Reading Achievement in Canada and the United States: Findings from the OECD Programme of International Student Assessment

Final Report

Learning Policy Directorate
Strategic Policy and Planning
Human Resources and Skills Development Canada

May 2004

SP-601-05-04E
(également disponible en français)

The views expressed in papers published by the Learning Policy Directorate, Strategic Policy and Planning are the authors' and do not necessarily reflect the opinions of Human Resources and Skills Development Canada or of the federal government.

■

This paper is available in French under the title *La capacité de lecture au Canada et aux Etats-Unis : Constatations issues du Programme international pour le suivi des acquis des élèves de l'OCDE.*
La version française du présent document est disponible sous le titre *La capacité de lecture au Canada et aux Etats-Unis : Constatations issues du Programme international pour le suivi des acquis des élèves de l'OCDE.*

■

Paper
ISBN : 0-662-37981-0
Cat. No.: HS28-3/2004E

PDF
ISBN : 0-662-37982-9
Cat. No.: HS28-3/2004E-PDF

HTML
ISBN : 0-662-37983-7
Cat. No.: HS28-3/2004E-HTML

■

General enquiries regarding the documents published by the Strategic Policy and Planning should be addressed to:

Human Resources Skills and
Development Canada
Publications Centre
140 Promenade du Portage, Phase IV, Level 0
Gatineau, Quebec, Canada
K1A 0J9

Facsimile: (819) 953-7260
http://www.hrsdc-rhdcc.gc.ca/sp-ps/arb-dgra

Si vous avez des questions concernant les documents publiés par la Direction générale de la politique sur l'apprentissage, veuillez communiquer avec :

Ressources humaines et Développement des compétences Canada
Centre des publications
140, Promenade du Portage, Phase IV, niveau 0
Gatineau (Québec) Canada
K1A 0J9

Télécopieur : (819) 953-7260
http://www.hrsdc-rhdcc.gc.ca/sp-ps/arb-dgra

Acknowledgements

The author is grateful for helpful comments on earlier drafts of this paper from Satya Brink and Urvashi Dhawan-Biswal. He is also appreciative of support received from the Social Sciences and Humanities Research Council for his position as the UNB Canada Research Chair in Human Development.

Table of Contents

List of Figures

List of Charts

1. Introduction

In the spring of 2000, Canada and the United States were among 32 countries that participated in an international comparative study of youth literacy skills, the Programme for International Student Assessment (PISA). PISA is a collaborative effort of member countries of the Organisation for Economic Cooperation and Development (OECD). Its aim is to assess how well 15-year-old youth are able to use the knowledge and skills they have acquired to meet the challenges facing them as they approach completion of their secondary schooling. PISA entails extensive testing of youth in their reading, mathematics and science literacy skills. It also includes questionnaires administered to students and school administrators aimed at collecting information on a wide range of family and school factors pertaining to the development of literacy skills. The content of the tests and questionnaires is developed by scientific experts from member countries and guided by the governments of participating countries based on their shared policy-driven interests. The PISA surveys are scheduled to be conducted every three years. PISA 2000 focused on reading literacy, with mathematics and science treated as minor domains, while in PISA 2003, mathematics was the major domain. PISA 2006 will emphasize scientific literacy, and then the cycle will be repeated, starting again in 2009.

The international findings of PISA were reported in *Knowledge and skills for life: First results from the OECD Programme for International Student Assessment (PISA) 2000.* The report provides comparisons of the performance of 15-year-olds in their literacy skills among the 32 countries, and an analysis of how literacy performance is related to students' family background and the schools they attend.

Canadian youth fared considerably better than their counterparts in the US. The average reading performance for Canadian youth was 534, compared with 504 for US youth. In mathematical and scientific literacy, the performance gaps were similar: the mean scores in mathematical literacy were 533 and 493 for Canada and the US respectively, while the scientific literacy scores were 529 and 499. The test scores for each of the PISA tests were scaled to have a mean of 500 and a standard deviation of 100 for the 28 OECD member countries that participated in PISA 2000. Thus, the Canada-US literacy gaps range from 30 to 40 points, or 30 to 40% of a standard deviation. This is a sizeable difference, equivalent to nearly one full year of schooling.[1]

These findings are consistent with earlier findings based on international comparative studies. For example, the difference between literacy scores in Canada and the US based on data from the 1994 International Adult Literacy Survey (IALS) and the National Adult Literacy Survey (NALS) for youth aged 16-25 were about 15% of a standard deviation for prose and document literacy, and 25% of a standard deviation for document literacy (Willms, 1999). The tests used in the IALS/NALS are in many respects comparable to

[1] In several of the countries that participated in PISA, 15-year old students spanned two grade levels by virtue of the month in which they were born. Willms (in press) estimated the "grade effect" for PISA reading scores with a multilevel analysis (students nested within schools nested within countries), using data for 12 countries where it was possible to distinguish between students who had likely repeated a grade from those who were on schedule in their school career. On average, the grade effect was 34.3 points (standard error = 3.5).

those used in PISA. These tests were designed to assess the knowledge and skills required in everyday life, rather the degree to which students had mastered a specific curriculum.

In contrast, the tests used in the Third International Mathematics and Science Study in 1994 (TIMSS) and 1999 (TIMSS-R) were designed to reflect what students were taught and learned in school. The TIMMS and TIMMS-R also differ from PISA in their sampling strategy. TIMMS and TIMSS-R select students at particular grade levels rather than students of a particular age. A particularly important comparison with respect to PISA results is the performance of students who were tested in TIMSS in grade 4 in 1994/95, and those who were tested in grade 8 in 1999. The majority of Canadian and US 15-year old students participating in PISA in 2000 were in grades 9 and 10 in the spring of 2000, and thus would have been in grades 8 and 9 in 1999, and in grades 4 and 5 in 1995. To a large extent, therefore, the grade cohorts of students tested in TIMSS overlap with the PISA age cohort of 2000. At grade 4 in 1994/95, Canadian students lagged behind US students in both mathematics and science: the average scores in mathematics were 532 and 565 for Canada and the US respectively (Mullis *et al.*, 1997), and 549 and 565 in science (Martin *et al.*, 1997). However, at grade 8 in 1999, Canadian students fared better than US students: the average grade 8 mathematics performance was 531 for Canada and 502 for the US, while the average science scores were 533 and 515 (National Center for Educational Statistics, 2003). The 1994/95 TIMSS results also showed a Canadian advantage at grade 8 in mathematics – the mean scores for Canada and the US were 527 and 500 respectively (Beaton *et al.*, 1996a) – but not in science, for which the differences were insignificant: 531 for Canada and 534 for the US (Beaton *et al.*, 1996b).

These results suggest that US children fare better than Canadian children in their early mathematical and science literacy development, at least through to grade 4. Thereafter, it seems that Canadian students make better progress. Note that the Canada/US differences observed in the 1999 TIMMS-R were 29 points for mathematics and 18 points for science, which are remarkably close to those observed in PISA for mathematics and science, especially given that the two studies used a different kind of test and a different sampling technique. Also, the differences between each country's average score and the international mean, set at 500 in both studies, is also remarkably similar, even though there was a different set of countries participating in each study.

There are many plausible explanations for the observed differences in literacy scores between Canada and the US. An important point is that these international surveys are cross-sectional, and provide estimates of the literacy skills at a particular age or grade level. The indicators represent the knowledge and skills that have been accumulated since birth, and as such reflect not only what has been learned at school, but also at home and in the community. They also reflect what is learned during the pre-school years, as well as the elementary, middle, and secondary school period. PISA can shed some light on why students' literacy outcomes differ in the two countries, as the data include considerable information about students' family backgrounds and their experiences in secondary school. PISA also provides a rich source of data for examining the distributions of student literacy skills within each county, and how this is related to students' background and the schools they attend.

The aims of this study are to:

- Examine the distribution of scores in Canada and the US, overall and at the student and school levels;
- Estimate the socioeconomic gradients associated with reading performance in Canada and the US, and examine the relationship between reading performance and socioeconomic status within and between schools;
- Examine the variation among schools in Canada and the US, and the relationships between reading performance and socioeconomic status; and
- Compare family and schooling inputs and the reading performance for different sectors of schools in Canada and the US.

The next section examines the distribution of student achievement in the two countries. The two sections that follow are concerned with the relationship between literacy skills and socioeconomic status, and the manner in which students from differing socioeconomic background are distributed within and among schools. Two devices are used to address these issues and explain their relevance to educational policy. One is the "socioeconomic gradient", which displays the relationship between literacy skills. The socioeconomic gradient for a schooling system can be partitioned into a *between-school gradient* that summarizes how average literacy skills for the country's schools are related to their average socioeconomic intake, and an *average within-school gradient* for the country's schools. The relative importance of these two components of the gradient has implications for the types of reform that are likely to be most effective. The second device is the "school profile", which shows the distribution of the average literacy skills for each school and their average socioeconomic composition. A number of research studies, including PISA, have shown that the average SES composition of a school has an effect on a student's achievement, over and above the effects associated with the student's own family SES. Thus, if a student of average SES were to attend a high SES school, his or her achievement would likely be higher than if he or she were to attend a low SES school. The results indicate important differences between Canada and the US in how students with differing socioeconomic backgrounds are distributed among schools, and therefore contextual effects particularly germane to educational policy. The fifth section examines the effects of particular school policy and practice variables, and the sixth section examines differences among rural, urban, and private sectors in each country. The final section provides a summary of the findings and discusses their policy implications.

2. The Distribution of Reading Scores in Canada and the US

Table 1 displays the results for Canada and the US, alongside the norms established by the OECD member countries. In PISA 2000 countries were required to sample at least 150 schools (if this number existed) at the first stage of a 2-stage stratified sampling design. At the second stage, 35 students were selected with equal probability from a list of the 15-year old students in each of the sampled schools. In most countries, therefore, the sample size comprised about 5,000 students. In the US, 3,700 students were assessed. In Canada, data were collected from a considerably larger sample, 29,461 students, in order to provide detailed information at the provincial level. Consequently, statistical estimates for Canada tend to be more accurate than those for the US.[2]

Table 1
Mean, standard deviation, and skewness on the combined reading literacy scale for Canada, US, and OECD countries (PISA 2000)

	Mean (SE)	Standard Deviation (SE)	Skewness (SE)
Canada	534 (1.6)	95 (1.0)	-0.26 (0.04)
United States	504 (7.0)	105 (2.7)	-0.24 (0.05)
OECD	500 (0.6)	100 (0.4)	-0.33 (0.01)

Note: The Canada-US difference in mean scores is 30 points, with a standard error of 7.2.

The results also indicate that reading performance in Canada is less variable than the US: the standard deviation of reading scores in Canada is 95, 5 points lower than the OECD standard deviation of 100, while in the US it is 5 points higher at 105. Also, in both Canada and the US, the results are skewed: the measure of skewness is -0.26 for Canada and -0.24 for the US. This measure indicates that there are disproportionately more students with very low scores relative to the mean than above it. However, the degree of skewness is less than that of all OECD countries.

The scaled scores in PISA were divided into five proficiency levels: level 5 (above 625), level 4 (553 to 625), level 3 (481 to 552), level 2 (408 to 480), and level 1 (335 to 407). Students at a particular level can typically answer about one half of the questions associated with that level, and can usually demonstrate the proficiencies associated with lower levels. Some students score below 335, the lower threshold for level 1. These students cannot be considered "illiterate"; however, they are likely to have serious deficiencies in their ability to use literacy in everyday activities.

[2] *Standard errors* reflect the degree of uncertainty in statistical estimates. For a particular sample statistic, one can infer that the corresponding population result would fall within a *confidence interval* of approximately plus or minus two standard errors of the sample statistic, in 95 out of 100 replications of the cases for different samples drawn from the same population. In PISA, because of the complex sample design, the standard errors are estimated using a procedure called Balanced Repeated Replicates (Rust & Rao, 1996).

Table 2 shows the percentages of students in Canada and the US who scored at each of the proficiency levels. About one-half of the students in each country scored at levels 2 and 3. However, there was a higher proportion of Canadian students scoring at levels 4 (27.7%) and 5 (16.8%) than in the US (22.3% and 9.5% respectively). Nearly 10% of Canadian students scored at level 1 or lower, while in the US 17.9% were at these levels.

In Canada, the threshold between levels 3 and 4 may be particularly important. Willms and Flanagan (2003) used data from the 1984 International Adult Literacy Survey (IALS) to examine the relationship between enrollment in post-secondary education and literacy scores for youth aged 19 to 25. The analysis included controls for age, sex, and the educational level of the respondents' parents. The odds of attending post-secondary education for youth who were in the bottom two quintiles of the literacy skill distribution were less than 20% of the odds for those in the top two quintiles. The odds for youth in the third quintile were about 63% of the odds for those in the top two quintiles. Although access to post-secondary has changed considerably over the past decade, their findings emphasize the importance of high literacy skills.

Table 2
Percentage of students at each level of proficiency on the combined reading literacy scale (PISA 2000)

	Canada		United States		OECD	
	%	(SE)	%	(SE)	%	(SE)
Level 5 (> 625)	16.8	(0.5)	12.2	(1.4)	9.5	(0.1)
Level 4 (553 to 625)	27.7	(0.6)	21.4	(1.4)	22.3	(0.2)
Level 3 (481 to 582)	28.0	(0.5)	27.4	(1.3)	28.7	(0.2)
Level 2 (408 to 480)	18.0	(0.4)	21.0	(1.2)	21.7	(0.2)
Level 1 (335 to 407)	7.2	(0.3)	11.5	(1.2)	11.9	(0.2)
Below level 1 (< 335)	2.4	(0.3)	6.4	(1.2)	6.0	(0.1)

3. Socioeconomic Gradients

Socioeconomic gradients depict *"the relationship between a social outcome and socioeconomic status for individuals in a specific community"* (Willms, 2003). The construct, socioeconomic status (SES), is defined as the relative position of a family or individual on an hierarchical social structure, based on their access to, or control over, wealth, prestige, and power (Mueller & Parcel, 1981). In many education and health surveys, it is operationalised as a composite measure of income, level of education, and occupational prestige (Dutton & Levine, 1989). Socioeconomic gradients are a useful tool for informing social policy because they call attention not only to the levels of performance for learning, behavioural, and health outcomes, but also to inequalities in outcomes associated with SES. A socioeconomic gradient is comprised of three components: the level, which is defined as the expected score on the outcome measure for a person with average SES; the slope, which indicates the extent of inequality attributable to SES; and the strength, which refers to how much individual scores vary above and below the gradient line.

Figure 1
Socioeconomic gradients for Canada and the United States

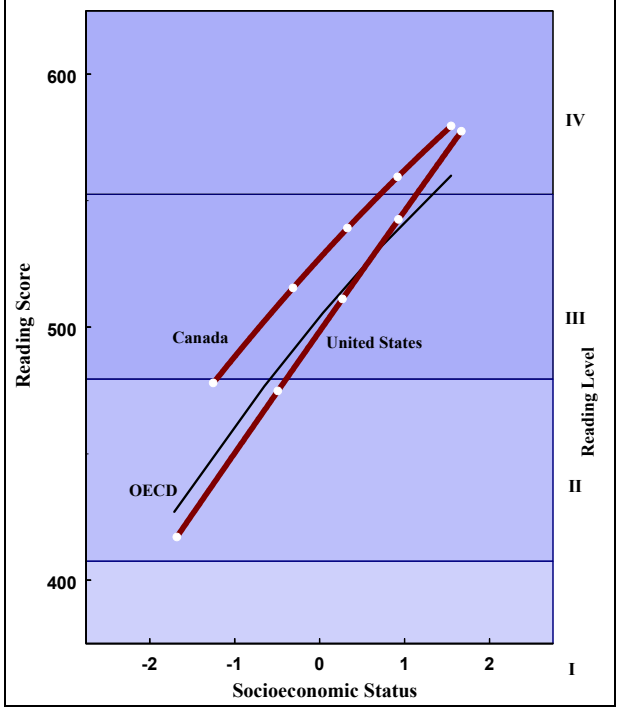


Figure 1 shows the socioeconomic gradients for Canada and the US alongside the pooled gradient for the 28 OECD countries that participated in PISA 2000.[3] The left-hand Y-axis is the PISA reading score scaled to have a mean of 500 and a standard deviation of 100 for OECD countries. The right-hand Y-axis indicates the reading levels. The X-axis is the measure of socioeconomic status developed for PISA, which describes students' economic, social, and cultural background. It was derived from data describing parental education and

[3] The socioeconomic gradients are derived with a simple linear regression within each country, regressing reading scores on the measure of socioeconomic status, and socioeconomic status squared:

$$Y_i = \beta_0 + \beta_1 SES_i + \beta_2 SES^2{}_i + r_i,$$

where Y_i is the outcome measure, reading performance, β_o is the intercept, β_1 and β_2 are regression coefficients pertaining to the slope of the gradient, and r_i are student-level residuals. A two-level multi-level model, with students nested within countries, yields virtually identical results, as the within-country sample sizes are relatively large. The quadratic term is included because the gradient is non-linear for Canada and for the overall OECD gradient. It was very small for the US, and not statistically significant. The average gradient across all OECD countries was estimated using a two-level multilevel statistical model, with students nested within countries (e.g., see Bryk & Raudenbush, 2002).

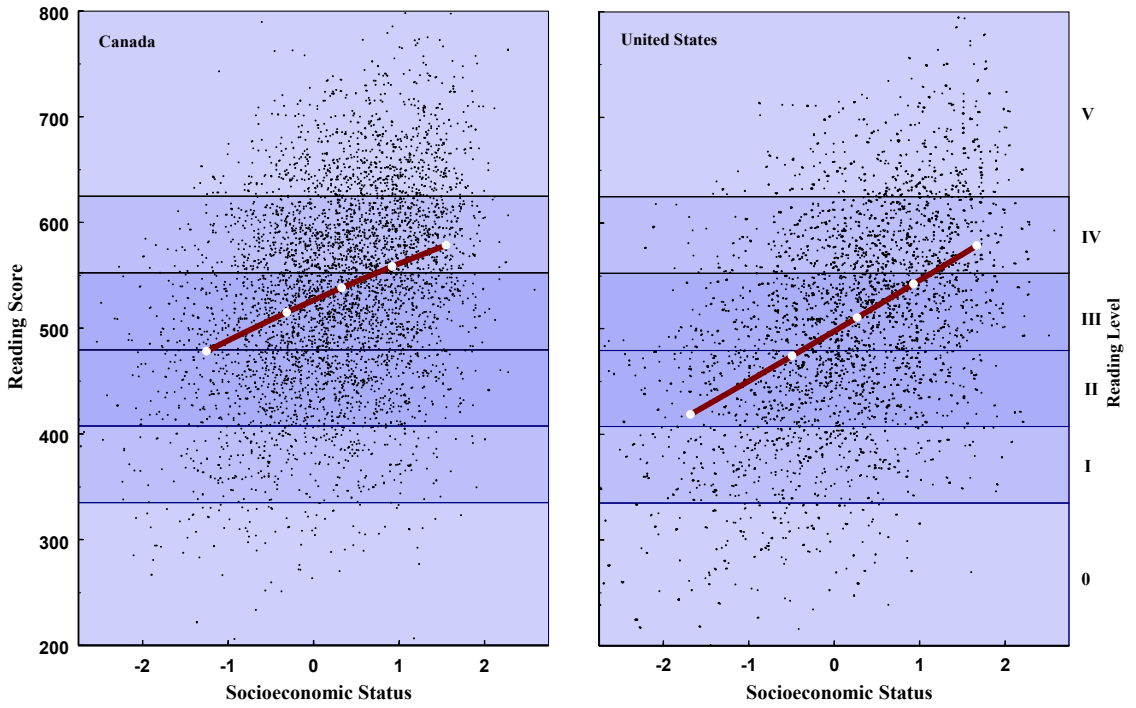
occupation, and the material, educational and cultural possessions in the home. It was scaled to have a mean of zero and a standard deviation of 1.0. For each country, the gradients are drawn from the 5th to the 95th percentiles of SES, and the small white dots on the gradient indicate the 5th, 25th, 50th, 75th, and 95th percentiles of SES. This is done to provide some indication of the range of SES in each country.

If we consider a hypothetical student with average SES (a score of zero), his or her expected reading score in the US would be 498, while in Canada it would be 527. This is the level of the gradient. The slope of the gradient is relatively steep in the US (47.9) while it is relatively gradual in Canada (36.9). This is perhaps the most striking difference evident in Figure 1. It shows that youth from relatively affluent backgrounds do not differ substantially in their performance in the two countries, whereas youth from low SES families fare much better in Canada then in the US. The *strength* of the gradient refers to how much individual scores vary above and below the gradient line. If the relationship is strong, then a considerable amount of the variation in the outcome measure is associated with SES, whereas a weak relationship indicates that relatively little of the variation is associated with SES. The most common measure of the strength of the relationship is a statistic called R-squared, which is the proportion of variance in the outcome measure explained by the predictor variable. The socioeconomic gradient for reading is stronger in the US than in Canada: 0.212 compared with 0.112. This difference is discussed further below, with reference to Figure 2.

Figure 2 displays the socioeconomic gradients separately for each country. These graphs also portray the scores of a representative sample of students. The results in Table 2 above indicated that over one-quarter of Canadian students, and nearly 40% of US students had scores at level 2 or lower. These graphs show that in both countries there are youth at all levels of SES with reading performance at these low levels. Although there is a disproportionate number of poor readers among low SES students, there is a substantial number of poor readers from average and high SES families. An important policy implication of this finding is that programs that are targeted towards youth from low SES families do not serve many youth who could benefit from assistance. Targeted programs need to be targeted on the basis of literacy performance, not SES.

The figure also shows that there are many youth in each country from poor backgrounds who have relatively high performance – at levels 3 and 4 for example. However, in both countries there are relatively few students from low SES families who had scores at level 5.

Figure 2
Variation in student reading performance in relation to socioeconomic status in Canada and the United States



In both Canada and the US, girls have considerably higher reading scores than boys. The difference is about 32 points in Canada and 29 points in the United States. Figure 3 shows the means scores and standard errors for each country. It is interesting to note that the average score for boys in Canada is comparable to that of girls in the United States. When tested for sex-by-SES interactions, in both countries these were not statistically significant. Thus, the slopes of the gradients for boys and girls are similar within each country. This suggests that the classroom and school factors that contribute to the level and slope of the gradient in each country probably have similar effects for boys and girls. The next section examines differences among schools in their reading performance.

Table 3
Differences between females and males in Canada and the US on the combined reading literacy scale (PISA 2000)

	Canada Mean (SE)	United States Mean (SE)	Difference
Females	551 (1.7)	518 (6.2)	-33 (6.4)
Males	519 (1.8)	490 (8.4)	-29 (8.5)
Difference	32 (1.6)	29 (4.1)	

4. School Profiles

Figure 3 shows school profiles for reading performance in Canada and the United States. These are scatter-plots of school mean reading achievement plotted against school mean SES. They are useful in that they indicate the range of school performance at varying levels of SES.[4] The graphs also show the between-school regression line for each country. The first panel of Figure 3 provides the results for Canada. It shows that there is a wide range in school performance at all levels of SES. At any particular level of SES, there is a range of about 120 points between the lowest- and highest-performing schools. More precisely, results of a multilevel analysis indicate that the standard deviation of the SES-adjusted school means is 30.6, and therefore about 95% of the schools would fall within 61.2 points (+ or – 2 SDs) of the regression line.

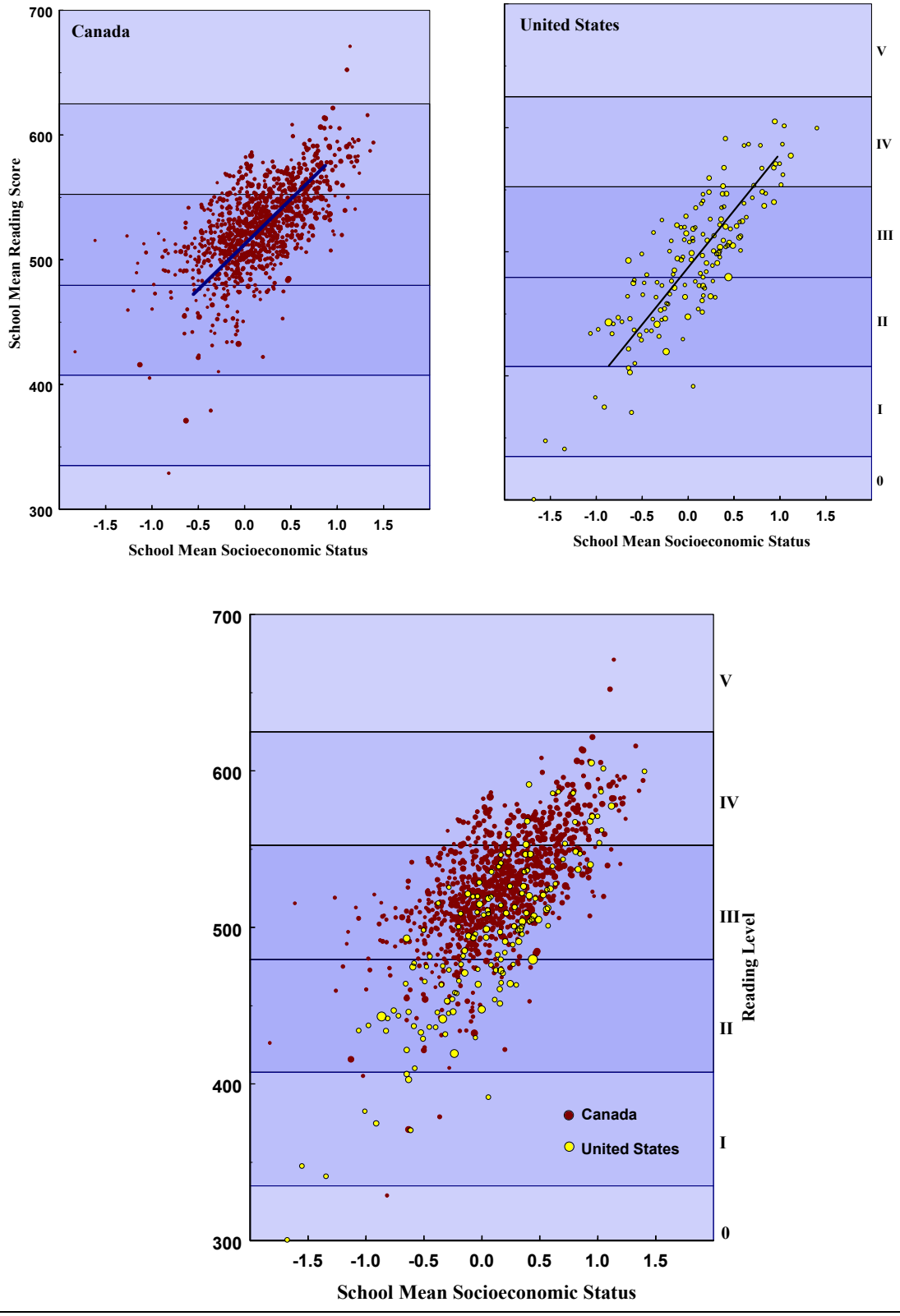
The results for the US, shown in the middle panel of Figure 3, also indicate that there is a wide range of performance at all levels of school mean SES, and the range is fairly consistent at each level of SES. The standard deviation of SES-adjusted school means is 26.6 points, and therefore there is a range of about 106 points between the lowest- and highest-performing schools.

The third panel in Figure 3 overlays the school profile for the US onto that of Canada. It shows that the Canadian advantage in reading performance is largely attributable to the performance of schools serving students of average and below-average SES. The results in Figure 3 also show that the relationship between school mean reading performance and school mean SES is steeper in the US than in Canada.

Socioeconomic gradients, like those presented in Figures 1 and 2, can be decomposed into within-school gradients and a between-school gradient. The distinction is important, because if the within-school gradients are relatively steep compared with the between-school gradient, it indicates that there are large inequalities among students within schools. This would call for classroom- and school-based policies that attempt to improve the performance of students within schools, particularly those from low SES backgrounds. For example, schools with steep gradients may wish to review the processes by which students are allocated to classrooms and school programs, as the segregation of students among classrooms within schools tends to result in steep gradients. Such schools might also try to strengthen their programs aimed at improving the performance of students with poor reading skills. In contrast, if the between-school gradient is relatively steep compared with the within-school gradients, it indicates that schools with low average SES intakes tend to fare poorly compared with schools that predominantly serve high SES students. This is often the case in systems where students are segregated among schools, either because of residential segregation or because of certain structural features of schools.

[4] The estimates of school mean reading achievement are estimated with a hierarchical linear regression model that differentially "shrinks" the estimate towards the grand mean (Raudenbush & Bryk, 2002). The shrunken estimates have been adjusted for sampling and measurement error. These provide a slightly conservative portrayal of the extent to which schools vary in their performance, but are considerably better than a description based on unadjusted means (see Raudenbush & Willms, 1995).

Figure 3
School profiles for reading performance in Canada and the United States



The decomposition of the overall SES gradient into within-school gradients and a between-school gradient can be expressed as a function of the between-school slope, the average within-school slope, and η^2, which is a measure of the extent of between-school SES segregation:

$$Overall\ Gradient\ Slope = \eta^2\ (Between\text{-}school\ Slope) + (1-\eta^2)(Within\text{-}school\ Slope)$$

where η^2 is the proportion of variation in SES that is between schools (Alwin, 1976). The index, η^2, can theoretically take on values between zero and one, or as a percentage between 0 and 100, but even in highly segregated school systems it is rarely above 0.6 or 60%. When η^2 is zero, there is no segregation among schools; that is, all schools have the same distribution of SES. Among countries that participated in PISA in 2000 and 2002, η^2 ranged from 11.6 (Norway) to 47.5 (Chile).

Table 4
Socioeconomic gradients on the combined reading literacy scale, and SES segregation for Canada and the US (PISA 2000)

	Canada	United States
	Estimate (SE)	Estimate (SE)
SES Gradient	36.5 (1.3)	47.8 (2.6)
Within-school Slope	27.8 (1.0)	28.9 (1.9)
Between-school Slope	72.5 (3.2)	91.8 (4.8)
SES Segregation Index (η^2)	19.5%	28.1%
Contextual Effect	44.9 (3.4)	63.4 (5.4)

Table 4 shows the decomposition of the SES gradient slope into within and between-school components. The average within-school slope is 27.8 for Canada, and 28.9 for the United States. This is a relatively small difference, and is not statistically significant. The two countries do differ significantly, however, in their between-school slopes: these are 72.5 for Canada and 91.8 for the United States. The SES segregation index, η^2, which is also shown in Table 4, is considerably larger in the United States (28.1%) than in Canada (19.5%). Generally, greater between-school SES segregation is associated with lower overall performance and steeper socioeconomic gradients, which is evident in these comparisons. The deleterious effects of segregation can be ameliorated through policies aimed at bolstering the achievement of schools with low average performance, or through policies that directly attempt to reduce SES segregation, such as redrawing school catchment boundaries or offering high status school programs in low SES areas to attract a representative mix of students.

Table 4 also provides an estimate of the "contextual effect" of school mean SES on students' reading achievement. In both Canada and the United States, the contextual effect is large and statistically significant. Consider a Canadian student who has an average SES (a score of zero on the international scale) and attends a school with a mean SES of 0.5 (see Figure 3). The expected reading score of that student would be about 45 points higher than a student with the same family SES who attended a school with a mean SES of -0.5. For a student in the United States the difference would be about 63 points.

The presence of large and statistically significant contextual effects suggest that the socioeconomic composition of the intake to a school has an effect of student performance over and above the effects associated with individuals students' family background. The contextual effect may be partially attributable to "peer effects"; for example, students in high SES schools may have higher expectations for performance, discuss homework with each other, and generally promote a culture conducive to learning. However, the effects may also be due to a differential allocation of resources. For example, high SES schools may be more likely to attract and retain talented and well-trained teachers and have higher levels of teaching resources such as well-equipped libraries and science laboratories. It is also likely that schools that predominantly serve high SES students have greater parental involvement, a better disciplinary climate, and stronger teacher-student relations. The next section examines the effects associated with these school-level factors.

5. Differences among Schools and the Effects of Policy and Practice

This section examines the extent to which differences between Canada and the US in their reading achievement are attributable to: (a) differences in the family backgrounds of students, (b) contextual effects, and (c) factors associated with school resources and school and classroom policy and practice. The analyses employ a two-level hierarchical linear model with students nested within schools. The first model, Model I, is essentially a "null model" as it does not include any student or school-level variables. However, it includes a dummy variable denoting whether the schools are Canadian or US schools. The coefficient for this variable is an estimate of the difference between the two countries in their (school-level) mean scores in reading achievement. The three models that follow extend this model to include other variables, which allows one to assess the extent to which the differences between Canada and the US are attributable to various factors measured in PISA.[5]

---

[5] The analyses in this section and in Section 6 employed a two-level hierarchical linear regression model, with students nested within schools. Model I is a null model, except that the dummy variable denoting whether the school was a Canadian or US school was included at the second level:

$$Y_{ij} = \beta_{0j} + \varepsilon_{ij} \quad \text{(student-level)}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(Cda_j) + u_{0j} \quad \text{(school-level)}$$

where Y_{ik} is reading performance for the i^{th} student in the j^{th} school. β_{oj} is the intercept for the j^{th} school, and ε_{ij} is a student-level residual. At the second level, the intercept, γ_{00}, is the mean score for US schools, and the coefficient, γ_{01}, is the estimate of the difference between the Canadian and US average. u_{0j} is a school-level residual. Table 5 reports the estimates of the γ's and their standard errors.

Model II extends the student level model to include the three pupil-level variables:

$$Y_{ij} = \beta_{0j} + \beta_{1j}Female_{1ij} + \beta_{2j}SES_{2ij} + \beta_{3j}Foreign-born_{3ij} + \varepsilon_{ijk} \quad \text{(student-level)}$$

β_1 to β_3 are the regression coefficients associated with the three student-level covariates, female, SES, and foreign-born. The analysis revealed that these parameters varied significantly among schools, and therefore they were treated as random coefficients. The school-level model included the dummy variable denoting country to test whether the slope of these parameters varied significantly among the two countries:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(Cda_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(Cda_j) + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}(Cda_j) + u_{2j}$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31}(Cda_j) + u_{3j} \quad \text{(school-level)}$$

In this case, there are eight γ's , which are reported in Table 5 with their standard errors.

Model III is identical to model II, except that two variables are added to the school-level intercept model. These variables are the mean SES of the school, and an interaction term for country-by-mean SES effects.

Model IV extends the model to include a large set of school-level variables described in Appendix A. In preliminary analyses, interaction terms for country differences were included also, but these were removed from the model as they were not statistically significant.

The estimate of the Canada-US difference and its standard error is presented in the first two columns of Table 5. The estimated difference is 31.7 points, with a standard error of 4.1. This is slightly larger than the 30-point difference observed in the first section of this report. This is because the hierarchical model estimates the weighted "mean of the school means", with each school mean weighted according to how accurately it was estimated with the available data. The mean of school means is generally different than the overall student-level mean. The school means vary significantly; the standard deviation of school means is 42.7.

Model II includes three variables describing students' characteristics: sex, family socioeconomic status, and whether the student was foreign-born. It also includes the interaction terms for each variable, which provide estimates of whether the effects of these variables differ between the two countries. The coefficient for "female" is 29.3, which indicates that on average females outperform males by an average of nearly 30 points. The interaction term (3.1 points) is not statistically significant, indicating that the magnitude of the sex difference is similar in both countries.

The estimated average SES slope is 33.5. The interaction term suggests that the slope is less steep in Canada than the US – a difference of 3.6. This difference is statistically significant at p less than 0.10, but not at p less than 0.05 ($p = 0.06$).

The difference between students born in the country and those who are foreign-born is 5.9 points, favoring those born in the country. On average, this is not statistically significant. However, the interaction term is large and statistically significant, indicating that the difference is much larger in Canada than in the US. This finding is important and calls for further analyses aimed at understanding the academic progress of immigrants in both countries.

In this analysis, the three variables describing student characteristics and family background were "centered" on the OECD means. This affects the estimates of the intercept and the Canada-US difference. One way to consider its effect is to imagine a group of 1000 students representative of all students in OECD countries. This group would comprise 506 females and 494 males, and 64 foreign-born students and 936 native-born students. On average their SES would be zero. The analysis essentially asks, "How well would this hypothetical group of students perform in reading in Canada and the United States?" The intercept for Model II (494.8) is therefore an adjusted mean indicating how well students in the United States would perform if their distribution of students were similar to that of all OECD countries. The estimate of the Canada-US difference (27.8 points) indicates that the adjusted mean for Canada is 27.8 points higher, or about 522.6. Note that the estimated Canada-US difference for Model II is about 4 points lower than that of Model I. This indicates that some of the Canadian advantage is attributable to the three factors describing students' characteristics and family background. Subsidiary analyses (results are not shown in the table) show that most of the 4-point difference is attributable mainly to differences between the countries in family SES.

Table 5
Estimates of regression coefficients and standard errors for models pertaining to differences between reading scores in the US and Canada (PISA 2000)

	Model I Unadjusted		Model II Sex, SES, and Foreign-born		Model III Sex, SES, Foreign-born, and School Mean SES		Model IV Sex, SES, Foreign-born, School Mean SES, and School and Classroom Factors	
	Effect	**(SE)**	**Effect**	**(SE)**	**Effect**	**(SE)**	**Effect**	**(SE)**
Intercept (US Mean)	493.2	(3.8)	494.8	(3.9)	489.6	(3.0)	480.6	(4.1)
Canada – US difference	31.7	(4.1)	27.8	(4.1)	24.9	(3.4)	28.7	(3.9)
Student-Level Variables								
Female			29.3	(3.2)	28.2	(3.2)	26.6	(3.2)
Canada – US difference			3.1	(3.6)	3.8	(3.6)	4.3	(3.6)
Socioeconomic Status (SES)			33.5	(1.8)	28.6	(1.9)	28.9	(1.9)
Canada – US difference			-3.6	(1.9)	-1.4	(2.1)	-1.3	(2.1)
Foreign-born			-5.9	(7.8)	-5.0	(19.7)	-4.5	(7.7)
Canada – US difference			-17.4	(8.6)	-19.7	(8.5)	-21.2	(8.5)
School Context								
School Mean SES					63.8	(5.7)	50.7	(5.9)
Canada – US difference					-18.0	(6.5)	-14.2	(6.5)
School Resources								
School Size (1 unit = 100 students)							0.9	(0.3)
School Size squared							-0.04	(0.02)
Student-Staff Teaching Ratio (1 unit = 1 student)							0.3	(0.5)
Quality of School Infrastructure							0.4	(0.4)
Students have access to computers at school (1 unit = 10 percent)							-2.2	(1.3)
Students' Use of Resources							2.7	(0.7)
School Administrators' Assessment of Teaching Staff							-0.6	(0.4)
Teachers received professional development (1 unit = 10 percent)							-0.2	(0.3)
Teachers with language arts major (1 unit = 10 percent)							1.9	(0.4)

Table 5
Estimates of regression coefficients and standard errors for models pertaining to differences between reading scores in the US and Canada (PISA 2000)

	Model I Unadjusted		Model II Sex, SES, and Foreign-born		Model III Sex, SES, Foreign-born, and School Mean SES		Model IV Sex, SES, Foreign-born, School Mean SES, and School and Classroom Factors	
School Policy and Practice								
Use of Formal Assessment							0.9	(0.4)
Teacher Morale and Commitment							0.5	(0.4)
Teacher Autonomy							0.9	(0.4)
Principal Autonomy							-0.2	(0.5)
Classroom Practice								
Use of Informal Assessment							-0.3	0.5
Student-Teacher Relations							2.4	(1.3)
Disciplinary Climate							2.7	(0.4)
Achievement Press							-0.4	(0.5)
Missing Data Dichotomous Indicators								
Data for SES			-10.9	(4.6)	-11.0	(4.5)	-11.3	(4.5)
Data for Foreign-born			-55.8	(6.8)	-54.7	(6.8)	-53.5	(6.8)
School Questionnaire Data							9.2	(6.4)
Variation Among Pupils and Schools								
Pupil Level (SD)	86.4		80.4		80.3		81.6	
School Level (SD)	42.7		33.5		27.4		24.2	
Variance Explained								
Pupil Level (%)	–		13.6		13.8		10.9	
School Level (%)	–		38.5		58.8		67.9	

Note. Analyses were based on data for 29,687 Canadian students in 1,117 schools, and 3,846 US students in 153 schools.

With multi-level models, the measure of the strength of the relationship (which when discussing gradients above was introduced as R-squared) has two components – one that pertains to the percentage of variance within schools, and another that pertains to the variance between schools that is explained with the model. The inclusion of the student-level variables in Model II accounted for about 14% of the student-level variation in reading performance, and about 39% of the variation among school means.

Model III includes the same set of background variables, as well as school mean SES and the corresponding Canada-US interaction term. This provides estimates of the "contextual effect" for the two countries. The estimate for the US is 63.8, suggesting that

a student with average characteristics (in the OECD sense) would perform 63.8 points higher if he or she attended a school with a mean SES of 0.5 rather a school with a mean SES of -0.5 (or generally one point higher in mean SES). This is a substantial effect – similar to one level of the reading scale. The estimated contextual effect for Canada is about 18 points lower, and the difference is statistically significant. These results show that in both countries there is a substantial advantage associated with attending a high SES school, even when account is taken of the students' individual family backgrounds.

Controlling for the mean SES of the school reduces the estimate of the Canada-US difference by 3 more points, to 24.9. This means that if we consider our hypothetical group of 1000 students, who are representative of all OECD students, and imagine that in both countries they attended schools with average SES intakes, then the difference in performance between the two countries would be 24.9 points, or about one-quarter of a standard deviation. We can see this graphically by returning to the third panel of Figure 3. If we consider schools of average SES (close to zero on the X-axis), the average score for US schools is about 490, while the average for Canadian schools is about 515.

The importance of school mean SES is also emphasized by the proportion of variance it explains. The variables in Model III account for about 59% percent of school-level variance – an increase of 20% over that obtained with Model II.

The last model in Table 5 extends Model III to include also a broad set of variables describing school resources, and school and classroom policy and practice. These variables are described in Appendix A. Most of these variables were scaled on a ten-point scale, ranging from zero to ten, such that if a school scored 3.5 on the scale, it would be at the 35th percentile among all OECD schools participating in the survey. Similarly, a school with a score of 7.6 would be at the 76th percentile. This allows us to interpret the estimates of the regression coefficients in a fairly straightforward way. For example, the estimated effect for "teacher-student relations" is 2.4. This suggests that the reading performance for a school at the 50th percentile on this scale was on average about 2.4 points higher than a school at the 40th percentile. Five of the variables were not scaled in this way, as their natural metric provided a direct interpretation. School size was scaled such that one unit represents an increase of 100 students. The student-staff teaching ratio was scaled such that 1 unit represents an increase of 1 student. Three variables were coded such that one unit represents an increase of ten percent; these include the percentage of students who had access to computers, the percentage of teachers who received professional development, and the percentage of teachers who had a language arts major. The model also includes a term for the square of school size, as its effect was non-linear.

The estimated coefficients for school size indicate that there is a curvilinear effect. The school size variable was centered on the OECD mean of 5.20, corresponding to a school size of 520. An increase of 100 students from the average is associated with an increase of less than one point in reading performance ((0.9*1.0)+(-0.04*1.0*1.0)), while an increase in school size of 200 students is associated with an increase in reading performance of about one and a half points ((0.9*2.0)+(-0.04*2.0*2.0)).

Among the other variables pertaining to school resources, the only two that were statistically significant were students' use of resources and the percentage of teachers with a major in language arts. Each one-point increase on the scale for students' use of resources is associated with an increase in reading performance of 2.7 points. An increase of 10 percent in the percentage of teachers with a language arts major is associated with an increase of 1.9 points. The effects associated with other school resource variables were not statistically significant.

Two of the school policy and practice variables were statistically significant: use of formal assessment and teacher autonomy. A one-point increase on the 10-point scales for these variables was associated with a 0.9 point increase in reading performance for each factor. Two of the classroom practice variables – teacher-student relations and disciplinary practice – were associated with increases of 2.4 and 2.7 points for each one-point increase on the respective scales.

The inclusion of the school policy and practice factors does not help explain differences between Canadian and US students in reading performance. In fact, the estimated Canada-US difference for Model IV is larger than that of Model III. This suggests that the Canadian advantage in reading performance is not attributable to Canadian students receiving a higher level of resources. It is noteworthy, though, that the estimated contextual effect is smaller when the school policy and practice factors are included in the models. This indicates that the contextual effect is to some extent mediated by these factors; that is, students tend to have better performance in high SES schools because these schools tend to have higher levels of school resources, and policies and practices that are conducive to higher performance. The set of school-level factors increased the proportion of school-level variance explained to 68%, and increase of about 9% over that obtained with Model III.

Overall, the analysis of school factors suggests that there is no single factor that contributes to the success or failure of a school; rather, there are several factors that each has a small but important effect. The most important factors in Canada and the US are: having teachers trained in the language arts, students' use of available resources, the use of formal assessment, teacher autonomy, positive teacher-student relations, and a strong disciplinary climate. The effects of these factors, however, are relatively small compared with the effect associated with the mean SES of the school.

6. Differences among School Sectors

This section examines differences among sectors of schools in Canada and the US. In preliminary analyses I attempted to cluster schools into different types, based on their scores on the most important school-level factors. However, the analysis did not yield clearly identifiable school types, and school mean SES tended to dominate the clustering. Over the past 25 years, there has been considerable interest in difference between the public and private sectors in their performance. However, the results of the previous section suggest that there is also considerable variation among public schools. Therefore, in analysis of sector differences, schools were divided into four sectors: *rural public schools* in areas with a population of less than 15,000 people; *town/small city public schools* in cities and towns with populations between 15,000 and 100,000 people; *large city public schools* in cities with populations greater than 100,000; and *private schools*.

Two sets of analyses were conducted. The first set compares the average levels of school mean SES and the average scaled scores for school resources, and for school policy and practice, among the four sectors within each country, and between Canada and the US. The second set of analyses compares the sectors in their reading performance. In these analyses, I extended the models presented in Table 5 by replacing the single dummy variable denoting country with separate dummy variables denoting the four sectors in Canada and the US. The results for both sets of analyses are presented in Table 6.

In both Canada and the US, as one would expect, the average SES of private schools is relatively high, and substantially higher than the OECD average. In the US, the mean is 0.64, and in Canada it is 0.58. The mean SES of rural schools in the US is -0.10, while in Canada it is -0.03. These means do not differ significantly from the OECD mean. In both countries, the mean SES of town/small cities is about 10 percent of a standard deviation above the OECD mean. This difference is statistically significant for Canada, but not for the US. The mean SES of schools in large cities in the US (-0.04) is close to the OECD mean; however, the mean SES of schools in large Canadian cities is considerably larger – about one-quarter of a standard deviation above the OECD mean. As the majority of private schools are in large cities, these sector differences suggest that the public/private divide along social class lines is not as great in Canada as in the US.

There are also considerable differences among the sectors in their average school size. The average school size of US rural schools was 321 students, while in Canada it was 412 students. These differ significantly from the OECD average of 520 students. The average school size among town/small city schools was about 750 in the US and nearly 900 in Canada. The US school size did not differ significantly from the OECD average, but the Canadian school size was significantly larger. Schools in large cities had large enrollments in both the US and Canada: 1128 and 948 respectively. US private schools were smaller in the US than in Canada: 366 compared with 490. These results also indicate that there are important differences between large city public schools and private schools in both countries, but these differences are greater in the US than in Canada.

Table 6
Differences among school sectors[1] in the US and Canada (PISA 2000)

	Mean Scores and Standard Errors[2]								Sector Effects for Reading (Canadian Towns/ Small Cities is the Reference Category)			
	Mean SES	**School Size**	**Teachers with Language Arts Major**	**Students' Use of Resources**	**Use of Formal Assessment**	**Teacher Autonomy**	**Teacher-Student Relations**	**Disciplinary Climate**	**Unadjusted Mean**	**Adjusted for Student SES**	**Adjusted for Student SES and School Mean SES**	**Adjustment based on Full Model**
United States												
Rural	-0.10 (0.05)	**321** (31)	73 (4)	**6.1** (0.2)	**7.8** (0.1)	5.5 (0.4)	6.1 (0.3)	5.0 (0.2)	**-39.5** (9.0)	**-34.2** (7.7)	**-25.1** (6.4)	**-28.6** (6.1)
Town/ Small City	0.11 (0.09)	755 (159)	**42** (12)	**6.6** (0.3)	**7.4** (0.5)	4.3 (0.7)	**7.5** (0.6)	**6.2** (0.5)	-13.7 (10.1)	-13.6 (7.7)	**-12.9** (5.7)	**-20.0** (5.3)
Large City	-0.04 (0.18)	1128 (282)	**99** (3)	5.6 (0.6)	**8.1** (0.3)	3.6 (1.1)	4.7 (0.9)	4.0 (0.6)	**-43.9** (9.6)	**-35.7** (7.8)	**-28.2** (6.1)	**-36.1** (6.6)
Private	**0.64** (0.11)	**366** (54)	**97** (2)	**7.4** (0.2)	**7.3** 0.5)	4.7 (0.6)	**8.3** (0.4)	**8.0** (0.7)	16.1 (15.3)	1.6 (10.8)	-11.8 (7.9)	**-20.2** (9.8)
Canada												
Rural	-0.03 (0.02)	**412** (142)	**61** (2)	**7.9** (0.1)	4.9 (0.1)	**5.9** (0.1)	**6.0** (0.1)	**4.1** (0.1)	**-9.6** (3.7)	-3.9 (3.2)	5.9 (3.2)	**10.6** (3.0)
Town/ Small City	**0.08** (0.03)	**886** (42)	71 (3)	**6.5** (0.1)	5.1 (0.2)	5.3 (0.2)	**5.6** (0.2)	**3.5** (0.2)	—	—	—	—
Large City	**0.25** (0.03)	**948** (30)	75 (2)	**7.1** (0.1)	4.9 (0.2)	**5.5** (0.2)	**6.0** (0.1)	**4.0** (0.1)	4.8 (4.3)	3.5 (3.7)	-3.0 (3.3)	-3.2 (2.9)
Private	**0.58** (0.05)	490 (43)	73 (3)	**7.5** (0.2)	**6.2** (0.3)	4.9 (0.3)	**7.0** (0.2)	5.2 (0.3)	**41.3** (6.9)	**28.1** (5.6)	4.3 (5.2)	7.2 (5.1)

Notes:

1. Schools were classified into four sectors:
   (a) Rural – public schools in areas with a population of less than 15,000 people;
   (b) Town/Small City – public schools in cities and towns with populations between 15,000 and 100,000;
   (c) Large City – public schools in cities with populations over 100,000; and
   (d) Private – all private schools.
2. Mean scores for each sector were compared with OECD international means: school mean SES (0.0); school size (520 students); teachers with language arts majors (72.9%); students' use of resources (5.0); use of formal assessment (5.0); teacher autonomy (5.0); teacher-student relations (5.0); and disciplinary climate (5.0). Differences that are statistically significant ($p < 0.05$) are in bold text.

The sectors also vary considerably in the percentage of teachers with language arts majors. In the US, nearly all teachers in large city and private schools are trained at this level (99% and 97% respectively), whereas in rural areas the percentage is 73%, and in towns and small cities it is only 42%. In Canada, the percentages are generally lower but less variable, ranging from 61 to 75%.

There are also important differences among the sectors on the five aspects of school and classroom processes. Recall that the measures of these processes were scaled on a 10-point scale, with the OECD median set at 5.0. Rural schools in the US scored fairly high on these measures – at or above 5.0 on all measures, and significantly above 5.0 on the measures of students' use of resources and use of formal assessment. Towns and small cities in the US scored higher than the OECD average on all measures except teacher autonomy, and particularly high on the use of formal assessment (7.4) and teacher-student relations (7.5). The profile for US large cities is not as strong. They scored high on the use of formal assessment (8.1), but did not differ significantly from the OECD average on the other four measures. Private schools in the US, like towns and small cities, scored significantly above OECD norms on four of the five measures. The exception was teacher autonomy, for which their score was close to the OECD median.

Rural schools in Canada scored above OECD norms on students' use of resources, teacher autonomy, and teacher-student relations, but below OECD norms on the measure of disciplinary climate. The profile for schools in towns and small cities was similar, although their score on teacher autonomy was not significantly above the OECD median. The profile for schools in large cities was also similar to that of Canadian rural schools. Private schools scored above norms in students' use of resources, use of formal assessment, and teacher-student relations. Their scores were close to OECD norms for teacher autonomy and disciplinary climate. Taken together, these results suggest that Canadian schools are somewhat more uniform in their quality than the US. They also suggest that poor classroom disciplinary climate is an important issue in many Canadian schools.

The last four columns of Table 6 present the results of regression analyses that fit the same model as those presented in Table 5, except that the dummy variable denoting US (versus Canada) was replaced with seven dummy variables denoting sector. The reference category is Canadian towns and small cities. The first model provides estimates of the unadjusted means. US rural schools and schools in large cities scored about 40 points lower than schools in Canadian towns and cities. These differences were statistically significant. The unadjusted mean scores of schools in towns and small cities and of private schools in the US did not differ significantly from those of Canadian schools in towns and small cities. Canadian rural schools scored about 10 points lower than the reference set of schools (a statistically significant difference), while schools in large cities scored about 5 points higher (a difference that was not statistically significant). Private schools in Canada scored about 41 points higher than those in towns and small cities, a difference which is statistically significant.

The second model introduces the controls for sex, socioeconomic status, and whether the student was foreign-born. These factors explain some, but not all of the variation among sectors. In the US, for example, the 40-point disadvantage associated with attending a rural or large city school diminished by about 5 points. The scores of students in US towns and small cities, and in US private schools, did not differ significantly from that of students in Canadian small towns and cities. Among the Canadian sectors, the results are remarkably uniform, although there is a private school advantage of about 28 points.

The third model controls for the mean socioeconomic status of the school, in addition to the previously mentioned controls. This model compares schools with similar socioeconomic intakes. The results for the US suggest that some of the disadvantage associated with rural and large city schools is attributable to their SES intake – the differences compared with the Canadian reference sector is only about 25 points for both sectors. US schools in towns and cities scored about 10 points lower than the Canadian reference sector, but the difference is not statistically significant for private schools. The Canadian results are remarkably consistent across the four sectors, with no statistically significant differences among them. This finding also emphasizes the uniformity of school quality among Canadian schools. After account is taken of the socioeconomic background of students and the socioeconomic intake of schools, there are no sector differences. Students have better performance on average if they attend high SES schools, but once this is taken into account, it does not matter whether they attend a private school or any type of public school.

The last model in Table 6 controls also for the eight most important school process variables presented in the left side of the table. This model essentially asks how well students in each sector would score if their schools had levels of these school processes comparable to OECD norms. The results for the US suggest that the performance of students in each of the sectors would be somewhat lower. The most significant difference is for US private schools, which have an adjusted mean that is 20 points lower than the Canadian reference category. In contrast, the Canadian results do not change much with this adjustment, and there is very little variation among the four sectors. The findings suggest that Canadian rural schools would score about 10 points higher than schools in towns and small cities.

7. Conclusions and Policy Implications

This study used data from PISA 2000 to examine differences in the performance of students in Canada and the US. It focused mainly on reading achievement and its relationship to family background and school factors, which was the focus of the PISA 2000 survey. Seven of the most important findings emerging from this study are:

(1) Canadian students have about a 30-point advantage over students in the US in their reading performance. This is a large and statistically significant advantage, equivalent to nearly one full year of schooling at this age. The findings suggest that the range of scores in Canada is narrower than in the US. However, in both countries the distribution of scores is negatively skewed, indicating that there is a disproportionate number of students with very low scores relative to the respective country mean scores. These findings are consistent with findings from other international studies.

The PISA data are cross-sectional, and therefore it is not possible to discern whether there is a particular age at which students in the US begin to fall behind their Canadian counterparts. Indeed, many of the differences observed in PISA could be attributable to the differences in children's environments during the early years, prior to entering school at age 5 or 6. The results for the Third International Mathematics Study, however, suggest that for mathematics and science, US students fared better than Canadian students through to the end of grade 4, and thereafter fell behind. The results for the TIMSS and PISA studies for students at the early secondary level are remarkably consistent, even though the studies employ a different sampling strategy and a different kind of test.

(2) Both countries have a high percentage of students at Level 3 and lower: about 56% in Canada and 68% in the US. Students at Level 3 are "capable of reading tasks of moderate complexity, such as locating multiple pieces of information, making links between different parts of a text, and relating it to everyday knowledge", whereas students at Level 4 are "capable of difficult reading tasks, such as locating embedded information, construing meaning from nuances of language and critically evaluating a text" (OECD, 2001, p. 46). Other analyses suggest that this distinction is important in Canada, as students achieving Level 4 literacy skills are much more likely to pursue further education.

An important policy issue for Canadian educators is that the full-time and part-time enrollment rate in higher education in 1999 for adults aged 18 to 29 was 17% in Canada and 20% in the US. This was not the case in 1994, when Canada had a higher enrollment rate than the US: 24% compared with 21% (National Center for Educational Statistics, 2003). Enrollment rates in post-secondary education are affected by several factors, including tuition costs, students' qualifications, and students' attitudes towards school. The PISA results call attention to an important inconsistency: in 1999-2000 there was a considerably higher percentage of Canadian students with literacy skills at Levels 4 and 5 than in the US, yet the

enrollment rate in higher education was lower in Canada. A recent study of student engagement, also based on PISA 2000 data, indicated that students' participation in school, as gauged by class attendance and truancy, was lower in Canada than in the US (Willms, 2003). We do not yet know how the combined effects of literacy skills and engagement, combined with other factors relevant to access to post-secondary, will affect enrollment rates over the next decade. However, the results in this report call for further research aimed at understanding the transition from school to higher education and the work force.

(3) The reading performance of youth from high socioeconomic backgrounds does not differ substantially between Canada and the US; however, youth from low socioeconomic status backgrounds fare markedly better in Canada than the US. Thus, Canada's advantage in reading performance is mainly attributable to its success for students from low SES backgrounds. In both countries there are many youth from low SES backgrounds who attain literacy scores at Levels 3 and 4; however, there are relatively few students from low SES families who attain scores at Level 5. We would expect, therefore, that there is a fairly strong relationship between enrollment in higher education and SES, as many low SES students do not have either the literacy skills or the financial means required for enrollment. Further research on this issue will be possible for Canada when data from the longitudinal component of PISA 2000 become available.

The analysis of socioeconomic gradients revealed that the difference in the extent of socioeconomic inequalities between Canada and the US stems mainly from difference between schools, not within them. In the US, there is a large number of low SES schools that do not fare well compared with comparable schools in Canada. However, the average level of socioeconomic inequalities within schools is about the same in the two countries.

(4) In both Canada and the US, girls have considerably higher scores than boys. The female advantage is about 30 points in both countries, and does not differ for youth from differing socioeconomic backgrounds. This is a sizeable difference, which is evident in all of the countries that participated in PISA 2000. International comparisons based on PISA 2000 data have indicated that females tend to do especially well compared with males on tasks requiring critical evaluation and the ability to relate text to personal experience, knowledge and ideas (OECD, 2002). Analyses of data from Canada's National Longitudinal Survey of Children and Youth (NLSCY) have shown that gender differences in language ability are evident when children enter kindergarten (Willms, 2002). It may be that a disproportionate number of boys fall "off-track" in their reading development during the primary grades, and subsequently develop negative attitudes towards reading. The PISA results call for further research and attention by policy makers on gender inequality.

(5) There is a large "contextual effect" associated with the average socioeconomic status of the school. This effect is stronger in the US than in Canada. Youth who attend high SES schools tend to have higher academic achievement than youth with comparable family backgrounds who are attending low SES schools. In both countries, there is a fairly high level of segregation along socioeconomic lines – the

index is 19.5 for Canada and 28.1 for the US. Residential segregation, especially in cities, contributes to the segregation of students along socioeconomic lines. This is because in many jurisdictions students are allocated to schools based on socially and economically homogeneous catchment areas. But in some jurisdictions, the exclusion of less advantaged students is exacerbated by educational policies (Lee, Groninger & Smith, 1994). Some examples include public funding for schools with selective admission criteria, open enrolment policies, language-immersion programs, and parental choice programs that do not make provisions to ensure marginalized groups are proportionally represented.

Tackling segregation stemming from such policies is difficult, because the middle class has a vested interest in maintaining segregated alternatives. Also, a case can be made that students learn more when they are in programs that are more closely matched to their interests and talents. Addressing the choice/segregation dilemma can best be tackled by ensuring that there are no "low-status" options. This requires the creation of school programs in disadvantaged areas that are equally attractive to those offered elsewhere. It also requires a concerted effort to improve schooling in less advantaged areas, which may require compensatory funding.

(6) Differences among schools in their performance are moderately related to the percentage of teachers with a major in language arts, students' use of school resources, the use of formal assessment, teacher autonomy, teacher-student relations, and the disciplinary climate of the classroom. Although these factors have small, independent effects on school performance within Canada and the US, they do not account for the difference in performance between them. The school-level factors examined in this study explain about 20% of the contextual effect in both countries. This suggests that at least some of the school policy and practice measures are correlated with the mean SES of the school; that is, high SES schools tend to have better resources and more positive school climates. The results also indicate that these factors have an effect independent of school mean SES. Therefore, if we consider the range of school performance among schools that predominantly serve low SES students, the success of the high performing schools would to some extent be attributable to these factors. However, there is no single factor associated with school effectiveness that can explain why some schools have higher performance than others – it is the cumulative effect of several school and classroom factors.

(7) After account is taken of students' family background, the average reading performance of Canadian schools in rural areas, towns and small cities, and large cities does not vary significantly across these sectors, and is comparable to private schools in the US. Canadian private schools have an advantage of about one-quarter of a standard deviation, but this is fully accounted for by the contextual effect of school mean SES. These findings have important implications for parents trying to decide whether it is worth the expense to send their child to a private school. In Canada, the effect is about one-quarter of a standard deviation, which is equivalent to about two-thirds of a year of schooling at this grade level. However, the results also show that public schools with comparable socioeconomic intakes have similar results. It is not the public/private distinction that is important; it is the socioeconomic intake of the school. In the US, the private school advantage is about 35 points over schools in rural areas and in cities, after accounting for students' family background. About one-half of this advantage is attributable to the average socioeconomic intake of the school.

Overall, the findings in this study indicate that the 30-point advantage of Canadian students over their US counterparts is not attributable to the school resource and process factors measured in PISA. These factors have important effects on student performance among schools within each country, but they do not account for the US-Canada achievement gap. One of the most prominent differences between the two schooling systems is that Canadian schools tend to have a more heterogeneous mix of students in terms of their socioeconomic status, and although there is considerable variation among schools in their outcomes in both countries, on average Canadian schools tend to be more uniform in their performance. The Canadian advantage appears to be mainly attributable to schools that achieve success with students from low socioeconomic families.

Bibliography

Alwin, D. F. (1976). Assessing school effects: Some identities. *Sociology of Education, 49*, 294-303.

Beaton, A.E., Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Kelly, D.L. and Smith, T.A. (1996a). *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study.* Boston, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Beaton, A.E., Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., Smith, T.A. and Kelly, D.L. (1996b). *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study.* Boston, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Dutton, D. B., & Levine, S. (1989). "Overview, methodological critique, and reformulation." In J. P. Bunker, D. S. Gomby, & B. H. Kehrer (Eds.) *Pathways to health* (pp. 29-69). Menlo Park, CA: The Henry J. Kaiser Family Foundation.

Lee, V. E. and Smith, J. B.(1993), "Effects of school restructuring on the achievement and engagement of middle-grade students." *Sociology of Education, 66*(July), 164-187.

Martin, M.O., Mullis, I.V.S., Beaton, A.E., Gonzalez, E.J., Smith, T.A., and Kelly, D.L. (1997). *Science Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study (TIMSS).* Boston, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Mueller, C. W., & Parcel, T. L. (1981). "Measures of socioeconomic status: Alternatives and recommendations." *Child Development, 52*, 13-30.

Mullis, I.V.S., Martin, M.O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L. and Smith, T.A. (1997). *Mathematics Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study (TIMSS).* Boston, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

National Center for Educational Statistics (2003). Comparative indicators of education in US and other G8 Countries: 2002. (Report 2003-026). Washington: US Department of Education, Institute of Education Statistics.

Organisation for Economic Cooperation and Development (2001). Knowledge and skills for life: First results from the OECD Programme for International Student Assessment (PISA) 2000. Paris: OECD. (http://www1.oecd.org/publications/e-book/9601141E.PDF)

Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (2nd ed).* Thousand Oaks, CA: Sage.

Raudenbush, S. W. & Willms, J. D. (1995). "The estimation of school effects." *Journal of Educational and Behavioral Statistics*, 20(4), 307-335.

Rust, K. F., & Rao, J.N.K. (1996). "Variance estimation for complex estimators in sample surveys." *Statistics in Medical Research, 5*, 381-397.

Willms, J. D. & Flanagan, P. (2002). *Programs to increase access to post-secondary education: A research report prepared for the Canada Millennium Scholarship Foundation.*

Willms, J. D. (1999). *Inequalities in literacy skills among youth in Canada and the United States*. (International Adult Literacy Survey No. 6). Ottawa, ON: Human Resources Development Canada and National Literacy Secretariat.

Willms, J. D. (2001). "Monitoring school performance for "standards-based reform"." *Evaluation and Research in Education, 14*, (3&4), 237-253.

Willms, J. D. (2002). *Vulnerable children: Findings from Canada's National Longitudinal Survey of Children and Youth.* Edmonton, AB: University of Alberta Press.

Willms, J. D. (2003). *Ten hypotheses about socioeconomic gradients and community differences in children's developmental outcomes.* Report prepared for Human Resources Development Canada. (http://www.hrdc-drhc.gc.ca/sp-ps/arb-dgra/publications/research/2002docs/SP-560-01-03/SP-560-01-03E.pdf)

Willms, J. D. (2003). *Student engagement at school: A sense of belonging and participation.* Paris: Organisation for Economic Cooperation and Development. (http://www1.oecd.org/publications/e-book/9603131E.PDF)

Appendix A: Measures of school resources, school policy and practice

The measures of school resources, school policy and practice were derived as follows:

School size was derived from the school administrators' report of the school enrolment. One unit on this scale represent 100 students.

Student–teaching staff ratio was defined as the number of full-time equivalent teachers divided by the number of students in the school. One unit on this variable represents a change of one student per teacher.

Quality of school infrastructure is a summary measure derived from school principals' reports of the extent to which the learning of 15-year-olds was hindered by (a) poor condition of buildings; (b) poor heating, cooling and/or lighting systems; (c) lack of instructional space (*e.g.,* classrooms); (d) lack of instructional material (*e.g.,* textbooks); (e) not enough computers for instruction,; (f) lack of instructional materials in the library; (g) lack of instructional multimedia equipment; and (h) inadequate science laboratory equipment. One unit on this scale represents 10 percentile points, with higher scores indicating a better quality of school infrastructure.

Students have access to computers is based on school principals' reports of the number of computers in the school available to students divided by the school enrolment.

Students' use of resources is a summary measure derived from students' reports of the extent to which they used the following resources in their school: (a) the school library, (b) computers (c) calculators, (d) the Internet, and (e) science laboratories. One unit on this scale represents 10 percentile points, with higher scores indicating more positive ratings of teaching staff.

School administrators' assessment of teaching staff was derived from school principals' reports of the extent to which the learning of 15-year-olds was hindered by: (a) low expectations of teachers; (b) poor student–teacher relations; (c) teacher turnover; (d) teachers not meeting individual student needs; (e) teacher absenteeism; (f) staff resisting change; (g) teachers being too strict with students; and (h) students not being encouraged to achieve their full potential. One unit on this scale represents 10 percentile points, with higher scores indicating more positive ratings of teaching staff.

Teachers received professional development was derived from school principals' reports of the percentage of their teachers were involved in professional development programmes. One unit on this scale represents 10 percentile points, with higher scores indicating more positive ratings of teaching staff.

Teachers with language arts major was derived from principals' reports of the percentage of their teachers who held a post-secondary level qualification with a major in the test language (e.g., English in most US and Canadian schools, and French in Canadian schools where French was the language of instruction). One unit on this scale represents 10 percentile points, with higher scores indicating more positive ratings of teaching staff.

Use of formal assessment was derived from school principals' reports on the frequency with which standardised tests were used, and on whether or not the assessments were used to monitor the school's progress from year to year and monitor the school's progress from year to year. One unit on this scale represents 10 percentile points, with higher scores indicating greater use of formal assessments.

Teacher morale and commitment was derived from school principals' reports on the extent to which they agreed with these statements concerning teacher morale and commitment: (a) the morale of teachers in this school is high; (b) teachers work with enthusiasm; (c) teachers take pride in this school; and (d) teachers value academic achievement. One unit on this scale represents 10 percentile points, with higher scores indicating a higher level of teacher morale and commitment.

Teacher autonomy was derived from a question asked of principals as to who had the main responsibility for: (a) hiring teachers; (b) firing teachers; (c) establishing teachers' starting salaries; (d) determining teachers' salary increases; (e) formulating the school budget; (f) deciding on budget allocations within the school; (g) establishing student disciplinary policies; (h) establishing student assessment policies; (i) approving students for admittance to school; (j) choosing which textbooks are used; (k) determining course content; and (l) deciding which courses are offered. This scale indicates the extent to which teachers had responsibility for these activities. One unit on this scale represents 10 percentile points, with higher scores indicating a higher level of teacher autonomy.

Principal autonomy was derived from the same question described above. In this case, the scale indicates the extent to which principals had responsibility for the various activities. One unit on this scale represents 10 percentile points, with higher scores indicating a higher level of principal autonomy.

Use of informal assessment was derived from school principals' reports on the frequency with which students were assessed using teacher-developed tests, teachers' judgemental ratings, student portfolios and student assignments/projects/homework, and on how frequently assessment information was formally communicated to parents and the school principal. One unit on this scale represents 10 percentile points, with higher scores indicating greater use of informal assessments.

Student–teacher relations was based on students' reports of the extent to which they agreed or disagreed with the following statements concerning student–teacher relations: (a) students get along well with teachers; (b) most teachers are interested in students' wellbeing; (c) most of my teachers really listen to what I have to say; (d) if I need extra help, I will receive it from my teachers; and (e) most of my teachers treat me fairly. The student scores were aggregated to the school level, and scaled such that one unit on the scale represents 10 percentile points, with higher scores indicating better student–teacher relations.

Disciplinary climate was based on students' reports of the extent to which they agreed or disagreed with the following statements concerning student–teacher relations: (a) the teacher has to wait a long time for students to quieten down; (b) students cannot work well; (c) students don't listen to what the teacher says; (d) students don't start working for a long time after the lesson begins; and (e) there is noise and disorder. The student scores were aggregated to the school level, and scaled such that one unit on the scale represents 10 percentile points, with higher scores indicating a more positive disciplinary climate.

Achievement press was based on students' reports of the extent to which they agreed or disagreed with the following statements concerning teachers' expectations: (a) the teacher wants students to work hard; (b) the teacher does not like it when students deliver careless work; (c) the teacher checks students' homework; and (d) students have a lot to learn. The student scores were aggregated to the school level, and scaled such that one unit on the scale represents 10 percentile points, with higher scores indicating greater press for academic achievement.