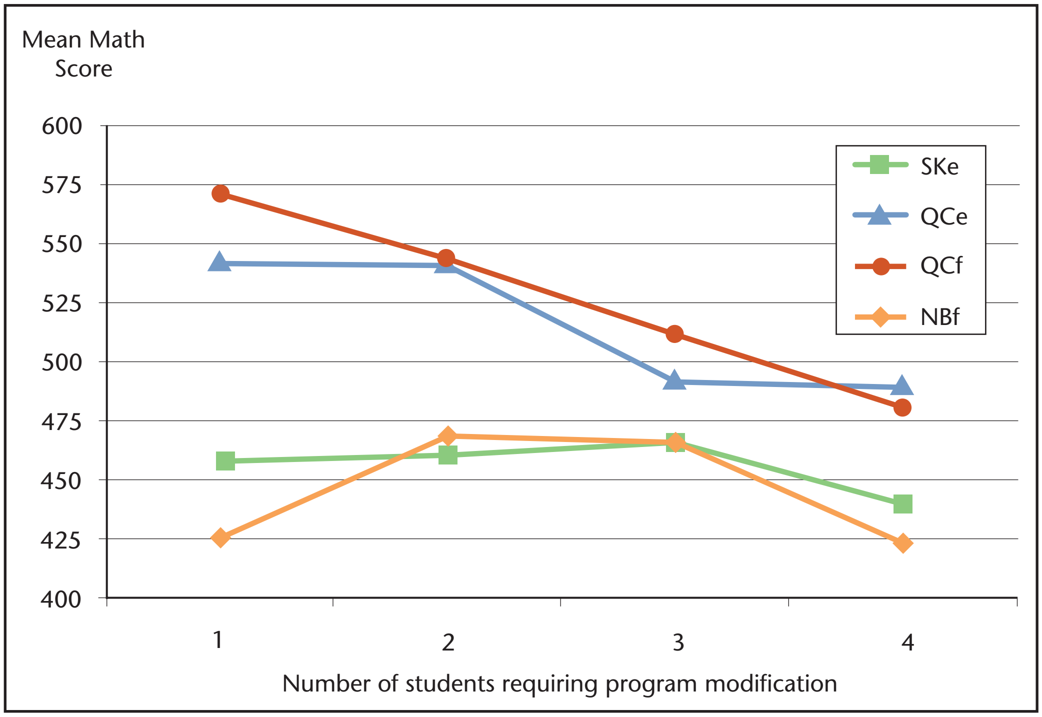
Chart 7.9 **Mean mathematics score by immigration status: Selected jurisdiction/language groups**



The number of students requiring program modification was measured by a question on the teacher questionnaire. The response was, thus, directly applicable to language arts. Aggregated to the school level, program modification was also taken as an indicator of student diversity for the school as a whole.

On average, the number of students reported as requiring program modification was similar for the four selected groups. Results for the selected jurisdictions are given in Chart 7.10. In this case, the group effect, the diversity effect, and the interaction effect are all statistically significant. The interaction effect is evident in the difference between the relatively linear pattern for the two Quebec groups and the non-linear pattern for Saskatchewan and New Brunswick English. The number of students requiring program modification has a larger effect on mathematics performance for Quebec students than for those in the other groups.

Chart 7.10 **Mean mathematics score by number of students requiring program modification: Selected jurisdiction/language groups**

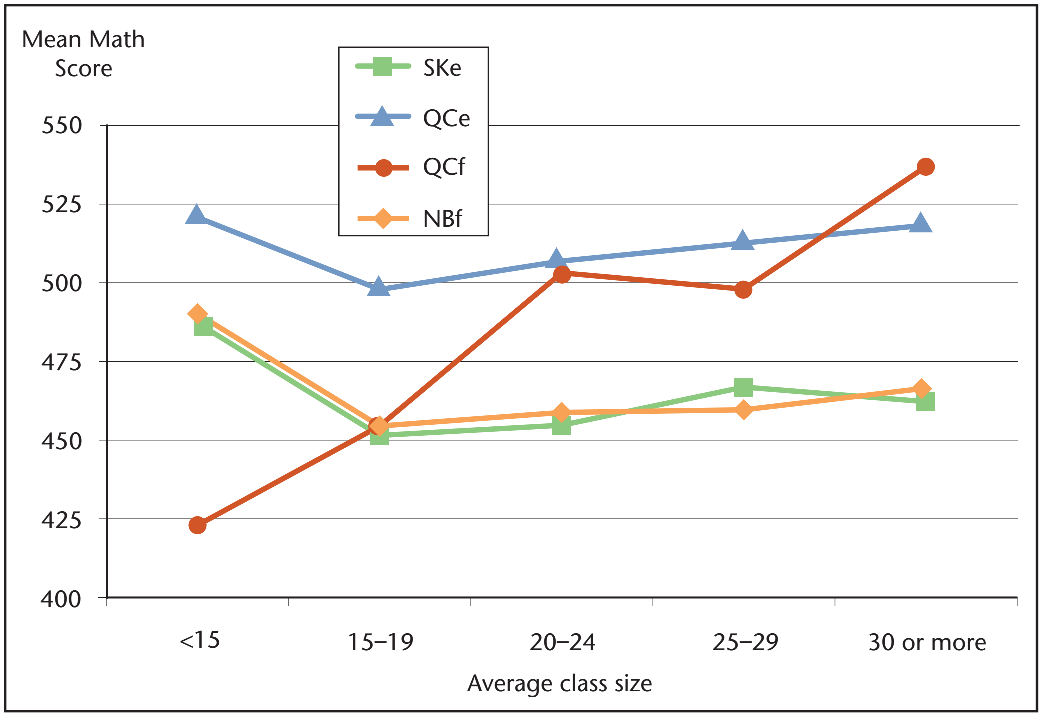


Class Size

Teachers reported average class sizes for language arts classes. Again, aggregated to the school, these numbers were taken as also representative of mathematics classes. Interpolating within the ranges reported, median class sizes were 21 for Saskatchewan, 22 for New Brunswick French, 23 for Quebec English, and 28 for Quebec French.

Overall, neither the class size effect nor the interaction effect is statistically significant in this case. More specific contrasts indicate that performance of students in class sizes greater than 30 is significantly higher than that for other size groups (Chart 7.11). This is obviously a function of the pattern for Quebec French, which shows a strong positive relationship between class size and mathematics performance and particularly high performance for those in the largest classes. For the other groups, the pattern of higher performance for class sizes less than 15 is consistent with the literature, although these differences are not statistically significant here.

Chart 7.11 **Mean mathematics score by class size: Selected jurisdiction/language groups**

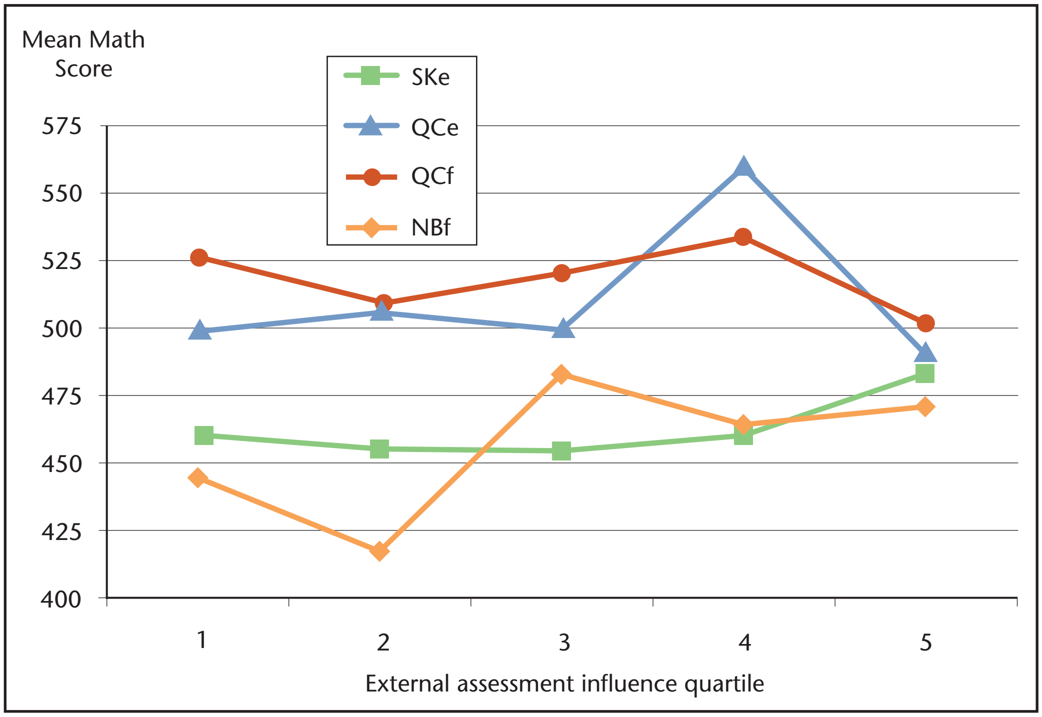


External Assessment Influence

The influence of external assessments on school programs was measured by a derived variable composed of a set of items about external assessments on the school questionnaire. In the narrowest sense, this should be a jurisdiction-level variable, as external assessments are mainly a matter of jurisdictional policy. This is evidenced by differences in the average value for this variable across the four selected groups (mean standard scores of –0.06, 0.15, –0.09 and 0.56 for SK, QCf, QCe and NBf respectively). Nevertheless, principals seem to have interpreted this in different ways for their own schools, as evidenced by variation in responses within jurisdictions. For example, for QCf, 22% of principals reported no influence, 14% a little influence, 38% more than a little influence, and 26% a lot of influence of external assessments that count toward student marks.

For purposes of presentation, the derived variable was divided into quintile ranges at a national level. Mathematics scores for students in schools within these quintiles were then computed for the selected groups. These results are given in Chart 7.12. In this case, the interaction between degree of influence and group is statistically significant, indicating that the effect of an external assessment differs among the groups. The chart shows a complex non-linear pattern, with those at the fourth quintile for Quebec French showing the strongest contrast with other groups. New Brunswick French shows higher performance for students in schools in the highest three quintiles.

CHART 7.12 **Mean mathematics score by external assessment influence: Selected jurisdiction/language groups**



Jurisdiction/Language Group Model Summary

This section gives a summary of the full model results for all of the jurisdiction/language groups and for a broader set of predictor variables than presented in the previous section. Tables 7.1 and 7.2 show the patterns of statistically significant model coefficients. Table A7.2 gives a complete summary of the coefficients and their standard errors.

One of the most obvious features of these results is the consistency of effects across groups. While not all effects are statistically significant across all groups, there are only a few instances in which differential effects are found for different groups. This indicates that the results can be treated as replications and that many of the predictor variables exert universal effects. Among the most consistent positive effects, for both mathematics and science, are the following:[14]

- grade level
- number of books in the home
- perceived performance on the PCAP test
- out-of-school reading/writing activities
- out-of-school entertainment activities

[14] It is important to note that the statistical power of the tests used in these models is considerably lower than that for the national models. This is especially true for school-level variables because of the small number of sampled schools. This is reflected in the relatively large standard errors for some of the effects (shown in Tables A7.2 and A7.3).

Homework time is also significantly positive in most cases, but is not as consistently so as the above variables. Taken together with a significant national effect and all the non-significant effects being in a positive direction (see Table A7.2), this is sufficient to indicate a relatively universal positive homework effect.

Negative effects are fewer and less consistent, though few show opposite directions for different groups. The strongest negative effects are for home language other than English or French and for fatalism. Some other effects that are statistically significant negatively for the bivariate and national models are attenuated in the group models. These include gender, public/private schools, and days absent.

Table 7.1 **Significant full-model coefficients by jurisdiction/language groups: Mathematics**

Variable	BC	AB	SK	MBe	MBf	ONe	ONf	QCe	QCf	NBe	NBf	NS	PE	NL
Demographics														
Gender														
Home language other			S–	S–					S–	S–	S–	S–		S–
Grade level		S+	S+	S+	S+	S+	S+	S+	S+	S+	S+	S+	S+	S+
Born in Canada		S–		S–			S+				S+	S–		
Number of books in the home	S+	S+	S+	S+	S+	S+	S+	S+	S+	S+	S+	S+	S+	S+
School demographics														
School size					S+							S+		S–
Public school	S–	X	X		X	X	X			X	X	X	X	X
Aboriginal students		S–	S–	S–										
Community size	S+													
Attitudes														
Enjoys school		S+						S+	S+					
Attributions														
Fatalism			S–		S–		S–		S–	S–	S–	S–		
Internal attributions of success and failure		S+			S–					S+	S+	S+		S+
Perceived performance on PCAP test	S+	S+	S+	S+	S+	S+	S+	S+	S+	S+	S+	S+	S+	S+
Out-of-school activities														
Reading/writing	S+	S+	S+	S+			S+				S+	S+	S+	
Entertainment	S+	S+			S+		S+	S+		S+	S+	S+	S+	S+
Academic/cultural activities				S–			S–	S–			S–			
Instructional climate														
Class size								S+	S+					
Number of students requiring program modification		S–		S–	S–						S–	S–	S–	S–
Disciplinary climate								S–		S–				
Time														
Days absent from school	S–	S–			S–	S+			S–				S–	
Time on homework	S+			S+		S+	S+	S+		S+	S–			
Total instructional days lost									S+	S–				
Influences on school programs														
External influence on school						S+								
Curriculum/teacher influence	S–													
External assessment influence				S+										S+

Key: S+ = significant positive relationship   S– = significant negative relationship   Blank = Relationship not significant   X = no data available

Table 7.2 **Significant full-model coefficients by jurisdiction/language groups: Science**

Variable	BC	AB	SK	MBe	MBf	ONe	ONf	QCe	QCf	NBe	NBf	NS	PE	NL
Demographics														
Gender										S+				
Home language other	S–	S+	S–	S–			S–		S–	S–	S–	S–	S–	S–
Grade level	S+	S–	S+	S+		S+	S+	S+	S+	S+	S+	S+	S+	S+
Born in Canada							S+							
Number of books in the home	S+	S+	S+	S+	S+	S+	S+	S+	S+	S+	S+	S+	S+	S+
School demographics														
School size					S+									
Public school	S–	X	X		X	X	X			X	X	X	X	X
Aboriginal students		S–	S–	S–	S+									
Community size	S+						S+	S–						
Attitudes														
Enjoys school	S+	S+										S+		
Attributions														
Fatalism			S–				S–				S–		S–	
Internal attributions of success and failure		S+						S+		S+		S+	S+	S+
Perceived performance on PCAP test	S+	S+	S+	S+	S+	S+	S+	S+	S+		S+	S+	S+	S+
Out-of-school activities														
Reading/writing	S+	S+	S+	S+	S+	S+	S+		S+	S+	S+	S+	S+	
Entertainment	S+	S+	S+	S+		S+	S+	S+		S+	S+			S+
Academic/cultural activities				S–	S+		S–	S–	S–		S–			
Instructional climate														
Class size					S+			S+	S+	S+		S+		
Number of students requiring program modification					S–								S–	
Disciplinary climate					S+		S+							
Time														
Days absent from school		S–												S–
Time on homework	S+					S+		S+						
Total instructional days lost														
Influences on school programs														
External influence on school			S+											
Curriculum/teacher influence	S–				S+				S–					
External assessment influence														

Key: S+ = significant positive relationship S– = significant negative relationship Blank = Relationship not significant X = no data available

8 DISCUSSION AND POLICY IMPLICATIONS

This study was based on an educational productivity model derived from human capital theory. Whereas the classic education/earnings function uses education indicators as predictors and economic indicators as outcomes, the educational productivity model uses education (in this case achievement in core subjects) as the outcome and a variety of student- and school-level indicators as predictors. Two specific models — the Carroll time model and the proximity model of Wang, Haertel and Walberg — were used in constructing the PCAP questionnaires, thus identifying the specific variables to be used. Many of these variables have some support as predictors of achievement, based on both the theoretical literature and the review of empirical research presented. Specific groups of variables were selected for study, based on the literature and on the research questions posed.

The study used data from PCAP 2007, which measured achievement in core subjects for Canadian 13-year-olds. Predictor variables were selected from student, teacher, and school questionnaires. Some of these variables were directly measured while others were derived from item sets within the questionnaires. The effects of these variables on mathematics and science achievement were examined on a bivariate basis and through hierarchical (student and school) models in which other variables were controlled. Both national results and selected results for jurisdictions and the two official language groups within some jurisdictions were given.

This section discusses the results in relation to the research questions and examines the policy implications of the major findings, both within the study and in relation to the literature reviewed.

It is important to reiterate that this study addresses a limited number of variables, especially in relation to teaching and learning strategies, because mathematics and science were minor domains in PCAP 2007. Most of the questionnaire items were therefore directed at the major domain — reading. The variables that could be used here were more generic ones, including student and school characteristics, attitudes, and out-of-school activities, along with a few classroom climate variables where inferences were made from responses that were mainly about reading. At the jurisdiction/language group level, the study is also limited by smaller sample sizes than was the case for reading.

What Does the Research Literature Tell Us?

The research consistently shows a positive effect of higher socioeconomic status on achievement. In large-scale assessments, this effect may be overestimated relative to effects for teaching and learning strategies. Both achievement and SES represent stable, cumulative features of the student's experience, whereas teaching and learning strategies are typically measured at a single point in time and may not reflect the student's long-term experience with schooling.

Variables related to SES such as having an immigrant background and a home language different from the language of the school generally contribute to lower achievement.

The research also indicates that time is important for achievement. However, since the macro-components of time (e.g., length of school year and day) are not highly variable across schools or jurisdictions in Canada, it is difficult to give a full picture of the effects of time. Components of time that are fairly easily measured — time lost in classrooms, homework, and absenteeism — have effects in the expected direction with more time contributing to higher achievement.

Positive attitudes are associated with higher achievement. Although the direction of causality is not clear, since achievement is the ultimate goal, it makes more sense to assume that better attitudes contribute to higher achievement than the other way around.

With respect to school characteristics, large-scale assessment results tend to show no differences or differences favouring larger schools and larger classes. However, other studies tend to show results in the expected direction. Students in private schools tend to do better than those in public schools, even after controlling for SES. However, other selection factors, such as underlying student abilities, are generally not controlled in examining this effect.

A classroom climate that presents an orderly, structured approach to teaching tends to have positive effects on achievement. However, the research in this area was not fully reviewed here. In any event, these effects were not investigated in any detail in this study because only limited data on classroom climate in mathematics and science were gathered in PCAP 2007.

More generally, the results support the Carroll and Wang-Haertel-Walberg models of school achievement, particularly the idea that time is important and that proximal factors are greater contributors than distal factors to achievement.

How Do Student Characteristics Influence Performance?

At a national level, the results may be summarized as follows:

- Grade level and number of books in the home (the measure of SES) are the most consistently positive predictors of achievement in both subjects.
- Speaking French as a home language is also a positive predictor. This is clearly a function of the high performance of Quebec francophone students as other francophone groups have much lower average scores. The complement of this is that speaking English at home is a negative indicator in the full model, as is speaking a language other than English or French.
- The gender effect favours males for both subjects for the full model but not the bivariate model. This indicates that the gender effect is suppressed by other variables when taken alone.
- Students born in Canada do less well in mathematics than those born outside of Canada. The opposite is true for science when taken alone but not with other variables controlled.

How Do School Characteristics Influence Performance?

- At a national level, the strongest school characteristic effect is public versus private schools, with students in private schools having higher scores. This effect is specific to the small number of jurisdictions with sufficient private schools to permit analysis.
- Schools having higher proportions of Aboriginal students have lower average scores. This effect is significantly reduced for mathematics when other factors are controlled.
- Schools in larger communities have higher scores in mathematics but not in science.
- Students in larger classes show higher achievement levels in the bivariate model. However, this effect is attenuated when other variables are controlled, suggesting that class size is related to other factors included in the full model. Specifically, class sizes are larger in larger schools and communities, both of which are positively related to achievement.

How Do Students' Attitudes and Attributions of Success and Failure Influence Performance?

- Enjoyment of school, enjoyment of reading and student perception of being a good reader are positive indicators of both mathematics and science performance.
- External attributions of success are a negative predictor of performance in the full model but not in the bivariate model. Again, this difference indicates an effect that is suppressed by other variables, most likely attitudes, in this case.
- Student perceptions of their performance on the PCAP test are positively related to achievement, although this is attenuated somewhat by other variables in the full model. The results for student perceptions of the fairness of the PCAP test are in the same direction but somewhat weaker.

How Do Reading Behaviours and Strategies Influence Mathematics and Science Performance?

This issue was investigated only in a limited way because exploratory analysis revealed few significant effects. However, enjoyment of reading and perceptions of being a good reader are both positive indicators of mathematics and science performance.

The ability to address the related question about the extent to which reading, mathematics, and science performance are inter-correlated was limited by the fact that the students who wrote the PCAP reading test were one group and the students who wrote the PCAP mathematics/science test were a different smaller group. Therefore, reading could be correlated with mathematics/science *only* at the school level. Nationally, the correlation between mathematics and science is moderate (0.61 at the student level and 0.64 at the school level). The correlations with reading at the school level are lower (0.38 for mathematics and 0.36 for science). At the jurisdictional/language group level, the correlations are variable but all significantly positive. While not very definitive, these results suggest that something other than a general ability trait is being measured by the three assessments.

What Internal and External Factors Influence School Programs?

The influence of external resources on school programs is positively related to performance in mathematics and is marginally so for science in the full model. Other sources of influence (specifically curriculum, teachers, and external assessment) show no significant effects.

Jurisdictional/Language Group Effects

The specific question posed for this study was *"How do provincial/territorial characteristics in practices, policies, and procedures such as school size, governance model (public/private), class size, large-scale assessment procedures, and diversity of student population influence mathematics and science performance?"*

The jurisdictional reference in this question was extended to include a breakdown of the two official language groups within some jurisdictions and was also pursued in greater depth than implied by the question. This was done mainly in an exploratory manner, in the hope that this would shed light on the analytical issues in comparing effects across jurisdictions.

Three different analytical approaches were used. First, models were developed in which jurisdiction/language groups were included as dummy variables. This yielded a set of coefficients corresponding directly to the difference between each jurisdiction and a reference group (Ontario English in this case). Adding the various predictor variables to that model allowed changes in these coefficients to be observed. Significant changes in the mathematics coefficients were found for three groups — Saskatchewan, New Brunswick French, and Nova Scotia English. Marginal changes were found for Manitoba English and New Brunswick English. This suggests that the differences between these jurisdictions and the reference groups are significantly mediated by the predictor variables.

This type of analysis does not shed much light on the question that is likely of greatest policy interest, that of "*Why do some jurisdictions have higher performance than others.*" It is understood that this question will be the subject of another project using the PCAP 2007 database. However, the question may be framed in terms of explaining the variance between jurisdictions using variables characteristic of jurisdictions, as implied by the research question posed for this study. The logically obvious way to do this is to aggregate some of the available variables to the jurisdictional level (or find some other variables that are characteristic of jurisdictions) and run a hierarchical model with jurisdiction as the highest level.

Given the structure of the data set, this requires a 3-level model with student, school, and jurisdiction as the levels. A preliminary exploration of the 3-level model revealed that it is technically feasible to develop and analyze such a model. However, that model is limited by the small number of units at Level 3 (15 in all including the language breakdowns). Effective hierarchical analysis requires that there be many more units at the highest level and also that the selected units be a random sample of all the units in the population. Neither of these requirements can be met with the PCAP data.

In any event, running the 3-level model revealed, first, that the level-3 effects account for only 3 percent of the total variance in the model. This makes it clearer that, despite their high policy interest, jurisdictional differences are much smaller than either student or school differences. The analysis also shows that adding student- and school level variables aggregated to the jurisdiction accounts for more of the variance than using the student- and school level variables at their "natural" or measured levels. However, none of the specific jurisdiction level variables are statistically significant. This is almost certainly related to the small number of level-3 units, which results in large standard errors at that level. Overall, the 3-level model was judged inappropriate as a means of answering the question of differences across jurisdictions. However, this model does suggest that jurisdictional differences may not be the most important policy issue emerging from the data.

Finally, a series of exploratory analyses was used to investigate more directly the issue of the equity of school systems within jurisdictions and the differential impact of some of the predictor variables on mathematics and science achievement.

On the equity question, the results show some differences in the variability of scores across jurisdictions at the level of both the student and the school. These differences mean that some jurisdictions have managed to reduce disparities among students and schools more than others. Differences are especially large for mathematics, where Quebec English stands out as having larger disparities in student scores and across its schools than other jurisdictions. The differences in school-level variance are more striking than those for student-level variance. For example, 30% of the mathematics variance is across schools in the Quebec English system but schools account for only 3% and 2% of the variance for Prince Edward Island and Yukon respectively. This is likely related to the relative diversity of school populations, although that point could not be pursued in detail in our study.

The next step was to conduct an exploratory analysis of bivariate effects by jurisdiction/ language group. A table of mathematics and science scores for each level of the categorical predictor variables and a table of correlations between scores and values for continuous predictor variables were produced as references for those who might be interested in specific results within jurisdictions. A more direct analysis (two-way ANOVA with graphical displays) of interactions between selected predictors and jurisdiction/ language groups (Saskatchewan, Quebec English, Quebec French, and New Brunswick French) with different means and variances and selected predictors was also conducted to illustrate how differential effects across jurisdictions might be explored.

Generally, this analysis shows some comparable trends and some interaction effects. As an example, the socioeconomic gradients for the four selected jurisdictions are similar, even though average scores for these groups are near the extremes. The private school effect is also similar across the four jurisdictions with sufficient private schools for analysis. The school-size effect is non-linear for all but Quebec French, where mathematics performance increases linearly with school size. The greatest variability in performance is found in schools in the 101–500 size range, which is the modal range for all except Quebec French, where schools are generally larger. The pattern for number of students requiring program modification (a within-school measure of diversity of the student population) is linearly negative for Quebec French and Quebec English and non-linear for the other two groups examined. In the latter cases, mathematics performance is lower for the lowest (none) and highest (more than 5) number of students requiring program modification, with higher performance for those in the mid-range. Class size is another example of a differential effect for Quebec French relative to others, with mathematics performance increasing with larger class sizes, whereas there is only a small effect for other jurisdictions.

Although by no means comprehensive, these results illustrate the complexity of conducting detailed differential analysis by jurisdiction and trying to make sense of the patterns found. This is especially complex when samples are small at the jurisdiction level, because the analysis results in large standard errors for specific categories. Analysis might be somewhat easier for reading than for mathematics/science because sample sizes are larger, and also because the questionnaires are more explicitly directed toward reading.

Finally, again mainly for reference purposes, a set of full models was run for each jurisdiction/language group using predictors selected on the basis of the national results, differential effects, and policy interest. For the most part, these results are consistent across jurisdictions, indicating that most of the variables act in the same direction for all jurisdictions. This, in turn, is indicative of effects that are more universal than differential. At the same time, these models are not capable of detecting non-linearities or differential average values for the predictors that might influence how a particular predictor acts to affect performance within a group.

Policy Issues and Implications

On the surface, the main policy issue driving large-scale assessments such as PCAP is a desire to report to the public on the performance of the provincial/territorial education systems in Canada. This is expressed in the statement "the Pan-Canadian Assessment Program (PCAP) is the CMEC's most recent commitment to informing Canadians on how well their education systems are meeting the needs of students and society" (CMEC, 2008).

Such studies, however, are clearly intended to go beyond a simple description of performance levels. An obvious comparative element exists in PCAP (and its predecessor SAIP) because all the main reports give direct comparisons of achievement levels for the separate (provincial and territorial) systems within Canada and for the two official language groups, which operate to a large extent as separate systems within jurisdictions. Also, as these studies have evolved, there is more and more interest in deriving added value by examining results from the questionnaires that accompany most such studies. Again, the initial interest was in comparing responses across jurisdictions. However, the main focus of most recent research based on large-scale assessments is on the factors that influence achievement.

As it happens, most of the factors identified in this and other studies are universal in nature, in the sense that they exert much the same influence in all jurisdictions. Thus, for example, it would be most unusual to find a jurisdiction in which the socioeconomic effect favours students of low socioeconomic status. Although some jurisdictions might show greater socioeconomic equity than others, the general pattern is the same for all — an observation also true for most other effects.

This is actually reassuring from both a scientific and a policy perspective because it indicates that it is possible to identify some factors on which there can be universal agreement. Although education systems are local, many factors influencing achievement are not local. The next question, therefore, is "What are these factors and what are their effects?"

The results for socioeconomic status and for other variables likely related to SES (such as private versus public schools, the proportion of Aboriginal students in a school, and the number of students requiring program modification) are generally consistent with the literature. It is clear that this remains one of the most challenging aspects of policy, related to both average achievement levels and to equity. While it is beyond the scope of this study to give prescriptions for overcoming socioeconomic disadvantage, a large body of literature and many policy initiatives can be found on this issue. To the extent that some of the other factors affecting achievement have a socioeconomic component, these might shed light on possible policy initiatives.

One example of this is the result for time. Although only a few components of time were measured in this study, the results for variables such as days absent from school and homework time are consistent with the literature. The possible socioeconomic link to absence and homework was not pursued in this study, but would be a useful area to investigate in studies more explicitly focused on socioeconomic effects. The question of whether ways should be found to provide more time for learning for those who are at socioeconomic disadvantage (or other forms of disadvantage for that matter) is of obvious policy importance. The research clearly points to this as potentially useful. However, research reviewed here does not answer the question of how this can be achieved.

The results for grade level are interesting because grade level relates to school starting age and is, hence, directly subject to policy influence. This seems not to have been examined in other studies and, indeed, can only be examined for age-based groups such as found in PCAP and PISA. On the surface, those in higher grades at age 13 would appear to have had more exposure to schooling than those in lower grades. On this assumption, the grade-level results are consistent with the time model. However, it is also possible that those in lower grades may have been "grade retained" and those in higher grades "grade-advanced." If this is so, then the grade-level results may represent an ability factor more than an exposure to schooling factor. Since there is no information on those factors in the database, it is not possible to draw a clear conclusion from the grade level results. This clearly deserves further investigation and might be a source of insight into differential jurisdictional or school policies that can shed light on differences in achievement.

The results for attitudes are also consistent with the literature in showing that positive attitudes are associated with higher achievement, even with other factors controlled. Of particular interest is that attitudes toward reading show positive effects for mathematics and science achievement. This most likely indicates that attitudes are generic traits. In this case, as with other variables in these models, it is not possible to draw a clear empirical conclusion on the direction of causation, and theory does not shed as much light on this as it does on other variables. Despite this, it is more plausible to argue that policies and practices designed to improve attitudes might, in turn, positively influence achievement than to simply assume that positive attitudes will follow from high achievement.

The results for out-of-school activities are obviously linked to the time model. This makes the results for entertainment puzzling. One would expect time spent on entertainment would detract from time on learning activities (the entertainment factor is composed of items on watching television and using computers for games or personal correspondence). The interesting question for further study is whether computer use may facilitate learning in mathematics and science or whether students who are good at mathematics and science are the ones who happen to make the most use of computers, whatever the purpose. Since the list of activities given in the questionnaire was not exhaustive, it is possible that students who spend time on those activities included under entertainment are doing so at the expense of even less-productive activities.

These variables are a bit like socioeconomic status in that they cannot be directly influenced by education policy. However, the link to time is important in that all these activities take time so that policies directed toward productive use of time should be encouraged.

Class size is a key aspect of policy, both because of strong pressures to keep class sizes small and because class size reduction is a costly initiative. Not much can be said about class size from these results other than that it seems to make little difference to mathematics or science achievement when other variables are controlled.[15] Other large scale assessments tend to give results favouring larger classes. However, this conflicts with the results of the few large-scale experiments available, which point to advantages for smaller classes. This may be related to lack of full control of extraneous variables in the survey research or it may be a function of the grade levels studied, as almost all of the experimental studies have been in the early grades. The main point that can be made is that the survey research does not point to any advantage of reduced class size and, hence, does not support major (and expensive) class-size reduction initiatives in the intermediate grades as a means of increasing achievement.[16]

The allocation and use of resources is one of the most obvious policy tools available to jurisdictions. Indeed, most high-level policy initiatives involve, either directly or indirectly, decisions on resource levels and on how resources are to be allocated. This issue has been quite controversial over the years, although there now seems to be some consensus that resource allocations and, particularly, resource use can have an impact. Except perhaps for the class-size results, this study sheds only limited light on resource issues. The influence of external resources on school programs has a positive effect, with the schools that reported a higher level of such influence having higher achievement. However, the question relates to the principal's perceptions of influence and not to the absolute level of resources available. That issue has not been investigated in large scale assessments in Canada, but should be a fruitful area to pursue in examining differences across jurisdictions.

[15] It should be reiterated that the class-size question was asked of language arts teachers, and hence is not a direct measure of class size in mathematics or science. The reading results for class size show higher achievement for larger classes in the full model.

[16] There may be reasons for reducing class size other than that of increasing achievement. These are not investigated in large-scale assessment research.

This brings us, finally to the question of differences among jurisdictions. As noted earlier, the comparative element in these assessments is typically limited to reports on differential achievement. Since such studies almost always show significant differences across jurisdictions, the next most obvious question is "Why do some jurisdictions do better than others?" While this question is crucial from a policy perspective, the research based on large-scale assessment rarely addresses the question directly. Some effort was made to examine jurisdictional and language differences in the hope that a more explicit study in the near future can pursue this issue in more detail.

The results presented here perhaps show why it is difficult to get a clear answer to the question of jurisdictional differences. These differences are actually quite small compared to the differences between students and between schools within jurisdictions. It might therefore be argued that the question of jurisdictional difference is much less important than that of the differences between students and, in some jurisdictions in particular, of the differences between schools. Also, the ability to analyze data aggregated to the jurisdictional level is limited both by the small number of units available and by the fact that the most important variables exert similar effects in all jurisdictions.

Assuming that the ultimate goal of all education policy is to increase average performance levels and to reduce inequalities in performance, this and other studies of this nature suggest that it is better to focus on students and schools than on jurisdictions. Nevertheless, the question of jurisdictional differences is crucial both because jurisdictions have ultimate responsibility for education and because looking to other jurisdictions is a common way of determining and defending large-scale policy initiatives.

This study has pointed to some possible ways of looking at this issue, but falls considerably short of a comprehensive analysis. The analysis has revealed some fundamental difficulties in conducting analysis at the jurisdictional level, where data are highly aggregated. These include 1) the small proportion of total score variance that is at this level, 2) similar small variations in values for the predictor variables, and 3) the small number of cases available for analysis. All of these would seem to preclude using 3 level models to investigate the issue. Furthermore, it is likely that important variations in jurisdictional policies and practices have been missed because the relevant data are not captured by the PCAP questionnaires. Examples would include curriculum content and coverage, resource allocations, and system-level elements of time.

Nevertheless, this issue is too important to ignore, and addressing the issue has the potential of adding considerable value to the PCAP assessments. The final suggestion is, therefore, that the question of jurisdictional differences be placed more explicitly on the PCAP research agenda, and that the initial study (using the PCAP 2007 data) concern itself with developing a clearer conceptual approach toward doing this and toward more fully investigating the data requirements and the analytical techniques required.

REFERENCES

Becker, G. (1993). *Human capital: A theoretical and empirical analysis with special reference to education*. Chicago, MI: University of Chicago Press.

Bloom, B. S. (1976). *Human characteristics and school learning*. New York, NY: McGraw-Hill.

Braun, H., Coley, R., Jia, Y., & Trepani, C. (2009). *What works in science instruction: A look at the eighth-grade science classroom*. Princeton, NJ: Educational Testing Service.

Bussière, P., Cartwright, F., & Knighton, 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*. Ottawa, ON: Statistics Canada.

Carroll, J. B. (1963). A model for school learning. *Teachers College Record*, 64, 8, 723–733.

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. (2002). *Student writing: The Canadian context*. (SAIP 2002, Writing III). Toronto, ON: Author.

Council of Ministers of Education, Canada. (2003). *Mathematics learning: The Canadian context*. (SAIP 2001, Mathematics III). Toronto, ON: Author.

Council of Ministers of Education, Canada. (2005). *Science III*, 2004. (SAIP 2004, Science III). Toronto, ON: 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, ON: Author.

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

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

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

Greenwald, R., Hedges, L. V., & Laine, R. (1996). The effect of school resources on student achievement. *Review of Educational Research*, 66, 3, 361–396.

Grissmer, D., Flanagan, J., Kawata, J., & Williamson, S. (2000). *Improving student achievement: What state NAEP scores tell us*. Santa Monica, CA: RAND Corporation.

Gunnarsson, V., Orazem, P. F., Sánchez, M., & Verdisco, A. (2004). *Does school decentralization raise student outcomes? Theory and evidence on the roles of school autonomy and community participation*. Unpublished paper, Department of Economics, Iowa State University. Retrieved, April 30, 2010, from **http://econ2.econ.iastate.edu/faculty/orazem/School%20Autonomy%20final.pdf**

Hanushek, E. A. (1996). A more complete picture of school resource policies. *Review of Educational Research*, 66, 3, 397–409.

Hawley, W. D., Rosenholtz, S., Goodstein, H. J., & Hasselbring, T. (1984). Good schools: What research says about improving student achievement. *Peabody Journal of Education*, 61, 4, 1–178.

Hiebert, J., & Grouws, D. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning*. Charlotte, NC: Information Age Publishing.

Irmsher, K. (1997). School Size. ERIC Digest #113. Retrieved April 30, 2010, from **https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/3335/ digest113.pdf?sequence=1**

Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, 30, 3, 420–437.

Leithwood, K., & Jantzi, D. (2009). A review of empirical evidence about school size effects: A policy perspective. *Review of Educational Research*, 79, 1, 464–490.

Ma, X., & Crocker, R. (2007). Provincial effects on reading achievement. *Alberta Journal of Educational Research*, 53, 1, 81–109.

Marzano, R. (2003). *What works in schools: Translating research into action*. Washington, DC: Association for Supervision and Curriculum Development.

Mincer, J. (1974). *Schooling, experience and earnings*. New York, NY: Columbia University Press.

Molnar, A., Smith, P., Zanonik, J., Palmer, A., Halbach, A., & Ehrle, K. (1999). Evaluating the SAGE program: A pilot program in targeted pupil-teacher reduction in Wisconsin. *Education Evaluation and Policy Analysis*, Special Issue, 21, 2, S165–177.

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. (2001). *Knowledge and skills for life: First results from PISA 2000*. Paris: Author.

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.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Second edition*. Thousand Oaks, CA: Sage.

Scheerens, J., & Bosker, R. (1997). *The foundations of educational effectiveness*. New York, NY: Elsevier.

Shepard, L. A., Smith, M. L. (1990). Synthesis of research on grade retention. *Educational Leadership*, 47, 8, 84–88.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75, 3, 417–453.

Slavin, R., Lake, C., & Groff, C. (2009). Effective programs in middle and high school mathematics: A best-evidence synthesis. *Review of Educational Research*, 79, 2, 839–911.

Stigler, W., & Hiebert, J. (1999). *The teaching gap: Best ideas for the world's teachers for improving education in the classroom*. New York, NY: Free Press.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 63, 3, 249–294.

White, K. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, 91, 461–481.

Willms, J. D. (2003). *Ten hypotheses about socioeconomic gradients and community differences in children's developmental outcomes: Final report*. Report prepared for the Applied Research Branch, Strategic Policy, Human Resources Development Canada, Ottawa, ON.

Willson, V. L. (1983). A Meta-analysis of the relationship between science achievement and science attitude: Kindergarten through college. *Journal of Research in Science Teaching*, 20, 9, 839–50.

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 A6.1 **Coefficients and standard errors for pan-Canadian models**

Variable	Source	Scale/categories	Mean	SD	Bivariate		Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8	
Mathematics																						
					coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE
Demographics																						
Gender	student	Male=1, Female=0	0.49		4.49	3.20	9.44	3.07	9.86	3.10	11.72	3.16	10.21	3.15	9.05	3.19	9.19	3.15	9.08	3.14	9.03	3.12
Home language English	student	English=1, not English=0	0.67		-7.58	4.63	-24.04	3.96	-15.62	3.98	-13.26	3.92	-16.47	3.87	-16.10	3.88	-15.57	3.95	-15.96	4.01	-17.27	4.12
Home language French	student	French=1, not French=0	0.2		12.97	4.17	24.04	3.96	15.62	3.98	13.26	3.92	16.47	3.87	16.10	3.88	15.57	3.95	15.96	4.01	17.27	4.12
Home language other	student	Other=1, not other=0	0.13		4.12	5.75	-19.53	6.37	-14.14	6.24	-11.74	6.15	-15.98	6.10	-15.61	6.10	-14.69	6.14	-15.85	6.14	-16.80	6.10
Grade level	student	1–5; grades 6 to 10	3.14	0.52	28.09	3.47	26.69	3.40	26.66	3.30	26.69	3.22	25.77	3.08	25.01	3.06	24.74	3.01	24.73	2.98	25.04	2.99
Immigration status	student	Born in Canada=1, not born in Canada=0	0.91		-20.60	6.28	-22.32	6.34	-18.03	6.38	-15.34	6.21	-14.64	5.97	-13.30	6.12	-13.04	6.15	-10.42	6.05	-11.00	5.97
Number of books in the home	student	1–5; 0–10 books to more than 200 books	3.36	1.21	20.69	1.36	21.31	1.35	20.80	1.34	14.19	1.33	13.66	1.28	13.03	1.34	12.91	1.34	12.46	1.37	12.48	1.36
School demographics																						
School size	school	1–4; < 100 to > 1000	2.57	0.82	10.33	2.89			-1.07	2.87	0.44	2.73	-0.04	2.77	-0.21	2.76	-1.60	2.88	-2.83	3.04	-3.08	3.02
Grade 8/secondary II enrolment		1–5; <25 to >100			1.39	1.59			0.08	1.64	0.37	1.56	0.62	1.49	0.78	1.49	0.24	1.56	0.36	1.50	0.27	1.49
Public school	school	Public=1, not public=0	0.92		-40.64	5.93			-21.74	5.87	-25.64	5.61	-26.35	5.50	-27.46	5.35	-25.58	5.54	-22.09	5.42	-25.71	5.47
Aboriginal students	school	1–5; 0 to more than 50%	1.99	0.83	-19.46	2.43			-14.33	2.27	-14.80	2.20	-12.98	2.16	-12.06	2.15	-11.22	2.14	-9.66	2.07	-8.60	2.05
ESL/FSL students	school	1–4; equal intervals of 25%	1.17	0.56	2.84	4.19			2.46	3.42	1.31	3.25	0.39	2.84	0.66	2.86	1.25	2.83	1.15	2.74	0.61	2.75
Community size	school	1–5; <5,000 to more than 500,000	3.19	1.53	9.24	1.53			5.87	1.64	4.88	1.58	4.45	1.54	3.99	1.53	4.13	1.55	3.37	1.49	3.17	1.47
Attitudes																						
Enjoys school	student	Derived from attitude to school scale	50	10	1.46	0.17					0.86	0.16	0.86	0.17	0.87	0.17	0.88	0.18	0.77	0.18	0.77	0.18
Sense of belonging to school		Derived from attitude to school scale	50	10	0.41	0.18					-0.14	0.17	-0.06	0.17	-0.05	0.17	-0.04	0.17	-0.10	0.17	-0.09	0.17
Enjoys reading	student	Derived from attitude to reading scale	50	10	2.51	0.18					1.39	0.20	1.30	0.20	1.07	0.24	1.08	0.24	1.05	0.23	1.06	0.23
Good reader	student	Derived from attitude to reading scale			2.03	0.18					1.28	0.18	1.00	0.19	0.89	0.19	0.90	0.20	0.94	0.19	0.94	0.19

Variable	Source	Scale/categories	Mean	SD	Bivariate		Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8	
					coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE
Attributions																						
External attributions of success	student	Derived from attribution scale	50	10	0.04	0.18							-1.15	0.18	-1.16	0.19	-1.15	0.18	-1.19	0.18	-1.20	0.18
External attributions of failure	student	Derived from attribution scale	50	10	0.15	0.18							0.43	0.17	0.35	0.17	0.35	0.17	0.29	0.16	0.29	0.16
Fatalism	student	Derived from attribution scale	50	10	-1.15	0.17							-0.09	0.19	-0.14	0.19	-0.14	0.19	-0.08	0.19	-0.08	0.19
Internal attributions of success and failure	student	Derived from attribution scale			0.00	0.19							-0.05	0.16	-0.07	0.16	-0.07	0.16	0.05	0.16	0.04	0.16
Perceived performance on PCAP test	student	1–3; very well to not at all well			47.82	2.89							34.90	2.99	34.10	2.98	34.20	2.98	32.82	2.94	32.80	2.93
Perceived fairness of PCAP test	student	1–3; very fair to not at all fair	1.8	0.61	22.21	2.87							7.74	2.92	7.95	2.92	7.93	2.92	8.56	2.88	8.56	2.87
Out-of-school activities																						
Reading/writing	student	Derived from time scale questions	50	10	2.26	0.15									0.52	0.19	0.51	0.19	0.20	0.20	0.20	0.20
Entertainment	student	Derived from time scale questions	50	10	1.03	0.18									0.79	0.17	0.79	0.17	0.81	0.17	0.80	0.17
Academic/cultural activities	student	Derived from time scale questions	50	10	0.70	0.18									0.03	0.17	0.02	0.17	-0.04	0.17	-0.05	0.17
Instructional climate																						
Class size	teacher	1–5; fewer than 15; 30 or more	3.72	1	7.82	2.83											2.90	2.97	2.57	2.91	3.20	2.85
Number of grades per class	teacher	1 to 3 or more	1.41	0.69	-6.39	4.31											-5.35	4.07	-4.37	4.01	-4.81	3.87
Number of students requiring program modification	teacher	1 to 5 or more	2.67	0.99	-11.45	3.15											-4.36	2.90	-3.69	2.80	-4.09	2.71
Disciplinary climate					-0.13	0.17											0.08	0.16	0.14	0.16	0.14	0.16
Time																						
Length of class periods	School	Minutes	55	16	0.27	0.14													0.04	0.15	0.08	0.15
Days absent from school	student	1–5; 0–2 days to more than 20 days	2.5	1.21	-11.66	1.38													-7.14	1.14	-7.16	1.14
Time on homework	student	1–5; less than 30 minutes to more than 3 hours	3.17	1.37	14.27	1.32													7.78	1.39	7.80	1.39
Total instructional days lost	teacher	Days: Sum over several categories	16	11	0.62	0.33													0.16	0.27	0.11	0.26
Influences on school programs																						
External influence on school	school	Derived from influence scale	50	10	5.03	2.44															6.36	2.13
Curriculum/teacher influence	school	Derived from influence scale	50	10	-0.83	2.85															0.00	2.42
External assessment influence	school	Derived from influ-ence scale	50	10	1.17	2.23															-0.73	1.91

Science

Variable	Source	Scale/categories	Mean	SD	Bivariate		Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8	
					coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE
Mathematics achievement	Math test	IRT scaled score	500	100																		
Science achievement	Science test	IRT scaled score	500	100																		
Student demographics																						
Gender	student	Male=1, Female=0	0.49		-1.07	3.08	4.12	2.89	4.30	2.89	11.61	2.85	7.30	2.97	5.89	3.07	5.80	3.06	6.23	3.08	6.18	3.08
Home language English	student	English=1, not English=0	0.67		20.35	3.92	-14.97	3.99	-8.49	4.12	-4.74	4.02	-9.33	4.06	-9.30	4.06	-8.64	4.03	-9.71	4.07	-11.61	4.16
Home language French	student	French=1, not French=0	0.2		9.96	4.16	14.97	3.99	8.49	4.12	4.74	4.02	9.33	4.06	9.30	4.06	8.64	4.03	9.71	4.07	11.61	4.16
Home language other	student	Other=1, not other=0	0.13		-29.72	4.73	-36.43	6.22	-30.79	6.27	-26.60	6.18	-30.91	6.04	-30.76	6.00	-29.71	5.99	-30.85	5.90	-32.22	6.06
Grade level	student	1–5; grades 6 to 10	3.14	0.52	28.99	3.39	27.82	3.23	28.07	3.08	27.88	3.01	27.52	2.95	26.84	2.91	26.70	2.85	26.74	2.68	26.81	2.69
Immigration status	student	Born in Canada=1, not born in Canada=0	0.91		14.96	5.55	0.75	6.34	2.66	6.41	5.25	6.17	5.35	5.99	6.26	6.04	6.45	6.09	8.27	6.06	7.86	6.03
Number of books in the home	student	1–5; 0–10 books to more than 200 books	3.36	1.21	21.05	1.33	20.28	1.35	19.73	1.35	11.71	1.41	11.28	1.37	10.99	1.42	10.87	1.41	10.56	1.41	10.54	1.40
School demographics																						
School size	school	1–4; < 100 to > 1000	2.57	0.82	9.00	2.60			0.56	2.59	2.41	2.53	2.01	2.49	1.97	2.49	0.21	2.48	-0.44	2.63	-0.46	2.61
Grade 8/secondary II enrolment		1–5; <25 to >100			1.33	1.42			1.73	1.42	2.04	1.36	2.22	1.33	2.43	1.32	1.86	1.37	1.92	1.36	1.69	1.38
Public school	school	Public=1, not public=0	0.92		-25.65	5.63			-13.62	5.29	-19.02	5.24	-18.75	5.10	-20.27	4.99	-18.87	5.14	-16.86	5.09	-19.43	4.96
Aboriginal students	school	1–5; 0 to more than 50%	1.99	0.83	-10.16	2.16			-8.69	2.14	-9.42	2.06	-8.35	2.09	-7.65	2.08	-6.52	2.12	-5.66	2.12	-4.92	2.10
ESL/FSL students	school	1–4; equal intervals of 25%	1.17	0.56	-9.11	3.09			-4.01	2.68	-4.98	2.47	-5.49	2.44	-5.26	2.39	-4.54	2.36	-4.50	2.39	-4.85	2.40
Community size	school	1–5; <5,000 to more than 500,000	3.19	1.53	1.54	1.45			0.94	1.41	-0.44	1.35	-0.57	1.32	-0.95	1.32	-0.94	1.36	-1.41	1.34	-1.48	1.34
Attitudes																						
Enjoys school	student	Derived from attitude to school scale	50	10	1.31	0.17					0.67	0.19	0.89	0.21	0.90	0.21	0.89	0.21	0.84	0.21	0.83	0.21
Sense of belonging to school		Derived from attitude to school scale	50	10	0.30	0.17					-0.35	0.17	-0.24	0.16	-0.19	0.16	-0.19	0.16	-0.22	0.16	-0.22	0.16
Enjoys reading	student	Derived from attitude to reading scale	50	10	2.70	0.16					1.60	0.18	1.59	0.18	1.36	0.22	1.37	0.22	1.37	0.22	1.37	0.22
Good reader	student	Derived from attitude to reading scale			2.77	0.17					1.99	0.18	1.74	0.17	1.65	0.17	1.66	0.17	1.68	0.18	1.69	0.18

Variable	Source	Scale/categories	Mean	SD	Bivariate		Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8	
					coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE
Attributions																						
External attributions of success	student	Derived from attribution scale	50	10	-0.20	0.17							-1.20	0.17	-1.17	0.17	-1.16	0.17	-1.19	0.17	-1.18	0.17
External attributions of failure	student	Derived from attribution scale	50	10	0.19	0.17							0.48	0.15	0.42	0.15	0.41	0.15	0.38	0.15	0.38	0.15
Fatalism	student	Derived from attribution scale	50	10	-1.38	0.17							-0.28	0.18	-0.33	0.18	-0.32	0.18	-0.28	0.18	-0.28	0.18
Internal attributions of success and failure	student	Derived from attribution scale			0.57	0.18							0.31	0.16	0.29	0.16	0.29	0.16	0.36	0.16	0.36	0.16
Perceived performance on PCAP test	student	1–3; very well to not at all well			32.04	3.02							19.90	3.06	19.40	3.08	19.48	3.08	18.88	3.09	18.90	3.08
		Reverse scored for analysis	2.12	0.55																		
Perceived fairness of PCAP test	student	1–3; very fair to not at all fair	1.8	0.61	10.51	2.65							1.21	2.71	1.24	2.70	1.09	2.71	1.38	2.68	1.37	2.68
Out-of-school activities																						
Reading/writing	student	Derived from time scale questions	50	10	2.11	0.17									0.48	0.22	0.47	0.22	0.27	0.22	0.26	0.22
Entertainment	student	Derived from time scale questions	50	10	1.01	0.17									0.75	0.16	0.76	0.16	0.77	0.16	0.76	0.16
Academic/cultural activities	student	Derived from time scale questions	50	10	0.30	0.17									-0.25	0.16	-0.26	0.16	-0.31	0.17	-0.31	0.17
Instructional climate																						
Class size	teacher	1–5; fewer than 15; 30 or more	3.72	1	7.21	2.46											4.50	2.59	4.31	2.53	0.30	1.90
Number of grades per class	teacher	1 to 3 or more	1.41	0.69	-4.61	3.90											-4.02	3.45	-3.57	3.39	-4.30	3.29
Number of students requiring program modification	teacher	1 to 5 or more	2.67	0.99	-11.72	2.81											-4.44	2.62	-3.97	2.56	-4.17	2.50
Disciplinary climate					-0.13	0.15											-0.05	0.15	-0.03	0.14	-0.02	0.14
Time																						
Length of class periods	School	Minutes	55	16	0.28	0.14													0.01	0.11	0.01	0.11
Days absent from school	student	1–5; 0–2 days to more than 20 days	2.5	1.21	-5.39	1.28													-2.40	1.15	-2.42	1.15
Time on homework	student	1-5; less than 30 minutes to more than 3 hours	3.17	1.37	10.17	1.32													5.32	1.36	5.43	1.36
Total instructional days lost	teacher	Days: Sum over several categories	16	11	0.49	0.29													0.07	0.25	0.05	0.25
Influences on school programs																						
External influence on school	school	Derived from influence scale	50	10	2.90	2.29															3.11	2.03
Curriculum/teacher influence	school	Derived from influence scale	50	10	-4.21	2.71															-2.62	2.41
External assessment influence	school	Derived from influence scale	50	10	2.31	2.21															0.30	1.90

Table A7.1 **Coefficients and standard errors for models with jurisdiction/language as variables**

Jurisdiction/Language			Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8	
			Student characteristics		School characteristics		Attitudes		Attributions		Out-of-school activities		Instructional climate		Time		Influences	
	coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE	coeff	SE
BC	-23.61	6.42	-23.32	6.06	-18.10	7.03	-16.36	6.75	-12.20	6.76	-11.00	6.65	-13.04	6.54	-13.82	7.06	-10.74	6.90
AB	-5.78	6.72	-5.38	6.25	-1.05	6.63	0.20	6.47	1.80	6.30	3.31	6.24	2.99	6.34	3.66	6.16	5.95	6.07
SK	-46.73	6.35	-38.47	5.92	-25.77	6.92	-24.67	6.51	-20.60	6.27	-19.96	6.24	-22.25	6.38	-19.65	6.31	-17.03	6.24
MBe	-27.31	6.88	-21.21	6.50	-11.99	7.14	-11.10	6.86	-8.38	6.72	-6.63	6.63	-8.92	6.72	-5.45	6.58	-2.25	6.48
MBf	-43.52	11.83	-36.79	11.50	-31.41	12.44	-26.09	11.79	-14.22	10.80	-16.26	10.90	-19.96	11.03	-17.51	10.69	-14.72	10.51
ONf	-41.98	6.85	-31.88	6.32	-36.54	6.57	-35.17	6.38	-31.15	6.13	-30.23	6.19	-32.57	6.51	-31.90	6.89	-29.47	6.83
QCe	0.61	9.44	6.69	8.97	-1.18	9.03	2.43	8.79	0.62	8.73	-0.44	8.55	0.14	8.59	-0.99	8.88	1.71	8.86
QCf	10.78	7.31	26.49	6.68	17.31	7.85	13.62	7.38	16.36	7.20	15.95	7.17	14.75	7.25	16.20	7.92	19.08	8.05
NBe	-47.00	6.95	-40.04	6.49	-34.05	6.87	-32.65	6.70	-29.80	6.58	-28.15	6.53	-27.43	6.59	-25.16	6.45	-21.62	6.43
NBf	-54.59	8.38	-29.89	7.68	-29.93	8.19	-32.55	7.94	-24.02	7.67	-21.98	7.69	-23.65	7.79	-19.24	7.74	-16.56	7.68
NS	-51.54	6.70	-41.24	6.34	-33.72	7.24	-30.93	6.99	-28.26	6.77	-27.49	6.71	-27.26	6.90	-26.37	6.84	-24.60	6.74
PE	-42.83	8.49	-38.09	7.98	-31.48	8.33	-30.06	8.25	-25.96	7.91	-26.03	7.77	-28.94	7.55	-24.55	7.38	-24.59	7.35
NL	-29.90	7.11	-20.62	6.64	-16.49	7.34	-12.21	7.17	-9.63	6.77	-9.10	6.72	-12.55	7.02	-13.92	7.11	-12.40	7.06
YK	-65.20	13.03	-53.71	13.85	-23.25	18.44	-22.36	18.10	-18.28	18.38	-15.28	18.33	-17.26	19.43	-18.99	18.75	-13.87	18.73

Note: Coefficients for this model represent the difference between mean scores for each group relative to the reference group, Ontario English. The latter group is thus omitted from the subsequent models. The coefficients for models 1 through 8 should be interpreted in terms of the change from the immediately previous model associated with the added variables.

Table A7.2 **Coefficients and standard errors by jurisdiction/language: Full models**

Mathematics: Full model																												
Variable	BC		AB		SK		MBe		MBf		ONe		ONf		QCe		QCf		NBe		NBf		NS		PE		NL	
	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE
Demographics																												
Gender	10.05	5.93	8.76	6.21	2.92	5.90	9.15	5.94	-4.91	10.14	11.44	6.60	11.24	6.43	13.26	9.91	-8.14	7.72	2.57	6.18	9.74	5.50	7.58	6.02	3.62	6.26	9.12	7.20
Home language other	4.70	8.95	1.43	8.48	-28.78	13.58	-22.78	10.67	-4.26	27.16	1.54	9.68	-25.64	13.70	-2.30	15.83	-35.01	15.20	-51.08	18.91	-60.60	10.36	-28.07	11.94	10.81	18.44	-49.80	24.74
Grade level	10.43	6.71	31.43	5.64	32.79	5.50	19.03	7.02	38.37	10.91	36.12	11.64	37.50	7.03	35.57	7.21	38.36	8.79	34.15	5.90	47.45	10.42	37.25	7.62	25.63	5.34	30.55	6.77
Born in Canada	-7.85	9.86	-21.62	10.65	-41.92	22.19	-37.07	14.51	38.57	54.46	-11.03	10.45	61.29	17.10	13.26	17.30	-15.83	16.59	-32.94	18.45	90.32	35.47	-42.00	19.92	17.29	23.47	-4.02	27.53
Number of books in home	13.23	2.40	11.69	2.78	13.78	2.62	15.12	2.70	15.23	2.50	16.67	2.95	15.66	2.49	12.96	4.00	8.94	2.69	13.82	2.50	18.16	2.73	11.61	2.89	9.65	2.78	20.61	3.61
School demographics																												
School size	-4.43	5.76	4.56	7.16	-2.01	8.07	8.93	5.83	43.88	12.57	-13.12	7.21	9.74	9.32	-2.41	7.68	5.81	5.24	1.29	6.36	-2.38	10.22	16.58	8.15	2.94	7.90	-25.21	8.82
Public school	-25.25	10.34		0.00		0.00	-21.10	14.69		0.00		0.00		0.00	-26.98	18.88	-15.73	12.04		0.00		0.00		0.00		0.00		0.00
Aboriginal students	2.03	6.35	-12.63	4.85	-17.55	3.53	-11.37	5.51	5.44	8.95	-7.49	7.51	-0.94	7.82	-16.84	11.63	4.73	6.39	1.87	6.48	-1.84	9.22	-8.94	9.51	-17.39	12.20	5.54	5.27
Community size	7.73	3.43	4.30	2.26	-0.23	2.86	0.85	2.53	4.65	3.24	1.20	3.19	3.82	3.59	-4.79	4.44	2.09	4.00	2.73	4.22	8.76	5.68	2.94	3.44	0.40	4.16	-1.09	4.63
Attitudes																												
Enjoys school	0.78	0.41	1.19	0.33	0.44	0.33	0.00	0.37	0.56	0.49	0.55	0.34	0.31	0.29	1.00	0.46	0.80	0.38	0.03	0.30	0.28	0.34	0.28	0.29	0.66	0.41	0.54	0.34
Attributions																												
Fatalism	-0.27	0.37	0.51	0.38	-1.16	0.32	-0.07	0.39	-1.77	0.85	0.02	0.36	-1.09	0.32	-0.41	0.39	-1.05	0.33	-0.81	0.28	-1.07	0.32	-0.78	0.30	-0.35	0.29	0.54	0.34
Internal attributions of success & failure	-0.07	0.37	0.91	0.36	0.11	0.35	-0.01	0.34	-1.33	0.59	0.02	0.33	-0.17	0.30	0.33	0.41	0.37	0.37	0.93	0.36	0.45	0.23	0.64	0.29	0.66	0.45	0.99	0.40
Perceived performance on PCAP test	41.46	5.56	46.53	5.59	26.81	5.28	29.18	5.73	55.28	9.33	35.22	5.80	38.22	5.89	27.76	7.78	30.79	7.92	30.34	5.49	30.09	5.72	38.21	5.93	24.55	4.16	38.80	7.43
Out-of-school activities																												
Reading/writing	0.77	0.34	1.17	0.33	1.06	0.34	1.05	0.40	0.24	0.72	0.68	0.37	0.91	0.37	0.54	0.56	0.81	0.43	0.62	0.34	1.64	0.33	1.14	0.43	1.20	0.34	0.08	0.40
Entertainment	0.99	0.36	0.89	0.31	0.63	0.32	0.60	0.31	1.27	0.54	0.70	0.37	1.10	0.34	1.08	0.47	0.50	0.38	1.22	0.32	0.93	0.30	0.94	0.30	1.79	0.40	1.95	0.31
Academic/cultural activities	-0.17	0.28	-0.07	0.32	-0.20	0.29	-0.86	0.36	0.45	0.48	-0.22	0.34	-1.06	0.36	-1.11	0.36	-0.29	0.37	-0.51	0.29	-0.91	0.29	-0.25	0.35	-0.23	0.37	0.52	0.35
Instructional climate																												
Class size	-2.83	4.65	-3.42	4.44	-0.12	4.37	4.70	6.13	-1.67	8.86	-0.45	8.26	1.18	3.91	13.85	7.07	16.31	6.75	1.04	6.20	1.27	5.16	4.60	7.39	-10.92	8.03	-1.43	5.98
Number of students requiring program modification	0.09	5.98	-9.48	4.08	-0.17	4.64	-12.76	5.28	-41.51	10.16	-4.81	5.46	-8.91	5.04	-7.80	9.56	-9.13	6.05	-8.91	5.52	-18.58	6.05	-15.57	5.77	-20.40	6.42	-12.30	6.27
Disciplinary climate	0.11	0.31	-0.46	0.32	0.06	0.28	0.10	0.29	1.15	0.63	0.48	0.34	0.37	0.32	-0.94	0.37	-0.06	0.27	-0.70	0.26	-0.31	0.29	0.02	0.29	0.08	0.32	-0.19	0.35
Time																												
Days absent from school	-9.19	2.81	-9.46	2.19	-3.01	2.72	-4.63	2.83	-13.49	4.28	7.70	2.99	-1.88	2.85	-3.21	3.33	-5.84	2.78	-4.35	2.36	-0.50	2.51	-3.50	2.36	-6.04	3.04	-3.56	2.96
Time on homework	11.15	2.72	4.87	2.53	-2.31	2.53	8.12	2.87	6.94	4.48	16.67	2.95	6.89	2.49	18.34	3.53	4.39	3.11	5.74	2.37	-5.27	2.18	4.16	2.44	0.96	2.42	2.59	2.93
Total instructional days lost	0.23	0.52	0.33	0.75	-0.71	0.39	0.12	0.60	-1.19	0.78	0.02	0.54	-0.22	0.55	-0.48	1.35	1.19	0.52	-1.57	0.61	0.18	0.82	-0.49	0.42	-0.01	1.04	-0.97	0.59
Influences on school programs																												
External influence on school	2.33	3.31	4.73	4.04	4.86	3.78	-2.98	4.96	8.16	6.23	12.24	5.04	-2.03	4.66	-5.23	8.40	-0.56	4.10	1.92	4.45	-8.10	7.24	-4.50	4.05	2.29	5.69	-2.39	4.86
Curriculum/teacher influence	-10.34	3.73	-8.66	4.84	2.21	3.44	4.51	4.27	7.08	6.58	4.99	5.91	0.58	4.29	-1.94	5.97	-2.44	3.05	1.88	4.99	3.07	4.82	-4.14	4.10	7.90	9.33	0.16	6.31
External assessment influence	-3.27	4.36	-3.25	4.55	4.97	4.00	12.94	5.00	-3.78	4.74	-0.65	3.89	1.42	4.66	3.25	7.30	-2.44	3.05	3.30	4.57	3.40	5.30	2.43	4.75	5.04	4.27	8.13	3.75

Note: Standard errors adjusted for finite population size. Yukon is omitted from these models because of small sample size.

Science: Full model																												
Variables	BC		AB		SK		MBe		MBf		ONe		ONf		QCe		QCf		NBe		NBf		NS		PE		NL	
	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE	Coeff	AdjSE
Demographics																												
Gender	0.33	6.10	8.99	6.71	-4.21	6.07	1.86	6.90	-8.71	11.18	0.12	6.69	8.67	7.72	-3.98	8.07	6.26	8.06	12.89	6.49	11.74	6.45	7.62	6.23	12.88	7.50	0.78	8.43
Home language other	-27.35	10.57	31.10	4.81	-50.80	13.12	-47.74	11.53	-22.71	34.19	-15.33	9.56	-61.22	14.27	-27.30	14.26	-48.21	15.25	-63.69	22.80	-54.63	14.55	-43.61	17.01	-74.05	27.25	-102.01	30.73
Grade level	18.30	6.21	-20.08	9.30	33.41	6.23	13.94	6.51	12.53	12.99	33.02	9.23	25.85	7.99	36.82	7.85	35.78	8.13	30.45	6.62	29.15	12.36	22.16	7.68	25.23	4.32	31.39	7.08
Born in Canada	16.11	10.94	12.28	10.82	-9.63	16.89	-4.91	12.37	-14.72	43.10	6.36	10.81	64.32	19.15	5.86	18.05	-26.53	15.98	-29.91	20.48	75.85	39.01	3.03	15.34	12.41	18.51	27.37	24.85
Number of books in the home	14.90	2.91	11.59	3.17	14.51	2.72	13.54	3.34	14.74	5.82	13.37	2.88	16.94	3.09	9.01	3.61	11.10	3.02	13.63	2.78	13.79	2.45	14.68	2.88	21.25	2.74	18.31	3.34
School demographics																												
School size	-9.99	5.44	0.91	6.18	-8.45	9.31	-5.31	5.57	47.71	18.59	-1.82	5.59	-0.18	7.97	11.30	6.94	2.26	4.27	7.95	4.99	-1.99	9.67	9.49	7.37	-15.25	8.61	6.36	10.63
Public school	-17.29	8.83		0.00		0.00	-7.49	12.71		0.00		0.00		0.00	-23.97	15.71	-8.57	12.72		0.00		0.00		0.00		0.00		0.00
Aboriginal students	2.08	5.54	-13.15	3.43	-10.74	3.94	-14.27	4.57	20.47	10.40	-2.05	6.42	-11.86	7.97	7.95	9.62	3.82	6.54	0.40	6.78	-2.97	9.97	-17.80	9.16	-2.82	7.11	0.23	6.58
Community size	8.86	3.53	-2.54	1.95	-2.29	3.16	-0.77	2.26	3.89	4.29	-2.48	2.87	6.42	3.24	-9.10	4.02	-4.67	4.82	-1.45	3.54	2.97	4.89	0.71	3.30	2.92	4.19	-0.38	5.86
Attitudes																												
Enjoys school	1.30	0.41	0.68	0.30	0.27	0.39	0.12	0.43	0.87	0.52	0.53	0.40	0.48	0.31	0.63	0.46	0.61	0.35	0.37	0.31	0.42	0.38	0.72	0.32	0.59	0.40	0.05	0.34
Attributions																												
Fatalism	-0.44	0.37	0.00	0.35	-1.33	0.34	-0.64	0.39	-0.94	0.68	-0.68	0.37	-1.36	0.38	-0.59	0.42	-0.41	0.36	-0.66	0.35	-1.01	0.30	-0.31	0.33	-0.91	0.35	-0.25	0.35
Internal attributions of success and failure	0.44	0.41	1.56	0.35	0.46	0.41	-0.23	0.42	-0.05	0.67	0.64	0.34	-0.39	0.34	1.41	0.37	0.13	0.37	0.92	0.30		0.00	1.23	0.33	1.20	0.44	1.39	0.40
Perceived performance on PCAP test	19.28	6.08	22.77	5.14	15.19	5.84	15.51	5.44	49.39	9.99	22.91	6.65	24.04	6.35	21.01	7.28	24.13	6.86	9.40	6.22	22.73	6.56	14.85	5.44	14.00	6.17	28.32	7.05
Out-of-school activities																												
Reading/writing	1.28	0.37	1.66	0.37	1.29	0.37	1.97	0.39	1.02	0.40	0.94	0.41	1.08	0.39	1.00	0.51	1.59	0.37	1.30	0.39	1.87	0.34	1.64	0.40	1.86	0.51	0.55	0.44
Entertainment	1.19	0.34	1.02	0.32	0.73	0.34	0.94	0.33	0.96	0.60	0.72	0.34	1.33	0.33	1.01	0.42	0.68	0.42	0.69	0.32	1.13	0.31	0.55	0.31	0.53	0.36	1.86	0.34
Academic/cultural activities	-0.37	0.29	-0.32	0.31	-0.46	0.31	-1.25	0.38	1.03	0.48	-0.61	0.32	-1.44	0.38	-1.07	0.42	-0.89	0.41	-0.67	0.35	-0.74	0.30	0.09	0.39	-0.20	0.25	0.20	0.39
Instructional climate																												
Class size	-5.32	4.65	1.73	3.78	2.17	4.28	0.55	5.11	34.54	11.99	5.22	6.58	6.71	4.57	12.70	5.91	22.15	6.33	13.16	6.33	4.36	4.76	18.54	6.38	-7.04	7.53	-1.68	6.81
Number of students requiring program modification	5.98	5.87	-4.56	3.71	-3.69	4.95	-6.76	4.71	-34.60	15.28	-6.86	4.96	-7.62	5.33	-11.53	7.76	-9.17	6.66	-4.62	5.65	-8.97	5.17	-8.40	5.06	-11.15	5.64	-5.41	5.62
Disciplinary climate	0.09	0.35	-0.20	0.32	0.05	0.27	-0.16	0.29	1.25	0.59	0.02	0.32	0.87	0.33	-0.45	0.40	0.22	0.34	-0.04	0.30	-0.32	0.30	0.35	0.34	0.39	0.45	-0.14	0.33
Time																												
Days absent from school	-0.58	2.70	-5.03	2.30	-3.81	2.59	-4.50	2.96	-0.08	4.79	-3.44	2.42	-0.37	2.92	-2.87	3.85	0.86	3.34	-0.67	2.59	-2.37	2.62	-2.47	2.79	-2.50	2.43	-7.85	2.72
Time on homework	6.07	2.61	3.78	2.71	-1.35	2.53	1.38	3.38	4.21	4.79	6.58	2.90	-1.97	2.70	8.37	4.00	-0.12	2.94	-0.63	2.26	-3.33	2.42	3.39	2.67	-6.20	3.25	0.62	2.69
Total instructional days lost	-0.78	0.51	-0.05	0.61	-0.90	0.48	0.50	0.44	-2.92	1.71	0.27	0.50	0.21	0.55	-0.44	0.97	0.71	0.47	-0.51	0.55	0.47	0.79	-0.27	0.47	-0.61	1.60	-0.13	0.72
Influences on school programs																												
External influence on school	-3.88	3.01	2.19	2.98	11.10	3.43	-1.74	4.29	8.39	11.24	9.13	5.19	3.81	4.09	-7.25	5.96	-1.52	4.12	1.53	3.97	-3.14	5.63	0.17	3.83	6.62	6.22	-9.32	5.14
Curriculum/teacher influence	-11.99	3.25	-3.78	3.86	6.04	3.26	4.39	3.84	15.00	7.36	0.34	5.42	1.75	3.87	5.56	5.40	-7.74	2.92	-2.00	5.25	0.09	4.58	-4.25	3.60	-7.48	8.58	-0.58	5.88
External assessment influence	2.44	3.80	-4.17	3.51	0.84	4.09	8.07	4.70	5.72	8.55	-1.78	3.94	4.41	5.50	-0.85	6.25	-2.30	4.31	-0.98	4.99	3.05	4.31	-0.33	4.09	-0.50	3.98	2.10	5.35

Note: Standard errors adjusted for finite population size. Yukon is omitted from these models because of small sample size.

Treatment of Missing Data

Almost all surveys have some missing data. Data may be missing because respondents to not complete a questionnaire or an item, or because the response is not readable or is outside the range for the item. In PCAP, there are no missing data for the performance indicators because only students who wrote the test are included in the data file.[17] However, for questionnaire variables, it is typical to find that 3% to 5% of the cases have missing data.

Several techniques are available for treating missing data. These include "pair-wise" or "list-wise" deletion of cases with missing values, replacement of missing values with the mean or median of the variable, nearest neighbour imputation, and single or multiple imputation using values predicted for the missing case through statistical modelling.

Case deletion is not a significant problem for descriptive statistics or simple correlations when the number of missing cases is small. Thus for PCAP, where less than 5% of data is missing for any one variable, missing data can generally be ignored for single-variable analysis. However, for statistical modelling, even if only a few cases are missing for any one variable, the number of missing cases increases exponentially as more variables are added to the model.

Mean or median replacement is a plausible technique where the number of missing cases is small. However, with larger amounts of missing data, this technique narrows the distribution of values and, thus, leads to bias in computing standard errors. For this reason, various imputation techniques have been devised that are intended to preserve both the central tendency and the standard error. The most common of these is statistical imputation, in which regression methods are used to estimate values for a particular variable of interest, based on other variables correlated with that variable. The missing values are then replaced by the predicted values.

Multiple imputation is a variation on that technique in which multiple estimates are created, separate data files produced with the estimated values at each stage, and the final analysis conducted on the separate files, with pooling of the standard errors at the end. While this technique is complex, it has the advantage of creating multiple "plausible values" for the missing values, thus capturing more of the variability inherent in the data. Most statistical packages, including SPSS and HLM (which are the packages used here), are capable of conducting multiple imputations and/or using these in subsequent analysis procedures to create composite estimates of parameters of interest.

Because multiple imputation has not previously been used in PCAP, it was proposed that the technique be explored in this study with a view to determining its feasibility and suitability for use in future PCAP assessments. Using the procedures available in SPSS, multiple imputations were therefore run on several variables, with a view to including the resulting files in the models to be run using HLM. What was found was unexpected,

[17] Of course, failure to write the test on the part of any student in the original sample is a form of missing data, which is captured by reporting the response rate. Adjustment for missing data is generally not appropriate for the outcome variables. However, non-response is an additional source of potential bias in the results.

in that the multiple imputation process did not yield imputed values for all of the missing data. Further exploration indicated that this was likely a consequence of patterns in the missing data. In particular, data missing for one variable were frequently missing for other variables used in the prediction equations, because respondents tended to omit questionnaire items in sections, where the sections correspond to scales. For example, on the assumption that the best predictor of attitudes is other attitudes, equations built with this in mind led to missing imputed values because no values were available for the predictors.[18]

This points to an underlying issue in multiple imputation, namely that it is appropriate only if data are missing completely at random (MCAR) or missing at random (MAR). The property of MCAR data is that the missing cases represent simply a random sample from the population. In this case, replacement of missing data by a set of random numbers with the same distribution as the non-missing data is sufficient to meet the imputation requirements. Data that are MAR have the missing data correlated with other variables in the data set so that it is possible to predict values using regression-type equations.

Multiple imputation is appropriate for data that are either MCAR or MAR but not for data missing in patterns as found in PCAP. From a modelling perspective, having to resort to case deletion to fill in missing values, even after the complex process of multiple imputation, is considered inappropriate because it does not fully address the problem of exponential increase in missing cases as the number of variables in the model increases. Overall, therefore, multiple imputation was considered to be not worth the effort in this case.

Nevertheless, since HLM is intolerant of missing data (particularly at level 2 of the model), some means had to be found to impute values or else a large number of cases would have had to be discarded before running the models. SPSS offers the usual range of alternative missing value replacement techniques, including replacement either by the series mean or median or by the mean or median of selected nearby values (a form of nearest neighbour replacement). In the end, the latter procedure was selected on the grounds that the nearest neighbours in the data file, sorted by school and jurisdiction, would be students in the same school or schools in the same jurisdiction. Hence the "neighbour" values would be, to some extent, predictors of the missing values. Also, this procedure yields a more favourable distribution than series mean or median replacement because some variation in the imputed values is created by averaging over selected cases rather than all cases.

Standard Error Estimation

PCAP, like most other large-scale assessments, employs a complex sampling design. In particular, stratification and clustering may be used to create a sample that has the desired properties (i.e., adequate samples of sub-populations or simplicity in selecting sampling units) but that does not meet the requirement of equal probability of selection of the sampled units. Thus, the assumptions typically used in computing standard errors for various statistics do not yield the best estimates of standard errors for complex samples.

[18] Predicted values for all cases could have been created by using the achievement measures as predictors because there are no missing data for these measures. However, this was judged inappropriate because the achievement variables were also the outcome variables to be predicted in the models.

> "**Bootstrap —** In statistics, bootstrapping is a modern, computer-intensive, general purpose approach to statistical inference, falling within a broader class of resampling methods." **/ Jackknife**: Jackknifing, which is similar to bootstrapping, is used in statistical inference to estimate the bias and standard error (variance) of a statistic, when a random sample of observations is used to calculate it. The basic idea behind the jackknife variance estimator lies in systematically recomputing the statistic estimate leaving out one or more observations at a time from the sample set. From this new set of replicates of the statistic, an estimate for the bias and an estimate for the variance of the statistic can be calculated./
> **BRR** = balanced repeated replication / IBM® **SPSS®** Statistics Standard, IBM SPSS Professional includes advanced statistical procedures to ensure the accuracy of your analyses and table features to help you better understand your data and easily report results.

Several techniques (Bootstrap, Jackknife, BRR) are now available for creating standard errors directly from the empirical data. This is typically done by repeated sub-sampling from the available sample and computing standard errors directly from the generated sampling distributions. The bootstrap method was used to compute standard errors for the reading scores in PCAP 2007, but was not used for other variables. It was proposed that, in this study, bootstrapping be explored as a means of estimating standard errors for all variables. Again, many standard statistical packages now provide bootstrap estimates. Because this is new in SPSS, one of the goals was to help users understand the SPSS procedure and to determine its strengths and limitations for more general application in PCAP.

Essentially, bootstrapping is one way of using the sample data to create empirical estimates of standard errors. Intuitively, this is similar to drawing separate random sub-samples from the available data and using the resulting sampling distribution of the statistic of interest to compute the standard error. Ordinarily, successive sub-sampling would result in sub-samples of a size much smaller than the initial sample. Bootstrapping overcomes this problem by successively selecting "with replacement" samples of a size that is the same as the original. If the original sample is large enough, replacement of every case as the procedure progresses ensures that many different, though overlapping, samples can be selected. The standard deviation of the sampling distribution of the statistic of interest is an empirical estimate of the standard error, which is free of assumptions about the nature of the sample.

Bootstrapping is now available for the first time in the most recent version (version 18) of SPSS. The SPSS procedure allows the analyst to select a large number of bootstrap samples (1,000 is the default) and to create composite estimates for various statistics and standard errors. This process is built into many of the common analytical procedures, which means that separate analysis of the bootstrap samples is not required.

As a first attempt at bootstrapping, mean mathematics and science scores and their bootstrap standard errors were produced for all categories of the variables used in this report for each of the jurisdictional/language groupings. Correlations and their standard errors between mathematics and science scores and the scores for the derived variables were also produced in the same way. The latter process was used to reduce the burden of dividing these continuous variables into categories (as was done in the *PCAP-13 2007*

Contextual Report). In each case, we used 100 samples because we found that this made little difference to the standard errors and required much less computer time than the default 1,000 samples.

These statistics are shown in Table B1 as a reference for those interested in examining detailed results within or across jurisdictions.

A long-standing issue in computing standard errors with weighted data in SPSS is that SPSS treats the sample size as if it were the full population size, thus yielding standard error estimates that are too low. The results presented in Tables B1 and B2 were therefore computed using unweighted data. This is not a problem for jurisdiction/language breakdowns because the weight within each group is constant (with some exceptions where a small number of minority-language students are included with the majority-language cases). However, unweighted data cannot be used when computing Canadian results because the weights vary widely across jurisdictions.

In the SAIP and PCAP studies, this was overcome by direct computations of the standard errors using correct sample sizes. However, this is not feasible for bootstrapped data because it would require manual computation for all the bootstrapped samples. Ordinarily, this can be addressed within SPSS by "normalizing" the weights so that the sum of the weights is equal to the sample size. Our exploration of this issue for the bootstrapping work, however, revealed that the bootstrapping procedure requires integer weights. In a normalized set of weights, many weights are less than 1, and rounding to the nearest integer obscures differences in the weights and creates zero weights. It was also found that a large amount of computing time is required to compute bootstrapped results with weights, presumably because the process is repeated many times for samples as large as the full population. Because no solution could be found for this within SPSS, none of the pan-Canadian results reported here use bootstrapped standard errors.

The modelling results could not be bootstrapped because that procedure is not available within the HLM program, and it is not feasible to run the models repeatedly. Although SPSS now also includes a "mixed models" procedure capable of running hierarchical models, that procedure also suffers from the weighting problem and proves to be extremely difficult to run with weighted data. It is possible that other statistical packages could be used to overcome this problem. It would be useful to explore this possibility before the analysis for PCAP 2010 commences. Alternatively, a more detailed examination of the differences between bootstrap standard errors and those computed using standard techniques could yield an adjustment factor (equivalent to the "design effect" sometimes used for complex samples) that might be used to avoid the need for using the bootstrap in all cases.

In several of the PCAP populations, the sample size is relatively small but happens to include all or most members of the population (a census or near-census). This is particularly true for several of the minority francophone population and is also the case for Prince Edward Island and Yukon. When the sample is close to a census, the standard error is smaller than is the case for the same samples drawn from a larger population (indeed, by definition, a census has a sampling error of zero).

In the SAIP and PCAP studies, a standard adjustment known as a "finite population adjustment" has been used in these cases. This adjustment is the factor shown below:

$$\sqrt{1 - \frac{n}{N}}$$

where n and N are the sample and population sizes respectively. In this report, all standard errors for jurisdiction/language groups are based on this adjustment. The adjustment is not required (although it does no harm because the result is close to 1) for the pan-Canadian results or in other cases where the sampling fraction is small. Again, the bootstrapping procedure cannot distinguish between the finite and infinite population situations so the adjustment has to be applied after completing that procedure. In practice, this is done by a simple computaiton in Excel.

TABLE B1 **Descriptive statistics by jurisdiction/language: Mathematics**

Student variables	Weight	61	48	16	18	6	190	9	21	144	8	4	14	4	9	5
Gender																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
Female	Mean	479	494	463	480	480	506	468	511	523	461	461	454	451	471	463
	SE	4.88	4.69	4.53	5.89	7.69	4.89	4.76	6.19	5.52	5.09	4.82	4.26	5.15	4.98	12.28
Male	Mean	487	501	461	480	474	510	473	510	513	463	458	461	448	482	442
	SE	4.72	4.78	4.20	4.58	10.57	4.68	4.83	6.48	5.56	4.99	4.73	4.45	5.35	5.61	12.64
Grade																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
Grade 6	Mean			309	460					352	413		460	383		
	SE			46.92	3.41					8.41	0.00		0.00	21.81		
Grade 7/secondary I	Mean	476	438	396	398	412	439	369	445	439	388	341	396	371	449	315
	SE	19.95	14.79	10.48	15.80	39.85	16.45	21.96	12.38	12.16	10.05	8.90	11.76	11.19	14.75	12.47
Grade 8/secondary II	Mean	480	489	463	479	464	508	465	512	524	459	456	459	451	470	454
	SE	3.47	3.66	3.54	4.67	6.78	4.34	3.86	5.05	3.56	4.28	3.73	3.46	4.23	3.89	10.68
Grade 9/secondary III	Mean	491	528	483	494	516	516	495	551	555	483	491	500	469	500	497
	SE	6.73	5.98	7.63	6.88	10.60	6.62	6.53	12.17	14.33	6.53	5.55	9.84	6.13	7.09	21.67
Grade 10/secondary IV	Mean	673			362			542	430	552						455
	SE	7.10			58.39			39.71	10.92	20.55						10.85
Were you born in Canada?																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
No	Mean	500	518	490	505	460	526	425	508	550	471	425	500	449	470	536
	SE	10.84	12.24	23.14	11.22	47.83	9.45	20.32	16.13	15.99	23.59	39.16	20.83	16.01	36.30	31.70
Yes	Mean	480	495	461	477	478	504	473	511	516	462	460	457	449	476	449
	SE	3.50	3.51	3.26	3.72	5.10	3.52	3.75	4.93	4.25	2.98	3.07	3.37	3.81	4.45	9.64
Other language spoken at home																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
No	Mean	479	495	463	483	480	505	473	513	521	465	464	459	449	479	452
	SE	3.31	3.65	3.30	3.69	5.76	3.73	3.55	5.19	3.49	3.48	3.46	3.22	3.24	3.78	10.06
Yes	Mean	496	510	441	464	450	518	447	493	492	404	405	434	450	416	493
	SE	8.31	8.89	16.01	10.38	25.50	7.63	11.39	16.22	12.27	14.90	9.54	15.28	15.83	22.34	22.81
About how many books are there in your home?																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
0–10	Mean	433	425	408	425	422	463	422	426	477	401	417	396	397	390	421
	SE	12.46	13.90	11.42	8.88	17.76	15.20	8.80	14.28	11.95	9.45	6.72	8.84	9.52	12.12	29.96
11–25	Mean	441	462	426	455	451	476	433	471	480	435	444	425	438	449	460
	SE	8.99	8.20	6.73	7.81	16.59	9.21	8.64	10.68	7.88	5.43	5.30	7.94	7.61	8.39	18.80
26–100	Mean	482	490	460	469	484	493	471	508	522	460	473	460	441	479	437
	SE	5.80	4.48	5.06	5.54	10.27	7.06	5.28	6.85	7.60	5.95	6.34	5.22	6.34	5.80	19.15

Student variables	Weight	61	48	16	18	6	190	9	21	144	8	4	14	4	9	5
101–200	Mean	491	507	475	509	478	519	484	514	542	476	481	456	470	501	438
	SE	6.20	6.42	6.24	6.81	13.15	7.67	7.33	10.17	8.01	7.58	8.98	6.08	6.40	8.09	19.56
more than 200	Mean	513	529	492	517	506	544	519	546	552	499	523	494	475	515	482
	SE	5.81	6.95	6.91	7.22	12.95	6.01	8.04	8.75	8.15	7.81	9.10	6.18	7.19	7.96	18.22
On average, how much time do you spend in total EACH WEEK on homework in all of your school subjects?																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
Less than 30 minutes	Mean	444	467	444	449	448	478	451	422	468	443	454	441	437	476	433
	SE	11.19	8.14	7.54	6.79	17.08	11.51	8.33	15.28	8.26	6.29	6.13	6.48	6.29	10.26	35.59
30 minutes to 1 hour	Mean	456	476	460	465	470	490	460	498	504	453	452	441	441	460	436
	SE	6.49	7.74	6.84	5.64	11.26	7.10	7.36	10.39	8.95	5.93	6.27	5.85	6.41	8.40	22.71
1–2 hours	Mean	477	490	467	480	477	497	459	493	525	464	465	462	452	470	453
	SE	5.78	5.51	6.28	7.03	12.96	6.65	7.38	8.56	7.39	6.42	8.06	6.14	10.58	7.62	10.38
2–3 hours	Mean	501	519	466	527	479	513	485	540	530	460	477	473	474	478	439
	SE	7.39	7.48	7.25	9.33	12.18	7.96	8.02	11.39	9.74	8.11	8.58	7.94	9.25	10.02	19.24
More than 3 hours	Mean	509	528	475	511	504	538	509	554	549	507	459	482	469	500	496
	SE	5.75	6.25	9.13	9.39	14.67	6.43	6.54	7.47	7.27	8.69	10.49	7.67	10.01	8.37	17.39
How many days have you been absent from school over this school year?																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
0–2 days	Mean	503	519	461	499	489	531	475	507	534	471	479	460	455	477	486
	SE	6.20	6.44	6.60	8.31	9.85	6.05	7.43	10.10	6.39	6.74	6.05	7.04	10.33	8.75	14.64
3–5 days	Mean	489	498	469	486	487	508	478	527	518	461	450	463	460	484	446
	SE	6.80	6.43	4.84	6.33	8.64	6.16	6.26	6.96	6.15	4.70	4.34	5.33	5.65	7.29	22.18
6–10 days	Mean	470	499	464	473	469	504	461	513	514	464	450	465	452	475	433
	SE	6.15	7.10	6.39	7.05	14.05	7.01	7.07	10.29	11.83	5.41	6.55	5.55	7.11	7.80	17.24
11–20 days	Mean	471	483	458	461	452	496	463	492	500	460	489	453	432	486	478
	SE	8.24	9.19	6.53	10.76	14.62	10.71	9.20	18.79	11.48	8.02	11.08	7.30	9.37	9.14	25.86
More than 20 days	Mean	444	469	439	449	471	459	461	457	456	438	389	425	416	443	462
	SE	13.16	9.11	14.43	11.79	31.45	11.97	16.11	17.63	20.84	13.19	16.12	9.27	11.64	10.63	32.36
How well do you feel you performed on the PCAP test you just completed?																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
not at all well	Mean	439	436	428	439	426	460	413	453	464	416	412	418	406	420	420
	SE	6.66	6.53	5.95	7.24	10.99	10.65	8.93	11.77	15.73	6.28	6.14	6.16	7.04	8.31	16.55
somewhat well	Mean	486	496	464	483	487	503	471	509	517	464	470	456	456	478	470
	SE	3.73	3.67	3.74	4.28	6.63	4.11	3.69	5.35	4.57	3.37	3.53	2.85	4.55	4.09	13.00
very well	Mean	541	565	494	523	552	549	520	547	558	506	483	530	487	531	463
	SE	9.56	7.22	8.13	10.34	19.59	6.93	9.83	10.18	9.55	9.01	11.01	10.03	10.76	11.19	38.41

School Variables	Weight	8.7	5.7	3.3	2.8	1.1	25.8	1.8	2.0	5.5	1.1	1.0	1.1	1.2	4.0	1.6
What is the total student enrolment in your school?																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
Less than 100	Mean	515	506	494	518	438	509	469	483	464	437	443	474	500	515	447
	SE	16.83	14.08	13.95	45.49	7.65	5.03	26.57	16.27	19.76	4.65	0.00	8.82	7.58	22.94	20.62
101 to 500	Mean	489	492	457	469	467	500	466	512	498	459	444	452	451	478	445
	SE	5.85	3.75	2.86	3.30	2.47	5.38	3.02	5.25	8.65	1.40	0.00	0.95	2.43	3.88	7.72
501 to 1,000	Mean	472	505	477	497	493	522	466	515	517	466	476	477	443	459	476
	SE	4.55	4.21	6.81	4.47	3.50	8.21	4.49	5.65	6.67	1.89	0.00	2.14	2.62	8.04	11.98
Greater than 1,000	Mean	493	518	541	498	564	508	534	499	533	466	497	434			
	SE	5.28	27.84	32.11	9.88	7.12	3.36	12.96	8.54	4.73	3.30	0.00	6.79			
Which of the following best represents the governing structure of your school?																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
private	Mean	520			538				539	552						
	SE	7.34			7.51				6.28	5.03						
public	Mean	477			472				495	499						
	SE	3.22			2.77				3.87	4.29						
Approximately what percentage of students in your school has Canadian Aboriginal ancestry (First Nations, Métis, Inuit)?																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
None	Mean	500	520	467	577		520	474	523	520	462	454	462	464	473	460
	SE	15.27	8.51	25.60	16.59	0.00	6.42	3.01	8.28	4.94	2.27	0.00	2.61	4.87	5.15	17.19
Less than 10%	Mean	490	503	474	492	477	504	465	511	519	461	471	458	446	483	422
	SE	4.12	3.58	3.51	3.78	2.74	4.30	3.91	4.40	5.10	1.22	0.00	0.97	1.80	5.85	15.48
10% to 25%	Mean	469	485	460	476	491	501	540	436		471	412	433		429	469
	SE	6.14	9.79	5.63	4.33	2.99	15.99	48.17	18.89	0.00	3.37	0.00	3.12	0.00	30.00	8.00
26% to 50%	Mean	479	431	440	466	418	472		442		427			462	435	422
	SE	24.86	19.52	9.56	7.12	5.49	13.24	0.00	17.13	0.00	5.59	0.00	0.00	11.33	21.70	16.27
More than 50%	Mean	491	430	376	388	454				462						
	SE	39.91	17.80	9.75	13.79	8.33	0.00	0.00	0.00	36.88	0.00	0.00	0.00	0.00	0.00	0.00
In what type of community, town, or city is your school located?																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
Rural community or small town (fewer than 5,000 people)	Mean	464	489	466	469	463	500	464	475	478	446	441	454	458	479	450
	SE	13.16	6.69	4.25	5.24	3.40	9.01	6.33	10.73	7.90	1.69	0.00	1.32	1.90	4.58	11.04
Medium-sized town (5,000 to 25,000 people)	Mean	468	470	454	464	398	521	458	520	516	480	465	462	436	468	454
	SE	7.89	6.76	5.25	6.56	13.31	12.45	3.86	7.10	8.60	1.84	0.00	2.23	3.12	8.78	9.68
Small city (25,000 to 100,000 people)	Mean	478	488	411	479	422	502	468	498	528	468	484	434	449	466	460
	SE	5.08	6.88	9.02	8.34	17.09	8.79	6.21	8.50	6.44	2.51	0.00	3.46	6.11	14.37	13.13

School Variables	Weight	8.7	5.7	3.3	2.8	1.1	25.8	1.8	2.0	5.5	1.1	1.0	1.1	1.2	4.0	1.6
Medium city (100,000 to 500,000 people)	Mean	487	532	470	492	454	494	477	492	517	457	483	476	428	476	
	SE	6.15	24.15	3.81	13.17	11.24	6.30	4.89	12.81	8.42	3.07	0.00	2.69	6.33	10.59	0.00
Large city (over 500,000 people)	Mean	503	515	467	490	488	522	484	520	535			446			
	SE	7.14	4.52	22.69	3.87	2.63	5.78	8.49	5.21	7.25	0.00	0.00	4.56	0.00	0.00	0.00
Total class size																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
15 or fewer	Mean	500	502	487	431	412	337	461	520	425	476	490	411	554	487	450
	SE	12.99	17.36	17.26	75.13	18.96	15.22	14.81	14.87	41.74	15.26	0.00	9.31	8.16	14.65	13.03
15–19	Mean	526	499	452	477	421	518	479	497	454	436	453	479	466	486	456
	SE	17.20	8.80	7.90	10.03	8.32	20.91	5.44	9.88	31.18	3.96	0.00	4.36	7.50	9.14	9.48
20–24	Mean	483	490	458	471	478	503	468	506	503	456	458	452	449	468	456
	SE	7.74	5.70	4.97	4.24	3.01	7.63	3.31	7.81	16.23	1.86	0.00	1.29	2.40	6.13	7.98
25–29	Mean	481	496	466	490	487	514	471	511	496	464	459	464	443	474	
	SE	3.59	4.23	4.18	3.49	3.20	4.14	5.01	5.91	7.15	1.53	0.00	1.63	2.25	6.98	0.00
30 or more	Mean	483	532	463	485	470	500	467	518	537	482	467	449		484	
	SE	8.15	8.99	8.64	12.30	6.88	8.76	11.18	6.49	3.86	6.18	0.00	3.71	0.00	9.40	0.00
Number of students requiring program modification																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
None	Mean	488	563	458	520	465	537	450	543	572	531	426	494	517	4.88	459
	SE	22.08	16.13	12.88	12.52	14.93	12.79	8.54	12.23	15.40	7.31	0.00	7.16	7.11	11.72	15.75
1–2	Mean	489	508	461	504	479	511	488	542	544	473	469	471	450	485	442
	SE	4.61	6.23	3.57	4.64	3.21	6.03	4.80	6.97	6.40	2.80	0.00	2.55	3.30	5.50	10.47
3–4	Mean	478	495	468	474	480	509	465	492	511	461	466	460	445	470	471
	SE	4.27	4.18	4.24	3.70	2.64	5.14	3.78	5.21	4.93	1.36	0.00	1.31	2.00	5.65	9.23
5 or more	Mean	488	486	440	448	338	495	459	490	481	448	423	448	461	445	415
	SE	14.75	6.06	9.18	6.80	13.17	8.36	6.44	7.27	13.45	2.08	0.00	1.46	15.34	15.52	20.06

Correlations: Student variables		BC	AB	SK	MBe	MBf	ONe	ONf	QCe	QCf	NBe	NBf	NS	PE	NL	YK
Disciplinary climate	corr	-0.03	-.071*	-0.01	-0.01	0.09	0.02	0.00	.160**	-0.08	-.076*	0.03	-0.02	0.00	-0.01	0.12
	SE	0.04	0.03	0.03	0.03	0.07	0.03	0.03	0.04	0.04	0.03	0.04	0.03	0.04	0.03	0.09
Enjoys school	corr	.201**	.194**	.108**	.125**	.222**	.123**	.117**	.210**	.204**	.132**	.112**	.155**	.193**	.161**	-0.01
	SE	0.04	0.03	0.04	0.04	0.06	0.03	0.04	0.04	0.03	0.03	0.03	0.03	0.04	0.04	0.08
Attribution of success and failure to luck	corr	-.121**	-.084**	-.194**	-.154**	-.204**	-.087*	-.199**	-.135**	-.253**	-.193**	-.187**	-.156**	-.136**	-.118**	-0.14
	SE	0.04	0.03	0.03	0.03	0.06	0.03	0.03	0.04	0.04	0.03	0.03	0.03	0.04	0.04	0.09
Internal attributions	corr	0.00	0.05	-0.01	-0.01	-.151*	-0.03	-0.06	-0.02	-0.07	.119**	0.04	.096**	.086*	0.08	0.08
	SE	0.03	0.03	0.03	0.04	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.03	0.04	0.04	0.11
Out-of-class reading/ writing	corr	.244**	.229**	.236**	.266**	0.12	.240**	.251**	.244**	.215**	.204**	.266**	.252**	.280**	.174**	0.13
	SE	0.03	0.03	0.03	0.03	0.05	0.03	0.03	0.04	0.04	0.03	0.03	0.03	0.03	0.04	0.12
Out-of-class entertainment	corr	.098**	.106**	.111**	.118**	0.12	.073*	.114**	.161**	0.07	.197**	.132**	.111**	.232**	.246**	0.15
	SE	0.03	0.03	0.03	0.03	0.06	0.03	0.03	0.04	0.04	0.03	0.04	0.03	0.03	0.03	0.10
Out-of-class academic	corr	.131**	.064*	0.06	0.05	-0.02	.115**	0.02	0.04	.084*	0.04	-0.01	.080*	0.07	.110**	0.04
	SE	0.03	0.03	0.03	0.04	0.06	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.11
Correlations: School variables		**BC**	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
External resources	Corr	-.024	.030	.055	.015	.009	.070	.061	-.049	-.114	-.011	.017	-.028	.045	-.028	-.030
	SE	.033	.034	.037	.034	.064	.038	.042	.044	.043	.035	.039	.038	.045	.040	.087
Curriculum/teacher	corr	-.050	-.096	.008	.016	-.072	.003	-.095	-.037	.045	.005	-.056	-.027	.044	-.007	-.034
	SE	.034	.031	.036	.038	.063	.043	.041	.045	.035	.037	.036	.036	.047	.037	.089
External assessment	corr	-.013	-.006	.071	.066	-.045	.041	.052	.018	-.018	.039	.122	-.026	.007	.062	.122
	SE	.030	.035	.032	.038	.070	.031	.038	.043	.038	.034	.039	.032	.050	.037	.087
Total days lost	corr	.087	.041	-.004	.038	-.076	.034	-.046	.042	.037	-.062	.107	-.090	.079	-.020	.119
	SE	.034	.032	.038	.036	.072	.036	.039	.042	.040	.037	.034	.031	.042	.041	.089

* significant at the 0.05 level
** significant at the 0.01 level

Table B2 **Descriptive statistics by jurisdiction/language: Science**

Student Variables	Weight	61	48	16	18	6	190	9	21	144	8	4	14	4	9	5
Gender																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
female	Mean	489	523	487	480	480	499	483	473	514	465	460	481	465	481	473
	SE	4.90	4.45	4.39	5.34	8.59	4.81	4.93	6.84	5.83	4.64	4.31	4.45	5.65	6.00	16.39
male	Mean	486	524	473	474	463	500	486	462	519	471	460	482	463	485	453
	SE	4.68	4.20	4.64	4.90	8.85	4.57	4.94	6.01	5.74	4.79	4.34	4.45	6.45	5.19	13.94
Grade																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
Grade 6	Mean			221	432					415	455	412	494	389		
	SE			31.07	46.88					13.57	0.00	184.73	0.00	42.77		
Grade 7/secondary I	Mean	456	479	416	411	436	444	398	401	442	395	391	431	396	445	412
	SE	17.77	13.41	10.94	15.19	45.24	17.88	21.12	13.55	13.30	11.32	12.98	12.57	10.69	14.01	18.21
Grade 8/secondary II	Mean	485	515	484	479	464	494	480	470	524	465	458	484	462	477	461
	SE	3.80	3.41	3.75	4.21	7.59	3.48	4.50	4.90	4.09	4.53	3.62	3.25	4.85	4.34	13.33
Grade 9/secondary III	Mean	497	552	493	480	500	520	504	496	535	489	478	506	490	506	513
	SE	7.56	6.29	7.94	7.33	11.68	6.57	6.66	12.88	12.19	6.08	5.59	9.36	6.65	7.68	19.16
Grade 10/secondary IV	Mean	660			545			498	382	543						
	SE	4.93			40.53			48.98	15.62	24.66						
Were you born in Canada?																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
no	Mean	465	502	458	454	493	485	413	455	543	472	417	474	433	446	471
	SE	10.73	11.36	20.21	10.85	37.70	8.31	16.34	17.57	15.90	22.88	33.67	23.40	23.02	34.97	38.14
yes	Mean	491	526	481	479	473	503	488	468	514	468	460	482	466	484	465
	SE	3.61	3.20	3.52	3.79	6.03	3.47	3.56	4.92	4.26	3.50	2.99	3.06	4.21	3.85	12.29
Other language spoken at home																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
no	Mean	495	528	484	485	475	505	491	471	521	471	464	484	468	488	468
	SE	3.54	3.28	3.38	3.82	5.70	3.68	3.86	4.72	4.48	3.53	3.11	3.06	4.12	3.68	12.40
yes	Mean	460	499	426	429	452	479	431	441	474	402	410	438	405	371	423
	SE	7.08	9.47	17.59	9.06	26.27	7.51	11.52	14.47	13.62	18.62	9.91	12.45	18.64	30.38	22.29
Books in home																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
0–10	Mean	434	446	424	405	427	451	433	412	473	417	424	412	385	397	422
	SE	12.09	15.60	12.03	11.09	16.50	13.30	9.45	13.59	10.74	10.28	6.43	12.67	10.77	10.86	21.32
11–25	Mean	436	497	448	456	433	463	456	441	480	435	450	443	445	450	466
	SE	11.12	8.33	8.34	9.28	14.95	7.87	7.92	10.22	8.35	6.45	5.76	8.07	10.01	7.67	20.08
26–100	Mean	483	507	473	473	471	488	487	463	523	467	467	480	453	488	455
	SE	5.25	5.28	6.08	5.73	10.49	5.83	6.11	7.50	7.42	5.99	6.04	5.30	6.35	5.08	29.83

Student Variables	Weight	61	48	16	18	6	190	9	21	144	8	4	14	4	9	5
101–200	Mean	507	539	495	497	482	515	486	471	530	482	483	488	482	511	471
	SE	5.94	5.85	6.84	7.76	12.05	5.92	8.17	8.62	8.80	7.39	7.82	6.13	8.62	7.99	28.90
more than 200	Mean	517	558	512	515	506	530	534	491	553	504	510	522	513	518	480
	SE	5.04	5.80	6.88	7.32	14.61	6.36	8.93	9.18	8.68	6.57	9.48	6.15	7.25	10.33	18.41
Weekly homework time																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
Less than 30 minutes	Mean	474	505	463	450	467	484	476	402	487	452	456	457	455	485	445
	SE	10.29	9.33	8.21	6.41	16.09	10.38	8.73	12.99	9.76	6.66	6.64	8.29	8.24	10.36	35.30
30 minutes to 1 hour	Mean	464	517	475	471	466	472	486	464	511	471	456	471	464	477	450
	SE	7.31	6.99	6.43	6.13	10.52	6.48	7.90	10.89	8.50	5.96	4.93	6.11	7.03	8.39	15.87
1–2 hours	Mean	477	513	482	480	470	500	471	455	516	469	454	485	457	471	455
	SE	5.73	5.65	5.88	7.08	12.75	6.55	6.02	8.23	6.64	6.21	6.80	6.11	9.00	8.77	20.34
2–3 hours	Mean	500	533	500	515	466	508	492	490	523	471	474	502	477	472	476
	SE	6.76	6.70	8.34	8.64	15.02	6.59	8.16	10.36	10.81	8.10	6.82	7.62	11.73	8.63	34.96
More than 3 hours	Mean	508	547	483	490	497	517	504	493	535	480	477	503	486	511	494
	SE	5.32	6.31	9.15	9.90	13.97	6.87	8.17	7.80	7.87	10.45	10.18	6.34	11.77	8.22	18.58
Days absent																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
0–2 days	Mean	485	526	476	481	477	504	481	462	526	473	473	481	467	490	488
	SE	6.15	6.90	6.70	7.21	12.95	6.54	7.18	8.26	6.75	6.64	6.50	7.93	10.80	8.60	17.90
3–5 days	Mean	488	530	483	493	480	502	495	476	511	466	455	490	473	498	453
	SE	5.39	5.37	6.03	6.24	9.80	5.62	6.90	8.50	6.11	5.82	4.43	5.09	6.69	7.77	24.35
6–10 days	Mean	492	524	486	472	452	500	478	477	517	475	454	488	465	475	457
	SE	7.61	6.82	6.53	6.59	14.48	7.31	7.72	8.71	12.03	6.30	6.75	5.36	7.97	8.81	17.07
11–20 days	Mean	491	523	484	456	502	499	476	459	505	464	477	472	446	480	483
	SE	9.46	8.75	6.23	11.24	20.11	9.82	8.29	16.09	14.00	8.57	9.86	8.39	10.91	8.94	19.78
More than 20 days	Mean	470	501	447	446	474	466	480	411	505	445	394	456	448	454	470
	SE	11.15	9.04	16.59	12.56	28.86	13.25	19.98	15.37	16.41	14.27	16.18	12.01	11.48	10.47	43.07
How well did you do on the PCAP test?																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
not at all well	Mean	466	486	455	449	421	468	452	433	478	441	423	456	423	445	467
	SE	6.86	7.76	6.55	6.56	11.83	10.10	8.84	11.28	12.31	8.04	6.88	7.53	9.27	9.99	21.27
somewhat well	Mean	486	522	484	479	485	497	482	462	515	468	467	481	473	482	465
	SE	4.18	3.92	3.97	4.45	6.72	3.84	3.82	5.20	4.39	4.16	3.38	3.27	4.61	4.68	14.98
very well	Mean	533	566	494	503	527	525	525	499	549	498	482	531	487	528	463
	SE	11.36	7.87	9.39	9.48	21.90	7.08	10.15	10.29	8.27	9.40	10.67	9.59	11.95	9.82	40.21

School Variables	Weight	8.70	5.70	3.30	2.80	1.1	25.8	1.80	2.00	5.50	1.10	1.00	1.10	1.20	4.00	1.6
Total student enrolment																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
Less than 100	Mean	514	530	499	538	394		460	435	459	456	443	524	482	468	460
	SE	19.22	12.26	15.63	30.24	12.33	0.00	26.65	17.33	20.11	5.16	0.00	4.80	10.54	19.99	44.64
101 to 500	Mean	494	522	478	475	472	494	483	463	508	460	452	477	470	483	454
	SE	7.23	3.52	2.89	3.27	2.40	4.55	2.79	5.76	6.44	1.28	0.00	1.12	2.54	4.02	10.68
501 to 1,000	Mean	485	530	487	481	491	490	483	471	517	473	466	495	455	484	489
	SE	4.89	4.61	6.95	5.17	4.04	5.61	6.05	4.58	6.70	1.95	0.00	1.87	3.04	7.05	8.07
Greater than 1,000	Mean	485	513	547	467	531	528	499	480	524	498	492	522			
	SE	5.64	24.61	45.51	9.62	17.40	6.94	14.16	9.04	5.32	3.27	0.00	10.98			
Governing structure of your school?																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
A private school	Mean	516			515				481	537						
	SE	6.46			7.20				5.92	5.42						
A public school	Mean	483	525	481	471	474	500	483	460	504	466	460	482	463	479	465
	SE	3.83	2.76	2.59	2.55	2.18	3.21	2.66	4.09	4.59	1.10	0.00	1.01	2.16	3.74	7.77
Approximately what percentage of students in your school has Canadian Aboriginal ancestry?																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
None	Mean	470	548	480	550		498	492	454	520	465	456	486	458	482	556
	SE	18.52	8.65	26.42	22.38	0.00	5.02	3.58	8.50	4.83	1.90	0.00	2.73	4.56	5.75	50.87
Less than 10%	Mean	492	527	489	493	466	502	473	474	514	467	469	484	464	489	458
	SE	4.27	3.06	3.35	3.28	3.01	4.17	4.02	4.14	5.17	1.55	0.00	1.08	2.29	5.95	16.24
10% to 25%	Mean	479	515	481	466	496	489	541	426		496	374	448		446	474
	SE	6.28	9.93	5.09	4.86	3.22	16.46	20.56	17.94		3.43		3.90		24.00	8.55
26% to 50%	Mean	496	499	463	467	445	521		435		417			463		
	SE	18.07	14.26	10.28	7.60	7.50	35.82		19.67	0.00	6.08			19.19	0.00	0.00
More than 50%	Mean	520	473	410	392	407				467					424	395
	SE	43.57	13.74	10.54	13.45	17.16	0.00			41.57					20.95	17.79

School Variables	Weight	8.70	5.70	3.30	2.80	1.1	25.8	1.80	2.00	5.50	1.10	1.00	1.10	1.20	4.00	1.6
In what type of community, town, or city is your school located?																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
Rural community or small town (fewer than 5,000 people)	Mean	479	530	485	479	451	501	476	450	501	453	445	477	474	481	463
	SE	15.58	6.84	4.88	5.00	4.59	8.86	7.15	9.74	9.84	1.54	0.00	1.40	2.60	4.33	16.96
Medium-size town (5,000 to 25,000 people)	Mean	483	518	474	462	410	520	481	485	511	488	468	487	447	484	476
	SE	9.59	6.56	6.14	7.84	10.58	11.08	4.75	7.64	8.46	1.98	0.00	2.47	3.82	7.70	8.79
Small city (25,000 to 100,000 people)	Mean	489	519	467	461	498	503	480	456	526	469	472	476	463	510	445
	SE	5.80	6.24	10.29	10.88	18.15	7.29	6.83	7.62	6.95	2.48	0.00	4.52	6.30	19.69	19.41
Medium city (100,000 to 500,000 people)	Mean	482	527	480	483	459	507	488	498	515	461	476	497	447	484	
	SE	5.56	26.06	4.03	13.32	7.03	6.13	5.27	10.85	8.69	3.36	0.00	2.76	6.99	13.62	0.00
Large city (over 500,000 people)	Mean	499	529	471	481	487	490	488	465	517			485			
	SE	7.25	4.27	24.20	4.06	2.30	5.62	7.25	4.83	9.16			5.07			
Total class size																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
15 or fewer	Mean	515	513	481	397	366	358	442	471	401	445	467	389	519	477	461
	SE	13.08	16.44	17.28	47.42	22.92	38.19	11.16	17.38	68.85	13.88	0.00	8.50	7.13	14.83	18.66
15–19	Mean	534	534	481	479	428	484	492	449	461	435	446	504	476	492	519
	SE	16.78	10.25	8.31	10.60	9.50	30.41	6.77	9.52	28.94	4.14	0.00	5.17	6.71	7.56	37.48
20–24	Mean	486	512	476	472	469	493	481	472	492	455	462	475	458	467	463
	SE	8.48	4.97	4.95	4.51	3.30	7.26	3.55	6.86	12.27	1.53	0.00	1.59	3.04	7.25	9.02
25–29	Mean	485	527	484	483	497	505	490	461	501	475	459	490	465	487	
	SE	4.10	3.95	4.41	3.67	3.47	4.20	5.34	5.31	7.49	1.61	0.00	1.25	2.92	6.59	
30 or more	Mean	483	551	478	468	452	495	483	484	533	482	471	489		521	
	SE	10.18	10.13	9.34	9.42	9.07	6.42	10.23	6.00	4.18	5.79	0.00	4.22		10.96	
Number of students requiring program modification																
		BC	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
None	Mean	503	563	468	518	431	542	482	490	579	459	458	483	498	496	452
	SE	24.84	12.97	15.19	9.12	21.60	14.58	8.15	7.27	18.39	4.50	0.00	7.30	8.89	13.90	12.56
1–2	Mean	489	529	486	498	475	508	494	493	533	475	468	492	461	486	445
	SE	4.83	5.77	3.47	4.77	3.77	5.98	5.01	7.12	6.31	2.63	0.00	2.85	4.38	6.42	13.60
3–4	Mean	486	526	477	466	478	497	480	454	510	469	462	483	462	479	493
	SE	4.69	3.92	4.27	3.90	2.56	4.87	3.85	4.87	4.59	1.30	0.00	1.43	2.32	4.91	8.93
5 or more	Mean	494	514	467	462	415	482	478	452	492	456	440	477	498	468	408
	SE	16.94	6.02	10.06	6.60	36.92	8.40	5.99	6.22	10.34	2.30	0.00	1.68	13.49	13.01	27.33

Correlations: Student variables		**BC**	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
Disciplinary climate	Corr	0.01	-0.03	0.00	-0.02	0.07	-0.01	0.04	-0.08	-0.02	-0.02	0.00	-0.01	0.03	0.00	0.09
	SE	0.04	0.03	0.04	0.04	0.05	0.03	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.10
Enjoys school	Corr	0.21	0.16	0.11	0.11	0.28	0.11	0.09	0.16	0.16	0.11	0.15	0.15	0.18	0.11	0.09
	SE	0.03	0.03	0.04	0.04	0.06	0.04	0.03	0.05	0.03	0.03	0.03	0.03	0.04	0.03	0.10
Attribution of success and failure to luck	Corr	-0.14	-0.11	-0.22	-0.16	-0.09	-0.14	-0.17	-0.12	-0.15	-0.16	-0.18	-0.13	-0.17	-0.08	0.02
	SE	0.03	0.03	0.03	0.03	0.06	0.03	0.03	0.04	0.04	0.04	0.03	0.03	0.04	0.03	0.08
Internal attributions to ability + versus luck –	Corr	0.05	0.14	0.03	0.02	-0.07	0.05	-0.05	0.11	-0.05	0.14	0.07	0.12	0.13	0.09	0.12
	SE	0.03	0.03	0.03	0.03	0.06	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.04	0.04	0.08
Out-of-class reading/writing	Corr	0.24	0.21	0.25	0.28	0.14	0.18	0.21	0.21	0.22	0.22	0.25	0.29	0.30	0.16	0.03
	SE	0.03	0.03	0.03	0.03	0.06	0.04	0.03	0.04	0.03	0.04	0.03	0.03	0.03	0.03	0.10
Out-of-class entertainment	Corr	0.11	0.13	0.11	0.13	0.09	0.09	0.13	0.16	0.11	0.14	0.16	0.06	0.12	0.21	0.12
	SE	0.03	0.03	0.03	0.03	0.05	0.04	0.04	0.04	0.04	0.03	0.04	0.03	0.04	0.03	0.11
Out-of-class academic work	Corr	0.07	0.02	0.05	0.02	-0.01	0.05	-0.04	0.00	0.02	0.01	-0.01	0.12	0.07	0.09	0.05
	SE	0.03	0.03	0.03	0.04	0.07	0.04	0.03	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.09
Correlations: School variables		**BC**	**AB**	**SK**	**MBe**	**MBf**	**ONe**	**ONf**	**QCe**	**QCf**	**NBe**	**NBf**	**NS**	**PE**	**NL**	**YK**
External resources	Corr	-.053	-.016	.083	.023	.005	.058	.074	-.069	-.077	-.035	.041	.005	.101	-.067	.126
	SE	0.03	0.03	0.03	0.03	0.02	0.04	0.02	0.03	0.03	0.01	0.00	0.01	0.02	0.03	0.07
Curriculum/teacher	Corr	-.071	-.050	.025	.005	-.022	-.049	-.079	.021	-.002	-.014	-.051	-.035	-.036	.003	.064
	SE	0.04	0.03	0.03	0.03	0.02	0.04	0.02	0.04	0.03	0.01	0.00	0.01	0.02	0.04	0.08
External assessment	Corr	.035	-.043	.025	.073	-.047	.050	.038	-.042	-.041	-.012	.084	-.030	.006	.012	.185
	SE	0.03	0.03	0.03	0.03	0.02	0.04	0.03	0.03	0.04	0.01	0.00	0.01	0.02	0.04	0.07
Days lost	Corr	.033	.009	-.027	.060	-.061	.062	-.002	.000	.020	.051	.063	-.062	.072	-.009	.025
	SE	0.03	0.03	0.03	0.03	0.02	0.03	0.02	0.03	0.04	0.01	0.00	0.01	0.02	0.03	0.07