Recent comparative studies of student achievement have sparked global debates about how best to design effective policies that raise the overall quality of learning while reducing disparities among students.

This report demonstrates a useful analytical tool, called the gradient, which represents the relationship between student learning outcomes and socioeconomic status. Using data from two recent international student assessments, this report explores ten key policy questions relevant to the educational performance of schools and schooling systems. It also shows how the gradient framework can be used to assess the likely effects of different policy interventions to reduce inequalities.

This publication highlights some of the main obstacles and opportunities facing policymakers in efforts to improve the quality and equity of national education systems.

**UNESCO Institute for Statistics**
P.O. Box 6128, Succursale Centre-Ville
Montreal, Quebec  H3C 3J7
Canada
http://www.uis.unesco.org

The UNESCO Institute for Statistics (UIS) is the statistical office of the United Nations Educational, Scientific and Cultural Organization (UNESCO) and is the UN depository for internationally comparable statistics in the fields of education, science and technology, and culture and communication.

Learning Divides: Ten Policy Questions About the Performance and Equity of Schools and Schooling Systems

UNESCO INSTITUTE for STATISTICS

UNESCO INSTITUTE for STATISTICS

UIS Working Paper No. 5

# Learning Divides:
## Ten Policy Questions About the Performance and Equity of Schools and Schooling Systems

# LEARNING DIVIDES:
# TEN POLICY QUESTIONS ABOUT THE PERFORMANCE AND EQUITY OF SCHOOLS AND SCHOOLING SYSTEMS

**By J. Douglas Willms**


**UNESCO Institute for Statistics, Montreal, 2006**

## UNESCO

The constitution of the United Nations Educational, Scientific and Cultural Organization (UNESCO) was adopted by 20 countries at the London Conference in November 1945 and entered into effect on 4 November 1946. The Organization currently has 190 Member States and six Associate Members.

The main objective of UNESCO is to contribute to peace and security in the world by promoting collaboration among nations through education, science, culture and communication in order to foster universal respect for justice, the rule of law, and the human rights and fundamental freedoms that are affirmed for the peoples of the world, without distinction of race, sex, language or religion, by the Charter of the United Nations.

To fulfill its mandate, UNESCO performs five principal functions: 1) prospective studies on education, science, culture and communication for tomorrow's world; 2) the advancement, transfer and sharing of knowledge through research, training and teaching activities; 3) standard-setting actions for the preparation and adoption of internal instruments and statutory recommendations; 4) expertise through technical co-operation to Member States for their development policies and projects; and 5) the exchange of specialised information.

UNESCO headquarters are located in Paris, France.

## UNESCO Institute for Statistics

The UNESCO Institute for Statistics (UIS) is the statistical office of UNESCO and is the UN depository for global statistics in the fields of education, science and technology, culture and communication.

The UIS was established in 1999. It was created to improve UNESCO's statistical programme and to develop and deliver the timely, accurate and policy-relevant statistics needed in today's increasingly complex and rapidly changing social, political and economic environments.

The UIS is based in Montreal, Canada.

UNESCO Institute for Statistics
P.O. Box 6128, Succursale Centre-Ville
Montreal, Quebec H3C 3J7
Canada

Tel: (1 514) 343-6880
Fax: (1 514) 343-6882
Email: publications@uis.unesco.org
http://www.uis.unesco.org

ISBN 92-9189-039-1



Ref: UIS/WP/06-02
Cover design: JCNicholls Design
Photo credits: ©UNESCO/M. Bario, M. Borg, G. Guit, D. Roger
Printed by: ICAO

# Table of Contents

# Introduction

## Socioeconomic gradients as an assessment framework

### PIRLS and PISA: International studies of students' reading literacy skills

The Progress in International Reading Literacy Study (PIRLS) was conducted in 2001 under the auspices of the International Association for the Evaluation of Educational Achievement (IEA). It was a large-scale collaborative effort, involving 35 countries around the world. Its aim was to assess the literacy skills of pupils in their fourth grade[1] of elementary school, using a comprehensive measure of early literacy skills. The study included surveys of students, parents, teachers and school administrators. Findings are presented in *PIRLS 2001 International Report* (Mullis, Martin, Gonzales and Kennedy, 2003).

The Programme for International Student Assessment (PISA) is a collaborative initiative of member countries of the Organisation for Economic Co-operation and Development (OECD) to assess the knowledge and life skills of 15-year-old youth as they approach the end of their compulsory period of schooling. PISA differs from other international assessments in that it emphasises the kinds of skills that students will need in their everyday lives as they approach post-secondary education and work in the knowledge society. Therefore, the literacy tests are primarily concerned with whether students can apply the knowledge they have learned at school, rather than the content of secondary school curricula that is common among countries.

The first PISA assessment was conducted in 2000 and included 28 OECD countries and four non-member countries. In 2002, another 14 non-OECD countries participated. The focus of PISA in 2000 and 2002 was on students' reading skills, with mathematics and science skills treated as minor domains. The assessment also collected extensive information on students' family background, including family structure, the education level and occupation of parents, and several aspects of social and cultural capital available to students. Findings from the 2000 assessment are presented in *Knowledge and Skills for Life* (OECD, 2001). The focus of the 2003 PISA assessment was on mathematics, with reading and science as minor domains, and in 2006 the focus is on science. The cycle will then be repeated in 2009, beginning with reading as the primary domain.

[1] The sample for each country was drawn from the highest grade with the most 9-year-olds, which was grade 4 in most countries.

Both assessment programmes are policy-oriented, designed and guided by an international steering committee to provide regular data pertaining to the most pressing policy issues confronting educational administrators and policymakers around the world. They include considerable information on the family and school factors that contribute to school performance in each country. However, most of the reports stemming from these programmes have emphasised differences among countries in their academic performance, even though the variation in student performance among countries is relatively small compared to that within each country.

In earlier work, my colleagues and I examined the relationship between student performance and the socioeconomic status (SES) of their families (e.g. see Willms and Somers, 2001, and the final chapter of *Knowledge and Skills for Life*). Not surprisingly, in every country there is a gradient in student performance associated with family socioeconomic status: youth from lower socioeconomic backgrounds have weaker literacy skills on average than those from more advantaged backgrounds. The results revealed that the strength of this relationship varies considerably among countries, suggesting that some are more successful than others in reducing the disparities associated with socioeconomic status. Moreover, there are some countries with high overall performance and relatively shallow gradients, demonstrating that it is possible to achieve relatively high levels of literacy while mitigating the effects of social disadvantage.

This paper has three aims. The first aim is to examine socioeconomic gradients in greater detail using data furnished by PIRLS and PISA, and discuss the implications of the findings for policy and practice in the participating countries. The second aim is to set out a general framework for analysing educational data that are collected in international, national and local studies. This is accomplished by setting out ten key policy questions that provide a more explicit link between educational indicators and practice. The third aim is to describe the statistical models used to address each question. To help maintain the flow of the discussion for the non-statistical reader, the statistical models are set out as footnotes within each section.

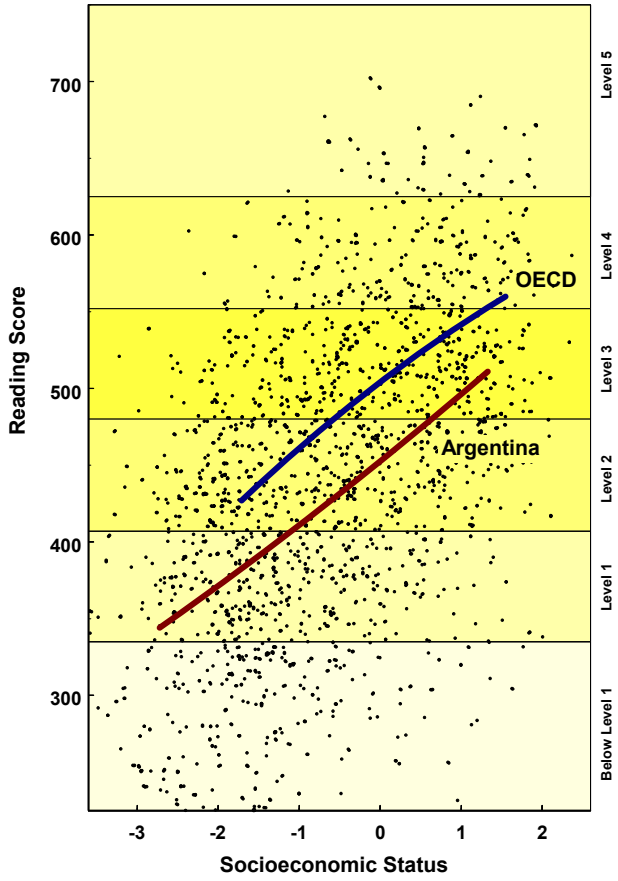
The remainder of this introductory section describes socioeconomic gradients and "school profiles", which together provide a useful summary of the relationship between student performance and socioeconomic status in a country, or smaller educational units. These two tools provide a framework for discerning how best to intervene to raise school performance and reduce inequalities. They also raise several questions concerning the performance and equity of schools and schooling systems. The ten sections that follow the introduction tackle some of these questions, using data from PIRLS and PISA. The report concludes with a summary of the findings and a discussion of how the answers to the ten questions are relevant to educational policies aimed at improving educational performance and reducing inequalities among groups of differing status.

## Socioeconomic gradients

A socioeconomic *gradient* describes the relationship between a social outcome and socioeconomic status for individuals in a specific jurisdiction, such as a school, a province or state, or a country (Willms, 2003a). The social outcome in this study is students' reading performance. Socioeconomic status (SES) is a sociological term that refers to the relative position of a family or individual in a hierarchical social structure, based on their access to, or control over, wealth, prestige and power (Mueller and Parcel, 1981). The key indicators of SES in most educational studies include the level of education of students' parents and the prestige of their parents' occupations. The PISA measure of socioeconomic status describes students' economic, social and cultural background. It was derived from data describing parental education, occupation as well as the material, educational and cultural possessions in the home. The PIRLS measure used in this study was based on parents' level of education, occupational status and family income.

**Figure 1** shows the socioeconomic gradient for reading performance in Argentina based on PISA. Argentina was chosen for this example because its gradient for both PIRLS and PISA is quite similar and its profile is comparable to that of many other non-OECD countries. The vertical axis has two scales: the left-hand scale is the continuous scale for reading performance, which has a mean of 500 and a standard deviation of 100 for all students in participating OECD countries. The right-hand axis depicts the five reading levels, which are described in *Knowledge and Skills for Life*. The horizontal axis is family SES, which is a statistical composite derived from analysis of five factors describing family background: the prestige of the parents' occupations, the level of education of the parents, a measure of wealth pertaining to material possessions in the home, a measure describing access to educational possessions in the home, and a measure of culturally-related possessions in the home, such as musical instruments or books of poetry. The five factors contribute approximately equally to the SES composite. The measure was scaled to have a mean of zero and a standard deviation of one at the student level for OECD countries.

**Figure 1. Socioeconomic gradient for Argentina**




*Source:* PISA, 2000-2002.

Socioeconomic gradients comprise three components: their level, slope and strength of the outcome-SES relationship.

a. The *level* of the gradient is defined as the expected score on the outcome measure for a person with average SES. The level of a gradient for a country (or for a province, state or school) is an indicator of its average performance, after taking account of students' socioeconomic status. The level of the Argentinean gradient is 452.

b. The *slope* of the gradient indicates the extent of inequality attributable to SES. Steeper gradients indicate a greater impact of SES on student performance – that is, more inequality – while more gradual gradients indicate a lower impact of SES – that is, less inequality. The Argentinean gradient is slightly curvilinear, with the slope increasing slightly with rising levels of SES. (Curvilinear gradients are discussed in a later section.) The slope of the Argentinean gradient is 42.6 (in the centre of the data), which indicates that the expected reading performance increases by 42.6 points for a one standard deviation increase in SES.

c. The *strength* of the gradient refers to the proportion of variance in the social outcome that is explained by SES. If the strength of 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 measure called R-squared, which for this example is 0.23.

The gradient line is drawn from the 5$^{th}$ to the 95$^{th}$ percentile of the SES scores for a particular population. For Argentina, the 5$^{th}$ and 95$^{th}$ percentiles are -2.72 and 1.33 respectively. Therefore, 90% of Argentinean students fall in this range. Students in Argentina on average have a lower SES than those in other OECD countries. The 5$^{th}$ and 95$^{th}$ percentiles for all OECD students are -1.71 and 1.55 respectively. For illustrative purposes, the graph also shows the reading performance and SES for a representative sample of 2,500 Argentinean students. These are the small black dots above and below the gradient line. They show that there is considerable variation in reading performance at all levels of SES.

Although the gradient line conveys considerable information about the distributions of reading performance and SES, and the relationship between them, it does not describe how these relationships vary within and between schools or among other jurisdictions within the country. Some of this information can be summarised with a "school profile", which will be described in the next section. Together, the two graphs provide a convenient summary of the most important relationships between reading performance and SES. They also provide a useful means for evaluating and comparing schools and a schooling system for setting standards and for thinking about the kinds of interventions that might be most beneficial.

The term "learning bar" is used in this report as a metaphor for the socioeconomic gradient. The central question facing most schools and countries is: "How can we raise and level the learning bar?" Increasing educational performance and reducing inequalities among students from differing socioeconomic backgrounds can be achieved in a number of ways. The approach that may work best depends on social and political issues, but it also depends on the distribution of student performance and SES within and among schools, and how these factors are related to and interact with school resources and various aspects of school policy and practice. The next section describes school profiles, again using data for Argentina as an example.

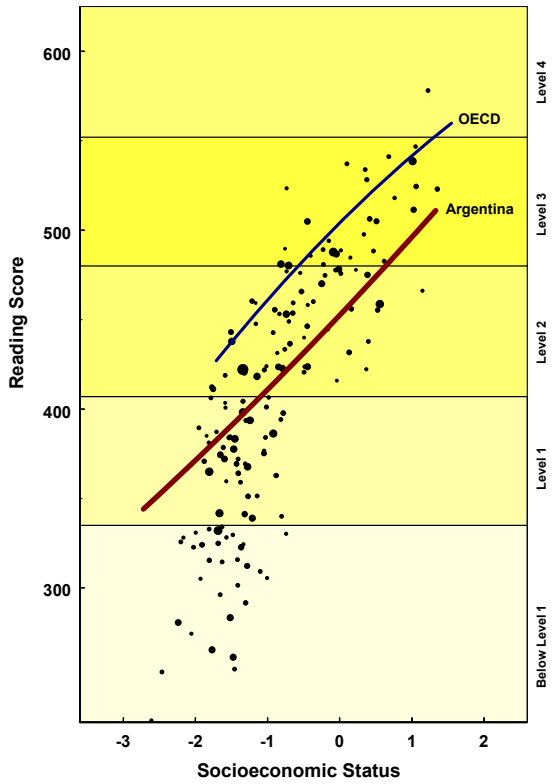
## School profiles

The term "school profile" is used in this report for graphs that display the relationship between school mean performance and school mean SES. **Figure 2** provides an example. It displays the relationship between average school performance in reading and school mean socioeconomic status for the 156 schools in Argentina that participated in PISA. In this graph, the dots represent schools rather than students. The size of the dots is proportional to school enrolment. The blue line is the SES gradient for OECD countries, while the red line is the SES gradient for Argentina, as in Figure 1.

A number of findings emerge from this analysis.

a. In many countries children in the bottom 20% of the income or SES distribution are considered to be living in poverty. The 20th percentile of SES for OECD countries is -0.82. Figure 2 shows that the range in school mean reading scores, among schools with mean SES scores above -0.82, is relatively small compared with those that have a mean SES below that level. Among the group of schools with a mean SES above -0.82, the range from the lowest to highest performing schools is about 80 points, which is consistent with the range in many OECD countries. Also, the graph indicates that many of the schools with a mean SES above -0.82 had average reading scores that were close to OECD norms.

**Figure 2. School profile for Argentina**




*Source:* PISA, 2000-2002.

b. However, there is a substantial number of very low SES schools in Argentina. More than one-half of the schools in the sample (57%) had a mean SES below -0.82. For these schools there is a substantial range in the mean reading scores, with the majority below Level 2.

The graphs displaying the socioeconomic gradient (Figure 1) and the school profiles (Figure 2) together provide a useful characterisation of the schooling system for policy purposes. The next section discusses five different types of policy interventions in the context of socioeconomic gradients, following the example of the learning bar for Argentina.

**Five types of policy interventions**

There are many different kinds of policy interventions that could be considered to raise and level the learning bar. To assess their potential impact on raising performance and reducing inequalities, they have been classified into five different types, which are described below.

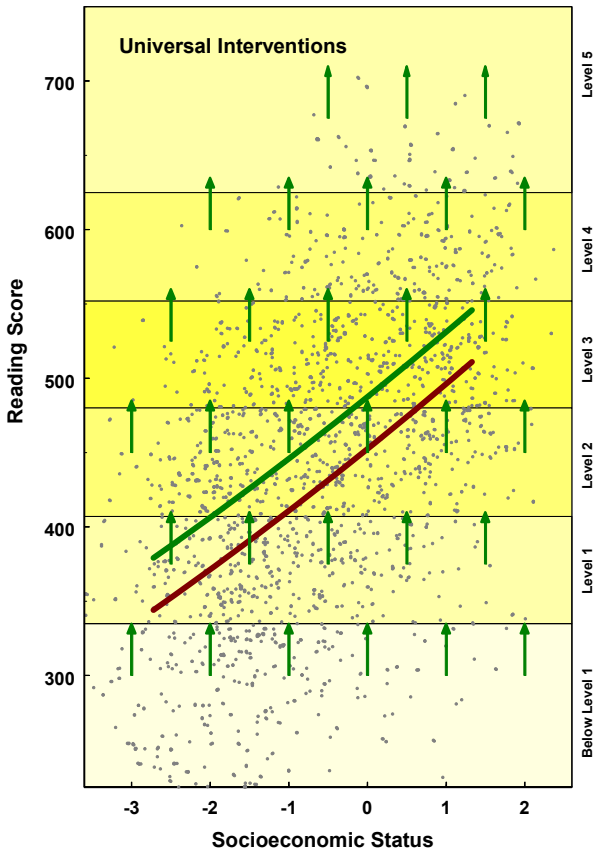
*Universal interventions* strive to increase the educational performance of all children through reforms that are applied equally across the schooling system. Generally they are aimed at altering the content and pace of the curriculum, improving instructional techniques or the learning environment in schools and classrooms. Some jurisdictions responded to PISA 2000 results by introducing major curriculum reforms, reducing class sizes, altering the age-at-entry to kindergarten, or increasing the time spent on reading instruction. These are all universal interventions.

Some universal interventions strive to improve children's learning environments by changing the structural features of schools. For example, a popular reform in the United States is to reduce school size, as there is some evidence that smaller schools have better teacher-student relations and fewer discipline problems. There has also been an effort to increase parents' involvement in their children's schooling in several ways. These include strategies to increase their involvement in school activities at home, the better use of parents as volunteers at school, and parent participation in school governance.

Most universal interventions, however, are directed at changing teacher practice. Teachers regularly receive in-service programmes pertaining to instructional approaches, assessment strategies and classroom management. Cooperative learning, the "whole language" approach to reading instruction, peer tutoring, portfolio assessment and positive behaviour management are examples that have received widespread attention over the past two decades. Perhaps the most prevalent universal intervention among OECD countries has been to increase the accountability of schools and schooling systems through the assessment of student performance. The underlying belief is that increased accountability will motivate administrators and teachers to improve the learning environment of schools and classrooms and provide better instruction.

**Figure 3** displays the effects of a universal intervention that bolstered the reading performance of all students by one-half of a standard deviation. This has the effect of raising the learning bar by 50 points but not levelling it. The overall mean performance is also increased by 50 points.

**Figure 3. Universal interventions**

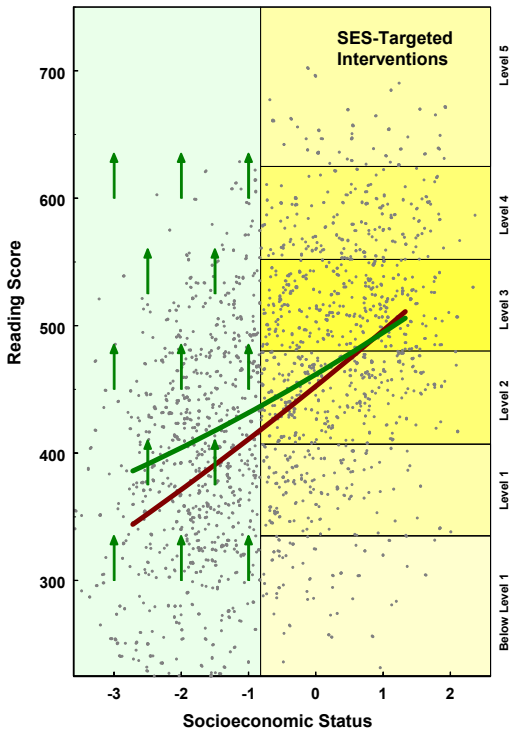



*Source:* PISA, 2000-2002.

*SES-targeted interventions* aim to improve the educational performance of students with low socioeconomic status by providing a specialised curriculum or additional instructional resources. The classic example is Head Start pre-school programmes for children from low socioeconomic backgrounds, but there is a wide range of programmes that target "at risk" children and youth. Some select students on the basis of a risk factor other than SES, such as whether the child is a new immigrant, member of an ethnic minority or living in a low-income community. The important distinction is that these programmes select children based on the family's SES or some other factor correlated with SES rather than on the cognitive ability of the child.

**Figure 4** provides an example of an SES-targeted intervention for Argentina. In this case, students with SES scores below -0.82 (on the OECD scale) receive an intervention that is successful in boosting their reading performance by 50 points. This has the effect of levelling the learning bar as the performance of low SES students is raised. However, the level of the gradient – the expected score for an average SES student – is raised only slightly by the intervention.

**Figure 4. SES-targeted interventions**




*Source:* PISA, 2000-2002.

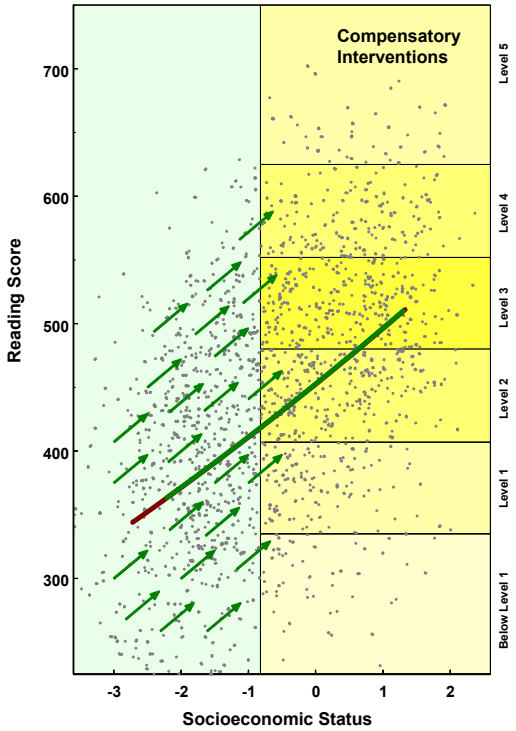
*Compensatory interventions* provide additional economic resources to students from low SES backgrounds. These could be considered a subset of SES-targeted interventions, as they target children from low SES families, rather than children with low cognitive performance. However, the emphasis is on improving the economic circumstances of children from poor families rather than providing a specialised curriculum or additional educational resources. The provision of transfer payments to poor families is a good example because it is one of the primary policy levers at the national level in many countries. The provision of free lunch programmes for children from poor families is another good example.

The distinction between compensatory interventions and other kinds is not always clear. For example, some jurisdictions, such as school districts, have compensatory funding formulas that allocate educational funds to schools differentially based on their SES. In some sense this is a compensatory

intervention, as it strives to compensate for the low SES of students in the targeted schools. However, within schools the funds might be used for SES-targeted, performance-targeted or universal interventions.

**Figure 5** shows the potential effect of a compensatory intervention for Argentina. In this example, we imagine a scenario whereby families with an SES below -0.82 were given transfer payments that were sufficient to increase their SES by 0.50 standard deviations. It is also assumed that the students in these families accordingly increased their academic performance so that it was consistent with other students of comparable SES in Argentina. This amounts to an increase in performance of about 22 points for students whose families received compensatory funding. The new gradient line (green line) is on top of the original line, indicating that the intervention did not raise or level the learning bar. The new gradient is shorter as the range of SES scores has been reduced. The unadjusted mean score for Argentina has increased by about 12 points as a result of the intervention, but the expected score of an average OECD student (i.e. the level of the gradient) remains the same.

**Figure 5. Compensatory interventions**



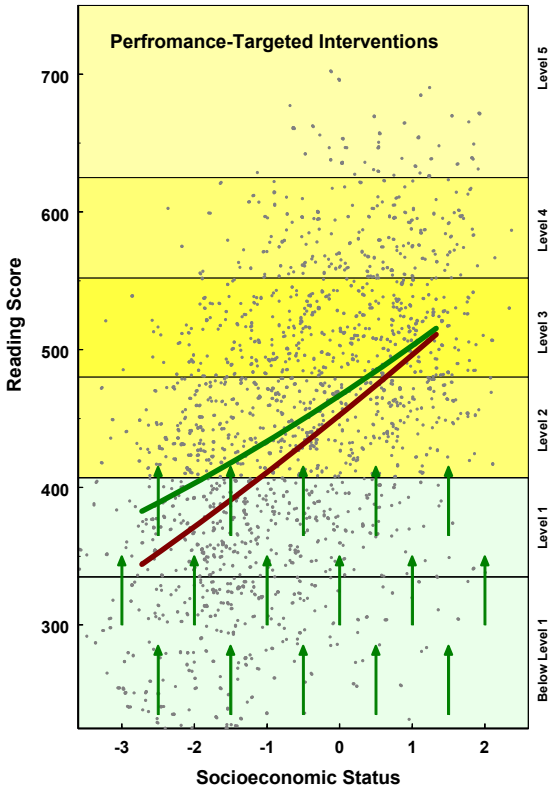

*Source:* PISA, 2000-2002.

*Performance-targeted interventions* provide a specialised curriculum or additional instructional resources for particular students based on their levels of academic performance. For example, in most schooling systems, students with special needs are provided with additional support through special education programmes. Generally these entail one-to-one or small group instruction from a specialist, either in the classroom or in a separate setting. Some schooling systems provide *early prevention* programmes that target children who are deemed at risk of school failure when they enter kindergarten or the first grade, while other systems provide *late prevention* or *recovery* programmes for children who fail to progress at a normal rate during the first few years of elementary school. Some performance-targeted programmes aim to improve children's capacity to learn by reducing maladaptive behaviour or improving self-esteem. These and other counselling and clinical programmes can also be placed in this category even though they are usually targeted towards children with certain behaviours rather than those with low academic performance. At the secondary school level, these programmes are often delivered in "alternative" schools. Some performance-targeted programmes aim to provide a modified curriculum for students with high academic performance or for gifted students.

More generally, programmes that track or stream students into different types of programmes can be considered performance-targeted interventions, because they strive to match curriculum and instruction to students' academic ability or performance. Grade repetition could be considered a performance-targeted intervention, because the decision to have a child repeat a grade is usually based mainly on school performance; however, in many cases grade repetition does not entail a modified curriculum or additional instructional resources and, therefore, would not fit the definition of a performance-targeted intervention.

**Figure 6** provides an example based on PISA data for Argentina. The arrows show that this intervention is targeted to students with low reading performance (in this case below Level 2). The hypothetical intervention has an effect of boosting the reading performance of the targeted youth by 50 points, which is one-half of a standard deviation. This would have the effect of raising and levelling the learning bar, as shown by the change from the PISA 2000 gradient (red line) to the new gradient (green line). An intervention that succeeded in improving the scores of low-performing students by this amount would increase the overall level of performance by about 21 points, and would substantially reduce inequalities along socioeconomic lines.

**Figure 6. Performance-targeted interventions**



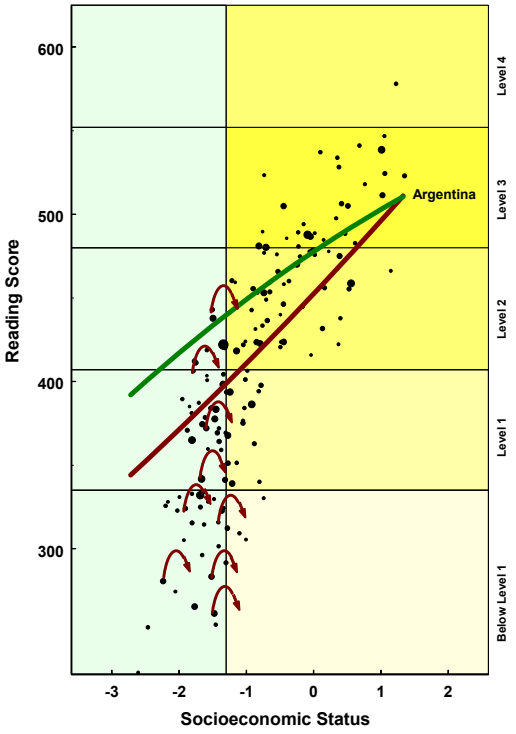

*Source:* PISA, 2000-2002.

*Inclusive interventions* strive to include marginalised students into mainstream schools and classrooms. Inclusive practices have often concentrated on including students with disabilities into regular classrooms rather than segregating them into special classes or schools. This report considers inclusive interventions more broadly to include reforms aimed at including any type of student who may be segregated, including students with disabilities, ethnic minorities and students from low SES families. Some inclusive interventions try to reduce between-school SES segregation by redrawing school catchment boundaries, for example, bussing students, amalgamating schools or creating magnet schools in low SES areas. Other inclusive interventions may attempt to reduce segregation among classes within schools, such as the integration of students with disabilities into the regular classroom.

An important policy issue is that programmes that track or stream students are usually at odds with efforts aimed at desegregating students. Similarly, programmes that provide parents with greater school choice through vouchers or open enrolment plans can result in greater segregation unless there are mandated quotas to ensure that students with disabilities, low SES or ethnic minorities are proportionally represented in all schools. Other kinds of special programmes, such as language immersion or charter schools, can also increase segregation. Segregation is also greater in systems where there is a strong private sector.

**Figure 7** provides an example for Argentina. In this case, we consider the effect on the learning bar if all students attending schools with an SES below -1.30 were reallocated to schools with SES scores above this threshold. This hypothetical intervention allows for very high SES schools, but ensures that there are no students attending very low SES schools. The result of the intervention is that students who were attending low SES schools receive a considerable boost in their scores – on average about 39 points. These changes in reading performance have a large effect on the socioeconomic gradient, which is evident in the figure by comparing the initial gradient (red line) with the gradient after the intervention (green line).

**Figure 7. Inclusive interventions**




*Source:* PISA, 2000-2002.

## PIRLS and PISA as tools to inform educational policy

A challenge for any international comparative study is to collect data that can provide strong indicators of student performance that can be used for international comparisons and routine monitoring, while also collecting data that can be used for research on the effects of educational policy and practice. One of the great strengths of both PIRLS and PISA is that they provide regular assessments of student performance alongside measures of student background. Therefore, countries can assess whether their "learning bar" is changing over time. Another important feature of these studies is that they enable countries to compare their results cross-nationally on key indicators describing

school resources and various structural and social features of schools and classrooms. As these studies progress, the core elements regarding performance and student background will be consistent across studies. However, new measures pertaining to different aspects of school and classroom policy and practice will be developed in an effort to increase our knowledge about "what works" in different kinds of educational systems.

The examples provided in the previous section are not presented to make a case that certain kinds of interventions are better than others. Rather, they are presented as a framework for thinking about what kinds of interventions might be most effective for particular countries and schooling systems. The argument is that certain kinds of interventions are appropriate for different schooling systems, depending on the relationship between performance and SES within and among schools, and how students with differing backgrounds are allocated to schools. Generally, the best "policy mix" for a country is likely to be some combination of two or more kinds of interventions. Socioeconomic gradients and school profiles are the starting points for a more thorough analysis of the relationships between schooling outcomes and the inputs and processes that affect these outcomes. The remainder of the report focuses on ten policy questions associated with raising and levelling the learning bar. They provide a framework for a set of analyses that can inform educational policy at the national and local levels.

# Question 1

# To what extent do countries and schools vary in their educational performance?

## Variation among students, schools and countries

The influential study of educational equity in the United States, *Equality of Educational Opportunity* (Coleman *et al.*, 1966), suggested that schools did not vary much in their outcomes after account was taken of students' family background. The report spawned two decades of debate about whether schools provided some "added value" to children's learning, over and above that associated with family background. The question was not whether more learning occurred at school than at home, but rather, whether some schools were more effective than others in producing high levels of educational achievement. Research during the 1980's and 1990's provided compelling evidence that schools do indeed vary in their outcomes, even when account is taken of differences in the family backgrounds and ability of students when they enter school (Rutter, 1983; Raudenbush and Bryk, 1986; Willms and Raudenbush, 1989).

In the course of this research, researchers were concerned with how best to estimate school effects, particularly how to integrate data at the individual and school levels (Burstein, 1980; Lau, 1979), and take into account measurement and sampling error (Aitkin and Longford, 1986; Goldstein, 1986; Raudenbush and Bryk, 1986). They developed powerful statistical models, which are now popularly called "hierarchical linear models", to estimate school effects. A school effect is defined as the effect on students' outcomes associated with attendance at a particular school, net of the effects associated with student family background and wider social and economic factors that lie outside the control of teachers or school administrators (Raudenbush and Willms, 1995).

Hierarchical linear models enable researchers to make comparisons among schools or other units, such as classrooms, school districts or countries, while taking account of measurement and sampling error. The most basic model, called a "null model", simply partitions the variation in a variable of interest into within- and between-group components. It also provides estimates of the extent to which units vary at each level. In the case of PIRLS and PISA, we can examine the extent of variation in literacy performance among countries, among schools within countries, and within schools.

**Table 1. Variation among students, schools and countries in reading performance**

| | PIRLS 2001 | | PISA 2000-2002 | |
|---|---|---|---|---|
| **Coefficients** | **Estimate** | **Standard error** | **Estimate** | **Standard error** |
| Grand Mean | 502 | (9.3) | 495 | (5.3) |
| **Variance components** | **Standard deviation** | **Variance (%)** | **Standard deviation** | **Variance (%)** |
| Students | 66.3 | 4,396 (48.2%) | 78.4 | 6,149 (60.7%) |
| Schools | 43.4 | 1,881 (20.6%) | 56.9 | 3,237 (31.9%) |
| Countries | 53.4 | 2,848 (31.2%) | 27.4 | 749 (7.4%) |

*Source:* PIRLS 2001 and PISA 2000.

**Table 1** presents results for a null model fitted to the data for reading performance in PIRLS and PISA. In PIRLS, the reading literacy scores were scaled to have a mean of 500 and a standard deviation of 100 at the student level, for *all* participating countries. The grand mean score estimated with the null model is 502. It is not exactly 500 because the multilevel model takes into account how accurately each school and country mean is estimated and weights each unit accordingly. Similarly, the grand mean for PISA is 495. Note that in PISA the scores were scaled to have a mean of 500 and a standard deviation of 100 at the student level for *all OECD countries*, not all countries, as in PIRLS. Thus, we cannot directly compare levels of scores across the two studies because of their different scales. Also, in the next section, when the scores for each country are displayed, the mean scores for countries that participated in both studies are lower in PISA than in PIRLS, but this is simply an artifact of the scaling method.

**PIRLS**

The results for the null model for PIRLS indicate that there is significant variation in reading performance at all three levels of the schooling system. The total variance (4396 + 1881 + 2848) is 9125, which corresponds to a standard deviation of 95.5.

At the student level, the variance is 4396 and the standard deviation is 66.3. We cannot infer directly from PIRLS the meaning of score points in terms of grade

equivalents; however, a rule-of-thumb with North American studies is that one grade level is about one standard deviation at the elementary level. For PIRLS, on average, the standard deviation within countries is about 79.7 points, while for the United States the standard deviation is 83.2 points. For discussion purposes, I will consider 80 points to be equivalent to about one grade level. Therefore, the average within-school standard deviation is roughly equivalent to about (66.3 divided by 80 =) 0.82 grade levels. This means that in a typical school, the scores of grade 4 students vary considerably: about two-thirds would have grade equivalent scores ranging from (3.9 – 0.8 =) 3.1 to (3.9 + 0.8 =) 4.7, while about 95% would have grade equivalent scores ranging from (3.9 – 1.6 =) 2.3 to (3.9 + 1.6 =) 5.5. However, we will see in the next section that the PIRLS reading scores are negatively skewed in nearly all countries. Therefore, the range of scores in the typical school likely ranges from well below a grade equivalent of 2.3 in most countries.

The variance at the school level is 1881, with a standard deviation of about 43.4 points. This represents substantial variation. It indicates that about 95% of schools have mean scores in a range from 416 to 588. Using the rough grade equivalent estimate of 80 points, the range of school mean scores spans more than two grade levels.

At the country level, the variance is 2848 and the standard deviation is 53.4. This is greater than the variance among schools within countries. It suggests that the range of scores is more than 200 points, or about two-and-a-half grade levels.

In percentage terms, 31.2% of the variation is among countries, while 20.6% is among schools within countries and 48.2% is among students within schools.

**PISA**

The results for PISA also indicate that there is considerable variation among schools and countries in the reading performance of students at age 15, although there is considerably less variation among countries. For PISA, only 7.4% of the variation in reading performance is among countries, while 31.9% is among schools within countries and 60.7% is among students within schools.

In PISA it is possible to approximately estimate a grade-equivalent score, because in some countries 15-year-old youth span two grades, by virtue of the cut-off date used for entry to primary school. In 12 OECD countries, it was possible to identify youth who were in either a lower grade or an upper grade, based on their birth date. For example, in the Czech Republic, most youth who were born between January and August 1984 were in grade 10 at the time of the PISA assessment, while the majority of those born between September and December 1984 were in grade 9. An estimate of the “grade effect” on PISA results in the Czech Republic can therefore be obtained by comparing the results of the youth in these two grades, excluding those who had not reached grade 9 or 10 on schedule, in most cases because they had been retained a grade.

Multilevel analyses (students nested within schools nested within countries) were conducted to estimate the grade effect and the maturity effect on PISA reading scores. For the 12 countries where 15-year-old youth spanned two grade levels, the grade effect was 34.3 points (standard error = 3.5). In this report 35 points on the PISA scale is considered to be roughly equivalent to one grade level.

At the student level, the variance among reading scores is 6149, corresponding to a standard deviation of 78.4. Thus, the range of scores for 15-year-olds in each school spans more than four grade levels above and below the mean. Within countries, the variance at the school level is 3237 and the standard deviation is 56.9. This suggests that there is considerable variation among schools within countries. On average, if we consider the range of two standard deviations above and below the mean, this is equivalent to about three full grade levels above and below the mean. Finally, the variance among countries is 749, with a standard deviation of 27.4. Thus, the average scores of the top-scoring countries are more than three full grade equivalents above the lowest performing countries.

### Variation among countries

**Table 2** presents summary statistics describing the distribution of scores for each country that participated in PIRLS and PISA.

**Table 2. Indicators of reading literacy**

| | PIRLS 2001 | | | PISA 2000-02 | | |
|---|---|---|---|---|---|---|
| | **Mean** | **Standard deviation** | **Skewness** | **Mean** | **Standard deviation** | **Skewness** |
| **OECD countries** | | | | | | |
| Australia | | | | **528.4** (3.5) | **101.7** (1.6) | **-.22** (.04) |
| Austria | | | | **507.1** (2.4) | 93.0 (1.6) | **-.38** (.05) |
| Belgium | | | | 507.1 (3.6) | **107.0** (2.4) | **-.50** (.05) |
| Canada | **544.1** (2.4) | **71.9** (2.4) | **-.23** (.11) | **534.3** (1.6) | 94.6 (1.0) | **-.26** (.04) |
| Czech Republic | **536.9** (2.3) | **64.6** (2.4) | **-.38** (.15) | **491.5** (2.4) | 96.3 (1.9) | **-.44** (.07) |
| Denmark | | | | 496.9 (2.4) | 98.0 (1.8) | **-.35** (.07) |
| Finland | | | | **546.5** (2.6) | **89.4** (2.6) | **-.45** (.14) |
| France | **525.2** (2.4) | **70.5** (2.4) | -.23 (.13) | 504.6 (2.7) | 91.8 (1.7) | **-.27** (.05) |
| Germany | **539.1** (1.9) | **67.3** (1.9) | **-.37** (.15) | **484.0** (2.5) | **111.3** (1.9) | **-.49** (.07) |
| Greece | **524.2** (3.6) | 73.3 (3.7) | **-.34** (.16) | **473.8** (5.0) | 97.1 (2.7) | **-.24** (.07) |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Hungary | **543.2** | (2.2) | **65.8** | (2.2) | **-.35** | (.12) | **479.9** | (4.0) | 93.8 | (2.1) | **-.24** | (.01) |
| Iceland | **512.4** | (1.2) | **74.7** | (1.3) | **-.28** | (.06) | **506.9** | (1.5) | 92.4 | (1.4) | **-.32** | (.04) |
| Ireland | | | | | | | **526.6** | (3.2) | 93.6 | (1.7) | **-.37** | (.04) |
| Italy | **540.7** | (2.4) | **71.0** | (2.4) | **-.36** | (.13) | **487.4** | (2.9) | 91.4 | (2.7) | **-.32** | (.11) |
| Japan | | | | | | | **522.3** | (5.2) | **85.7** | (3.0) | **-.50** | (.07) |
| Korea | | | | | | | **524.8** | (2.4) | **69.5** | (1.6) | **-.44** | (.06) |
| Luxemburg | | | | | | | **441.0** | (1.5) | **100.5** | (1.5) | **-.48** | (.08) |
| Mexico | | | | | | | **422.1** | (3.3) | **85.9** | (2.1) | .08 | (.06) |
| Netherlands | **554.2** | (2.4) | **57.3** | (2.4) | -.20 | (.18) | **531.9** | (3.4) | **88.6** | (2.7) | **-.39** | (.07) |
| New Zealand | **528.8** | (3.7) | **93.4** | (3.8) | **-.38** | (.14) | **528.8** | (2.8) | **108.2** | (2.0) | **-.36** | (.05) |
| Norway | 499.2 | (2.9) | 81.1 | (3.0) | **-.47** | (.13) | **505.4** | (2.8) | **103.6** | (1.7) | **-.44** | (.04) |
| Poland | | | | | | | **479.1** | (4.5) | 99.8 | (3.1) | **-.32** | (.07) |
| Portugal | | | | | | | **470.1** | (4.5) | 97.1 | (1.8) | **-.23** | (.05) |
| Slovak Republic | **518.1** | (2.8) | **70.2** | (2.7) | **-.51** | (.16) | | | | | | |
| Spain | | | | | | | **492.7** | (2.7) | **84.8** | (1.2) | **-.32** | (.04) |
| Sweden | **561.0** | (2.2) | **65.8** | (2.2) | **-.36** | (.12) | **516.3** | (2.2) | **92.2** | (1.2) | **-.31** | (.04) |
| Switzerland | | | | | | | **494.5** | (4.2) | **102.1** | (2.0) | **-.29** | (.04) |
| Turkey | **449.4** | (3.5) | 86.2 | (3.5) | -.17 | (.12) | | | | | | |
| United Kingdom | | | | | | | **523.5** | (2.6) | **100.5** | (1.5) | **-.22** | (.04) |
| England | **552.9** | (3.5) | 86.5 | (3.5) | **-.38** | (.14) | | | | | | |
| Scotland | **528.2** | (3.5) | 84.2 | (3.5) | **-.36** | (.17) | | | | | | |
| United States | **542.2** | (3.8) | 83.2 | (3.8) | **-.53** | (.15) | 504.3 | (7.0) | **104.7** | (2.7) | **-0.24** | (0.05) |
| **Non-OECD countries** | | | | | | | | | | | | |
| Albania | | | | | | | **348.9** | (3.3) | **99.4** | (1.9) | **-.13** | (.06) |
| Argentina | **419.5** | (5.9) | **95.6** | (5.9) | -.13 | (.18) | **418.2** | (9.9) | **108.5** | (3.4) | **-.21** | (.09) |
| Belize | **326.8** | (5.1) | **105.6** | (5.1) | .16 | (.13) | | | | | | |
| Brazil | | | | | | | **396.1** | (3.1) | **86.2** | (1.9) | .05 | (.06) |
| Bulgaria | **550.5** | (3.8) | 82.5 | (3.8) | **-.55** | (.22) | **430.4** | (4.9) | **101.6** | (3.0) | -.13 | (.07) |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Chile | | | | | | **409.6** (3.6) | **89.7** (1.7) | **-.12** (.06) |
| Colombia | **422.4** (4.3) | 80.5 (4.4) | -.08 (.18) | | | | | |
| Cyprus | **494.0** (2.9) | 81.3 (2.8) | **-.31** (.12) | | | | | |
| Hong Kong (SAR of China) | **527.9** (3.1) | **62.8** (3.1) | **-.47** (.18) | | | **525.4** (2.9) | **83.9** (2.4) | **-.62** (.07) |
| Indonesia | | | | | | **370.5** (4.0) | **72.5** (2.5) | -.05 (.05) |
| Iran, Islamic Rep | **413.8** (4.4) | **92.2** (4.3) | -.10 (.11) | | | | | |
| Israel | **508.9** (2.9) | **93.6** (2.8) | **-.47** (.12) | | | **452.8** (8.4) | **109.3** (4.0) | **-.29** (.07) |
| Kuwait | **396.5** (4.3) | **89.0** (4.2) | -.20 (.15) | | | | | |
| Latvia | **544.6** (2.3) | **61.5** (2.2) | -.26 (.13) | | | **457.9** (5.3) | **102.2** (2.3) | **-.21** (.06) |
| Liechtenstein | | | | | | **482.3** (4.1) | 96.3 (3.9) | **-.35** (.15) |
| Lithuania | **543.4** (2.6) | **64.3** (2.6) | **-.32** (.14) | | | | | |
| Macedonia, The Former Yugoslav Republic of | **441.6** (4.8) | **103.1** (4.8) | **-.28** (.13) | | | **372.5** (1.9) | 93.6 (1.2) | **-.11** (.05) |
| Moldova, Rep | **491.7** (4.1) | 75.2 (4.1) | -.25 (.18) | | | | | |
| Morocco | **349.5** (9.9) | **114.9** (9.9) | .20 (.36) | | | | | |
| Peru | | | | | | **326.8** (4.4) | 96.0 (2.2) | .11 (.07) |
| Romania | **511.7** (4.7) | **89.8** (4.7) | **-.41** (.19) | | | | | |
| Russian Federation | **527.9** (4.3) | **66.4** (4.4) | -.51 (.42) | | | **461.7** (4.2) | 92.0 (1.8) | **-.13** (.04) |
| Singapore | **527.9** (5.2) | 91.8 (5.3) | **-.71** (.24) | | | | | |
| Slovenia | 501.5 (1.9) | **71.7** (2.0) | **-.32** (.12) | | | | | |
| Thailand | | | | | | **430.3** (3.1) | **76.6** (1.7) | -.09 (.06) |

**Note.** Standard errors are indicated in parentheses. Means that differ significantly from the international mean (500) are indicated with bold text. Standard deviations that differ from the average standard deviation across countries (79.7 for PIRLS, 94.8 for PISA) are indicated with bold text. The distributions that are significantly asymmetrical (i.e. skewness significantly greater than or less than zero) are also indicated with bold text.

*Source:* IEA PIRLS 2001 and OECD PISA 2000.

## PIRLS

All OECD countries, except Turkey, scored above the international average in PIRLS. Among the non-OECD countries, eight countries/territories scored significantly above the international average: Bulgaria; Hong Kong (Special Administrative Region of China), China; Israel; Latvia; Lithuania; Romania; the Russian Federation; and Singapore. Nine non-OECD countries scored significantly below the international average, while the mean score for Slovenia did not differ significantly from the international average. The average standard deviation for PIRLS countries was 79.7. The spread of scores for New Zealand and eight of the 18 non-OECD countries was significantly greater than the average of all PIRLS countries. In contrast, the standard deviations were significantly less than the PIRLS average in 10 of the OECD countries and only five of the non-OECD countries. The third column of Table 2 indicates the skewness of the distribution of scores for each country. Skewness pertains to the extent to which a distribution is assymetrical. Distributions that are negatively skewed have low scores that extend further below the mean than the high scores extend above it; the reverse is the case for positively skewed distributions. The scores are significantly negatively skewed in 14 of the 17 OECD countries and in nine of the 18 non-OECD countries.

Taken together, the mean reading score (at the country level) for OECD countries is 529.4, the average standard deviation is 74.5 and the average skewness is -0.35. In contrast, the average mean score for the non-OECD countries is 472.2, the standard deviation is 84.6 and the skewness is -0.28. The results suggest that the distributions of scores for wealthier countries are higher than those of lower-income countries. However, an important aspect of their advantage is that they tend to have fewer students with very low scores, thereby having a high mean score but with less spread.

The results also reveal that countries with high mean scores tend to have smaller standard deviations: the correlation is -0.73. Also, countries with high mean scores have distributions that are less skewed: the correlation is -0.77. Therefore, countries with higher scores are more homogeneous in their achievement, with fewer students with very low scores.

## PISA

The average of the means for OECD countries is 501.1, compared with 420.2 for non-OECD countries. However, the standard deviation of scores in OECD countries (95.2) is similar to that of non-OECD countries (93.4). As in PIRLS, the average skewness of scores in OECD countries (-0.33) is considerably larger than that of non-OECD countries (-0.16).

The correlation between country means and standard devaitions for PISA is 0.07, which is considerably different for that oberved for PIRLS. The means scores are negatively correlated with the skewness though; the correlation is -0.74, which is similar to that of PIRLS.

## Policy implications

"Do schools vary in their educational performance?" The answer is unequivocal: there are large and statistically significant differences among schools in their performance within and among countries. At grade 4, countries with the highest mean scores tend to be more homogeneous in their achievement and have fewer students with very low scores. At age 15, the differences among schools and among countries are larger in absolute terms, and as at grade 4, the most successful countries are those with fewer students with very low scores. In other words, the countries with the highest scores tend to achieve their high performance not simply by raising all students' scores uniformly but also by reducing the number of children with very low scores.

One way to consider these results is that there is a critical transition from "learning-to-read" to "reading-to-learn". For most students this happens at about age 8 or 9, typically by the end of the third grade. If children are not able to read with ease and understand what they are reading when they enter fourth grade, they are less able to take advantage of the learning opportunities that lie ahead. A critical indicator for countries therefore is the percentage of children that are able to make this transition successfully.

## Question 2

## Is there a significant relationship between reading performance and socioeconomic status?

Socioeconomic gradients in children's outcomes are evident soon after birth. For example, children born into high SES families tend to have a better "temperament" than those born into low SES families (Japel, Normand, Tremblay and Willms, 2002; Sameroff, Seifer and Elias; 1982). SES gradients are also evident during the pre-school period for both behavioural and cognitive outcomes (Hertzman and Weins, 1996; Willms, 2002). Researchers who have tracked children's early vocabulary development, which is an important precursor to reading skills, have found that growth trajectories differ for children from differing socioeconomic backgrounds (Hart and Risely, 1995). When children enter school, the gradient is well established for both cognitive skills and behaviour (Willms, 2002, 2003). During the primary and middle school years, children are less likely to do well in academic pursuits or be engaged in curricular and extra-curricular school activities if their parents have low incomes, low levels of education or are unemployed or working in low-prestige occupations (Datcher, 1982; Finn and Rock, 1997; Johnson, Crosnoe and Elder, 2001; Voelkl, 1995). Children from low SES backgrounds are also more prone to leaving school early (Cairns, Cairns and Neckerman, 1989; Crane, 1991; Ensminger and Slusarcick, 1992; Janosz et al., 1997; Rumberger, 1995), and are less likely to successfully enter the labour market or pursue post-secondary training (Raudenbush and Kasim, 1998).

The early educational research on gradients was primarily concerned with the extent to which individuals' occupational attainment was determined by the socioeconomic positions of their parents and the extent to which educational attainment mediated that relationship (Bielby, 1981; Sewell and Hauser, 1975). During the late 1970's and 1980's, researchers began to question whether gradients could be altered through educational policy and reform (Heath, 1990; McPherson and Willms, 1986, 1987). Another strand of research was concerned with whether the effects of schools and school programmes varied for children of differing ethnic backgrounds and family circumstances (Gamoran, 1990, 1985; Raudenbush and Willms, 1995). The main focus of the research on gradients has been on the correlation between achievement and SES (White, 1982; Sirin, 2005), without attention to the three components of gradients – their level, slope and strength.

The most basic hypothesis about the slopes of SES gradients is that there is a significant bivariate relationship between social outcomes and SES. In the case of school reading performance, the hypotheses can be stated as: "There is a significant relationship between student reading performance and SES." In the simplest case, this hypothesis can be tested for a continuous outcome measure, such as reading performance using ordinary least squares regression analysis.[2] For example, the slope indicting the magnitude of the relationship between reading performance and SES for Argentina in PISA *(see Figure 1)* was 47.7, which is statistically significant ($p < 0.05$).

## Socioeconomic gradients for PIRLS and PISA

**Tables 3** and **4** provide the estimates of the socioeconomic gradients for PIRLS and PISA respectively. These are displayed graphically in **Figures 8** and **9**. In many countries, the relationship between reading performance and SES is curvilinear and therefore a second term denoting the square of SES was included in the model. Therefore, some gradients appear as curved lines in the two figures. The results show clearly that the levels and the slopes vary considerably among countries, for grade 4 students assessed in PIRLS and for 15-year-old students assessed in PISA.

---

[2] The ordinary least squares regression is given by:

$$Y_i = \beta_0 + \beta_1 X_i + r_i \qquad \textbf{(1)}$$

where $\boldsymbol{Y_i}$ is a student's reading score, and $\boldsymbol{X_i}$ is the student's SES. The intercept, $\boldsymbol{ß_0}$, is the expected outcome score for a student who has a score of zero on $\boldsymbol{X_i}$. The coefficient, $\boldsymbol{ß_1}$, is the slope of the socioeconomic gradient. The parameters, $\boldsymbol{r_i}$, are the student-level residuals; that is, the deviation of students' scores from the regression line. The *strength* of the gradient, called $R^2$, is the proportion of variance in the outcome measure explained by SES; it is the difference between the variance in $\boldsymbol{Y_i}$ and the variance of the residuals, expressed as a fraction of the variance in $\boldsymbol{Y_i}$.

The gradient hypothesis is:

$$H_0: \beta_1 = 0$$
$$H_1: \beta_1 \neq 0 \qquad \textbf{(2)}$$

The statistical significance of $\boldsymbol{ß_i}$ depends on the magnitude of the standard error, and is assessed with a t-test with n-1 degrees of freedom. With both PIRLS and PISA, due to the nature of the sampling design and the use of multiple reading booklets, the calculation of the standard error requires special programming to use the replicate design weights (jackknife weights for PIRLS, and balanced repeated replicate weights for PISA). Thus, the standard errors cannot be obtained directly from most standard statistical packages.

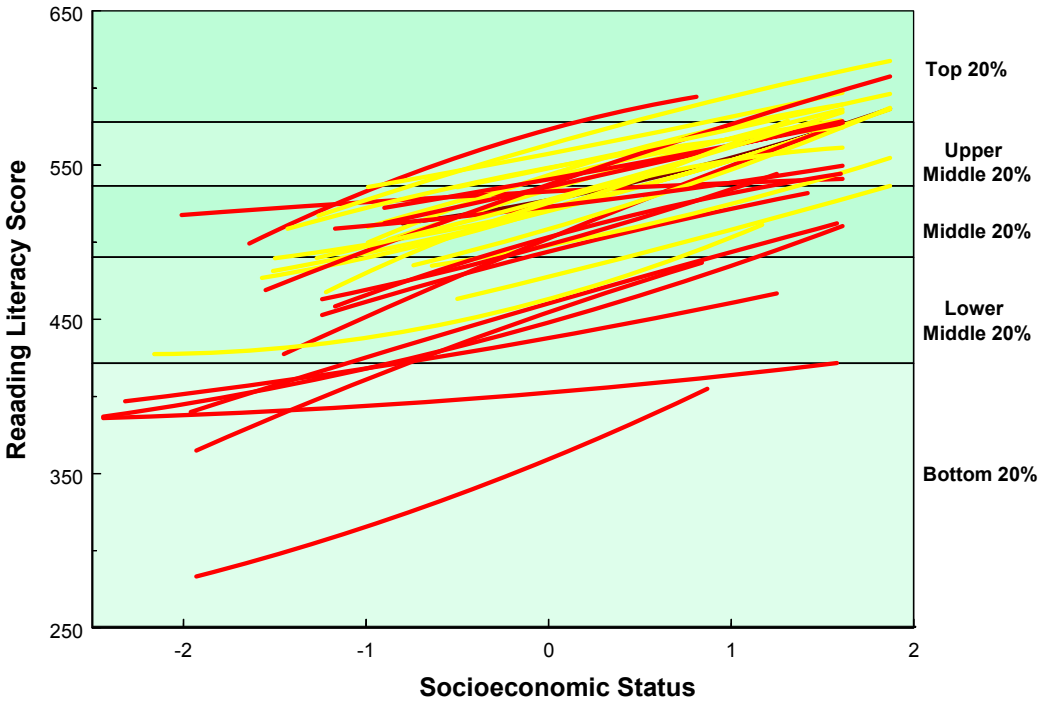
## Table 3. Gradient specifications for reading performance, PIRLS

| | SES-adjusted mean (standard error) | | SES slope (standard error) | | SES-squared slope (standard error) | | R-squared (%) |
|---|---|---|---|---|---|---|---|
| **OECD countries** | | | | | | | |
| Canada | **527.6** | (2.1) | **22.3** | (2.6) | **5.2** | (1.5) | **12.4** |
| Czech Republic | **542.1** | (2.3) | **27.4** | (2.4) | | | **9.8** |
| France | **522.7** | (2.09) | **27.7** | (1.7) | **3.5** | (1.3) | **16.7** |
| Germany | **533.8** | (1.6) | **34.8** | (1.4) | | | **14.9** |
| Greece | **520.0** | (4.1) | **30.2** | (2.0) | | | **16.4** |
| Hungary | **543.2** | (2.0) | **37.8** | (2.4) | **-5.2** | (2.1) | **20.8** |
| Iceland | **496.9** | (1.7) | **24.3** | (2.9) | | | **9.2** |
| Italy | **548.7** | (2.2) | **22.1** | (1.4) | **-3.9** | (1.5) | **8.9** |
| Netherlands | **558.4** | (2.2) | **22.7** | (2.1) | | | **11.6** |
| New Zealand | **508.8** | (4.3) | **33.3** | (6.0) | | | **14.0** |
| Norway | **477.8** | (3.5) | **30.3** | (4.9) | | | **9.8** |
| Slovak Republic | **526.5** | (2.5) | **36.7** | (3.1) | **-9.0** | (2.7) | **17.4** |
| Sweden | **548.6** | (2.3) | **23.1** | (2.8) | | | **8.9** |
| Turkey | **462.1** | (3.5) | **32.2** | (2.6) | **8.0** | (1.6) | **11.0** |
| England | **563.7** | (3.2) | **33.9** | (3.1) | | | **12.7** |
| Scotland | **530.5** | (4.1) | **32.8** | (2.7) | | | **12.9** |
| **Non-OECD countries** | | | | | | | |
| Argentina | **446.6** | (5.3) | **33.3** | (2.8) | **3.5** | (1.5) | **11.3** |
| Belize | **359.0** | (8.2) | **50.2** | (7.2) | | | **9.2** |
| Bulgaria | **572.8** | (3.4) | **30.9** | (3.0) | **-8.2** | (3.6) | **14.8** |
| Colombia | **437.5** | (6.8) | **21.1** | (5.7) | | | **6.2** |
| Cyprus | **493.0** | (3.5) | **29.8** | (2.3) | | | **6.4** |
| Hong Kong (SAR of China) | **533.1** | (2.8) | **5.9** | (1.7) | | | **2.8** |
| Iran, Islamic Rep | **455.2** | (4.7) | **39.9** | (2.4) | **-3.0** | (1.5) | **17.7** |
| Israel | **501.0** | (4.5) | **48.9** | (3.3) | | | **11.8** |
| Kuwait | **403.2** | (4.1) | **8.8** | (1.6) | | | **2.0** |
| Latvia | **541.5** | (2.5) | **20.2** | (2.2) | | | **8.7** |
| Lithuania | **536.1** | (2.6) | **27.9** | (2.9) | | | **9.8** |
| Macedonia, The Former Yugoslav Republic of | **459.0** | (5.4) | **35.1** | (2.6) | | | **11.0** |
| Moldova, Rep | **497.8** | (4.2) | **31.4** | (3.3) | | | **10.9** |
| Romania | **521.1** | (3.8) | **37.0** | (3.6) | **6.0** | (1.8) | **11.3** |
| Russian Federation | **523.5** | (4.2) | **13.4** | (2.5) | | | **3.5** |
| Singapore | **537.4** | (4.2) | **41.0** | (2.9) | | | **23.6** |
| Slovenia | **502.3** | (1.9) | **32.7** | (1.8) | **-3.4** | (1.6) | **13.8** |

**Note.** Coefficients that are statistically significant are indicated with bold text. Data on the socioeconomic background of students in the United States and Morocco were not available.

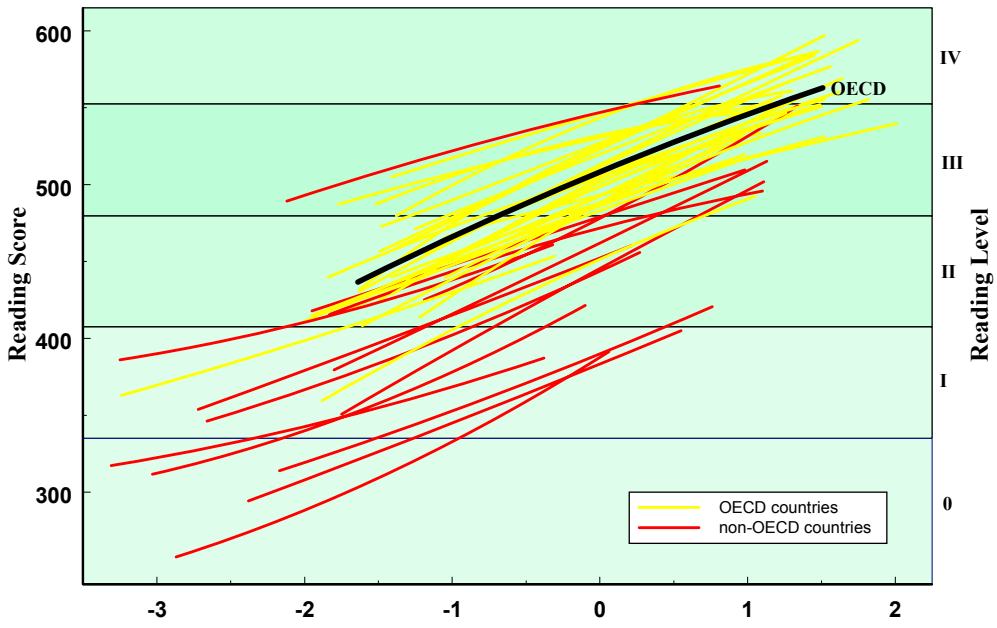
*Source:* IEA PIRLS 2001.

**Figure 8. Variation among countries in their socioeconomic gradients**




*Source:* PIRLS, 2001.

**Figure 9. Variation among countries in their socioeconomic gradients**




*Source:* PISA, 2000-2002.

**Table 4. Gradient specifications for reading performance, PISA**

| | SES-adjusted mean (standard error) | | SES slope (standard error) | | SES-squared slope (standard error) | | R-squared (%) |
|---|---|---|---|---|---|---|---|
| **OECD countries** | | | | | | | |
| Australia | **513.4** | (3.1) | **46.1** | (2.4) | | | **17.2** |
| Austria | **507.6** | (2.6) | **41.4** | (2.3) | **-6.0** | (1.6) | **13.7** |
| Belgium | **520.1** | (2.9) | **47.7** | (2.3) | **-10.5** | (1.5) | **22.4** |
| Canada | **527.3** | (1.5) | **36.9** | (1.3) | **-2.0** | (1.0) | **11.2** |
| Czech Republic | **501.5** | (2.2) | **49.3** | (2.2) | **-5.0** | (1.8) | **21.8** |
| Denmark | **498.3** | (2.3) | **41.9** | (2.1) | **-4.6** | (1.6) | **16.8** |
| Finland | **544.4** | (2.1) | **29.8** | (2.4) | | | **8.9** |
| France | **511.1** | (2.4) | **47.8** | (2.2) | | | **22.8** |
| Germany | **473.1** | (3.6) | **59.9** | (3.4) | | | **21.9** |
| Greece | **484.5** | (4.1) | **37.5** | (3.1) | | | **15.7** |
| Hungary | **486.5** | (3.5) | **53.9** | (2.9) | | | **26.2** |
| Iceland | **492.0** | (2.1) | **23.9** | (2.3) | | | **7.2** |
| Ireland | **526.9** | (2.9) | **37.6** | (2.2) | | | **13.7** |
| Italy | **485.1** | (3.1) | **31.6** | (2.3) | | | **11.0** |
| Japan | **530.3** | (4.2) | **23.7** | (2.9) | **-6.6** | (2.5) | **8.1** |
| Korea | **532.4** | (2.2) | **23.1** | (2.3) | | | **9.3** |
| Luxemburg | **448.5** | (2.1) | **45.6** | (1.7) | **-2.7** | (1.2) | **25.5** |
| Mexico | **459.5** | (3.1) | **35.1** | (2.5) | **2.1** | (0.8) | **22.5** |
| Netherlands | **533.9** | (3.1) | **38.0** | (2.6) | **-4.8** | (2.1) | **14.6** |
| New Zealand | **523.2** | (2.5) | **44.9** | (2.3) | | | **17.4** |
| Norway | **488.3** | (3.1) | **41.7** | (1.8) | **-2.6** | (1.3) | **13.6** |
| Poland | **495.0** | (4.4) | **37.8** | (3.3) | | | **16.8** |
| Portugal | **487.1** | (3.8) | **40.6** | (2.1) | | | **20.4** |
| Spain | **504.3** | (2.2) | **31.9** | (1.5) | **-2.9** | (1.1) | **16.6** |
| Sweden | **504.0** | (2.0) | **35.9** | (1.9 | | | **10.9** |
| Switzerland | **499.9** | (3.5) | **49.3** | (2.2) | **-6.3** | (1.4) | **19.2** |
| United Kingdom | **519.9** | (2.3) | **48.8** | (1.9) | | | **20.3** |
| United States | **498.0** | (3.1) | **47.8** | (3.0) | | | **21.1** |
| **Non-OECD countries** | | | | | | | |
| Albania | **381.8** | (3.2) | **39.3** | (3.0) | | | **18.2** |
| Argentina | **453.7** | (8.7) | **41.2** | (3.4) | | | **22.5** |
| Brazil | **433.8** | (3.3) | **38.8** | (2.7) | **4.1** | (0.9) | **20.1** |
| Bulgaria | **448.5** | (3.7) | **53.1** | (4.4) | | | **23.5** |
| Chile | **441.7** | (3.3) | **40.9** | (2.2) | **2.3** | (1.2) | **23.4** |
| Hong Kong | **545.7** | (3.1) | **27.5** | (3.9) | | | **9.9** |
| Indonesia | **419.1** | (7.4) | **33.2** | (5.6) | **3.1** | (1.2) | **10.7** |
| Israel | **453.6** | (6.1) | **46.7** | (4.3) | | | **17.1** |
| Latvia | **471.6** | (5.0) | **29.0** | (3.1) | **-4.5** | (1.9) | **10.2** |
| Liechtenstein | **480.9** | (5.3) | **47.4** | (6.3) | | | **18.2** |
| Macedonia | **391.4** | (1.6) | **35.7** | (1.8) | | | **20.0** |
| Peru | **383.1** | (4.3) | **50.0** | (3.6) | **4.4** | (1.4) | **23.8** |
| Russian Federation | **479.9** | (3.1) | **33.0** | (2.7) | | | **13.4** |
| Thailand | **466.5** | (4.4) | **32.1** | (3.7) | **4.6** | (1.1) | **9.9** |

*Source:* OECD PISA 2000-2002.

## Within- and between-school gradients

The national gradient for a country can be decomposed into a within-school and a between-school gradient. The relative strength of these relationships has several implications for educational policy which will be discussed later. One can estimate the average within-school gradient and the between-school gradient within a multilevel framework.[3] In this case, the "gradient hypothesis" pertains to the average within-school slope rather than to the overall slope for a country.

**Table 5** displays estimates for three separate three-level hierarchical linear models of reading performance for the PIRLS countries. **Table 6** displays the corresponding results for PISA 2001. The first model in each table is the null model, which was discussed above. It has no student or school-level variables; it simply partitions the variation in student performance into the components of student-, school- and country-level variation. The second model is the SES-

---

[3] In a two-level multilevel framework, a separate regression equation is fit to the data for each school:

$$Y_i = \beta_0 + \beta_1 X_i + r_i$$ Within-school equation **(3)**

where $Y_i$ is a student's reading score, and $X_i$ is his or her score on the SES measure. The intercept, $ß_0$, is the expected reading score for a student who has a score of zero on $X_i$. In most multilevel models, $X_i$ is "centred" on a particular value, such as the national mean, so that a value of zero on X refers to a hypothetical student with particular characteristics. In these analyses, SES is centred on the OECD international mean, and thus the $ß_0$ for each school is the level of gradient for an average OECD student. The parameter $ß_1$ is the slope of the socioeconomic gradient. It is an estimate of the expected change in the outcome score $Y_i$ for a one-unit change in $X_i$. The parameters, $r_i$, are the residuals; that is, the deviation of students' scores from the regression line. The strength of the gradient, as gauged by the proportion of variance in the outcome measure explained by SES (i.e., $R^2$), is the difference between the variance in $Y_i$ and the variance of the residuals expressed as a fraction of the variance in $Y_i$.

With j schools, one can write j such equations:

$$Y_{ij} = \beta_{oj} + \beta_{1j} X_{ij} + r_{ij}$$ A set of within-school equations **(4)**

where the subscript j has been added to each element of equation 1. Therefore, one now has j different $ß_0$'s, one for each school, and j different $ß_1$'s. The $ß_{0j}$'s are the levels of the socioeconomic gradients, and the $ß_{1j}$'s are the slopes of the socioeconomic gradients for the set of schools. The $ß_{0j}$'s are regressed at the second level on the school mean SES:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} \overline{X}_{j\bullet} + u_{0j}$$ Among-school equation for levels of the gradients **(5)**

where $\gamma_{00}$ is the mean of the adjusted school means $\gamma_{01}$ is the "compositional effect" associated with the mean SES of the school, and $\gamma_{01}$ is the deviation of each school's mean from the grand mean. Similarly, the slopes of the gradients vary among schools, and can be expressed as an average slope plus a deviation from the average slope:

$$\beta_{1j} = \gamma_{10} + u_{1j}$$ Among-school equation for slopes of the gradients **(6)**

where $\gamma_{10}$ is the mean of the within-school slopes, and $u_{1j}$ is the deviation of each school's slope from the mean slope.

The "gradient hypothesis" is that the average within-school socioeconomic gradient across the schools is statistically significant; that is, that $\gamma_{10}$ is significantly different from zero:

$$H_0: \gamma_{10} = 0$$ Gradient Hypothesis for within-school slopes **(7)**

$$H_1: \gamma_{10} \neq 0$$

This is assessed with a t-test with j-1 degrees of freedom. Note that in this case, the slopes were allowed to vary; that is, there is a different slope for each school.

gradient model which includes SES and SES-squared at the student level, and school mean SES at the school level. The third model, which is discussed later, is identical to the SES gradient model, except that it also includes the standard deviation of SES within each school. In estimating these models, the coefficients for SES and SES-squared were allowed to vary among schools within each country and among countries. Similarly, the estimates of the effects of school mean SES and the standard deviation of SES were allowed to vary among countries.

**Table 5. Hierarchical linear models describing reading performance among students, schools and countries, PIRLS**

| | Null model | | SES gradient model | | Relative deprivation model | |
|---|---|---|---|---|---|---|
| **Coefficients** | ***Esti-mate*** | ***(SE)*** | ***Esti-mate*** | ***(SE)*** | ***Esti-mate*** | ***(SE)*** |
| Intercept | **502.3** | (9.3) | **508.6** | (6.7) | **508.8** | (6.7) |
| School mean SES | | | **36.4** | (3.2) | **36.3** | (3.2) |
| School SD of SES | | | | | -7.7 | (4.6) |
| SES (student level) | | | **19.2** | (1.4) | **19.2** | (1.4) |
| SES-squared (student level) | | | -0.1 | (0.5) | **-0.1** | (0.5) |
| **Variance components** | ***Variance*** | | ***Variance*** | | ***Variance*** | |
| **Student** | **4,396** | | **4,101** | | **4,101** | |
| **School** | | | | | | |
| Intercepts | **1,881** | | **1,128** | | **1,116** | |
| SES slopes | | | **64** | | **64** | |
| SES-Squared Slopes | | | **21** | | **21** | |
| **Country** | | | | | | |
| Intercepts | **2,848** | | **1,345** | | **1,308** | |
| School mean SES | | | **294** | | **281** | |
| School SD of SES | | | | | **349** | |
| SES slopes | | | **53** | | **53** | |
| SES-squared slopes | | | **5.7** | | **5.8** | |

**Note:** Figures in bold text are statistically significant ($p < 0.05$).

*Source:* PIRLS 2001.

The estimate for student-level SES in the SES gradient model is the average within-school SES slope. It is 19.2 for PIRLS, indicating that on average for the PIRLS countries, students' scores are about 20 points higher for each one-unit increase in SES. The slopes within schools tend to be more gradual than the overall slopes within countries, which on average are about 29 points for each one-unit increase in SES. The variance components associated with the average within-school slopes are statistically significant at both the school and country levels. This indicates that the average OECD slope of 20.4 varies among countries and among schools within countries. The results for indiviudal countries

are dicussed later. The effects associated with the other variables in the model pertain to the remaining hypotheses and are also discussed below. At this point, it is suffice to note that there are significant socioeconomic gradients in every country, which vary markedly among countries, and that the average within-school gradient is statistically significant and varies among countries and schools within countries.

For the PISA study, the results are remarkably similar. The average within-school slope is 20.2, indicating that on average for the PISA countries students' scores are about 20 points higher for each one-unit increase in SES. As with PIRLS, the slopes within schools tend to be more gradual than the overall slopes within countries, which on average for PISA countries are about 40 points for each one-unit increase in SES. The variance components associated with the average within-school slopes are also statistically significant at both the school and country levels.

**Table 6. Hierarchical linear models describing reading performance among students, schools and countries, PISA**

| | Null model | | SES gradient model | | Relative deprivation model | |
|---|---|---|---|---|---|---|
| **Coefficients** | ***Estimate*** | ***(SE)*** | ***Estimate*** | ***(SE)*** | ***Estimate*** | ***(SE)*** |
| Intercept | **495** | (5.3) | **502.7** | (4.2) | **502.3** | (4.2) |
| School mean SES | | | **70.0** | (6.6) | **69.6** | (6.5) |
| School SD of SES | | | | | -4.4 | (5.7) |
| SES (student level) | | | **20.4** | (2.1) | **20.5** | (2.1) |
| SES-squared (student level) | | | **-2.4** | (0.4) | **-2.4** | (0.4) |
| **Variance components** | ***Variance*** | | ***Variance*** | | ***Variance*** | |
| **Student** | **6,149** | | **5,644** | | **5,644** | |
| **School** | | | | | | |
| Intercepts | **3,237** | | **1207** | | **1,200** | |
| SES slopes | | | **86** | | **86** | |
| SES-squared slopes | | | **34** | | **34** | |
| **Country** | | | | | | |
| Intercepts | **749** | | **465** | | **457** | |
| School mean SES | | | **1,114** | | **1,087** | |
| School SD of SES | | | | | **467** | |
| SES slopes | | | **117** | | **116** | |
| SES-squared slopes | | | **2.4** | | **2.1** | |

**Note:** Figures in bold text are statistically significant ($p < 0.05$).

*Source:* PISA 2000-2002.

## Policy implications

The question of whether there is a significant relationship between reading performance and socieoconomic status is clearly addressed: there is a statistically significant relationship in every country, for both PIRLS and PISA. After the release of international reports, attention is usually focused on the average levels of performance of a country or, simply, its rank order among other countries. The problem is that the broad policy community in lower-income countries may attribute poor performance to the poor socioeconomic circumstances of their populations. Conversely, some countries may celebrate high performance, even though their students may not be performing that well when account is taken of their advantaged circumstances.

Two findings of this analysis are especially poignant for lower-income countries. First, there is wide variation among countries in the level of their gradients, which is evident as early as grade 4. Indeed, the variation among non-OECD countries is even greater than among OECD countries. We do not know the extent to which these gradients are well established when children enter school, but it is likely that many of the observed differences are attributable to children's experiences before they enter school (Young, 2000). This would call for greater investments in early childhood provision. In the first instance, though, we require national assessments of children's skills when they enter school, such as those being conducted in Jordan (Al-Hassan, 2005).

Second, children from high SES backgrounds in low SES countries did not fare as well on average as their counterparts in high SES countries. This is evident when one examines the gradients in Figures 8 and 9. The gradients for non-OECD countries are close to being parallel; the differences among countries in their performance are unform across levels of SES. One might not have expected this to be the case, as many children from advantaged backgrounds in low SES countries attend private schools (Willms and Somers, 2002) and their efforts at school are supplemented with tutoring and after-school programmes (Bray, 1999).

# Question 3

## To what extent do schools vary in their outcomes, after account is taken of students' SES and the mean SES of the school?

### Schools make a difference

This is the question that was the focus of educational research on school effects during the 1980's and 1990's. The hypothesis can be stated as: "Schools vary in their outcomes, even after taking account of students' SES and the mean SES of the school."[4] Raudenbush and Willms (1995) distinguished between two types of school effects: *Type A* effects, which refer to the expected score of a student with average SES (or some other set of background characteristics) in each school, and *Type B* effects, which refer to the expected score of a student with average SES in a school after controlling for the mean SES of the school.[5] They argued that Type A effects were of primary interest to parents, as they would typically want to choose a school that conferred an advantage to their child, including any advantages associated with the mean SES of the school. In contrast, teachers and school administrators would be primarily interested in Type B effects, as they would want to know how well their school fared in its performance compared with other schools that had similar socioeconomic intake.

In the first instance, one can ask whether schools simply vary in their average performance, without consideration of the SES of their students. This is estimated in a multilevel framework with a "null model", which simply partitions the variation in student performance into within-school and among-school components. These results were presented in the previous section in Tables 5 and 6 for PIRLS and PISA respectively. The null model is a three-level multilevel model that partitioned the variation in reading scores into three components: among students within schools, among schools within countries and among countries. The model was then extended to the "SES gradient model" by including SES at the individual level and school mean SES at the school level.

---

[4] This hypothesis simply holds that the variance of the $u_{0j}$'s in equation 5, which is referred to as tau ($\tau$), is greater than zero:

$$H_0: Var(u_{0j}) = \tau_0 = 0$$
$$H_1: \tau_0 > 0$$

Hypothesis of school effects **(8)**

[5] The variance of the Type A and Type B effects can be estimated by estimating separate models, with and without school mean SES in the model (i.e. equation 5 in footnote 4).

When student-level SES and school mean SES are introduced into the SES gradient model, the variation among schools within countries and among countries is reduced. This is explained by the following factors. The estimates of variance components in Table 5 indicate that the SES variables in the SES gradient model for PIRLS explain about 6.7% of the variation in students' scores within schools (i.e. it decreases from 4396 to 4101). The SES factors also explain about 40.0% of the variance among schools (i.e. the variance decreases from 4396 to 4101) and about 52.8% of the variance among countries (i.e. a decease from 2848 to 1345). The results for PISA are quite similar: the SES factors explain about 8.2% of the variation in students' scores within schools, 62.7% of the variation among schools within countries and 37.8% of the variation in country means. The question underlying the analysis is whether the remaining variance among schools within countries, or among countries, is statistically significant (see footnote 5). The findings indicate that for both sets of data the variance remaining at the school level (1345 for PIRLS and 1207 for PISA) is statistically significant, and we can conclude that schools vary in their performance within countries, even after taking into account students' SES and the mean SES of the school. The remaining variance at the country level (1345 for PIRLS and 465 for PISA) is also significantly greater than zero, and therefore, we can claim that while some of the variation in country performance is attributable to SES, countries vary in their performance even after this is taken into account.

Figure 2 in the introductory section provided a profile of school performance in Argentina. It showed that schools in Argentina varied considerably in their performance, at every level of school mean SES. An important indicator pertaining to the success of a schooling system is the extent to which schools vary in their performance. For all PIRLS countries, the variation among schools within countries for the SES gradient model is 1128 *(see Table 5)*. The square root of this figure, 33.6, is the standard deviation of the SES-adjusted school means. This is an indictor of the extent to which schools vary in their performance after taking account of SES at the individual and schools levels. On average, then, within PIRLS countries about 95% of schools fall within about plus or minus (2 x 33.6 =) 67 points of the school profile line. Therefore, the range between the best and poorest performing schools is about 134 points or about one-and-a-half grade levels. The comparable figure for PISA countries is about 140 points, roughly equivalent to four grade levels.

**Policy implications**

While it now seems to be an obvious point that "schools make a difference", the findings from PIRLS and PISA emphasise how much they differ. Imagine a child of average SES who is attending a school with average SES intake. "To what extent would the child's level of performance differ if he or she attended a high-performing school rather than a low-performing school?" These results suggest that on average, for countries that participated in PIRLS, by the end of grade 4

the child would have an advantage of about one-and-a-half grade levels if he or she attended one of the best-performing schools in the country rather than one of the worst-performing. By age 15, as gauged by PISA results, the difference would amount to more than four grade levels. One should be cautious in this grade-level interpretation, as the PISA tests were not designed to specifically reflect middle and high school curricula. However, even if one uses an "effect size" metric, the results indicate that the difference in performance between the best and worst performing schools ranged by more than a full standard deviation.

At the country level, the partitioning of gradients into within- and between-school components, and the estimation of variance among schools, are important as they convey information about how best to intervene. If much of the variance in student achievement is within schools rather than between schools, then it is more appropriate to emphasise within-school interventions. For example, at the elementary level this could entail a three-tier intervention programme as advocated by the *Committee on the Prevention of Reading Difficulties in Young Children* (Snow, Burns and Griffin, 1999). However, if more of the variance in student achievement was between schools rather than within them, efforts might be directed towards inclusive interventions aimed at reducing segregation or whole-school reforms, such as "Accelerated Schools" (Levin, 1987) or "Success for All" schools (Madden, Slavin, Karweit and Livermon, 1989).

# Question 4

## Is the relationship between student achievement and socioeconomic status weaker at higher levels of SES?

### The hypothesis of diminishing returns

This question can be framed as "the hypothesis of diminishing returns" which holds that there are weaker effects on social outcomes above some SES thresholds (Willms, 2003a). One might predict, for example, that above a certain level of SES there would be little or no increase in students' academic achievement associated with SES. This is the case for certain health outcomes in the United States – at incomes below US$ 20,000, increases in income are associated with markedly better health outcomes; but above that threshold, there are diminishing returns (Epelbaum, 1990; House *et al.*, 1990; Mirowsky and Hu, 1996; Rogot *et al.*, 1992). Boyle and Willms (1999) and Wolfson *et al.* (1999) have reported similar findings for health outcomes in Canada, but the curvilinearity is less pronounced. This hypothesis is relevant to educational policy, because if it is possible to identify an SES threshold where the gradient begins to level off, then compensatory policies, such as transfer payments to the poor, stand to be effective in raising and levelling the socioeconomic gradient.

The results in Table 5 indicate that, on average for the countries that participated in PIRLS, there is *not* a statistically significant curvilinear relationship with SES. [6] However, the curvilinear component does vary significantly among countries, indicating diminishing returns for increasing SES in some countries, while the relationship is linear or with *increasing* returns in others. An inspection of the country-by-country results in Table 3 confirms this. There are diminishing returns for SES in a few countries (Bulgaria, Hungary, Islamic Republic of Iran, Italy,

---

[6] The hypothesis of diminishing returns can be tested by adding a quadratic term for SES into the model. In the example, this is tested separately for each country *(see Table 2)*. The hypothesis can also be tested in a multilevel framework, with the quadratic term added to the within-community model:

$$Y_{ij} = \beta_{oj} + \beta_{1j} X_{ij} + \beta_{2j} X_{ij}^{2} + r_{ij}$$ Within-community equations with curvilinear term **(9)**

where the ß$_{2j}$'s capture the curvilinear effects. These can be expressed as an average effect, $\gamma_{20}$, and the deviation of each community from the average effect:

$$\beta_{2j} = \gamma_{20} + u_{2j}$$ Among-community equation for curvilinear gradients **(10)**

where $\gamma_{20}$ is the mean of the curvilinear effects, and $u_{2j}$ is the deviation of each community's curvilinear effect from the mean effect. One can test whether the mean effect is statistically significant:

$$H_0: \gamma_{20} = 0$$
$$H_1: \gamma_{20} \neq 0$$ Hypothesis of Diminishing Returns **(11)**

This is assessed with a t-test with j-1 degrees of freedom.

Slovak Republic and Slovenia), while there are increasing returns in others (Argentina, Canada, France, Romania and Turkey). However, in all cases the extent of curvilinearity is slight and not sufficiently strong to identify a low SES threshold. This is evident in the graphical display of the gradients in Figure 8.

The results presented in Table 6 provide a test of the diminishing return hypothesis at the school level for PISA. They show that the average within-school slope is also curvilinear, consistent with the diminishing returns hypothesis: the coefficient for the SES-squared term is -2.4. However, this effect varies significantly among schools within countries (the variance is 34) and therefore there is likely to be some schools where the hypothesis does not hold. The results also indicate that the curvilinear effect varies significantly among countries (the variance is 2.4), as shown in Figure 9.

The results in Table 4 are consistent with the diminishing returns hypothesis for 11 of 28 OECD countries: Austria, Belgium, Canada, Czech Republic, Denmark, Japan, Luxemburg, Netherlands, Norway, Spain and Switzerland. This is also the case for Latvia. In one OECD country, Mexico, and in five of the non-OECD countries, there is some evidence of "increasing returns" with higher levels of SES.

**Policy implications**

Willms and Somers (2001) also found increasing returns for SES for reading and mathematics achievements of grade 3 and 4 students in several Latin American countries. They suggested that there was a premium associated with parents having completed secondary school. However, their results – and the PISA results reported here – may be attributable to a "floor effect" on the test. The results in this report for PIRLS suggest that this may be the case. Alternatively, it may be that some countries are less successful than others in enabling low SES children to achieve sufficient literacy skills by the end of the grade 4 to benefit as much from instruction in later grades. One would then observe increasing returns for SES at age 15 in PISA. With either scenario, the results call for increased monitoring at a younger age, preferably when children enter school or even earlier.

The results also indicate that we cannot identify a low SES threshold below which the majority of children have difficulties in learning to read. If this were the case, the findings would call for interventions targeted specifically toward low SES children. Rather, there are many children who are in families with average or above-average SES who have low reading scores.

# Question 5

## Do socioeconomic gradients converge at higher levels of SES?

### The hypothesis of converging gradients

Research based on the International Adult Literacy Study found that the literacy skills of youth aged 16 to 25 years tended to be similar across countries for those with high levels of parental education. However, for youth whose parents had relatively low levels of education, the average skill levels varied markedly among countries. This pattern was also evident among states in the United States and among provinces in Canada. Willms (2003b) provides a summary of this research. More generally, this question can be framed as "the hypothesis of converging gradients": the variation among communities in their social outcomes deceases with increasing levels of SES. The hypothesis has important implications for educational policy as it suggests that students from high SES backgrounds tend to fare well in their literacy skills in most schools, whereas those from lower SES backgrounds may vary considerably in their skills, depending on the school they attend.

**Figures 10** and **11** display the relationship between the levels of the socioeconomic gradients and their slopes for PIRLS and PISA respectively. For OECD countries in both PIRLS and PISA, there are significant negative correlations between SES-adjusted levels of peformance and the SES slopes: -0.16 for PIRLS and -0.31 for PISA. The same applies for the non-OECD countries, with negative correlations of -0.14 and -0.38 for PIRLS and PISA resepctively. These findings provide strong evidence of converging gradients. They also suggest that the extent to which gradients converge increases as students progress through the schooling system.

The hypothesis can also be tested with respect to schools within countries.[7] The three-level multilevel analysis for the SES gradient model found that on average, within PISA countries, the correlation between levels of the gradients and the

[7] The model described by equations 4, 5 and 6 above are fitted to the data to test this hypothesis. The estimation of this model entails estimation also of the variance of the gradient levels [i.e. Var ($u_{0j}$)], the variance of the gradient slopes [i.e. Var ($u_{1j}$)] and the covariance of the levels and slopes [i.e. Cov ($u_{0j,}u_{1j}$)]. In the first instance, one tests whether the slopes vary significantly among communities:

$H_0: Var(u_{1j}) = \tau_1 = 0$ Hypothesis of variable slopes **(12)**

$H_1: \tau_1 > 0$

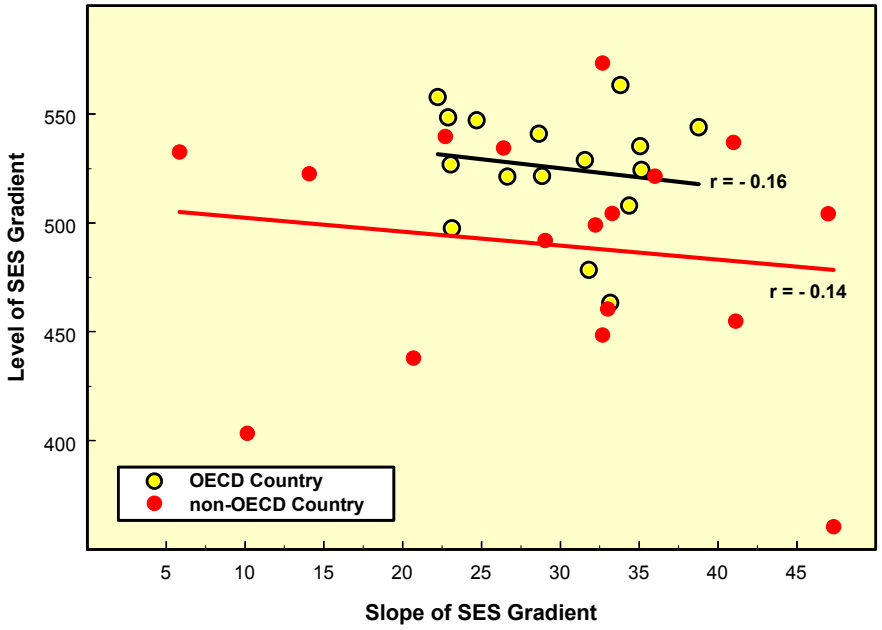
Assuming the slopes vary significantly among communities, one then wants to test also whether the covariance between levels and slopes is statistically significant:

$H_0: Cov(u_{0j}, u_{1j}) = \tau_{01} = 0$ Hypothesis of converging gradients **(13)**

$H_1: \tau_{01} > 0$

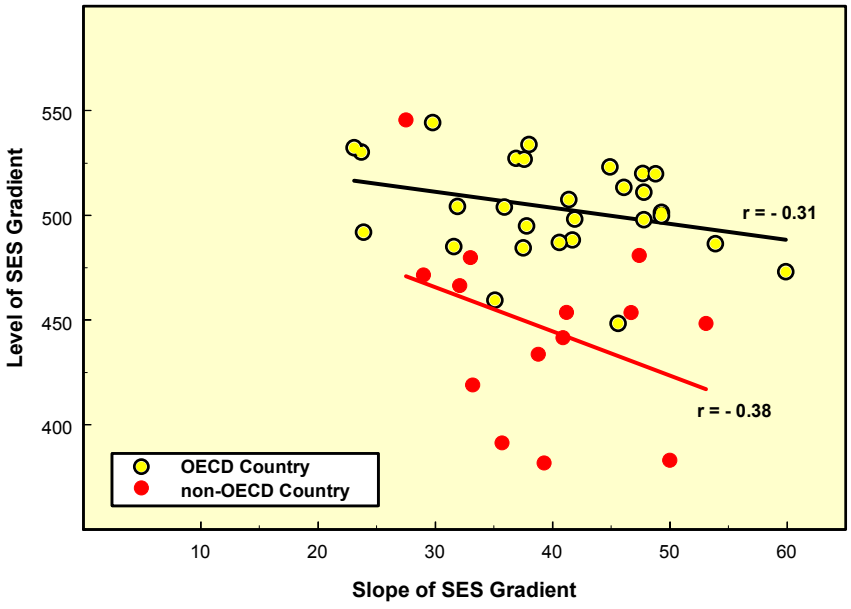
within-school slopes was -0.21 (not shown in Table 2). This finding confirms that gradients not only converge at the country level but also for schools within countries. It provides strong evidence that successful schools tend to be those that are successful in bolstering the performance of students from less-advantaged backgrounds.

**Figure 10. The relationship between levels and slopes of socioeconomic gradients**




*Source:* PIRLS, 2001.

**Figure 11. The relationship between levels and slopes of socioeconomic gradients**




*Source:* PISA, 2000-2002.

## Policy implications

The results indicate that, in most countries, schools with high levels of performance tend to have strong performance for their students from low SES backgrounds; that is, the schools with more equitable results tend be those with the best results.

There are at least two plausible processes that might explain why socioeconomic gradients converge at higher levels of socioeconomic status. One is that schools with higher levels of literacy performance are those that are successful in bolstering the skills of their low SES students. For example, successful schools may be those that place a greater emphasis on educational equity, which is reflected in various school policies and practices. With this explanation, less-successful schools provide positive learning experiences for children from high SES backgrounds but not for those from poor backgrounds. The effect of within-school tracking into different types of educational programmes is consistent with this explanation (Carbonaro, 2005). Another explanation is that when schooling is of poor quality, it affects all children more or less equally, but high SES children tend to succeed anyway because their efforts at school are supplemented with out-of-school learning opportunities. This might include, for example, greater investments by parents in helping them with assignments or in hiring tutors when they are falling behind. Also, there is considerable evidence that high SES children learn more during vacation periods than low SES students (Cooper, Nye, Charlton, Lindsay and Greathouse, 1996).

It is likely that both mechanisms operate to produce converging gradients. From a policy perspective, the first explanation would call for measures that ensured that all schools had internal offerings of uniform quality and programmes that enabled low SES children to succeed. The second explanation would call for measures to ensure low SES children had extra help outside of school hours, with augmentative instruction and summer learning programmes.

# Question 6

## Are there school composition effects on students' academic achievement?

### Composition *versus* contextual effects

The hypothesis of a socioeconomic gradient, which underlies Question 2, holds that there is a relationship between school performance and SES at the individual level. We found that there was a significant relationship in every country that participated in PIRLS and PISA and, indeed, within most schools within countries. The hypothesis underlying the presence of a composition effect is that there is an additional effect associated with the average level of socioeconomic status of schools. If the compositional effect is positive, it indicates that students attending high SES schools tend to have better scores than those attending low SES schools, even after taking account of students' SES at the individual level. This hypothesis is important to the study of schooling systems because it is directly relevant to issues concerning the manner in which students are allocated to schools, classrooms and instructional groups.

The academic literature has often used the term "contextual effect" to refer to the effect on student outcomes associated with the demographic characteristics of a school's composition, especially the mean SES of a school (e.g. Alexander and Eckland, 1975; Bryk and Driscoll, 1988; Willms, 1986). However, some authors have suggested that mean SES, or various other classroom- or school-level aggregates describing student composition, are a proxy for "peer effects" (Robertson and Symons, 1996; Zimmer and Toma, 1997). This has created some confusion about the role of composition effects, as other researchers have challenged the validity of the research, noting that aggregate measures of school composition are inadequate as a proxy for peer effects, and any attempt to infer a causal relationship is threatened by selection bias and an underspecified statistical model (e.g. Nechyba, McEwan and Older-Aguilar, 2004). Willms (2004a) suggests distinguishing between school or classroom "composition", which is generally defined with aggregates of factors, such as SES *versus* school and classroom "context", which refers to the environment in which teaching and learning takes place. School and classroom context includes factors describing the physical features of the learning environment and its culture, such as material resources, interactions among peers, the relationships between teachers and students, the disciplinary climate of the classroom and the norms for academic success. This distinction is important because a description of the relationship between school composition and academic performance is useful for evaluation purposes and for understanding how best to raise and level the gradient. However, one must be cautious about inferring causation (see also Alexander, Fennessey, McDill and D'Amico, 1979).

In most education systems, schools vary to some extent in their average SES due to residential segregation, especially in large cities. Though in many systems, students are further segregated along socioeconomic lines due to private schools or public schools with selective admission criteria. Students can also be segregated when they choose schools based on particular programmes, such as language immersion or baccalaureate programmes. A number of studies have shown that, when students are segregated according to their SES, those from advantaged backgrounds tend to do better, while those from disadvantaged backgrounds do worse (Brookover *et al.*, 1978; Henderson, Mieszkowski and Sauvageau, 1978; Rumberger and Willms, 1992; Shavit and Williams, 1985; Willms, 1986).

The early research on compositional effects maintained that it stemmed primarily from peer effects; when bright and motivated students work together, they learn from each other and set higher standards for performance. This is likely to be one source of the compositional effect; however, schools with high SES intakes also tend to have several other advantages associated with their context: on average they are more likely to attract and retain talented and motivated teachers and are more likely to have greater support from parents, fewer disciplinary problems and generally an atmosphere conducive to learning (Willms, 1986; Willms and Somers, 2001; OECD, 2001). Teachers in high SES schools are more likely to find it easier to set and maintain high standards for performance, and teach the curriculum at a faster pace. Finally, schools with higher SES intakes may also be more likely to have smaller class sizes and better teaching resources.

## School composition effects in PIRLS and PISA

The coefficients for school mean SES in the SES gradient model in Tables 5 and 6 are estimates of the "compositional effect". For PIRLS it is 36.4 and for PISA it is 70.0, and for both samples the coefficient is statistcially significant ($p < 0.05$).[8] This indicates that for grade 4 students in PIRLS student performance increases 36.4 points with each one-unit increase in school mean SES. For 15-year-old students in PISA, the increase is 70 points. For example, for the PIRLS sample, if we consider two hypothetical students whose family SES were at the international mean (i.e. an SES score of zero), and one student attended a school with an average SES of -0.5 while the other attended a school with an average SES of 0.5, the latter student in the high SES school would score, on average, about 36 points higher than the student in the low SES school.

[8] The coefficient for school mean SES in equation 5, $\gamma_{01,}$ is the "compositional effect". The hypothesis of a compositional effect is simply that this coefficient is significantly different from zero:

$$H_0: \gamma_{01} = 0$$
$$H_1: \gamma_{01} \neq 0$$

Hypothesis of a compositional effect **(14)**

This is assessed with a t-test with j-1 degrees of freedom.

The compositional effect is statistically significant in every country in PISA (Willms, 2004a) but varies considerably. The results in Tables 5 and 6 include estimates of the extent to which the effect varies among countries: the variance is 294 for PIRLS and 1114 for PISA (or a standard deviation of about 17 points for PIRLS and 33 points for PISA).

Willms (2003a) has referred to the compositional effect as "double jeopardy", because students from low SES families tend to be disadvantaged because of their circumstances at home, but when they are also segregated into low SES schools, they tend to fare even worse. One might also pose the question of triple jeopardy: "Is the compositional effect stronger for low SES students than for high SES students?" We might also ask: "Is the compositional effect stronger for males than for females, or for foreign-born students than for students born in the country?" These questions can be tested for explicitly in the multilevel framework by including a cross-level interaction term[9].

A test of the triple jeopardy hypothesis for all PIRLS countries combined found evidence of an interactive effect for low SES students. Considering the two hypothetical students described above, the bonus for a high SES student attending a high SES school was about 37 points, while the bonus for a low SES student was only about 35 points. For PISA, the coefficients were significant for the effects of school mean SES on both the SES slopes and the curvilinear component of SES (i.e. SES-squared). However, the coefficient for the interaction with the linear effect was positive, while the coefficient for the SES-squared term was negative. Taken together, with consideration of the range of SES covered by students in OECD countries, the two effects more or less sum to zero. The results are summarised graphically in **Figure 12**. For this graph, low SES schools were defined as schools with mean SES that were in the lowest 25% of OECD schools (i.e. schools with a school mean SES below -0.52). The gradient lines show the average performance for students in low SES schools and those in high SES schools. The lines are drawn to cover the range of SES from the 5$^{th}$ to the 95$^{th}$ percentile for each group (as for other gradients in this report). As one would expect, the graph shows that it is mostly low SES students that attend low SES schools and high SES students that attend high SES schools. However, there is a fair degree of overlap in the SES of students attending both types of schools – from about -1.3 to 0.5. This graph shows the importance of the compositional effect: students in that range of SES have markedly lower scores if they attended a low SES school. The results also show

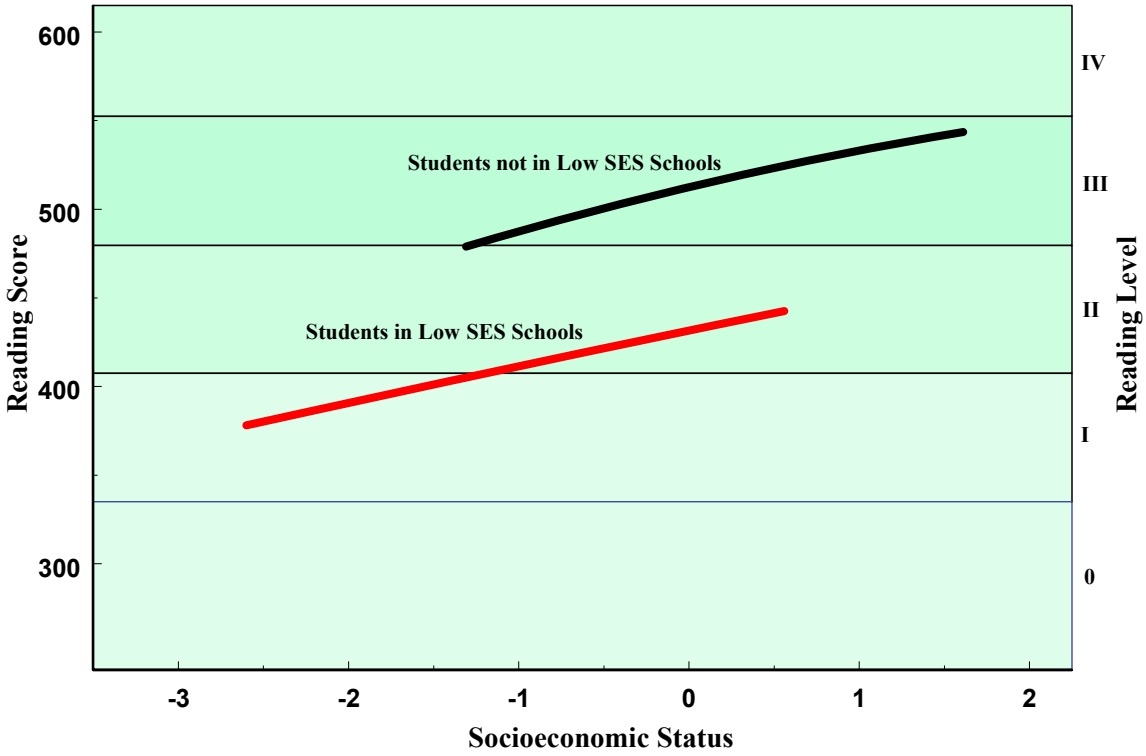
[9] The hypothesis of triple jeopardy is that the compositional effect varies depending on the characteristics of an individual, such as his or her level of SES, sex, or immigrant status. This is tested as a "cross-level interaction" in the multilevel model. For example, for the question about the interaction of the compositional effect with individual SES, the slopes of the gradients are regressed on the school mean SES:

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\overline{X}_{\bullet j} + u_{1j}$$ Hypothesis of Triple Jeopardy **(15)**

Similarly, one can include in the level-1 model dummy variables indicating whether a student is male or female or foreign-born.

that the lines are close to being parallel; thus, there is a weak "triple jeopardy" effect.

**Figure 12. Socioeconomic gradients for low and high SES students indicate the magnitude of school composition effects.**




*Source:* PISA 2000-2002.

For PISA, the "triple jeopardy" effect was statistically significant for boys and for foreign-born students. Overall boys scored about 26 points lower than girls, but if they were also in low SES schools, they were further disadvantaged by about 4 points. Similarly, students who were foreign-born on average scored about 22 points lower than those born in the country, but if they were also in low SES schools, they were further disadvantaged by about 8 points.

An important relationship is that the "compositional effect" is equivalent to the difference between the between-school and the within-school slopes. The multilevel model yields estimates of within-school slope and the compositional effect. The between-school slope is then simply the sum of these two coefficients. Also, the slope of the overall socioeconomic gradients for a country can be decomposed into a within-school and a between-school slope (Alwin, 1976). The decomposition is a function of the between-school slope, the average

within-school slope and $\eta^2$, which is a measure of the extent of between-school SES segregation:

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

The index $\eta^2$ is the proportion of variation in SES that is between schools. It can theoretically take on values between zero and one, but even in highly segregated school systems it is seldom above 0.6. When $\eta^2$ is zero, there is no segregation among schools; that is, all schools have the same SES distribution. Among countries that participated in PISA, $\eta^2$ ranged from 0.116 (Norway) to 0.475 (Chile). The index is calculated by estimating a multilevel null model for SES, which provides estimates of the proportions of variance in SES that are within and between schools.

## Policy implications

An understanding of the between-school and between-classroom segregation of students with differing ability and SES is fundamental to understanding the educational performance of the school system. The schooling systems that have the best results – meaning high and equitable student performance – with very few exceptions have low levels of between-school segregation. When students are segregated into different kinds of programmes as they progress through school, the gap tends to increase and overall levels of performance become worse. This is evident if we compare the PIRLS and PISA results for Bulgaria, Germany, Hungary, Latvia and the Russian Federation, which have highly selective school systems after age 10, with the results for Canada, New Zealand and Sweden, which have more inclusive systems through to the end of secondary school.[10] At the grade 4 level, the former group of countries had relatively high average levels of achievement on the PIRLS reading test, with scores ranging from 528 to 550. These are comparable to the results for Canada (544), New Zealand (529) and Sweden (561). However, the PISA results for the more selective systems were all below the OECD mean, with results ranging from 430 to 487, well below those of Canada (534), New Zealand (529) and Sweden (516). A notable exception is Hong Kong (Special Administrative Region of China): its scores in PIRLS and PISA were comparable to those of New Zealand, even though its system is quite selective ($\eta^2$ = 0.24 for PISA).

[10] The $\eta^2$ values for these countries at age 15, based on PISA 2000 results, were as follows: Bulgaria (0.36), Canada (0.19), Czech Republic (0.27), Germany (0.26), Hungary (0.45), Italy (0.27), Russian Federation (0.25), New Zealand (0.19) and Sweden (0.12) (see Willms, 2004).

In countries with large compositional effects, there are two basic strategies for raising and levelling the socioeconomic gradient. One is through reforms aimed at bolstering the achievement levels of low SES schools. This is difficult to achieve because when low SES or low ability students are concentrated in particular schools, it is difficult to maintain high expectations, establish a positive disciplinary climate and attract and retain talented teachers. The other strategy is through inclusive reforms aimed at decreasing socioeconomic segregation between schools. This is usually difficult to achieve politically, as high SES families have a vested interest in maintaining a selective school system.

# Question 7

## Do schools with more homogeneous intakes have better school performance?

### The hypothesis of homogeneous communities

The hypothesis underlying this question holds that: "Communities that are relatively homogeneous in their socioeconomic status tend to have superior social outcomes than those that are relatively heterogeneous in their socioeconomic status." The hypothesis stems mainly from research in health outcomes. Wilkinson (1992, 1996, 2000) found that the life expectancy of a country was related more closely to the extent of income inequality in the country than to its absolute level of income. Analyses of adults' life expectancy and mortality rates across U.S. states and cities have also supported this hypothesis (Kaplan *et al*., 1996; Kennedy, Kawashi and Prothrow-Stith, 1996; Lynch *et al.*, 1998). Similar findings have been reported for local authorities in the United Kingdom (Ben-Schlomo, White and Marmot, 1996).

This hypothesis has not received much attention in educational research, although it underlies many educational policies. In schooling terms, the hypothesis holds that schools with more homogeneous SES intakes have better school performance than those with more heterogeneous intakes. The argument that students will have better performance if taught in homogeneous groups is used to support policies, such as the tracking of students in academically- and vocationally-oriented schools and school programmes. This argument also underlies the practice of having students repeat a grade when they have fallen significantly behind their peers in performance.

The hypothesis is tested in the three-level hierarchical analysis by including a measure of the standard deviation of SES for each school.[11] Schools with a large standard deviation are more heterogeneous than those with a small standard deviation. The findings show that on average across PIRLS countries the effect is

---

[11] The second-level model for the ß$_{0j}$'s (equation 5) is extended to include a measure of the range or standard deviation of SES within each of the j schools:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\,\overline{X}_{j\bullet} + \gamma_{02}\,X_{SD_j} + u_{0j}$$ Among-school equation for gradient levels **(16)**

where $\gamma_{02}$ is the "relative deprivation effect" associated with the standard deviation of SES of the school. The hypothesis of relative deprivation is simply that this coefficient is significantly different from zero:

$$H_0: \gamma_{02} = 0$$
$$H_1: \gamma_{02} \neq 0$$ Hypothesis of relative deprivation **(17)**

This is assessed with a t-test with j-1 degrees of freedom.

small (-7.7) and not statistically significant. The average within-school standard deviation of SES for PIRLS is 0.76, and most schools range from 0.51 ($10^{th}$ percentile) to 1.01 ($90^{th}$ percentile). An increase in the heterogeneity of a school by 0.1 standard deviations in SES is associated with an increase in reading performance of less than one point on the international scale. In PISA, the effect is also small (-4.4 for a one standard deviation increase in the within-school standard deviation). For example, an increase in the heterogeneity of a SES from, say, a standard deviation of 0.70 to 0.80 would be associated with a decrease in reading performance of less than one-half of one point (0.10 times -4.4). The results also indicate that the effect varies significantly among countries, suggesting that in some countries the effect of increased heterogeneity is positive while in other countries it is negative.

**Policy implications**

These findings suggest that students in schools with more homogeneous intakes fare slightly better than those in schools with heterogeneous intakes. However, the effects are very small and not significant in educational terms. These null findings are important, though, because they suggest that policies aimed at reducing school heterogeneity through policies such as streaming or tracking are unlikely to increase literacy performance.

# Question 8

## Is the variation within and among schools attributable to levels of school resources and to school and classroom policy and practice?

### Why schools differ in their "added value"

One of the key findings of PIRLS and PISA presented above is that countries vary substantially in their levels of student performance and in the relationship between student performance and socioeconomic status. Successful countries tend to be those that have not only rasied the learning bar, but have also levelled it. Within countries, schools also vary signifcantly in their performance. Some of this variation is attributable to the family background of students entering the school, but some of it is also related to certain structural features of schools and schooling systems, and to the policies and practices of school administators and teachers. In other words there is an "added value" associated with attending a particular school.

Over the past three decades, eduational researchers have exerted considerable effort towards understanding *why* schools differ in their added value. The research has suggested that, while levels of material and human resources are important, there are measureable aspects of the learning climate of schools that also affect student performance. These include, for example, teacher-student relations, parental involvement, the disciplinary climate of the school and teachers' attitudes (Bryk, Lee and Smith, 1990; Ho and Willms, 1996; Pallas, 1988). Student performance tends to be higher in schools where teachers: have high expectations for all students to achieve; provide positive feedback and opportunities for success; practice team teaching and cooperative learning; and develop a classroom atmosphere where students know the rules and consider them to be fair (Anderson, 1985; Lee and Smith, 1993). Generally, when teachers have positive attitudes and high morale, they use time and resources in ways that promote better schooling outcomes (Dreeben and Gamoran, 1986; Plewis, 1991).

Despite the efforts of educational researchers to uncover the critical aspects of school policy and practice that affect school performance, much of the emphasis in international studies is on the mean performance scores for each country. While it is helpful to identify countries that are faring either particularly well or poorly, they do not provide much guidance for educators or policymakers. Those concerned with raising and levelling the learning bar want to know what factors are most strongly related to educational outcomes, where to intervene and what types of students or schools should be targeted.

There are three major limitations of PIRLS and PISA for identifying the important factors that contribute to the success of schools and schooling systems. One is that the study designs were cross-sectional and did not entail the random assignment of students to treatment and control schools or classrooms. The second limitation is that the reading results in PIRLS and PISA represent the cummulative effects of all factors that bear on a child's literacy development from birth or even conception. It is tempting to infer that strong PISA results are related mainly to the learning climate of secondary schools or that strong PIRLS results are mainly associated with children's learning experience during primary school, but it may be that most of the differences among schools were evident when the children entered primary school. The third major limitation with PIRLS and PISA is that the variables describing school policy and practice were collected and reported only at the student and school levels. Research on school effects that has partitioned variance in outcomes into student, classroom and school components indicates that there is more variation among classrooms within schools than there is among schools (Willms, 2001). Thus, a significant shortcoming of these PIRLS and PISA studies is the lack of data at the classroom level.

With these three limitations in mind, we cannot make strong causal inferences about the relationships when interpreting these results. We also might expect that the design lacks statistical power for identifying the effects of particular factors. However, as with socioeconomic gradients, the descriptive account of the factors associated with student performance can inform policymakers as to what may be a reasonable course of action, especially when the cross-sectional findings are supported by a number of smaller-scale research studies.

**The hypothesis of policy and practice mediators**

The hypothesis is that "variation within and among schools is attributable to levels of school resources, and to school and classroom policy and practice". One approach to the study of socioeconomic gradients presumes that students from high SES backgrounds have better school performance because their parents invest more in their education (Becker and Tomes, 1986; Coleman, 1988). For example, the argument would be that high SES parents are more likely to read to their children, discuss school projects or help them with their homework. When their children are falling behind, high SES parents would be more likely to seek assistance for their child, either at or outside the school. More generally, this approach would presume that high SES parents are more likely than low SES parents to provide a home atmosphere conducive to learning and maintain strong relationships with school personnel to ensure their child succeeds. Research in this vein aims to unpack the SES gradient by identifying family practices and processes that affect children's learning and overall development.

Another approach to the study of SES gradients emphasises the roles of schools, neighbourhoods and local communities in shaping and directing the actions of children and their parents (Coleman, 1988). For example, the amount of time parents invest in their child's learning may depend on the norms of their community, and the policies and practices of their child's school. Moreover, school policies and practices are likely to directly affect children's learning, as they concern the nature of the learning environment and affect children's motivation and effort. Research in this vein, as it applies to schooling outcomes, would attempt to explain socioeconomic gradients by analysing the effects associated with the structural and organizational features of schools, as well as school and classroom policy and practice.

Both approaches entail a search for *mediators* of socioeconomic gradients. A mediator is a factor that is directly influenced by SES and helps to explain why there is a relationship between SES and the schooling outcome (Kraemer, Stice, Kazdin, Offord and Kupfer, 2001). For example, family SES may be associated with the likelihood that a parent reads to their child on a regular basis. Also, research on children's development has shown that reading to a child has a strong positive influence on the development of reading skills (Greaney, 1986; Cook and Willms, 2002). Given evidence of these two relationships, one would claim that reading to a child *mediated* the socioeconomic gradient.

Multilevel models provide a powerful tool for integrating these two dominant approaches to the study of socioeconomic gradients.[12] First, as was shown above, it allows one to decompose the gradient into within- and between-school gradients. Second, it allows one to model separately the effects of family and school-level factors. Third, it allows one to examine interactions between family and school factors. Following the example above, it may be that schools with

---

[12] Individual-level mediators describing family practices can be added to the level 1 model:

$$Y_{ij} = \beta_{oj} + \beta_{1j} X_{1ij} + \beta_{2j} X_{2ij} + r_{ij} \qquad \textbf{(20)}$$

where $\boldsymbol{X_{1ij}}$ is SES, $\boldsymbol{X_{2ij}}$ is the potentially mediating factor. The $\boldsymbol{ß_{2j}}$'s can be expressed as an average effect across all communities, $\gamma_{20}$, plus the deviation of each community's $\boldsymbol{ß_{2j}}$ from the average, $\boldsymbol{U_{2j}}$ (as per equation 4). The primary criterion for a mediator is that it be related to the outcome, even after controlling for SES:

$$H_0: \gamma_{20} = 0$$

$$H_1: \gamma_{20} \neq 0$$

Hypothesis of individual-level mediator **(21)**

The hypothesis of a community-level mediator is tested by adding the potential mediator to the second-level equation for the $\boldsymbol{ß_{0j}}$'s:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j}$$

Hypothesis of community-level mediator **(22)**

where $\gamma_{00}$ is the mean of the *adjusted* community means, and $\boldsymbol{u_{0j}}$ is the deviation from each community's adjusted mean from the grand mean. One is primarily interested in the magnitude and statistical significance of the coefficient $\gamma_{10}$:

$$H_0: \gamma_{01} = 0$$

$$H_1: \gamma_{01} \neq 0$$

strong programmes for parental involvement encourage and support parents to read to their child on a regular basis. A test of this interactive effect can also be incorporated into a multilevel model.

### Findings from PIRLS and PISA

**Table 7** presents results pertaining to the relationship of reading performance to sets of variables constructed from the PIRLS and PISA surveys that describe school and classroom policy and practice. The details of the variables are presented in **Appendix A**.

**Table 7. The relationship between reading performance and school resources, school policy and classroom practice**

| | PIRLS 2001 | | | PISA 2000-02 | |
|---|---|---|---|---|---|
| | Effect | (SE) | | Effect | (SE) |
| **Student-level variables** | | | | | |
| **Female** | **16.3** [sc] | (1.2) | | **25.1** [sc] | (2.0) |
| **Socioeconomic status** | **20.4** [sc] | (1.3) | | **23.1** [sc] | (1.9) |
| **SES-squared** | 0.3 [sc] | (0.5) | | **-1.5** [sc] | (0.3) |
| **Foreign-born** | **-27.5** [sc] | (2.2) | | **-26.1** [sc] | (2.8) |
| **School resources** | | | | | |
| **Quality of school infrastructure** | 0.6 [c] | (0.4) | | 0.6 [c] | (0.4) |
| **Student-to-teaching staff ratio (STR)** | | | | **2.1** [c] | (0.5) |
| **STR-squared** | | | | **-0.13** [c] | (.03) |
| **Class size less than 20** | **-5.3** [c] | (2.1) | | | |
| **Class size greater than 30** | **10.2** [c] | (2.6) | | | |
| **Teachers have tertiary level qualifications in language arts** | 2.4 [c] | (2.4) | | **2.5** [c] | (0.6) |
| **Teacher experience** | **0.35** [c] | (0.1) | | | |
| **Schools have at least one computer for every 10 students** | | | | -1.7 | (2.1) |
| **Teachers receive professional development** | | | | -0.9 [c] | (0.2) |
| **School policy and practice** | | | | | |
| **Teacher factors affecting school climate** | | | | 0.4 [c] | (0.3) |
| **Principal autonomy** | | | | 0.9 [c] | (0.4) |
| **Teacher autonomy** | | | | -0.4 [c] | (0.5) |
| **Use of formal assessment** | | | | 0.1 [c] | (0.3) |
| **Time for reading instruction** | 0.0 | (0.2) | | | |
| **Classroom practice** | | | | | |
| **Teacher morale and commitment** | | | | **0.9** [c] | (0.3) |
| **Disciplinary climate** | **1.1** [c] | (0.3) | | **3.5** [c] | (0.5) |
| **Students' use of resources** | | | | **4.7** [sc] | (0.9) |
| **Parental support** | **2.1** [c] | (0.4) | | | |
| **Achievement press** | | | | 0.3 [c] | (0.5) |
| **Student-teacher relations** | | | | **3.3** [c] | (0.5) |
| **Use of informal assessment** | | | | -0.4 [c] | (0.3) |

*Source:* PIRLS 2001.

The first part of the model presented in Table 7 includes the student demographic variables, sex, SES and whether the student was foreign-born. The variable for sex was coded zero for males and one for females, such that the coefficient is an estimate of the difference in performance between males and females. The estimate for PIRLS is 16.3 points and for PISA it is 25.8 points. The analysis also revealed that this effect varies significantly among schools within countries and between countries. This means that, on average, the sex difference for PIRLS is about 16 points, but in some schools it is significantly larger and in others is significantly smaller. The same point also applies to countries; the difference between males and females varies significantly.

The second section of the table includes school-level variables pertaining to school resources. The coefficients indicate the effects associated with each factor, after controlling for student demographic factors and other factors in the model. These results indicate the average within-country relationship, and thus, they are not heavily influenced by the results for a particular country or a few schools within any country.

For PIRLS, only the effects of class size and teacher experience were statistically significant. The effects of class size are in the opposite direction than one might expect. They indicate that large class sizes, greater than 30 children, have better performance on average than classes in the range of 20 to 30. Similarly, class sizes smaller than 20 have slightly worse performance than those in the 20 to 30 range. This finding could be an artefact of rural and urban schools, although the analysis does control for family background factors. The effect associated with teacher experience, which was measured in years of experience is 0.35. This is also a relatively small effect, suggesting that average performance increases by only about 3.5 points with each 10 years of additional experience.

For PISA, three of the five school resource factors were statistically significant. The effects of the student-to-staff teaching ratio is difficult to interpret because of the curvilinear term. The results are shown graphically in **Figure 13**. They indicate that there is relatively little impact of the student-to-staff teaching ratio in the range from 10 to 25 but thereafter performance begins to decline. The results also suggest that there is a slight increase in student performance associated with increasing student-to-staff teaching ratio from 10 to about 18 or 20. This result may have been caused by the influence of schools serving disadvantaged students that have particularly small student-to-staff ratios. The measure of student-to-staff teaching ratio should not be confused with class size. This measure is the ratio of the total number of students in the school to the number of full-time equivalent teaching staff in that school. In many schools, the teaching staff includes the school administrator, librarian, and special education teachers. As such, it is closely related to the costs associated with educating each student, as staff wages are by far the largest component of the overall costs of education.

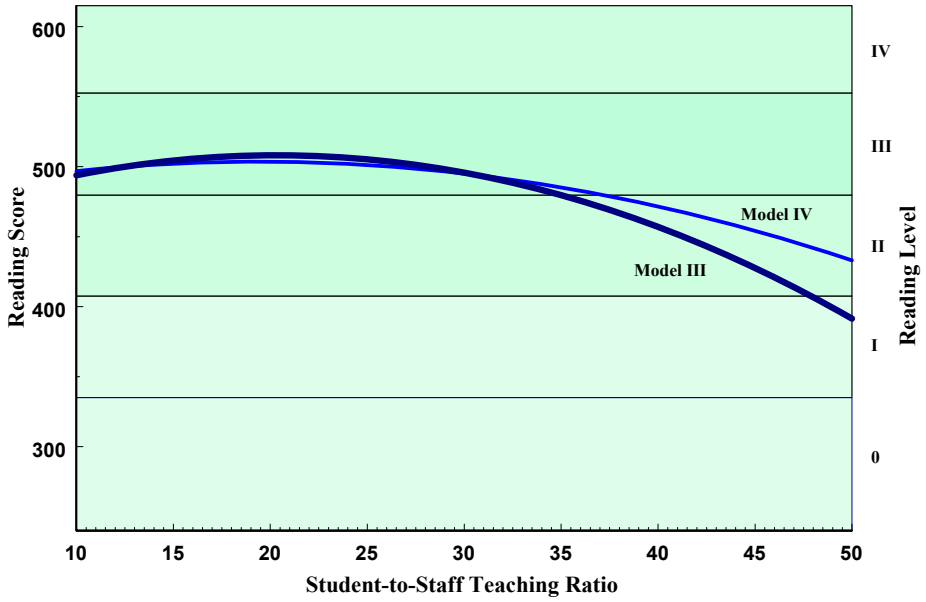
Students had better performance in schools where there was a higher percentage of teachers with tertiary-level qualifications in the test language. The effect of increasing the percentage of teachers qualified at this level by 10% is associated with an increase in performance of 2.5 points.

There was a negative effect associated with professional development: a 10% increase in the percentage of teachers receiving professional development was associated with a decrease in reading performance of 0.9 points.

The measures of school policy and practice did not yield significant findings in either study. However, the analysis indicates that these effects vary significantly among countries.

For PIRLS, the effects of the disciplinary climate of the classroom and parental support were both statistically significant. The results suggest that a one-point increase on the ten-point scales for these factors was associated with increases of 1.1 and 2.1 points in reading performance for disciplinary climate and parental support respectively.

**Figure 13. The relationship between reading performance and the student-to-staff teaching ratio.**




*Source:* PISA 2000-2002.

For PISA, four classroom practice variables emerged as statistically significant. A one-point increase on the teacher morale and commitment scale was associated with a 0.9 point increase in student performance. One-point increases on the disciplinary climate scale and the teacher-student relations scale were associated with increases in student performance of 3.5 and 3.3 points, respectively. A one point (ten percentile) increase in students' use of resources was associated with an increase in student performance of 4.7 points. Note that the effect associated with the quality of the school infrastructure was small and not statistically significant. Together these two findings suggest that the absolute level of resources is less important than the extent to which students use available resources.

**Policy implications**

Recent reviews of the literature on school effectiveness have stressed the importance of the effective use of class time and an approach to teaching that is structured and adaptive (Slavin, 1994; Nowacek, McKinney and Hallahan, 1990; Scheerens, 1992). The curriculum is also important, especially the extent of coverage and the content and pace of the curriculum (Alexander, 1982; Barr and Dreeben, 1983; Dreeben and Gamoran, 1986; Lee and Bryk, 1989). Other research has emphasised the context of the classroom and school, particularly the formal and informal mechanisms governing selection into particular schools and school programmes (Gamoran, 1986, 1990; Slavin, 1990; Willms, 1986).

The findings emanating from PIRLS and PISA are consistent with these general findings. But they also emphasise the importance of parental support, a strong school disciplinary climate and positive student-teacher relations.

# Question 9

## Do school resources and classroom practice differ in their effects in rural and urban schools?

### Differential effects in UNESCO's OREALC study

Findings from UNESCO's OREALC study of literacy skills in twelve Latin American countries also provide evidence that socioeconomic gradients vary considerably among countries (Willms and Somers, 2001). One of the most striking findings emerging from this research was Cuba's remarkable success. Detailed multilevel analyses of the OREALC data revealed that its success was not attributable to any single factor; rather to several factors, including universal daycare, greater parental involvement at home in educational activities such as reading to the child, smaller class sizes, higher levels of school and classroom material resources, better trained teachers, greater parental involvement in school, a strong classroom disciplinary climate and relatively few multi-grade or ability-grouped classes.

Within countries, there were also large differences in school performance among the private, urban public and rural sectors, and among schools within each of these sectors. Differences *among* sectors were mainly associated with material and human resources (e.g. smaller classes, more material resources, higher levels of teacher training), while variation among schools *within* sectors was mainly associated with schooling policies and practices (e.g. disciplinary climate, parental involvement). In addressing the issue of the effects of school resources, as well as school and classroom processes on student performance, we can ask whether the "resource-between" and "processes-within" results for OREALC also hold for PISA and PIRLS.

### Findings from PIRLS

The hypothesis is that school resources differ considerably *between* rural and urban schools and that these differences account for some of the urban-rural gap. However, *within* the urban and rural sectors, school policy and practice exert a stronger influence. This hypothesis was tested with the PIRLS data. The results are presented in **Table 8**.

The first row of the table provides estimates of the gap between urban and rural schools and between suburban and rural schools. The gaps, on average across the participating countries, are 23 and 16 points respectively. The second model controls for socioeconomic status, the student's sex and whether the student was foreign-born. The gaps reduce to 17.1 and 12.8 points for urban-rural and suburban-rural respectively. These variables account for 27% of the variation among schools. The third model includes the set of school resource variables presented in Table 7. These variables reduce the urban-rural and suburban-rural gaps to 14.1 and 10.7 points respectively, and account for an additional 4% of

the variance among schools. The fourth model replaces the school resource variables with the two classroom practice variables, parental support and disciplinary climate. These two factors reduce the gaps to 16.6 and 13.1 points for urban-rural and suburban-rural respectively and explain an additional 5% of the variation among schools.

The findings for both of these models are consistent with the resources-between and processes-within hypothesis, but they do not provide strong support for it. Rather, the evidence suggests that both resources, as well as policy and practice, weigh in to the explanation of the rural-urban divide. The final model in Table 8 includes factors related to demographics, school resources, policy and practice. Together these variables reduce the urban-rural and suburban-rural gaps to 14.3 and 11.4 point respectively, which is about one-third of the observed unadjusted gap.

**Table 8. The urban-rural divide**

| | Urban schools | | Suburban schools | | |
|---|---|---|---|---|---|
| | Effect | (SE) | Effect | (SE) | Variance (% explained) |
| **Unadjusted difference (vs. rural schools)** | **23.0** | (4.6) | **16.3**[sc] | (3.4) | 1,674 (0%) |
| **Adjusted for student background (SB) (SES, sex, and foreign-born)** | **17.1** | (4.0) | **12.8** [sc] | (3.2) | 1,221 (27%) |
| **Adjusted for school resources (SR), controlling for SB** | **14.1** | (3.7) | **10.7** | (3.1) | 1,153 (31%) |
| **Adjusted for school and classroom policy and practice (PP), controlling for SB** | **16.6** | (3.6) | **13.1** | (3.0) | 1,142 (32%) |
| **Adjusted for SB, SR, and PP** | **14.3** | (3.4) | **11.4** | (3.0) | 1,093 (35%) |

*Source:* PIRLS 2001.

## Policy implications

In responding to the findings of studies like PIRLS and PISA, many countries invest more resources in particular sectors of the schooling system or target resources to particular groups. However, these decisions are not usually based on detailed analyses of the data that could provide evidence to guide these decisions. An important consideration is whether children in different sectors or status groups differ in their access to particular educational resources. For example, with either the PIRLS or PISA data, we can establish cut-off scores for each of the resources and policy variables and ask: “What percentage of students in rural and urban communities are in schools with: small class sizes, high levels of classroom resources, teachers with specialised training in their subject area, positive student-teacher relations, etc.?” We could ask similar questions with respect to minority and non-minority students and high and low SES students. A more detailed multilevel analysis could help discern whether any particular school resource or classroom practice variables could explain some of the inequalities in school performance.

# Question 10

## Do school systems with less socioeconomic segregation have better performance and fewer inequalities?

### Findings from PISA

One of the most important findings of the results presented above is that there is a large compositional effect in every country. This finding is evident in the results for both PIRLS and PISA data. When students from low SES families attend schools with a low mean SES, they tend to have worse performance than they would have if they had attended schools with a high SES or heterogeneous intake. Students from high SES families tend to have better performance when they attend relatively high SES schools, but the effect is not as pronounced as it is for low SES students, especially for boys.

The "hypothesis of social inclusion" is that school systems with less socioeconomic segregation have better performance and fewer inequalities between high- and low-performing students; in other words, a higher and more level learning bar. This hypothesis is examined with the OECD PISA data.

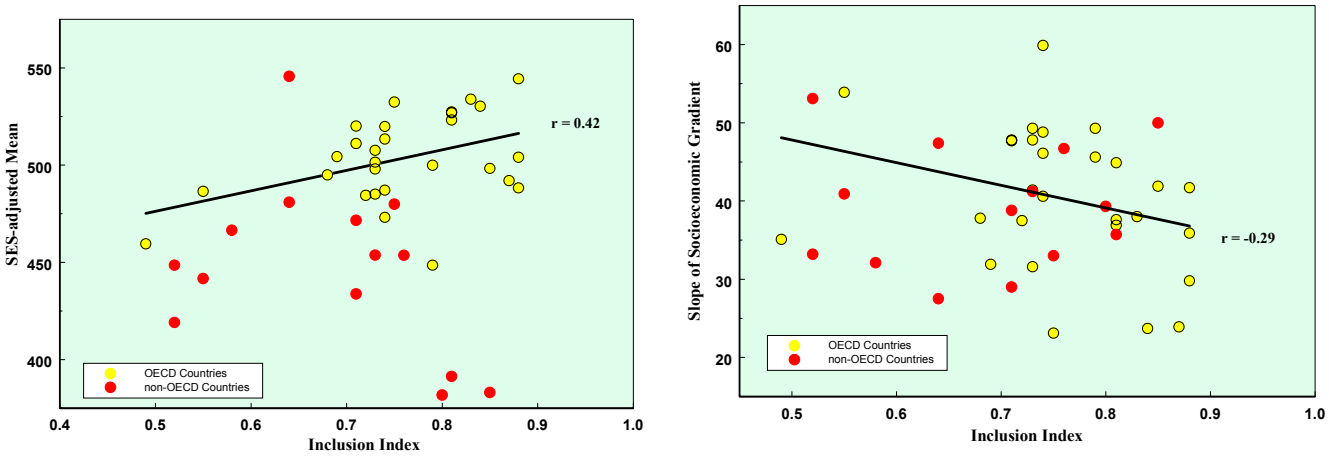
The overall socioeconomic gradients were presented in Tables 3 and 4. The section also noted that the overall socioeconomic gradient is comprised of separate gradients for each school and a gradient associated with the relationship between performance and SES at the school level. The overall gradient can be partitioned into an average within-school gradient and a between-school gradient. The decomposition is a function of the between-school slope, the average within-school slope, and $\eta^2$, which is the proportion of variation in SES that is between schools. The statistic, $\eta^2$, can be considered a measure of SES segregation (Willms and Paterson, 1995), which theoretically can range from zero for a completely desegregated system in which the distribution of SES is the same in every school to one for a system in which students within schools have the same SES score, but the schools vary in their average SES. The term, 1 - $\eta^2$, can be considered an index of social inclusion, which would range from zero for a segregated schooling system to one for a fully desegregated schooling system. Earlier it was noted that the overall gradient is related to the within- and between-school gradients through the segregation and inclusion indices:

$$\beta_t = \eta^2 \beta_b + (1-\eta^2)\beta_w ,$$

where $\beta_t$ is the overall gradient, $\beta_b$ is the between-school gradient, and $\beta_w$ is the average within-school gradient.

The left panel of **Figure 14** shows the relationship at the country-level between the level of the socioeconomic gradients (i.e. the SES-adjusted mean) and the inclusion index for each country. This relationship is positive (r = 0.42 for OECD countries), indicating that countries with greater social inclusion tend to have higher performance. The relationship between the slope of the socioeconomic gradient and social inclusion, shown in the left panel of Figure 14, is negative (r = -0.29 for OECD countries). Countries with greater social inclusion tend to have more gradual slopes. Taken together, these results suggest that more inclusive schooling systems have both higher levels of performance and fewer disparities among students of differing socioeconomic backgrounds.

**Figure 14. The relationship between the level and slope of the socioeconomic gradient and the inclusion index**




*Source:* PISA 2000-2002.

## Policy implications

In some countries socioeconomic segregation can be deeply entrenched due to economic divides between urban and rural areas and residential segregation in the cities. However, segregation can also stem from educational policies that stream children into certain kinds of programmes early in their school careers. For example, the reading performance and the slopes of the socioeconomic gradients in the Swiss French and Italian cantons stem partly from the allocation of students into academically-oriented and general academic programmes (Willms, 2003a). A comparison of school performance in Canada and the United States revealed that the average within-group gradients in the two countries are remarkably similar. The school profiles are also similar, except that the United States has a number of schools of very low SES and very low performance. It is these schools, where low SES students have been segregated, which account for most of the difference in the average reading scores between the two countries (Willms, 2004b). A detailed analysis of socioeconomic segregation within all countries that participated in PISA is presented by Willms (2004).

# Summary and discussion of findings

This report attempts to make a case for using socioeconomic gradients as a framework for educational assessment. The framework sets out ten key policy questions that are relevant to most school systems. The questions are addressed using data from two large-scale international assessments – the Progress in International Reading Literacy Study (PIRLS), which was conducted in 2001 under the auspices of the International Association for the Evaluation of Educational Achievement (IEA), and the Programme for International Student Assessment (PISA), which was conducted from 2000 to 2002 by the Organisation for Economic Co-operation and Development in cooperation with member countries. The analyses focus on the socioeconomic gradients associated with reading literacy in each country. A gradient is simply the relationship between some valued social outcome and socioeconomic status. In the case of literacy performance, the interest is in how literacy performance is related to students' family background. The argument underlying this analysis of socioeconomic gradients is that the distribution of student performance and socioeconomic status, and the relationship between them at the school and student levels, has implications for where educators and policymakers might place their emphasis in shaping educational policy.

The report sets out ten key questions relevant to the educational performance and equity of schools and schooling systems. The analyses of data from PIRLS and PISA are used to address these questions at a macro-level and to demonstrate how they might be relevant to analyses conducted within countries or other jurisdictions. The findings are discussed below.

**(1) Countries differ substantially in their average levels of reading performance.**

The results for PIRLS and PISA indicated that countries differ substantially in their educational performance. The differences among countries in PIRLS, which assessed students in grade 4, was considerably less than in PISA, which assessed students at age 15. The distribution of scores for many countries is skewed, with a disproportionate number of students scoring at the two lowest levels of performance. Countries with the highest average scores tend to be those that have reduced the number of very low scoring students.

**(2) In every country that participated in PIRLS and PISA, there is a significant relationship between reading performance and socioeconomic status.**

Although there is a significant SES gradient in every country, these relationships differ considerably among countries. The results indicate that the variation among non-OECD countries is greater than that among OECD countries. Also, the results suggest that children from high SES backgrounds in non-OECD countries did not fare as well in their reading performance as their counterparts in OECD countries.

**(3) Schools make a difference: in every country there is significant variation among schools in their performance, even after taking into account the SES of students' families and the mean SES of the schools they attend. Countries also vary significantly in their performance, even after SES is taken into account.**

For the countries participating in PIRLS, about 31% of the variation in school performance was among countries, 21% was among schools within countries and 48% was among students within schools. The individual-level and school-level measures of SES accounted for about 7% of the variation among students within schools, 40% of the variation among schools within countries and 53% of the variation among countries. For OECD countries, about 7% of the variation in reading performance was among countries, 32% was among schools within countries and 61% was among students within schools. The individual-level and school-level measures of SES accounted for about 8% of the variation among students within schools, 63% of the variation among schools within countries and 38% of the variation among countries. The results for both studies also showed that the slopes of the gradients differ among countries, and the average within-school slopes vary among countries and among schools within countries.

These results have implications for how administrators report results for schools based on monitoring systems. In some countries, adminstrators report the average scores for individual schools in league tables, without taking account of the socioeconomic background of students attending each school. Although such reports are useful in that they show the extent of variation among schools in their performance, the findings are often used to infer that certain schools are better than others in terms of the rate at which students are learning. When used in this way, such comparisons are unfair to school administartors, teachers and students. These findings show that about 40% to 60% of the variation in school performance is attributable to differences in student intake. Moreover, other studies have shown that, when one also takes into account a measure of student performance or ability when students enter school, the prior measure together with SES explain an even larger proportion of the variance than SES alone (Willms and Kerckhoff, 1995). Ideally, the best measure for assessing schools in

their performance is a measure of student growth based on measures taken on at least three occasions (Willett, 1988; Willms, 2001).

**(4) In some countries, the relationship between reading performance and socioeconomic status was weaker at higher levels of socioeconomic status. However, in others, particularly non-OECD countries, the relationship was stronger at higher levels of socioeconomic status.**

In all cases of diminishing returns, the change in the slope of the gradient was relatively slight. If it were strong, it would be feasible to identify a low SES threshold below which performance falls off considerably. This cut-point could be used to concentrate reform efforts on bostering the performance of low SES students through SES-targeted interventions or through compensatory interventions for low SES families. But this is not the case: although the gradient levels off in some countries, the bend is slight and it is not feasible to identify a low SES threshold.

In many non-OECD countries, the gradient gets steeper as SES increases, indicating increasing returns. It is likely that many low SES students in these countries do not make the crtitical transition from learning-to-read to reading-to-learn during the primary grades and then are unable to benefit fully from their schooling experiences in later years. This finding calls for more detailed studies of children's growth trajectories in their emergeing literacy skills during the primary grades.

**(5) Successful schools tend to be those that bolster the performance of students from less advantaged backgrounds. Similarly, countries that have the highest levels of performance tend to be those that are successful in not only raising the learning bar, but also levelling it.**

These findings provide strong evidence that strong school performance and equity can go hand in hand. Although there are examples of schools and countries where this is not the case, the weight of the evidence from the PIRLS and PISA results is that it is possible to achieve both equity and excellence. This applies not only to countries but also to individual schools. For schools or countries that have relatively steep SES gradients, these results suggest that SES-targeted or a combination of SES- and performance-targeted interventions may be most effective in raising and levelling the learning bar.

**(6) In all countries there is a school "compositional effect" associated with the mean SES of the school. The average level of socioeconomic status of a community has an effect on social outcomes over and above the effects associated with individuals' socioeconomic status.**

This finding is perhaps the most important for educational policy in most countries, and calls for more detailed study within countries. In many countries there is a significant number of low SES schools, where the average SES of the school is below the 20$^{th}$ percentile for all OECD students (-0.82). Children attending these schools tend to be at a significant disadvantage than students with similar family backgrounds that have been integrated into schools serving more heterogeneous populations. In countries with high levels of SES segregation, polices that aimed to reduce SES segregation through compensatory reforms would likely bring considerable gains in raising and levelling the learning bar.

**(7) The findings indicate that schools with a heterogeneous intake of students, in terms of their family SES, have equally high performance as those with a homogeneous intake.**

This finding is important in that it shows that it is not necessarily advantageous to have a homogeneous student intake. Schools with heterogeneous intakes on average tend to perform as well as those with homogeneous intakes.

**(8) The effects of school mean SES are to some extent mediated by school-level factors. The most important factors explaining reading performance in PIRLS were teacher experience, the disciplinary climate of the classroom and parental support. In PISA they were student-to-staff teaching ratio, the proportion of teachers with tertiary-level qualifications, students' use of resources, teacher morale and commitment, the disciplinary climate of the classroom and teacher-student relations. The results do not support the popular belief that smaller class sizes, or lower teacher-student ratios, yield better results. In PIRLS, the children in large classes fared slightly better than those in classes with 20 to 30 students, while in PISA the average performance was fairly even over that range.**

These findings provide some direction about why schools vary in their performance among countries. However, they do not suggest that any single factor is all-important; rather, it is a combination of factors that together make a difference to school performance. The analyses also provide evidence that the effects of the most important school-level factors vary among countries. Consequently, it is not possible to identify a small set of factors that jointly explain why some countries perform better than others; the best "policy mix" for one country is not likely to be the same as that of any other country.

**(9) The differences between urban and rural sectors are associated with material and human resources, such as smaller classes, better quality material resources and higher levels of teacher training, and various aspects of school and classroom policy and practice**.

The study found only moderate support for the hypothesis that it is material resources that account for differences *between* sectors, while school policy and practices explain differences *within* sectors. The results call for a more detailed country-by-country analysis documenting how the learning divides among sectors differ among countries, and how these are related to larger macro-economic factors.

**(10) Countries with high levels of segregation along socioeconomic lines tend to have lower overall performance and greater disparities in performance between students from high and low socioeconomic backgrounds.**

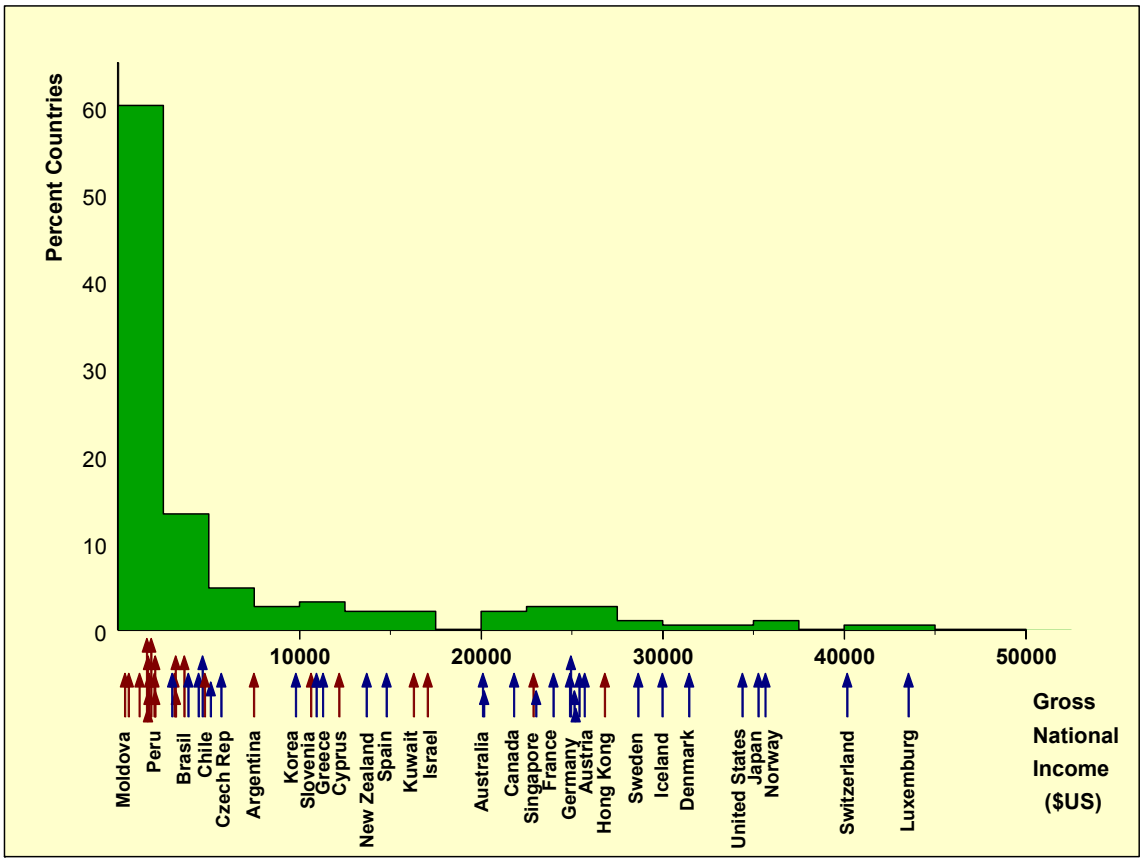
The results of PISA and PIRLS provide strong evidence that schooling systems can be highly inclusive and yield high literacy performance. There are a few exceptions, such as Hong Kong (Special Administrative Region of China), which has high overall performance despite a high degree of segregation. Yet generally, the results suggest that the top-scoring countries have low levels of SES segregation. The results also suggest that many of the countries with selective school systems have relatively high performance at the grade 4 level, before children are segregated into different schools and school programmes, but then fall well behind international standards at higher levels of schooling.

### Generalisability of results

The countries that participated in PIRLS and PISA tend to be relatively high-income countries. **Figure 15** shows the distribution of gross national income per capita (GNI)[13] for all countries that participated in PIRLS and PISA, with OECD countries indicated with blue arrows and non-OECD countries with red arrows.

[13] See http://web.worldbank.org/WBSITE/EXTERNAL/DATASTATISTICS/0,,contentMDK:20420458~menuPK:64133156~pagePK:64133150~piPK:64133175~theSitePK:239419,00.html.

**Figure 15. Gross national income per capita of countries participating in PIRLS and PISA (OECD in blue, non-OECD in red)**




*Source:* World Bank, 2000 data.

The average GNI (unweighted) per capita for the OECD countries is US$ 20,891, which is well above the world average of $US 5,220. The average GNI for the non-OECD countries that participated in the two studies was US$ 6,215. The figure indicates that 17 of the 25 non-OECD countries have a GNI below the world average. As the sample of countries is small and does not include any sub-Saharan African countries, we cannot claim that the participating non-OECD countries are representative of the substantial number of low- and middle-income countries.

## Concluding remarks

The analysis of gradients is a useful policy device because it sets an agenda for educational change that emphasises both excellence and equity – *the aim of educational policy is to raise and level the learning bar*. Gradients also provide a means to assess the likely effects of different kinds of interventions. Many of the international studies of student acheivement, as well as the school monitoring systems within countries, have collected information on students' family background and the important school and classroom processes relevant to school achievement. However, the links between data collection, analsyses and the reports that inform policy decisions have arguably been weak.

This paper argues that we can learn a great deal about a schooling system through analyses conducted in a gradient framework. The first step is in producing a portrait of the schooling system with a display of socioeconomic gradients and a school profile. This is followed by a more detailed analysis that examines: (1) the extent to which schools vary in their outcomes within the school system; (2) the extent that schools vary in their socioeconomic gradients; (3) whether gradients have a particular pattern, such as converging gradients or gradients with diminishing or increasing returns; (4) the level of between-school SES segregation and whether there are strong compositional effects associated with mean SES (and the same questions when ethnic segregation is an issue); (5) the variation within and among schools attributable to levels of school resources, as well as school and classroom policy and practice; and (6) and the extent of variation among sectors (e.g. rural versus urban, public versus private) in student outcomes and in the relevant school resource and classroom process factors. This kind of analysis can provide some indication as to what may be the most appropriate kind of intervention to raise performance and reduce inequities.

For many countries, especially those with high levels of SES segregation, an initial approach is to take measures to increase social inclusion. One of the most direct ways to level the learning bar, and to some extent raise it, is to reduce the number of children concentrated in very low SES schools. The school profile provides an indication of the number of schools with very low SES, which together with the inclusion index, indicates the need for this kind of intervention.

The analyses also suggest that countries or schooling systems with relatively steep gradients may pursue SES-targeted interventions in concert with performance-targeted and universal interventions. Countries or schooling systems with relatively gradual gradients are likely to achieve better results with some combination of performance-targeted and universal interventions. Performance-targeted interventions can be put in place to improve the results of low-performing schools or to improve the skills of low-performing students within schools.

For all schooling systems, a well-planned set of universal interventions is an essential component of a policy plan aimed at raising and levelling the learning bar. PIRLS and PISA provide some direction in this respect, as they include measures of both schooling resources and the within- and between-school processes that affect student performance. These measures can be supplemented with local efforts to monitor schooling outcomes and processes to discern what aspects of school policy and practice deserve the most attention.

When considering the information furnished by PIRLS and PISA, policy analysts tend to focus their attention on the schooling system, particularly on the features of the primary and secondary schools. This is natural as PIRLS is an assessment at grade 4 and PISA is an assessment of students at age 15. Indeed, the analyses pertaining to school effectiveness presented in this report are based on data describing school offerings at the late primary or secondary levels. However, PIRLS and PISA are not assessments of what youth learned during their previous year at school or even during their primary or secondary school years. They are an indication of the learning and skill development that has occurred since birth. A country's results on PIRLS and PISA also depend on the quality of care and stimulation provided to children during infancy and the pre-school years, and on the opportunities children have to learn both in school and at home during the elementary and secondary school years.

PIRLS and PISA provide a means to assess the performance of students in a consistent framework. For example, when data for reading performance are collected as part of PISA 2009, it will be possible to discern whether the bar has been raised over this nine-year period. The essential question then is, "How can we raise and level the bar?" or, specifically, "How can we improve the learning, behaviour and health outcomes of our youth, while reducing inequalities associated with family background?"[14] Raising and leveling the learning bar requires a long-term view and a broad perspective. For some countries, this may mean taking measures to safeguard the healthy development of babies or improving early childhood education. For others, it may mean social reforms that enable families to provide better care for the children, combined with efforts to reduce poverty, increase social inclusion and improve school offerings.

[14] The Canadian Research Institute for Social Policy (CRISP) at the University of New Brunswick has launched a five-year research programme, with major funding from the Canadian Research and the Social Sciences and Humanities Research Council, aimed at addressing this question. The Canadian Institute for Advanced Research (CIAR) and the New Brunswick Department of Education have also supported this research. The programme of research is being carried out by a national network of 30 scholars dedicated to research on the development of children and youth. The work is organised around five themes: (1) safeguarding the healthy development of infants; (2) strengthening early childhood education; (3) improving schools and local communities; (4) reducing segregation and the effects associated with poverty; and (5) creating a family-enabling society.

## References

Aitkin, M. and N. Longford (1986). "Statistical modeling issues in school effectiveness studies." *Journal of the Royal Statistical Society, Series A.,* 149(1): 1-43.

Al-Hassan, Suha (2005). Education reform for the knowledge economy (ERfKE) project. Phase III of the learning readiness assessment (national level). Unpublished manuscript. Jordan: The Hashemite University.

Alexander, K.L. and B.K. Eckland (1975). "Contextual effects in the high school attainment process." *American Sociological Review,* 4: 402-416.

Alexander, K.L., J. Fennessey, E.L. McDill and R.J. D'Amico (1979). "School SES influences – Composition or context?" *Sociology of Education,* 52: 222-237.

Alexander, K.L. (1982). "Curricula and coursework: A surprise ending to a familiar story." *American Sociological Review, 47*(October): 626-640.

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

Anderson, C.S. (1985). "The investigation of school climate." In G. R. Austin and H. Garber (Eds.), *Research on exemplary schools* (pp. 97-126). Orlando: Academic Press.

Barr, R.D. and R. Dreeben (1983). *How schools work.* Chicago: University of Chicago Press.

Becker, G.S. and N. Tomes (1986). "Human capital and the rise and fall of families." *Journal of Labor Economics,* 4(3): S1-39.

Ben-Schlomo, Y., I.R. White and M. Marmot (1996). "Does the Variation in the Socioeconomic Characteristics of an Area Affect Mortality?" *British Medical Journal*, 312(20): 1013-1014.

Bielby, W.T. (1981). "Models of status attainment." *Social Stratification and Mobility,* 1: 3-26.

Boyle, M. and J.D. Willms (1999). "Place effects for areas defined by administrative boundaries." *American Journal of Epidemiology,* 149(6): 577-585.

Bray, M. (1999). "The shadow education system: Private tutoring and its implication for planners." Fundamentals of Education Planning – 61. Paris: UNESCO. (http://www.unesco.org/iiep).

Brookover, W.B., J.H. Schweitzer, J.M. Schneider, C.H. Beady, P.K. Flood and J.M. Wisenbaker (1978). "Elementary school social climate and school achievement." *American Educational Research Journal,* 15(2): 301-318.

Bryk, A.S. and M.E. Driscoll (1988). *The high school as community: Contextual influences and consequences for students and teachers*. Madison: National Center on Effective Secondary Schools (ED 302 539).

Bryk, A.S., V.E. Lee and J.B. Smith (1990). "High school organization and its effects on teachers and students: An interpretative summary of the research." In W. H. Clune and J. F. Witte (Eds.), *Choice and control in American education. Volume 1: the theory of choice and control in education.* London: Falmer Press.

Burstein, L. (1980). "Issues in the aggregation of data." *Review in Research in Education*. D. C. Berliner. Washington, D. C.: American Research Association.

Cairns, R., B. Cairns and H. Neckerman (1989). "Early school dropout: Configurations and determinants." *Child Development,* 60(6): 1437-1452.

Carbonaro, W. (2005). "Tracking, students' effort, and academic achievement." *Sociology of Education,* 78*:* 27-49.

Coleman, J.S. (1988). "Social capital in the creation of human capital." *American Journal of Sociology,* 94: 95-120.

Cook, C. and J.D. Willms (2002). "Balancing work and family life." In J. D. Willms (Ed.), *Vulnerable children: Findings from Canada's National Longitudinal Survey of Children and Youth.* Edmonton, AB: University of Alberta Press*.*

Cooper, H., B. Nye, K. Charlton, J. Lindsay and S. Greathouse (1996). "The effects of summer vacation on achievement test scores: A narrative and meta-analytic review." *Review of Educational Research,* 66: 227-268.

Crane, J. (1991). "The epidemic theory of ghettos and neighbourhood effects on dropping out and teenage child bearing." *American Journal of Sociology,* 96(5): 1226-1259.

Datcher, L. (1982). "Effects on community and background on achievement." *Review of Economics and Statistics,* 64(1): 32-41.

Dreeben, R., and A. Gamoran (1986). "Race, instruction, and learning." *American Sociological Review,* 51(5): 660-669.

Duncan, G.J. and J. Brooks-Gunn (1997). *Consequences of growing up poor*. New York: Russell Sage Foundation.

Ensminger, M. and A. Slusarcick (1992). “Paths to high school graduation or dropout: A longitudinal study of a first grade cohort.” *Sociology of Education,* 65(2): 95-113.

Epelbaum, M. (1990). “Sociomonetary patterns and specifications.” *Social Science Research,* 19(4): 322-347.

Finn, J.D. and D.A. Rock (1997). “Academic success among students at risk for failure.” *Journal of Applied Psychology*, 82(2): 221-234.

Gamoran, A. (1986). Instructional and institutional effects of ability grouping. *Sociology of Education,* 59(October), 185-198.

Gamoran, A. (1990). “Instructional organizational practices that affect equity.” In H. P. Baptiste, H. C. Waxman, J. Walker de Felix and J. E. Anderson (Eds.), *Leadership, equity, and school effectiveness.* Newbury Park, CA: Sage.

Goldstein, H. (1986). “Multilevel mixed linear model analysis using iterative generalized least squares.” Biometrika, 71(1): 43-56.

Greaney, V. (1986). “Parental influences on reading.” *Reading Teacher,* 39(8): 813-818.

Hart, B and T.R. Risely (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Brookes.

Heath, A. (1990). “Class inequalities in education in the twentieth century.” *Journal of the Royal Statistical Society, Series A,* 153(1): 1-16.

Henderson, V., P. Mieszkowski and Y. Sauvageau (1978). “Peer group effects and educational production functions.” *Journal of Public Economics,* 10(1): 97-106.

Hertzman, C. and M. Weins (1996). “Child development and long-term outcomes: A population health perspective and summary of successful interventions.” *Social Science Medicine,* 43(7): 1083-1095.

Ho, S.-C. and J.D. Willms (1996). “The effects of parental involvement on eighth grade achievement.” *Sociology of Education,* 69: 126-141.

House, J., R. Kessler, R. Herzog, R.P. Mero, A.M. Kinney and M.J. Breslow (1990). “Age, socioeconomic status, and health.” *The Milbank Quarterly,* 68(3): 383-411.

Janosz, M., M. LeBlanc, B. Boulerice and R.E. Tremblay (1997). “Disentangling the weight of school dropout predictors: a test on two longitudinal samples.” *Journal of Youth and Adolescence,* 26(6): 733-762.

Japel, C., C. Normand, R. Tremblay and J.D. Willms (2002). "Identifying vulnerable children at an early age." In J. D. Willms (Ed.). *Vulnerable Children: Findings from Canada's National Longitudinal Study of Children and Youth* (pp. 105-20). Edmonton: University of Alberta Press.

Johnson, M.K., R. Crosnoe and G.H. Elder, Jr. (2001). "Students' Attachment and Academic Engagement: The Role of Race and Ethnicity." *Sociology of Education,* 74(4): 318-340.

Kaplan, G.A., E.R Pamuk, J.W. Lynch, R.D. Cohen and J.L. Balfour (1996). "Inequality in income and mortality in the United States: Analysis of mortality and potential pathways." *British Medical Journal,* 312(7037): 999-1003.

Kennedy, B.P., I. Kawashi and D. Prothrow-Stith (1996). "Income distribution and mortality: Cross sectional ecological study of the Robin Hood index in the United States." *British Medical Journal,* 312: 1004-1007.

Kraemer, H.C., E. Stice, A. Kazdin, D. Offord and D. Kupfer (2001). "How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors." *American Journal of Psychiatry*, 158: 848-856.

Lau, L.J. (1979). "Educational production functions." *Economic Dimensions of Education.* Washington DC: National Academy of Education.

Lee, V.E. And A.S. Bryk (1989). "A Multilevel Model of the Social Distribution of High School Achievement". *Sociology of Education,* 62: 172-192.

Lee, V.E. and J. Smith (1993). "Effects of school restructuring on the achievement and engagement of middle-grade students." *Sociology of Education,* 66(3): 164-187.

Levin, H.M. (1987). "Accelerated schools for disadvantaged students." *Educational Leadership*, 44(6): 19-21.

Lynch, J., G.A. Kaplan, E.R. Pamuk, R.D. Cohen, K.E. Heck, J.L. Balfour and I.H. Yen (1998). "Income inequality and mortality in metropolitan areas of the United States." *American Journal of Public Health*, 88(7): 1074-1080.

Madden, N.A., R.E. Slavin, N.L. Karweit and B.J. Livermon (1989). "Restructuring the urban elementary school." *Educational Leadership*, 46(5): 13-18.

McPherson, A.F. and J.D. Willms (1986). "Certification, class conflict, religion, and community: A socio-historical explanation of the effectiveness of contemporary schools." In A. C. Kerckhoff (Ed.), *Research in sociology of education and socialization,* Volume 6 (pp. 227-302). Greenwich, CT: JAI Press.

McPherson, A.F. and J.D. Willms (1987). "Equalization and improvement: Some effects of comprehensive reorganization in Scotland." *Sociology,* 21(4): 509-539.

Mirowsky, J. and P. Hu (1996). "Physical impairment and the diminishing effects of income." *Social Forces*, 74(3): 1073-1096.

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

Mullis, I.V.S., M.O. Martin, E.J. Gonzalez and A.M. Kennedy (2003). *PIRLS 2001 International Report: IEA's Study of Reading Literacy Achievement in Primary Schools.* Chesnut Hill, MA: Boston College.

Nechyba, T., P. McEwan and D. Older-Aguilar (2004). *The impact of family and community resources on student outcomes: An assessment of the international literature with implications for New Zealand.* New Zealand Ministry of Education.

Nowacek, E., J. McKinney and D. Hallahan (1990). "Instructional behaviors of more and less beginning regular and special educators." *Exceptional Children,* 57: 140-149.

Organisation for Economic Co-operation and Development (2001). *Knowledge and skills for life: First results from the OECD programme for international student assessment (PISA) 2000*. Paris: OECD.

Pallas, A. (1988). "School climate in American high schools." *Teachers College Record,* 89: 541-553.

Plewis, I. (1991). "Using multilevel models to link educational progress with curriculum coverage." In S. W. Raudenbush and J. D. Willms (Eds.), *Schools, classrooms, and pupils: International studies of schooling from a multilevel perspective* (pp. 149-166). San Diego: Academic Press.

Raudenbush, S.W. and A.S. Bryk (1986). "A hierarchial model for studying school effects." *Sociology of Education,* 59: 1-17.

Raudenbush, S.W. and R.M. Kasim (1998). "Cognitive skill and economic inequality: Findings from the National Adult Literacy Study." *Harvard Educational Review,* 68(1): 33-79.

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

Robertson, D. and J. Symons (1996). "Do Peer Groups Matter? Peer Groups Versus Schooling Effects in Academic Attainment." Discussion Paper No. 311, Centre for Economic Performance, London School of Economics.

Rogot, E., P.D. Sorlie and N.J. Johnson (1992). "Life expectancy by employment status, income, and education in the National Longitudinal Mortality Study." *Public Health Report*, 107: 457–461.

Rumberger, R.W. and J.D. Willms (1992). "The impact of racial and ethnic segregation on the achievement gap in California high schools." *Educational Evaluation and Policy Analysis,* 14(4): 377-396.

Rumberger, R.W. (1995). "Dropping out of middle school: a multilevel analysis of students and schools." *American Educational Research Journal,* 32: 583-625.

Rutter, M. (1983). "Schools effects on pupil progress: Research findings and policy implications." *Child Development*, 54(1): 1-29.

Sameroff, A.J., R. Seifer and P.K. Elias (1982). "Sociocultural variability in infant temperament ratings." *Child Development,* 53: 164-173.

Scheerens, J. (1992). *Effective schooling: Research, theory, and practice.* London: Cassell.

Sewell, D.H. and R.M. Hauser (1975). *Education, occupation and earnings.* New York: Academic Press.

Shavit, Y. and R.A. Williams (1985). "Ability grouping and contextual determinants of educational expectations in Israel." *American Sociological Review,* 50(1): 62-73.

Snow, C.E., M.S. Burns and P. Griffin (1999). *"*Preventing reading difficulties in young children." In *Reading research: Anthology: The why? of reading instruction.* Consortium on Reading Excellence (CORE), Novato, CA, US: Arena Press.

Sirin, S.R. (2005). "Socioeconomic status and academic achievement: A meta-analytic review of research 1990-2000." *Review of Educational Research*, 75(3): 417-453.

Slavin, R.E. (1990). "Achievement effects of ability grouping in secondary schools: A best evidence synthesis." *Review of Educational Research*, 60: 471-499.

Slavin, R.E. (1994). "Quality, appropriateness, incentive, and time: a model of instructional effectiveness." *International Journal of Educational Research,* 21(2): 141-158.

Voelkl, K. (1995). "School warmth, student participation, and achievement." *Journal of Experimental Education,* 63(2): 127-138.

White, K.R. (1982). "The relation between socioeconomic status and academic achievement." *Psychological Bulletin,* 91(3): 461-481.

Wilkinson, R.G. (1992). "Income distribution and life expectancy." *British Medical Journal,* 304: 165-168.

Wilkinson, R.G.. (1996). *Unhealthy societies: The afflictions of inequality*. London: Routledge.

Wilkinson, R.G. (2000). "The need for an interdisciplinary perspective on the social determinants of health." *Health Economics* (9): 581-3.

Willett, J.B. (1988). "Questions and answers in the measurement of change." In E. Z. Rothkopf (Ed.), *Review of research in education, Volume 15* (pp. 345-422). Washington, DC: American Educational Research Association.

Willms, J.D. and A.C. Kerckhoff (1995). "The challenge of developing new social indicators." *Educational Evaluation and Policy Analysis,* 17(1): 113-131.

Willms, J.D. and S.W. Raudenbush (1989). "A longitudinal hierarchical linear model for estimating school effects and their stability." *Journal of Educational Measurement*, 26(3): 209-232.

Willms, J.D. and M-A. Somers (2001). "Family, Classrooms, and School Effects on Children's Educational Outcomes in Latin America." *School Effectiveness and School Improvement,* 12(4): 409-445.

Willms. JD. (1986). "Social class segregation and its relationship to pupils' examination results in Scotland." *American Sociological Review*, 51: 224-241.

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

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

Willms, J.D. (2003a). *Ten hypotheses about socioeconomic gradients and community differences in children's developmental outcomes.* Report prepared for Human Resources Development Canada.

Willms, J.D. (2003b). "Literacy proficiency of youth: evidence of converging socioeconomic gradients." *International Journal of Educational Research,* 39(3): 247-252.

Willms, J.D. (2004a, October). *Compositional effects on student outcomes*. Paper presented at the Jacob Foundation conference on educational influences.

Willms, J.D. (2004b, October). *Reading Achievement in Canada and the United States: Findings from the OECD Programme for International Student Assessment*. Human Resources and Skills Development Canada.

Wolfson, M., G. Kaplan, J. Lynch, N. Ross, E. Backlund, H. Gravelle and R.G. Wilkinson (1999). "Relation between income inequality and mortality. Empirical demonstration." *British Medical Journal*, 319: 953-957.

Young, M.E. (Ed.) (2000.). *From Early Child Development to Human Development: Investing in Our Children's Future.* Washington, DC: The World Bank.

Zimmer, R.W. and E.F. Toma (1997). *Peer effects in private and public schools: Across country empirical analysis.* University of Kentucky.

# APPENDIX A

# VARIABLES USED TO DESCRIBE SCHOOL AND CLASSROOM POLICY AND PRACTICE

## School resources, policy and practice in PIRLS

Ten school-level variables pertaining to school resources, policy and practice were constructed from the PIRLS data, emanating from the teacher and school administrator surveys. They include:

***Quality of school infrastructure*** is a summary measure derived from school administrators' reports of the extent to which they felt that the school's capacity to provide instruction was affected by a shortage or inadequacy of: (a) instructional staff; (b) teachers qualified to teach reading; (c) instructional materials; (d) supplies; (e) school buildings and grounds; (f) heating, cooling and lighting systems; (g) instructional space; (h) special equipment for physically disabled children; (i) computers for instructional purposes; (j) computer software for instructional purposes; (k) computer support staff; (l) library books; and/or (m) audio-visual resources. One unit on this scale represents 10 percentile points, with higher scores indicating a better quality of school infrastructure.

***Class size*** was based on teachers' reports of class size. Two variables were constructed indicating class sizes below 20 and above 30, with 20 to 30 as the reference category.

***Teachers have a teaching certificate*** is a dichotomous indicator of whether the teacher had a teaching certificate.

***Time devoted to reading instruction*** was based on teachers' estimates of the time they devoted to reading instruction in a typical week. One unit on this scale indicates one hour of teaching.

***Disciplinary climate*** was based on principals' assessments on whether each of the following was a problem in their school: (a) classroom disturbance; (b) cheating; (c) profanity; (d) vandalism; (e) theft; and/or (f) intimidation or verbal abuse among students. One unit on this scale represents 10 percentile points, with higher scores indicating a more positive disciplinary climate.

***Parental support*** was based on the principals' overall assessments of whether their school could be characterised as having parental support for student achievement, as well as four items indicating the percentage of students whose parents would: (a) volunteer regularly to help in the

classroom or another part of the school; (b) attend teacher-parent conferences; (c) attend cultural, sporting, or social events at the school; and/or (d) do fundraising and other support activities for the school. One unit on this scale represents 10 percentile points, with higher scores indicating a greater parental support.

**School policy and practice in PISA**

The variables described below differ slightly, mainly in terms of scale, from those constructed by Willms for Chapter 8 of the international report on PISA (OECD 2001). These modifications do not result in any substantively significant differences in the relationships with reading performance.

**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 materials (e.g. textbooks); (e) not enough computers for instruction; (f) lack of instructional materials in the library; (g) lack of instructional materials in the library; and/or (h) inadequate science laboratory equipment. One unit on this scale represents 10 percentile points, with higher scores indicating a better quality of school infrastructure.

**Student-to-teaching-staff ratio** was defined as the number of full-time equivalent teachers divided by the number of students in the school. The model also included the square of this variable to capture any curvilinear effect associated with staff resources. For some analyses, an indicator describing the percentage of students in schools with student-to-staff ratios below 25 is used to describe this aspect of the schooling system.

**Teachers have tertiary-level qualifications in language** indicated the percentage of teachers in the school who were trained to the university level with a specialisation in a test language. One unit on this scale indicates 10 percentile points.

**Schools have at least one computer for every 10 students** was based on a question asked of the school administrator about how many computers were available to students. These data were used with total school enrolment to estimate the percentage of students who had computers. The variable is a dummy variable, such that the coefficient indicates the difference in performance between schools with computers available at this level and those which do not.

**Teachers receive professional development** was derived from a question asked of school administrators about the percentage of teachers who had received professional development in the previous three months. One unit on this scale indicates 10 percentile points.

**Students' use of resources** was derived from a question asked of students: "At your school, how often do you use … (a) school library; (b) computers; (c) calculators; (d) internet; and (e) science laboratories?" One unit on this scale represents 10 percentile points, with higher scores indicating greater use of resources.

**Teacher factors affecting school climate** 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 favourable attitudes and behaviours.

**Principal 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 principals had responsibility for these activities. One unit on this scale represents 10 percentile points, with higher scores indicating a higher level of principal autonomy.

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

**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. 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.

**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 tells students they can do better; (c) the teacher does not like it when students deliver careless work; (d) the teacher checks students' homework; and (e) 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.

**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' well-being; (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.

**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.