Quality and Equality in Brazilian Secondary Schools:  
A Multilevel Cross-National School Effects Study

Valerie E. Lee  
University of Michigan, USA  
velee@umich.edu

Creso Franco  
Pontifical Catholic University of Rio de Janeiro, Brazil  
creso@edu.puc-rio.br

Angela Albernaz  
Brazilian National Bank for Economic and Social Development  
angela@bndes.gov.br

Second Revision August 30, 2006

We gratefully acknowledge partial support for this work from the Brazilian National Research Council and the Ford Foundation. An earlier version of this paper was presented at the Annual Meeting of the American Educational Research Association, San Diego, CA, April 2004.
Quality and Equality in Brazilian Secondary Schools:  
A Multilevel Cross-National School Effects Study

Abstract

This study identifies features of Brazilian secondary schools linked to both high achievement (quality) and an equitable social distribution of achievement (equality). We use data from a current cross-national study, PISA [Programme for International Student Assessment], collected in 2000 in 32 countries by OECD. We focus on the reading achievement of 4,893 15-year-olds in 324 Brazilian secondary schools, compared to similar-sized samples of counterpart students and schools in PISA countries in Latin America (Mexico), Europe (Portugal, Spain) and the United States. Using multilevel methods (HLM), both appropriate for such analyses and fitting the PISA design well, we constructed almost identical models for separate analyses within each country. Quality and inequality in students' achievement in reading in each school were adjusted for students' retention and remedial history, gender, and students' and schools' SES. We estimated school effects with measures of school structure, academic and social organization. Though somewhat different across countries, we found school effects on quality and equality based on the resources upon which schools may draw, average absenteeism, academic press, and teacher/student relationships. Policy implications are emphasized.
**Introduction**

There is broad agreement that comparative studies in education should address, either directly or indirectly, questions of concrete educational and social policy (Husen and Postlethwaite 1985). We locate our research within such a cross-national and policy-relevant context, by focusing on both quality and equality in and between schools. Good secondary schools simultaneously evidence *quality* -- high levels of student achievement -- and *equality* -- an equitable social distribution of achievement across the social background of the school's students (Bryk, Lee, and Holland 1993; Lee and Bryk 1989; Raudenbush and Bryk 2002). Though educational contexts vary substantially between countries, this definition of "good schools" would surely be accepted by all countries. The means to develop and sustain good schools might, however, differ considerably across national education systems. This definition guides this study, which centers on secondary schools and students in Brazil. Though we identify the structural and organizational features of schools linked to achievement and its equitable distribution in Brazil, we embed the Brazilian focus within comparisons in both developed and developing countries. We compare Brazil to a developing country in Latin America: Mexico. We compare schools in these two countries to linguistic counterparts in two developed countries: Portugal, and Spain, and to schools and students in another developed country: USA. The analysis in all these countries is viewed through a cross-national lens that contrasts educational research findings based on family background and school resources.

We draw our comparisons from a recent and rich cross-national data source: Program for International Student Assessment [PISA], collected in 2000 by the Organization for Economic Cooperation and Development [OECD] in 32 OECD and non-OECD countries (OECD 2001). PISA-2000 focused on assessing the reading literacy achievement of 15-year-olds. We use separate analytic models in each country, using similar designs and methods, rather than a single analytic model including all five countries. Based on our definition of "good schools," our outcomes are (1) reading achievement (quality), and (2) the gradient between achievement and socioeconomic status [SES] (equality). Though educational contexts and sources of inequality in school outcomes vary substantially across countries, we assume that all countries share the goal of increasing both educational quality and equality.
Background

The Context of Cross-National School Effects Studies

What counts differs across countries. There is some agreement in international educational research that school factors are more strongly associated with achievement in developing countries, whereas family background is more important in developing countries (Fuller 1987; Heyneman 1976; Heyneman and Loxley 1983). Despite agreement on "what counts" for achievement in societies differentiated by levels of economic and social development, there is less agreement about why these differences occur. Heyneman and colleagues (1976, 1983) suggest that disparities in the measurement and distribution of such constructs across national contexts may explain, in whole or in part, this pattern of findings. Whereas there is substantial variation in basic physical facilities devoted to education in non-industrialized countries, almost all schools in developed nations have such facilities (Heyneman 1976).

Not only do the basic mechanisms of selection and background effects vary across societies based on levels of industrialization and economic and social development, but in single countries these mechanisms may vary across time. As a country develops, both opportunities for education and actual attainment typically expand (Blossfeld and Shavit 1993). To test this hypothesis, Shavit and Kraus (1990) investigated whether basic educational opportunities changed over time in one country (Israel) that simultaneously experienced profound social, demographic, and economic change. The authors were surprised at how stable the educational stratification process was over two decades of considerable social change in Israel in the mid-20th century, at least in this small but socially heterogeneous country.

Substantive or methodological explanation? The distribution of family background is varied in developed countries, whereas most children in under-developed countries live in poverty (Heyneman and Loxley 1983). In developed countries almost all children attend school, regardless of family background. This would result in less variation in social background for children attending school (especially secondary school) in developing than developed countries, where almost all adolescents are in school. Such differences across national contexts suggest a statistical as well as a substantive explanation: more variability in measures results in stronger effects. Another statistical explanation relates to selectivity (Heyneman and Loxley 1983; Mare 1993). In developing countries, where participation in even basic education is often far from universal, there is selectivity in which children actually either attend school or progress to higher levels of schooling among the least advantaged.
children. Differential selectivity would be more extreme at the secondary level, even if elementary school attendance were almost universal. If selectivity is substantial, the variation in the socioeconomic background of students who proceed beyond basic education would be constrained -- again restricting researchers' capacity to detect differential effects.

Such differences across national contexts led Heyneman and Loxley (1983) to question whether educational paradigms developed in research in industrialized societies would be consistently applicable to schooling in non-industrialized societies. Their cautions about cross-national comparisons were developed with elementary school children, and differential selectivity by level suggests that their conclusions may not generalize to secondary schools. Reflecting such cautions about cross-national comparisons was a study that identified social inequality in school outcomes from sources that differed widely across national contexts (Fuller and Clark 1994). Thus, it is unclear whether general conclusions about background or resources as the major determinants of educational outcomes across educational contexts are real, whether they are artifacts of differing conditions of education across countries with varied economic and social structures, or whether the same conclusions would apply at the elementary and secondary levels. As different social structures might lead to conclusions that may be either artifactual, statistical, or substantive, it may be more appropriate to analyze data separately for each country with some standardization of methodology and measures. We pursue this approach in this study.

Considering the link between social background and educational attainment across countries, Mare (1993) focused on methodological issues, including the importance of omitted variables that are systematically, but differentially, linked to children's educational transitions. At higher levels, selectivity is quite prevalent in countries where secondary education is a scarce commodity. His suggestions for improving analyses that estimate social background effects on educational attainment include transition models involving differential restriction to higher levels of educational attainment, opportunity models in labor markets, and models that focus on differences among several generations (to account for changes in social, economic, and educational conditions in individual countries over time). One solution Mare proposed for comparing relationships between background characteristics and educational outcomes across countries was to employ multilevel statistical models. We follow Mare’s suggestion in this study.

**Do schools matter?** Though most cross-national studies of educational attainment focus on differential effects of family background, it has become increasingly common to also consider the schools where students are educated. A seminal U.S. school effects study (Coleman et al. 1966) concluded that resource differences between schools were unimportant for student achievement
compared to social background. As described above, studies in developing countries typically demonstrate that school resources (what Fuller [1987] calls "material school inputs") are important determinants of students' academic success. These conclusions would lend support to the issues raised by Heynman (1976; 1983). However, there are much richer ways of characterizing schools than just resources. Some recent school effect studies in the U.S. and the U.K. highlight schools' social organizations and instructional practices (e.g., Barr and Dreeben 1983; Bryk et al. 1993; Lee and Bryk 1989; Rutter et al. 1979). Although there are surely relatively underresourced schools in these countries, the average resource base is generally much higher than in developing countries.

More recent data and methodological advances. Earlier cross-national studies focused more on attainment than achievement, mainly because it was easier to measure. Consideration of other outcomes (mostly achievement) as indicators of school effectiveness arose from data developed by the IEA¹ (Buchmann 2002). Using IEA data [SIMMS], Stevenson and Baker (1991) explored how control over curriculum (at the state, province, or local level) influenced what was taught in classrooms in 15 developed countries. Another study using IEA data [TIMMS] explored middle-school students' educational aspirations in 12 countries, linking it to the peer and parental influence (Buchmann and Dalton 2002). More recently, studies of school achievement in several sub-Saharan African countries were facilitated by data from the Southern African Consortium for Monitoring Education Quality (SACMEQ)². A study using SACMEQ data compared school effectiveness in 14 African countries (Lee, Zuze, and Ross 2005). That study identified many features of schools consistently linked with student achievement: composition, human and fiscal resources, and organizational characteristics.

These cross-national studies -- although focusing on different issues, countries, and questions -- use similar analytic models in analyses that are separate by country. Though most used regression methods, the most recent (Lee et al. 2005) used the multilevel methods suggested by Mare (1993). Regardless of analysis approach, methodological issues described here question the validity of earlier substantive conclusions about the relative importance of school resources and social background in international studies. These concerns have led us to take a somewhat cautious approach to cross-national research. Rather than making international comparisons between countries with widely varying social and cultural conditions, it seems preferable to compare countries with less extreme differences, as did Stevenson and Baker (1991) and Lee et al. (2005), or at least not to compare countries with widely varying levels of development (Buchmann and Dalton 2002).

Considering these issues carefully, we chose a middle ground. Though international comparisons are useful, we suggest that comparisons between countries be carefully considered. Here we compare educational contexts in both developing and developed countries, but our comparisons do not include
countries with very low levels of economic development. Rather, even our less developed countries (Brazil, Mexico) are more developed than countries in sub-Saharan Africa, the Middle East, or south Asia. Moreover, cross-national comparisons link countries with historical, linguistic, and/or hemispheric ties.

**Design counts.** The design and methodology of school effects studies have serious implications for what these studies find. A basic question in identifying characteristics of schools linked to student achievement is whether, after controlling for students' family background, there actually are residual school effects. The Coleman et al. (1966) study concluded that schools do not make much difference; rather, student performance was found to be almost entirely dependent on family background. School effects researchers have spent the last four decades trying to disprove that conclusion.

One issue is the **outcome measure.** Attempts to identify school effects on student achievement should account for differences in family background. Another issue is the **unit of analysis.** Though school effects are captured with school-level measures, achievement (and family background) are measured on students. Historically, school effects studies have skirted this issue by considering achievement as a school aggregate, rather than as a measure characterizing students. Thus, achievement was typically averaged across the school's students. However, much (usually most) of the variance in achievement is between students in the same school, rather than between schools. Using aggregates as outcomes discards most of the variance in achievement. Of the 60 school-effect studies in developing countries reviewed by Fuller (1987), all but one used school aggregated achievement as the outcome. School effects considered in these studies fell into a few categories: expenditures, material inputs, teacher quality, teaching practices, classroom organization, and school management. Family background was typically not considered.

**Measuring family background.** Though there is general agreement that students' family background must be accounted for in studies exploring school effects on educational outcomes, some question whether residual school effects remain after taking background into account. Even with agreement about the need to account for family background, how it is measured is inconsistent across countries and studies (Buchmann 2002). If background is improperly or weakly specified in school effects studies, the school effects may be biased (usually over estimated). Research linking SES to educational outcomes was developed within a theoretical tradition, primarily sociological, in seminal studies of status attainment (Buchmann 2002). How SES is (or should be) measured varies considerably across countries, as does the construct itself. U.S. social researchers have standardized SES as three components: parental education, parental occupational prestige, and family income.
School Effects in Brazil

Clearly, how the components of SES would be distributed and measured varies somewhat between industrialized and non-industrialized countries. Fortunately, cross-national educational studies from IEA, PISA, or SACMEQ have helped standardize the family background construct by measuring the components similarly across all countries that participate in such studies. Buchmann (2002) identified several goals for refining cross-national studies of school effects: (1) examining school effects net of family effects, (2) exploring links between family background and school outcomes in their own right, and (3) exploring achievement distribution both within and across country contexts. Despite these laudable goals, the relative importance of the several components of family background differs between countries with dissimilar social structures and industrial and economic development (Fuller and Clark 1994).

Controlling for student ability. Of the 60 studies Fuller (1987) reviewed, only one controlled for student ability (i.e., a pretest/posttest design). Almost all studies were cross-sectional. Research exploring student learning rather than achievement status is rare in developing countries. Of the several studies linking family background and achievement reviewed by Buchmann (2002), all used IEA cross-sectional data and, thus, focused on achievement status. Sadly, the many rich cross-national sources of educational data from IEA and even the more recent data from PISA and SACMEQ continue to use cross-sectional designs. However, these data sources have well developed measures of achievement, rich family background variables, and solid information about schools and schooling. But uniformly they lack measures of students' academic status as they enter the schools whose effects researchers wish to estimate. Researchers using cross-sectional data often try, sometimes with varying success, to capture proxy measures of student ability. We revisit this important issue later.

Studying School Effects

Two phases. Among the several issues related to design and measurement in cross-national studies that attempt to estimate school effects, the early studies used school aggregated achievement as outcomes, taking family background into account by including school averages of measures that actually capture characteristics of each school's students. These "Phase I" cross-national school effects studies were conducted before more appropriate analytic methods were available. Most studies reviewed by Fuller (1987) and Buchmann (2002) are of this type. In the last two decades, "Phase II" multilevel school effects have appeared. They depend on the availability of both more appropriate statistical methods and of hierarchically structured data sets. The Phase-II approach is summarized by Raudenbush and Willms (1991:1): "In nearly all countries, children learn in classrooms situated within
schools which are regulated by districts or educational authorities... within a legal and political framework constructed at the national level."

**Unit(s) of analysis.** A primary question in educational research concerns the unit of analysis. Educational outcomes accrue to individuals; instruction is implemented in classrooms; basic interventions are typically implemented in schools. Within a program evaluation format, Haney (1980) asked: Should researchers use the student as the unit of analysis, as educational interventions typically aim to influence student learning, or should the research be focused at the level where the intervention occurs (typically schools or classrooms)? Phase II school effects researchers would unanimously respond: "Both!" Studies with this orientation require educational data structured as schooling is: students nested in classes and/or schools. Most commonly, cross-national data have a two-level structure: students nested in schools.

**Multilevel methodology.** Fortunately, multilevel statistical software (e.g., Hierarchical Linear Modeling, or HLM) is now available to conduct Phase-II school-effects studies. The nested structure of the schooling process, and often the structure of the most appropriate data to capture this structure, may be capitalized upon by HLM and a multilevel approach. Such studies involve "a search for statistical associations between school factors, on the one hand, and student-level variables on the other hand" (Lee and Bryk 1989: 173). Most school-effects studies target secondary schools, exploring how academic achievement and equity are influenced by such factors as school size (Lee and Smith 1997), curriculum structure (Lee, Croninger, and Smith 1997); school restructuring (Lee 2001), teachers' attitudes (Lee and Smith 1996), and sector (Bryk et al. 1993, Lee and Bryk 1989). A dichotomous outcome such as dropping out may be explored as a function of school characteristics (Lee and Burkam 2003).

**Sources of educational inequality.** Here we focus on educational quality and equality in and between schools, seeking to identify school features that influence them. A inequitable social distribution of school outcomes, especially achievement, derives from many sources, which differ widely across national contexts (Fuller and Clark 1994). Inequality exists both between students in the same school and between schools. Between-school sources might include whether the school is public or private; how resources are allocated; whether students' access to secondary school is restricted (e.g., by exam scores); what curriculum and courses are available; whether some students drop out of school before secondary school; and/or whether school attendance is linked to residential locations that are usually economically stratified; as well as other explicit educational policies. Other sources of inequality occur within schools: whether students are retained in grade if their skills and behaviors are deemed inadequate to advance; whether low-scoring students are
channeled into remedial courses; whether the school's curriculum is differentiated into tracks or more- and less-demanding courses. Virtually all these sources of inequality -- which involve decisions by nations, states, schools, families, and students -- are associated with students' family background. These factors represent major mechanisms for creating within-school inequality, though educational systems may also stratify students' experiences in more nuanced ways.

Ideally, a school-effects study that investigates both quality and equality would have information on sources of inequality measured both within and between schools. Almost never are cross-national data available on all sources. However, some information is often available to capture within-school and between-school inequality. In this study, we explore sources of between-school inequality by considering school average SES, by several measures of school structure, academic and social organization. Within schools, we consider remedial coursework, grade repetition, students' gender, and SES.

Education in Brazil

Organization. Several sources of educational inequality typify Brazil, although the country has made major strides in expanding access in the last decade. The public schools serve mainly low-income families. Public school funding is mainly tied to states' and cities' budgets; thus, schools in relatively affluent locations (e.g., Sao Paulo, Rio de Janeiro) are better funded than in poor states (Bahia, Pernambuco) and rural areas. Brazilian secondary schools have an academic focus. Curriculum differentiation is rare, because non-academic course are not offered. Retention in grade and remedial placement are the most common educational responses to low student performance. Low-track courses that are common in U.S. are unavailable in Brazilian secondary schools.

Brazilian schooling has four levels: elementary (ideally serving ages 7 to 10), lower secondary (typically, enrolling ages 11-14), upper secondary (ages 15-17, if no retention occurs), and university (after upper secondary). Attendance of children aged 7-14 at the elementary and lower secondary levels is almost universal (97 percent), but upper secondary schools enroll lower proportions (only 82 percent of the 15-17 age cohort in Brazil attends school, half at the lower secondary level due to the prevalence of retention). From 1995 to 2001, enrollment in upper secondary school increased steadily from 5.3 to 8.4 million students (Franco, Albernaz, and Ortigao 2002). At all levels, education is offered in shifts: half-day schooling is ubiquitous in Brazil (as well as throughout Latin America and much of the developing world). Almost all Brazilian secondary schools offer morning and afternoon sessions; some run three sessions. Teachers typically move between schools by shifts or even within shifts.
Bolsa-Escola. During the 1990s, several local governments in Brazil adopted Targeted Conditional Transfers programs collectively called "Bolsa-Escola." These programs awarded cash grants to poor families with school-age children, on the condition that the children enroll in school and attend regularly (typically 85 percent attendance). In 2001, Bolsa-Escola was expanded nationwide, with the aim of increasing school enrollment, attendance, and education attainment for children from low-income families. To be eligible for Bolsa-Escola, families must have: (a) at least one school-aged child, (b) per capita income no more than half the minimum wage, and (c) children enrolled in school, with no more than two absences/month. The program's subsidy is the equivalent of US$5.00/month/child, with a maximum of US$15.00 per family. Payments, in the form of bank cards from the Ministry of Education, go directly to mothers. In 2003, four million families were enrolled in the program.

Evaluations of Bolsa-Escola indicate that it has helped reduce school dropout (World Bank 1997) and increase promotion rates (Sant'Anna and Moraes 1997). Income among the poorest ten percent of Brazilians also increased significantly (Arajo and Nascimento 2001), surely due at least in part to Bolsa Escola. In 2004, all cash transfer programs were merged into the broader Bolsa-Familia Program by the Brazilian government. For families with children aged 7-15, however, the new program still requires school enrollment and 85 percent attendance as a condition of cash transfer. These effects suggest that this social intervention is promising as a means to promote both educational equity and school attendance in developing countries.

Research Questions

Three theoretical frameworks guide our study of school quality and equality. First, we make cross-national comparisons of the relative importance of students' social background and schools' resources in links to student achievement, conceptualized within the theory first raised by Heyneman and colleagues (1976, 1983) and enriched by other sociologists interested in international education (e.g., Baker, Buchmann, Fuller, Mare, and Shavit). Second, we follow the conceptual, structural, and methodological guidelines developed in Phase-II school effects studies. Third, we shine our "school effects lens" on questions of concrete educational and social policy, as Husen and Postlethwaite (1985) urged. Although we focus on a single country -- Brazil -- we provide parallel analyses in four other countries that are not arbitrarily chosen and, we argue, make reasonable comparisons. Our study aims to identify characteristics of secondary schools in five countries that are associated with both quality and equality. We organize the study around four research questions. Though our focus is in Brazil, our cross-national study also explores these questions in comparison countries selected on two criteria: (1)}
each participated in PISA 2000, and (2) the countries, for reasons spelled out below, serve as reasonable comparisons for investigating these questions in Brazilian schools. Outcomes include both quality and equality, which we define with students' reading achievement and its equitable distribution by SES.

Question 1: Background and achievement. How are Brazilian secondary school students' social and academic background characteristics linked to their reading achievement? How do these relationships compare to counterpart students in Mexico, Portugal, Spain, and the U.S.? What might explain these differences?

Question 2: Educational quality. Which features of the organization and structure of Brazilian secondary schools are associated with school average reading achievement? Are these characteristics the same in comparison countries' schools? If not, what might account for differences across countries?

Question 3: Educational equality. Which features of Brazilian secondary schools' organization and structure are associated with an equitable distribution of achievement by students' SES within schools? Are these school features the same in the comparison countries? What might explain cross-national differences?

Question 4: Background and resources. How do the effects of students' social background and schools' resources on quality and equality compare between developed countries (Portugal, Spain, the U.S.) and developing countries (Brazil, Mexico)?

Method

Data

About PISA. The data for this research come from the 2000 PISA study, a collaborative effort to collect internationally comparable evidence on educational performance in 28 OECD member countries and four non-member countries (including Brazil). PISA's goal, to provide evidence for collaboration and policy dialogue to define and operationalize educational goals (OECD 2001a), coincides well with our aims. Each country that participated in PISA 2000 was required to commit substantial resources to the data collection effort. However, OECD staff standardized, and kept tight control over (a) the procedures for sampling schools and students, (b) collecting survey data from
students and principals in sampled schools, and (c) testing students. The majority of OECD members are developed countries (including most European countries and many of the most developed countries in Asia and in North America).

The design of PISA is cross-sectional. That is, data that are quite comparable across countries were collected toward the end of the 1999-2000 school year in each participating country. When schools open and close, of course, differs in northern and southern hemisphere countries. School years typically end in June in the northern hemisphere and in December in the southern hemisphere. Data collection was timed accordingly. Although PISA collected some assessment data in mathematics and science, the major focus of the 2000 data collection was on reading achievement. In subsequent 3-year PISA cycles, the focus will be on mathematics (2003) and science (2006). Our study uses the PISA measure of reading achievement as our dependent measure. Testing was conducted in the major language of each country, and considerable efforts were expended to make surveys and assessments comparable across language groups.

The focus of PISA is on secondary schools. Because different nations organize their schooling systems somewhat differently, the decision was to focus on students of a particular age (15 years old) rather than a particular grade. This age was chosen because it was assumed that in most countries, 15-year-olds who had progressed in school without repetition would be in their first year of upper secondary school. However, both because national educational systems are organized somewhat differently and because grade repetition is more common in some countries than others, the grade levels of 15-year-olds varied considerably both in the same countries and across countries.

Cross-national studies of education have become increasingly common over the last few decades. Countries that decide to participate engage in considerable discussion about what should be explored, what skills should be tested, how students and schools should be sampled within each country, and how survey and assessment instruments should be worded and organized to best capture education in each country. Despite their growing popularity, there is also considerable controversy about such studies, not only about the structure of data to be collected but also the comparisons that should be made across countries. A major value of using PISA data to study secondary schooling in Brazil is, of course, the ability to make comparisons between Brazil and other relevant countries. We argue that comparisons of education in a developing country like Brazil with many of the developed countries in OECD is of limited use. In fact, of the 32 countries participating in PISA 2000, Brazil's students scored dead last. The only other Latin American country, Mexico, has achievement test scores that were also close to the bottom.
Choosing comparison countries. Brazil was our focus for three reasons. First, one of our team was part of the original PISA design and data collection in Brazil, and two team members conducted analyses of the Brazilian PISA data for the Brazilian Ministry of Education. Second, another team member has participated in collaborative research during several visits to Brazil. Third, our team’s keen interest in educational equity and in identifying characteristics of schools associated with the equitable distribution of achievement suggested a school effects study. Although we wanted to make use of these excellent data for learning more about Brazilian schools, we felt it would capitalize best on PISA by making comparisons between Brazilian schooling and that in other countries which were included in the PISA 2000 study. Our logic in selecting countries' educational systems to include in our study is based on three criteria: (1) economic development; (2) continent location; and (3) linguistic similarity.

The choice of Mexico as a comparison country was obvious. Mexico, the only other Latin American country included in PISA 2000, is an OECD member. Mexico's average literacy achievement ranked just above Brazil's but lower than almost all other PISA 2000 participating countries. We searched for developed countries that would be reasonable comparisons to these two countries. Portugal was a natural choice, based on history and language. The link between Portugal and Brazil dates to the 16th century, when Portugal colonized Brazil. They are also the only countries in PISA that share the Portuguese language (students were tested in that language). We decided to include Spain, as it shared similar historical and linguistic links with Mexico. We included the United States as a comparison country mainly as a benchmark. Much of the body of research about school effects has been conducted in U.S. secondary schools. Our choice of comparison countries to Brazil among these included in PISA 2000 was purposive. All sampled countries except Brazil were OECD members.

Samples and sampling. PISA staff constructed stratified sampling frames for each country, including a minimum of 150 schools. School samples were stratified by type, size, region, urbanicity, and sector. Within each school selected for participation in PISA, 35 15-year-old students were randomly drawn (fewer only if there were less than 35 students of this age). Response rates were high; PISA required rates of at least 85 percent (sometimes achieved by including replacement schools -- OECD 2001b). PISA data include school and overall student weights, which are required in all analyses to adjust for the differing sampling probabilities for school and student selection. We use weights in all analyses in this study. Students in schools were selected randomly, so we used no within-school weighting in our multilevel analyses.
Our analytic samples include all students and schools in the PISA data files in each country. Our total sample contains almost 25,000 15-year-old students enrolled in almost 1,000 schools. In Brazil, this sample contains 4,893 students in 324 schools. Our Mexican sample includes 4,600 students in 183 schools. Portugal's PISA sample contains 4,585 students in 149 schools. The Spanish sample includes 6,214 students in 185 schools. The U.S. sample contains 3,846 students in 153 U.S. schools. With weights, school and student samples in each country are nationally representative. Results are generalizable to secondary schools and their 15-year-old students in these countries.

PISA data come from three sources: (1) direct assessments of student achievement; (2) surveys of students; and (3) surveys of school principals. As PISA did not survey teachers, the information we use to characterize schools is from two sources: (1) reports from principals about the physical, academic, and social characteristics of their schools; and (2) reports from students about their schools, which we aggregate to the school level.

Measures

Our measures characterize both students and schools. We provide brief descriptions of the variables used in this study, but more detailed descriptions of the PISA measures and the procedures used to construct composite measures are provided in Appendix 1. The measures described in the text were retained in our final analytic models. However, we considered many other school characteristics which we subsequently deleted due to non-significance in any country we considered. Those measures also listed in Appendix 1 but not described in detail. Of course, our ability to capture constructs of interest was limited to measures available in PISA.

Variables describing students. Our dependent variable of focus is reading achievement, which measures "the ability to understand, use, and reflect on written texts in order to achieve one's goals, to develop one's knowledge and potential, and to participate effectively in society" (OECD 2001a:21). Students were allowed two hours to complete a paper-and-pencil test. Test items took the form of multiple-choice, short answer, and extended responses. Items were drawn from a wide range of cultures and languages, which is appropriate for any international comparative study (OECD 2001b). Scores were equated using Item Response (IRT) methods, as not all students were asked to complete the same items.

Our student-level independent variables are of two types. One type is a series of measures we used as proxies for ability, to capture students’ academic status as they began upper secondary school. Two are dummy-coded indicators (scored 0 or 1), that indicate whether students were delayed for one year or for more years from the expected grade (both compared to students in the expected grade) as
consequence of repeating grades, or starting schooling late, or “stopping out”. Another dummy variable indicates whether students reported participating in remedial classes in the previous three years. Two student-level independent variables are included as demographic controls. The first is a rich composite measure of socioeconomic status, or SES, including parents' occupational prestige, parents' education, family wealth, and household cultural possessions. This measure of SES is standardized across OECD countries (M=0, SD=1). We also include a statistical control for student gender (females coded 1, males 0).

Variables describing schools. We created and explored a wide range measure of school composition, structure, and organization. Except when mentioned, these are standardized (M = 0, SD = 1). Our descriptive results present such variables standardized across the entire OECD sample, but in separate-by-country analyses we re-standardized them within each country. All are described in Appendix 1. School composition we capture with a school aggregate of SES. We included two measures of school structure focusing on resources. One measures the schools' physical infrastructure (reports from principals about building conditions and crowding); another captures pedagogical resources (the only relevant PISA measure was drawn from reports by students of their access to science laboratories).

We consider both the academic and social organization of schools. School academic organization is captured with a single measure of school academic press, where students report how hard their teachers push them (and help them) to succeed. More numerous are our measures of school social organization. One composite focuses on principals' reports of problems due to student absenteeism (from school and class). Another composite captures teachers' commitment. Two composites measure how school members relate to one another. One is an aggregate of students' reports of the degree to which they feel a sense of belonging in the school. We also include a composite measure of teachers' collective responsibility for student learning.

School-level measures we considered but subsequently deleted from our analytic models include a dummy-coded indicator of school sector (coded 1 for private schools, 0 for public schools), several measures of computer availability and usage, measures of autonomy (of schools, teachers), a measure of problems in teacher-student relations, measures of school climate (disciplinary, teaching, student), principal reports of problems due to teacher supply: shortages, weak qualifications, or lack of certification. More detail on all measures is provided in Appendix 1.
Analytic Method

Using HLM. As both the PISA data structure and our research questions are multilevel, our multivariate analyses use Hierarchical Linear Models [HLM] (Raudenbush and Bryk 2002). The use of HLM is strongly recommended by PISA for analysis of their data, and most analyses in PISA reports have used HLM (OECD 2001a; 2001b). Parallel within-school HLMs in each of the five countries address Research Question 1; parallel two-level HLM school-effects models address Research Questions 2 and 3. Comparative results for Level-1 and Level 2 HLMs address Research Question 4. All analyses were weighted at the school level to adjust for sample stratification and oversampling, using weights normalized to a mean of one in each country to assure accurate tests of statistical significance.

We conducted separate HLM analyses in each of the five countries included in this study. The descriptive information we report about the countries describes countries as they compare to one another. However, in our multivariate and multilevel analyses within each country, we standardized all continuous variables as z-scores (M = 0, SD = 1) within each country. However, we preserved the dependent variable – reading achievement -- in its original PISA metric. This IRT-equated score was standardized to M= 500, SD = 100 for the 28 OECD countries in PISA.

Structure of HLM our analyses. Using HLM in a school-effects format, we conducted our analyses in three steps, separately for each country. In Step 1 (fully unconditional model), we partitioned the variance in reading achievement into its between-school and within-school components. The proportion of total variance that lies systematically between schools, or intraclass correlation (ICC), is one indicator of inequality between schools in a country. Moreover, it is only that proportion of variance that lies between schools that may be influenced by school effects. In Step 1 we also estimate the reliability of reading achievement in each country.5

Step 2 of our HLM analyses (unconditional at Level 2) estimates within-school models. Here we address Research Question 1 by investigating how the student-level independent variables in our analyses -- retention history, remedial course enrollment, SES, and gender, are associated with reading achievement. Student-level independent variables may be considered in two ways -- fixed or random effects. For fixed effects (typically included as statistical controls), we centered these variables around the grand mean for each country. By "fixed," we mean that between-school variances in the relationship of these variables to the outcome are fixed to zero. We also estimate random effects, whereby the relationship between independent variables and the dependent measure are allowed to vary between schools, which become dependent variables at Level 2 and may be modeled as measures of between-school equity. These independent variables are centered around their respective school
means. We considered SES as a random effect, so we centered it around the mean SES in each school. The relationship between SES and reading literacy achievement responds to Research Question 1.

Thus, our analyses investigate two dependent variables as functions of school effects: (1) school average reading literacy, adjusted for retention history, remedial class enrollment, SES, and gender; and (2) the relationship between SES and reading achievement in each school adjusted for retention, remedial and gender. By our definition, "good" schools simultaneously evidence both high average achievement (quality) and a modest relationship between achievement and SES (equality). In Step 3 of our HLM analyses, we search for school effects that are both positively associated with high average achievement (Research Question 2) and negatively associated with SES achievement slope (Research Question 3). The major results of our study are found in our between-school (step 3) HLM models, although results addressing Research Question 4 and in HLM steps 2 and 3.

Model-building strategy. Our within-school HLM models, separate in each country, contain two outcomes: average achievement and the SES/achievement slope. With so many measures of potential school effects, however, we needed a strategy that simultaneously recognized not only differences within countries but also allowed for comparison between countries. In the same conceptual groupings in which we described the school variables above (but also including those we subsequently deleted from models), we began our search for school effects, estimated in full HLM models.

One by one, we deleted variables from our models that did not achieve at least a .10 probability level in at least one country. We ran a large number of HLMs, as each time a variable was deleted the coefficients for other variables changed somewhat. Our semi-final models (what we called "saturated models") retained school-level variables in the analyses for each country that were statistically significant in at least one country. Our final HLM school-effects models also retain these variables, but we deleted them in each country if they were not statistically associated with either average reading achievement or the SES/achievement slope in that country. Thus, our final HLM models are not completely identical across countries. Readers interested in the results from the fully parallel saturated models may find them in Appendix 2.

Results

Descriptions of Secondary Education in Five Countries

Students. The academic and social background characteristics of 15-year-old students in each country are reported in Table 1. Average reading achievement in Brazil is lowest of the 32 PISA
countries (403); it compares unfavorably to the OECD countries' average (500) and to averages in Mexico (423); Portugal (472); Spain (493); and USA (504), all of whom are OECD members. Correspondingly, Brazilian and Mexican students' SES is over 1 SD below the PISA countries' mean, Portugal's .41 SD below, Spain's .24 SD below, USA's .16 SD above. The proportion of Brazilian students below the appropriate grade level is quite high (42 percent are one or two grade below the expected grade), compared to 14 percent in Mexico, 20 percent in Portugal, 4 percent in USA, and 2 percent in Spain. Over 40 percent of students in Brazil, Mexico, and Portugal reported having engaged in remedial coursework within the three past years, compared to fewer that 30 percent in Spain and the USA.

Brazilian and Mexican secondary-school students, although more academically and socially disadvantaged than comparison counterparts, were actually somewhat more select within their countries. This is because fewer adolescents in Brazil and Mexico (71 percent and 60 percent) attend secondary school than comparison countries (85-94 percent). If 15-year-olds who have already left school or who were not in secondary school were included in the Brazilian and Mexican samples, achievement differences would likely be every greater.

-------------------------
Insert Table 1 about here
-------------------------

Which types of students are retained in grade? As variables capturing students' retention histories serve as proxy measures of ability, we explored mean differences between retained and non-retained students in each country (Table 2). Achievement differences in effect-size units (standard deviations) between the two types of students are very large: .76 SD in Brazil; .97 SD in Mexico; 1.36 SD in Portugal; 1.24 SD in Spain; 1.52 SD in the U.S. That they are relatively smaller (although still large) in Brazil may be explained by the fact that so many students are retained (48 percent of Brazilian students). SES differences are also quite large between retained and non-retained students, indicating that less advantaged students are more often retained. Differences in SES by the relation status of students range from .59 SD in Brazil to 1.18 SD in the U.S.; all very large differences. This phenomenon seems to reflect the pattern described by Heyneman and colleagues (1976, 1983). Girls are somewhat less likely to be retained in all countries; largest in the U.S. (a difference of 22 percent) but closer to equal in other countries (albeit an 8 percent advantage).
Schools. Differences among secondary schools in the five countries are displayed in Table 3. Unsurprisingly, average school SES follows the same cross-country pattern as student SES: very low in Brazil and Mexico, with Portugal falling well below the OECD average but relatively higher than the two Latin American countries. Spain and the U.S., although higher, are still slightly below the OECD mean. The standard deviation of within-school SES is a measure of social mix within each school. The greater the standard deviation of the within-school SES, the wider is the social diversity within schools. By this measure, social mix within schools is much lower in Mexico and Brazil than the other countries. As described earlier, in many Latin American countries, private schools are attended by almost all students whose parents can afford to pay the tuitions. This phenomenon would exacerbate between-school social inequalities, shown to be large in the two Latin American countries. About half of all Spanish schools in the sample are private, about a fourth of schools in Brazil, about a sixth in Mexico and the U.S., and only about 10 percent in Portugal. In some countries (Spain, Portugal) private schools are subsidized at least in part by the state; however, in Brazil, Mexico, and the U.S. they are funded by tuitions paid by parents. Resource differences follow the same pattern as average SES; the least affluent countries have the lowest resources --measured in terms of both physical infrastructure and pedagogy.

How do the descriptive results in Tables 1-3 to compare the countries on the criteria on which we selected them (development, location, and language)? Economic development seems to be the most differentiating factor, as the less developed countries (Brazil, Mexico) show lower reading achievement and SES, lower proportions of students attending secondary school, and schools that are simultaneously lower average SES and less socially mixed. The number of countries considered in this study is modest, so developmental matching coincides with the locational matching (i.e., Brazil and Mexico are in Latin America, Spain and Portugal in Europe). Language matching indicates that Brazil and Portugal are more similar (particularly in the high proportions of grade repetition) than Mexico and Spain. Among our developed countries, Portugal's students are schools are less socially advantaged than those in Spain and the U.S., although more advantaged than in the Latin American countries. The measures that capture school social and academic organization demonstrate few
discernable patterns associated with our country selection criteria. These descriptive statistics suggest that the variables measuring school characteristics could have somewhat different meanings in different countries.

How respondents interpret the meaning of questionnaire items may differ somewhat across cultural contexts. This suggests that separate analyses for each country are preferable to an overall analysis with all countries taken together, an approach also suggested by most cross-national educational researchers. Clearly, with such differences in school average SES between countries and with substantial differences in variability between schools in average SES within countries, these measures must be taken into account when estimating school effects that focus on organizational and structural differences between schools in the five countries we study.

**Multilevel Models**

**Step 1: Partitioning variance.** Results of fully-unconditional HLM models for each country are displayed in Table 4. The average number of students per school (within-school sample sizes) varies somewhat between countries, with the largest in Spain (34) and Portugal (31), still quite large in U.S. and Mexico (25), and smaller in Brazil (15). Within-school parameters of interest -- average achievement and the SES/achievement slope -- are estimated more reliably with larger within-school sample sizes. Fortunately, the dependent variable is very reliably estimated in all countries (from .83 in the U.S. to .92 in Mexico). Especially important also in Table 4 are the intraclass correlations (ICCs) -- the proportions of the total variance in reading achievement that lies systematically between secondary schools. These proportions are quite high in the Latin American countries (.52 in Mexico, .42 in Brazil); somewhat lower ICCs are in the U.S. (.38) and Portugal (.35). Noteworthy is the very low ICC in Spain (.19).

---

*Summary of fully unconditional HLM models.* The outcome is quite reliably measured in all countries. The ICCs can be seen as indicators of inequality between schools in these countries – quite high in Brazil and Mexico; quite low in Spain. Of course, substantial variability exists between students within schools in all countries (sigma-squared in Table 4). The ICCs in all five countries are adequate to model in 2-level HLM models and, they provide some indication of the possible success of our school effects investigations in each country.
Step 2: Within-school HLM models. Research Question 1, which focuses on how social and academic background differences among students in each country that are associated with reading achievement, is addressed in within-school HLM models in each country, shown in Table 5. Three dummy-coded variables serve as proxy controls for ability: whether a student was retained once or more than once, both compared to not having been retained, and whether or not the student reported having taken remedial courses in the last three years. We also control for gender (females compared to males). These variables, estimated as fixed effects, are centered around the population mean in each country. The effects of SES, which we have converted to a z-score (M=0, SD=1) in each country, is estimated as a random effect and centered around the mean in each school. Our within-school models in each country thus have two measures which become outcomes at Level-2: reading achievement adjusted for retention, remediation, gender, and SES (our measure of quality) and the adjusted SES/achievement slope (our measure of equality).

The intercept term represents the average school achievement for students in each school. These intercepts, which vary significantly between schools in all countries, show the same general ranking of countries as was evident in Table 1: Brazil and Mexico very low, Portugal somewhat higher, and Spain and the U.S. a full SD above Brazil. Although SES is strongly and positively related to achievement within the schools in each country, within-school SES effects are considerably larger in Spain, the U.S., and Portugal, than in the two Latin American countries. This finding is consistent with the descriptive information in Table 3 that average SES is considerably more variable between schools in Brazil and Mexico. Much of the social differentiation in educational outcomes in Brazil and Mexico is between schools, whereas in Spain and the U.S. there is more social differentiation within schools. This provides additional support for the Heyneman (1976) hypothesis, even across this modest number of countries.

The strong negative effects of the retention indicators reflect the substantial descriptive differences between students by grade level shown in Table 2. Students who were ever retained (or who started school later) score almost 1 SD below their counterparts in appropriate grade levels. Remedial class experience is also negatively related to achievement -- more so in the non-Latin American countries. Substantial gender differences in reading achievement favoring female students typify all countries.
The bottom panel of Table 5 provides useful information about the between-school variance in our outcomes (1) the intercept, which is adjusted school average achievement, SES, and ability; and (2) the school average SES/achievement slope. In all five countries, the adjusted intercept varies considerably between schools. Variance in the SES/achievement slope also varies across schools in all countries except Spain. Note, however, that the variance in this outcome is consistently smaller than the intercept. This suggests that identifying significant school effects on SES/achievement slopes will be more difficult than on the intercept. The magnitude of the variance in these within-school SES/slopes is about twice as large in the U.S. as in Brazil, Mexico, or Portugal; the SES/achievement slope is much smaller in Spain. Residual within-school variances, in the last line of Table 5, have decreased as compared to the sigma-squared presented in Table 4. For example, the within-school model has explained about 16 percent of the within-school variance in reading achievement in Brazil.  

Summary of within-school models. Results of these within-school HLM models indicate a clear and consistent response to Research Question 1. Secondary school students' social backgrounds (SES, gender) and their academic backgrounds (retention and remediation) are strongly linked to their achievement in reading. Students from upper-SES backgrounds and with no repetition and/or remediation history achieve at higher levels in reading, as do female students. SES effects within schools are somewhat lower in Mexico and Brazil than in the more developed countries, provide evidence for Research Question 4. Remediation and repetition effects are, however, stronger in these countries. This suggests much of the social and academic differentiation in these countries is between rather than within schools.

Step 3: Between-school HLM models. A major aim of this study is to identify school effects that are associated with both quality (school average achievement, Research Question 2) and equality (the SES/achievement slope, Research Question 3). The results of our Level-2 HLM models are displayed in Table 6. Here we present the "refined" models, including only statistically significant gamma coefficients representing school effects in the Level-2 HLM models within each country. These models are all estimated simultaneously with the Level-1 models shown in Table 5; results of the Level-1 models have changed only very little. School analyses represents an important feature of schools' social context.  

The fully saturated models, where all gamma coefficients for each school measure were retained (i.e., identical models in each country), are shown in Appendix 2. These Level-2 HLM analyses model both outcomes simultaneously: adjusted school achievement (the quality indicator) and the SES/achievement slope (the equality indicator). We remind readers of our ideal pattern of "good" school characteristics described earlier: they are simultaneously
related positively to average achievement and negatively to the SES/achievement slope. We structure most of our discussion here separately for each country (i.e., looking vertically at the results in Table 6). Later, we discuss school effects across countries (a horizontal view of Table 6).

**School effects in Brazil.** Given the large descriptive differences in average SES and social mix across countries from Table 3, it is not surprising that average SES is strongly related to average reading achievement in Brazil. Taking average SES into account in fact explained away private/public school achievement differences in all the countries included in this analysis, an important finding to which we return in the discussion. The availability of pedagogical resources is associated positively with average achievement, but school infrastructure not. Student achievement is higher in Brazilian schools with strong academic press and schools where collective responsibility is more prevalent (i.e., when students feel their teachers really care about them and their learning).

Only one feature of Brazilian secondary schools is related to both the quality and equality parameters. In Brazilian schools where student absenteeism is low, students both achieve at higher levels and achievement is more equitably distributed by student SES. We consider this finding quite important, for three reasons. First, it reflects the ideal pattern for good schools. Second, control over absenteeism is usually within the policy domain of each school. Third, our findings relate directly to major social programs in Brazil—*Bolsa-Escola* and *Bolsa Familia*.

**School effects in Mexico.** Beyond average school SES, which is strongly and positively associated with average achievement, school effects in Mexican secondary schools capture features of their social organizations. Achievement is higher in Mexican schools where students report a strong sense of belonging (this construct is more important in Mexico than in the other countries). Belonging is also associated with the SES/achievement slope, but not in a productive way. In Mexican secondary schools where more students feel they belong, achievement is more unequally distributed by SES.

**School effects in Portugal.** Although Portuguese students achieve at higher levels in schools of higher average SES, this relationship is somewhat weaker in Portugal than in the other countries. Resources are important in Portuguese schools; achievement is higher in schools with better infrastructures and more pedagogical resources. Although academic press is not associated with higher achievement in Portuguese schools, achievement in that country is more equitably distributed in schools where all students are pressed to work hard.
School effects in Spain. Recall from Table 4 that the proportion of variance in achievement between Spanish schools is low (19 percent), well below the other countries in this study. Our ability to identify school effects is therefore constrained in this country. Only two school effects, associated with teachers’ positive behaviors towards their students and with their profession, are linked to higher achievement in Spanish schools. There are no school effects associated with equality, however.

School effects in the U.S. Among the five countries considered in this study, the relationship between school SES and average achievement is strongest among the U.S. secondary schools in PISA. Beyond this, our analyses have identified few other school effects. Students’ sense of belonging is negatively related to average achievement; meaning that schools where students feel more comfortable have weaker academic outcomes. This measure is positively associated with achievement in Mexico and Portugal. Achievement in U.S. schools is positively associated with more pedagogical resources (also the case is to Brazil and Portugal). None of the school effects considered here are linked with the equitable distribution of achievement in either U.S. or Spanish secondary schools.

Summary of school-effects models. There are few similarities in the pattern of school effects in our cross-national study of Brazil and four comparison countries. In a study such as PISA, which was carefully designed to bring comparability of measures across national boundaries (including the dependent variable), we found, consistent results about composition. However, less consistent were the cross-national patterns about the structure and organization of secondary schools that influence student achievement and its equitable distribution. Achievement is strongly influenced by the types of students who attend the schools, as measured by the socioeconomic status of their families, and by the aggregate social composition of their schools in all five countries. Mapping students to schools is strongly determined by socioeconomic status in all countries. The consistency of the finding across countries with differing levels of development contradicts the hypothesis posed by Heyneman (1976; Heyneman and Loxley 1983), although a five-country comparison offers only a modest test of the hypothesis.

In Brazil we identified a single school characteristic – principals’ statements of problems posed by student absenteeism -- that is linked simultaneously to quality and equality. Not only is it important to keep students in school and from dropping out, but in Brazil it is particularly important that students attend school and class regularly. Infrastructure resources and/or pedagogical resources were shown to be important for school quality in Portugal, Brazil, and the U.S. The non-finding of effects related to resources in Spain is an expected result for countries with low intra-class correlations. Not finding a resource-related effect for Mexico was unexpected, given the high intraclass correlation in Mexico. School effects that capture social organization, particularly those
characterizing how teachers and students relate to one another, were shown to influence achievement in several countries. Moreover, students' descriptions of how comfortable they feel in school is positively linked to achievement in two countries (and negatively in one). Students' description of the academic press of their schools is somewhat important, although not consistently across countries. In Brazil, students' achievement at a higher level in schools that press academic pursuits more; in Portugal student achievement is more equitably distributed in schools where teachers press students to achieve. These findings are logical and important.

Results in the lower panel of Table 6 provide evidence that a large proportion of between-school variance in the intercept (average achievement) has been explained. In the Brazilian and Mexican samples, our school effects models have explained about 80 percent of the between-school variance in average achievement. In Spain, although our models were less successful, 66 percent of the between-school variance has been explained. However, these is also evidence in the bottom panel of Table 6 that significant residual variance remains to be explained in the parameters measuring both quality and equality. This is indicated by the statistically significant chi-squared statistics for both intercepts and slopes across countries.

Differential effects by national development. Research Question 4 explored whether our findings provide support for the hypotheses posed by Heyneman (1976, 1983) about differential educational effects due to countries' levels of economic and social development. The developmental levels of the countries in this study vary, with Brazil and Mexico lower, Portugal in the middle, and Spain and the U.S. higher (an indicator is average SES from Table 1). At least one explanation offered to explain the Heyneman hypotheses was statistical, in that variability in students' social background was less in developing than developed countries, whereas resource variability was larger in developed countries. However, in our separate-by-country analyses, we standardized all measures within each country, resulting in equivalent variability on all continuous measures across countries. The hypothesis of differential effects of social background on educational outcomes was not confirmed. Both within schools (Table 5) and between schools (Table 6), SES was strongly related to reading achievement in all countries.

Average SES measures school social composition here, although others have interpreted it a proxy measure of the resources schools may draw upon (e.g., Lee and Bryk 1989). Our study included two measures of school resources that Fuller (1987) called "material school inputs:" physical infrastructure and pedagogical resources. Both are drawn from principal reports, and neither has particularly strong effects. Although our only pedagogical resource measure mentions science labs, and our outcome is reading achievement, we suggest that science laboratories in secondary schools
represent a stronger resource base for teaching (unfortunately this was the only measure capturing this construct in PISA). With these caveats, we found no pattern for school resource effects on achievement between developed and developing countries. Seen broadly, our findings provide no empirical support for the Heyneman hypothesis. This could flow from our decision to analyze the data separately by country and to employ constant variances in these measures (but not in the outcome), or from the modest number of countries considered here. The PISA data do not represent the full range of countries based on economic development, so such a test of the hypothesis may not be possible with PISA.

**Discussion**

**Quality and Equality, Reconsidered**

**Revisiting the study's purpose.** One of our aims has been to identify features associated with "good" secondary schools in Brazil and several comparison countries. We defined "good" schools as those that simultaneously exhibit high average achievement and a distribution of that achievement that is socially equitable across students of different social backgrounds. In this case, we captured social background as socioeconomic status, or SES. The study used data from PISA, a valuable source of international information about secondary education and achievement in reading literacy for 15-year-old students. Our focus is on Brazilian secondary schools, partly because these schools operate within a society that is simultaneously strongly committed to improving its education system, and that has one of the most inequitable distributions of wealth in the world. Our study of Brazilian schools locates them within an international comparative framework that was carefully chosen on several important criteria.

**Findings seen in the context of quality and equality.** One finding from this study is somewhat unsettling. We identified several school effects that are important for improving either quality or inequality in several countries, but almost no school effects that satisfied both criteria for "good schools." We identified several important features of the academic and social organizations in Brazil's and comparison countries' schools that are linked to educational quality. In only one instance did we identify a feature of secondary schools linked to both quality and equality: low absenteeism in Brazilian schools. School resources (both infrastructure and pedagogical) were associated with achievement in Brazil, Portugal, and the U.S. School effects associated with schools' social organizations were also evident. These include students' sense of belonging (in Mexico, Portugal, and the U.S.), teachers' collective responsibility for student learning (in Brazil and Spain), and teachers' commitment to their profession (in Spain). Schools effects on equality were more illusive.
Achievement is distributed more equitably in Brazilian secondary schools where attendance is higher, in Portuguese schools where academic press is stronger, and in Mexican schools where students' sense of belonging is weaker but a stronger sense of belonging is associated with higher average achievement in that country.

We admit to a tension in constructing this study and interpreting its results. On the one hand, our team's interest in and knowledge about education in Brazil is strong, motivating our focus on that country. As Brazil has been included in few international educational studies (none from IEA), we were excited to explore thoroughly the Brazilian PISA data. On the other hand, we wanted to capitalize on a major strength of PISA -- its multi-country focus. Carefully constructed measures and scrupulously supervised data collection across countries argues strongly for cross-national comparisons. Our compromise was to focus on Brazil and compare our results in parallel models with relatively small and carefully chosen set of other countries. Thus, our study is both comparative (we use the same analytic models in several countries included in PISA) and focused (we shine our brightest "analytic light" on Brazil). Perhaps a more complete comparative context for Brazil would be the full set of Latin American countries. However, in the 2000 PISA data, Brazil and Mexico were the only such countries included. Even in later PISA data collections, the full set of Latin American countries was not included (although data from Argentina, Chile, Peru, and Uruguay have been collected in subsequent PISA cohorts).

Policy Implications

Reducing absenteeism is important. Throughout the world, there is a general push to provide schools with more local autonomy to make decisions, to move important decisions about education away from national and state education agencies to those that are made at local levels (even or the school level). Many of our findings have policy relevance for schools (rather than for education agencies more removed from the schools). For example, we suggest that schools can actually control absenteeism. This control could be linked to a new generation of active welfare policy, such as established in Bolsa-Escola and the new Bolsa-Família. Our findings provide evidence to support this kind of “active welfare,” in which recipients of the welfare policy must make an active contribution, such as enrolling their children in schools and assuring their attendance. However, since the beginning of 2004 the Brazilian government has suspended monitoring the families’ responsibility for their children attendance to school and officials have mentioned that the Bolsa-Escola program will be assessed on a new basis in the future (Folha de São Paulo, July, 6, 2004). Relaxing control over absenteeism in the context of welfare programs in Brazil may be unwise given our findings.
In some U.S. settings where absenteeism historically has been very high, local school officials (including teachers) have given much attention to this problem. If teachers are responsible for relatively small groups of students in an advisory role, they can discover early in the school day when their advisees do not appear at school. Teachers in some schools have been freed up to make personal phone calls to the homes of absent students (or to the workplaces of the absent students' parents). Sometimes, teachers even offer to bring absent students to school. Schools can also make parents aware of how important school attendance is, to discourage them from holding older children out of school to care for younger siblings.

Problems of absenteeism are usually seen as evidence of a breakdown of discipline and order, so that they are often addressed by individuals for individuals. Here we have considered absenteeism as organizational problem, and we suggest that it evoke an organizational response. Our results suggest that this is especially important in Brazil. Rather than simply exhorting students to come to school and class, we suggest that schools make substantial efforts to make this happen. Active social policy, with family incentives to keep children in school, seems reasonable.

Policies that involve teachers. Academic press is another school effect that can be implemented locally. As measured in PISA, this measure taps students' perceptions about whether their teachers press them to work hard, encourage them to expend effort on their work, require high-quality work, and inform students that the educational enterprise is a vast and important one in which students and teachers are engaged together. Rather than as an individual student's perception, we see academic press as a feature of schools' academic organizations, at the intersection of teachers and their students. There is empirical evidence that in order for academic press to be maximally important, it must be coupled with academic support (Lee and Smith 1999). Other school effects identified in this study also center on teachers and their relationships with their students -- responsibility and commitment. Our findings suggest that in some countries students flourish academically when their teachers care about them, press them to work hard, and assume responsibility for their learning. Student achievement is also higher when teachers are committed to their profession.

Schools can work to be better places for their students. But how do schools, through their principals, encourage teachers to care about their students in this way, to push them hard, and to assume responsibility for their learning? Teaching is not a high-paid profession in any of the countries included in this study. Although students' sense of belonging, as measured by PISA, does not reflect directly on teachers, it does suggest that students who "belong" are those who feel comfortable in the school, are accepted by their peers, and do not feel out of place. Because so much time students spend in school is spent in classrooms, teachers surely have control over this. When they see evidence of
cruelty among students, or of some students being excluded from daily activities, teachers must step in and take action. They must make sure that the school environment is an accepting and comfortable place for students to be.

Principals can create communities of learning, where teachers are urged to do their best, where students are assumed to treat one another and their teachers with respect, and where learning is the goal for everyone. There is a rich strain of research that focuses on the development of professional communities in schools (e.g., Louis, Marks, and Kruse 1996). Much can be accomplished by individuals in schools to create communities of learning for everyone.

Important Non-Findings

Statistical power. It is always dangerous to claim non-findings as findings in any study. The major danger in such claims relates to Type II error rates and statistical power. That is, if studies do not have sufficient statistical power to find effects, then they may not claim a non-finding as important. However, in this study, there is little problem of statistical power. In each country there were large numbers of schools (from 149 schools in Portugal to 324 schools in Brazil, see Table 3). Moreover, there were substantial student sample sizes in each country (from 6,214 in Spain to 3,846 in the U.S., see Table 1). Important to estimating HLM parameters in two-level models are within-school sample sizes. These also were reasonable compared to many other school effects studies familiar (from 15.1 students/school in Brazil to 33.6 students/school in Spain -- Table 4). We conclude that statistical power in this study was sufficient to allow us to draw substantive conclusions about non-findings.

What wasn't important? Perhaps the most important non-finding is that public/private school comparisons dropped from statistical significance in four of the five countries in models that included statistical controls for average school SES and academic press (all but the U.S.). We have considerable evidence of the importance of private school attendance in many of these countries, particularly at the secondary level. However, once other important characteristics of schools that we know to be associated with private school organization (e.g., composition, resources, relationships, academic press) private/public school comparisons dropped to non-statistical significance. This suggests that school composition, resources, the way they organize academics and relationships is perhaps more important than whether they are private or public. However, in other studies based on Brazilian datasets, the private sector indicator was positively and significantly related to achievement, even after controlling for school composition and several other school resources and organizational features (e.g., Albernaz, Ferreira and Franco 2003; Franco, Albernaz and Ortigão 2002). To possibly explain these divergent results, we remind readers that the standard deviation of reading achievement in Brazil is
smaller than in Portugal and the U.S. (Table 1), despite an intraclass correlation (ICC) that is larger in Brazil than in these countries. The heavy emphasis in Brazilian schools on grammar may not be consistent with PISA’s emphasis on diversity of reading styles and reading materials, precluding private school students showing their advantage over students in public schools (Franco 2002).

A glance at the end of Appendix 1, where we list the school effects we considered but subsequently excluded, suggests other important non-findings. Not important in this study were measures of school discipline, teacher qualifications, computer availability and use, various measures of school climate defined by principals that involve teachers or students, and even the quality of resources. The list is quite long, and the constructs that they measure have been shown to be important in other school effects studies. Why are these measures not important here? It is difficult to separate substantive reasons from the possibility of measurement problems associated with validity. This is particularly important when we consider that most school effects research uses school factors measured by survey responses from students or teachers, whereas PISA relies mostly on questionnaire responses from principals to capture school characteristics.

How Do Schools Accommodate Low-Performing Students?

The discussion has so far centered on school effects. However, the results in Table 5 indicate how characteristics of students influence achievement within schools. Because of the PISA cross-sectional structure, we made use of students' educational histories to capture (in part) their academic background. Every school in virtually every country faces a dilemma: how to simultaneously serve well their most motivated and able students and also provide a solid education to students with low commitment to school, low performance, and/or weak academic backgrounds. In every country in this study one response to this challenge is retention (or grade repetition); another is remedial classes. In Brazil, almost half of the students in secondary school have a history of repetition (one or two years). Though much lower, these figures are still noteworthy in Mexico and Portugal. In Brazil, Mexico, and Portugal, half the students have taken remedial classes (compared to a quarter in Spain and the U.S.).

These practices are responses to low student performance; our results suggest that they are not solutions. Students with a history of remediation and/or repetition (particularly multiple repetition) achieve at much lower levels than their counterparts in each country without such histories. Some national educational systems (e.g., the U.S.) respond to this challenge by substantially differentiating the curriculum. Other systems (e.g., Germany) respond with very different types of secondary schools for students of different demonstrated abilities and aspirations. Clearly, how schools serve students of different abilities, academic experiences, and aspirations for the future is a major
challenge. The evidence here is that the responses of repetition and remediation are not necessarily successful. Appropriate responses can surely occur both between and within schools. However, the characterization of responses represents the essence of how schools (and national education systems) are challenged to address quality and inequality simultaneously. In one sense, this is the focus of our study. In another, it is a broader issue than we can address here.

Are Cross-National Studies Useful?

Structuring comparative studies. One of us is a relative newcomer to cross-national research, having concentrated many research efforts on studying school effects in U.S. schools. Two of our team have devoted much research effort to studying Brazilian education. As stated, we came together because we share two passions: a fundamental interest in educational quality and equality as we have defined it here, and an interest in the appropriate methodology for conducting such studies. The definition of "good schools" seems valid cross-nationally. Thus, we felt that the conduct of a cross-national study made sense in the context of (1) a fundamental focus on one country, and (2) the choice of reasonable comparison countries.

Our conclusions here are mixed. Access to a current cross-national study like PISA, extremely well conceived and carefully executed, represents an excellent resource for educational researchers interested in cross-national study. Moreover, the design of PISA is well suited to the type of analyses with which we are both familiar and skilled: HLM. However, many of the original analyses using PISA data and presented in the "Knowledge and Skills for Life" report (OECD 2001a) used HLM within a framework that we regard with some suspicion. That is, many school effects analyses in that report were constructed as three-level HLMs, with students nested in schools, which were nested in countries. Analyses that use the country as a separate unit of analysis for cross-national studies seem misguided in most instances.\(^{14}\)

The educational systems of the Scandinavian countries, for example, or of such reasonably highly developed Asian countries as Japan or Korea, seem to share very little with school systems in Latin America. Fortunately, subsequent PISA data collections include other Latin American countries which represent reasonable comparisons for Brazil. Thus, cross-national educational research is reasonable, as long as the countries to be compared are somewhat similar to one another, in terms of either language, educational system design, or histories of economic and social development.
The Value of a Different Design

Despite the strong value of PISA data, PISA is far from perfect for conducting the kind of school effects analysis that focuses on quality and equality -- our purpose in this study. PISA is very strong in terms of how it measures achievement, and in sampling and data collection designs. PISA has adhered to the highest standards of international studies of education. However, we "curb our enthusiasm" for PISA in two important ways.

First, we distinguish between achievement, which measures academic status at a particular time point, and learning, which captures growth in achievement over time. If researchers want to identify the characteristics of schools that are associated with student learning, then they need measures of achievement over at least two time points -- when students enter a school and when they depart. Schools can only influence that portion of achievement (or achievement gain) that accrues over the time students are educated a particular school. This important distinction suggests that to estimate school effects accurately, data collection designs need to be longitudinal. Researcher would need data on students (achievement, in particular) when they enter and when they leave the school. Although substantial resources would be needed to follow the same students over time, there would be enormous value in doing so (and some conceptual costs of not doing this). The distinction between learning and achievement is an important one, in educational research.

Second, PISA lacks important educational information --from teachers. It would be difficult to design a study that included teachers. Would you only sample teachers that were attached to students as they entered the school? In secondary schools, this could include several teachers. Moreover, in most secondary schools students change teachers for each subject, and in most cases each year. This would complicate data collection from teachers. However, the major activity of all schooling occurs in classrooms, namely instruction. It is difficult to characterize schools that are of high quality and also equitable without information about instruction. We used information about teachers and teaching drawn from both principals and students in PISA. Moreover, several of the school effects identified in this study actually focus on teachers. Teachers are at the heart of the schooling enterprise, and educational datasets without that source of information source are inherently incomplete.

Thus we recommend that PISA enrich the pattern of cross-national studies and consider a different design in two ways. First, the same students should be followed over their lives in secondary school (i.e., a longitudinal design with at least two time points, maybe more). Second, large surveys of education need to include data from many (or all) teachers in each school. We appreciate PISA; much about the study is valuable. However, the study of school effects we have presented here is
compromised by PISA's cross-sectional design and the absence of data from teachers as an independent data source.
Technical Notes

1. IEA-sponsored studies include the First and Second International Math Studies (FIMS, SIMS), the First and Second International Science Study (FISS, SISS), and the Third International Math and Science Study (TIMSS). Until recently, with the development of PISA, IEA studies have dominated cross-national research in education. Both IEA-sponsored studies and PISA follow cross-sectional designs, with one exception: SIMS.

2. Quite recently, under the sponsorship of UNESCO, the International Institute for Educational Planning (IIEP) has collected a second set of educational data in 15 sub-Saharan African countries that have formed the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ). PISA was a model for that data collection effort, although the focus in SACMEQ is on a particular grade level (6th), regardless of students' age.

3. We recognize that our analyses do not represent a solid test of the Heyneman and Loxley (1903) hypothesis. Although we do make comparisons between more and less developed countries, and the range of comparisons is modest. Rather, we use the conceptualization to help frame our analyses and interpretations.

4. As grade retention is common in Brazil, original sampling indicated that there were non-trivial numbers of 15-year-old students in grades below lower-secondary school (7th grade). As PISA study’s purpose is to study secondary schooling across the world, a decision to exclude schools below lower secondary in Brazil was made. Thus, generalizability in Brazil is only to 15-year old students in lower- or upper-secondary schools.

5. The HLM methodology is now quite familiar among researchers concerned with school effects research, both within the U.S. and internationally. As we assume the most readers are familiar with HLM, we do not include equations to represent each of the three HLM levels. Readers who wish more information about the methodology should consult Raudenbush and Bryk (2002). For a pedagogical treatment of a school effects study in more detail, see Lee and Bryk (1989).

6. This is not surprising, as many of the school-level measures are relatively highly correlated. Of course, correlation patterns also vary across countries. Our model-building and model-refining process was simultaneously systematic, time-consuming, and theory-driven.

7. These figures for retention in Brazil are actually underestimates. Recall that all Brazilian 15-year-olds who were not at least in the seventh grade were not included in the study, to conform to the PISA directive that data collection was in secondary schools. Some Brazilian 15-year-olds are still in elementary school.

8. Effect sizes were computed by simple differences divided by the SD of achievement across OECD countries (100). Effect sizes over .5 SD are considered large; .3-.5 SD are medium; .1-.3 SD are small; and those below .1 SD are trivial (Rosenthal and Rosnow 1984).

9. Although gender is not a focus of this study, we know that there are differential dropout rates by gender in Mexico; more boys leave school early. Thus, the boys who are left in school would be a relatively more select group than the girls. This would explain these seemingly
unusual results. The large gender differences favoring females in PISA suggest that this may be a fruitful research direction with these data.

10. We computed the percent of within-school variance explained as follows, using relevant numbers from Tables 4 and 5 for Brazil: 
   \[ \frac{4294 - 3615.2}{4292} = 0.158 \]
   A comparable figure for Portugal is: 
   \[ \frac{5840 - 4043.08}{5840} = 0.308 \]
   Our within-school models have explained only modest proportions of the within-school variance in reading achievement. If PISA had a longitudinal design, and thus included a prior measure of reading achievement, these proportions would be much larger.

11. Although the proportion of secondary schools that are private rather than public varies considerably across these five countries (from 54 percent in Spain to 17 percent in the U.S., and 23 percent in Brazil --see Table 3), we found that whether the school was private or public dropped to statistical non-significance in all five countries once we took average school SES into account. Accordingly, we deleted the private school indicator from our analyses.

12. We computed the percent of between school variance explained as follows, using relevant numbers from Tables 4 and 6 for Brazil: 
   \[ \frac{3101 - 631.7}{3101} = 0.796 \]
   A comparable calculation for Spain is: 
   \[ \frac{1296 - 439.9}{1296} = 0.661 \]

13. Unfortunately, PISA collected no data about students' experiences with the high-school curriculum. Many studies of school effects have reported important findings about the structure of the curriculum. For example, secondary schools in the U.S. were shown to be more effective and more equitable if they offered a rather narrow curriculum composed of mostly academic courses (Lee and Bryk 1989; Lee and Burkam 2003; Lee, Croninger, and Smith 1997).

14. An exception is a recent study that employed almost all PISA countries in a study that examined how national educational policies about vocational education, social mix, and repetition influenced both achievement and its equitable distribution by SES and average school SES (Franco, Lee, and Satyro 2005). In this 3-level HLM study, students were nested in schools, which were nested in countries. Although the country-level data were not available from PISA, they were readily available from World Bank and UNESCO sources.
References


Table 1: Academic and Social Background Characteristics of 15-Year-Old Students in Brazil and Four Comparison Countries

<table>
<thead>
<tr>
<th></th>
<th>BRAZIL</th>
<th>MEXICO</th>
<th>PORTUGAL</th>
<th>SPAIN</th>
<th>USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Sample Size&lt;sup&gt;a&lt;/sup&gt;</td>
<td>4893</td>
<td>4600</td>
<td>4585</td>
<td>6214</td>
<td>3846</td>
</tr>
<tr>
<td>Reading Achievement&lt;sup&gt;b&lt;/sup&gt;</td>
<td>402.9</td>
<td>423.0</td>
<td>472.4</td>
<td>493.4</td>
<td>503.5</td>
</tr>
<tr>
<td>Female</td>
<td>0.54</td>
<td>0.50</td>
<td>0.52</td>
<td>0.51</td>
<td>0.52</td>
</tr>
<tr>
<td>Remedial Classes&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.42</td>
<td>0.43</td>
<td>0.49</td>
<td>0.21</td>
<td>0.27</td>
</tr>
<tr>
<td>Retained Only One Year</td>
<td>0.26</td>
<td>0.11</td>
<td>0.13</td>
<td>0.02</td>
<td>0.04</td>
</tr>
<tr>
<td>Retained Two Years</td>
<td>0.16</td>
<td>0.03</td>
<td>0.07</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SES&lt;sup&gt;d&lt;/sup&gt;</td>
<td>-1.23</td>
<td>-1.24</td>
<td>-0.41</td>
<td>-0.24</td>
<td>0.16</td>
</tr>
<tr>
<td>SD in Reading Achievement</td>
<td>89.06</td>
<td>88.96</td>
<td>97.15</td>
<td>85.53</td>
<td>103.55</td>
</tr>
<tr>
<td>School Enrollment in Secondary Education&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.71</td>
<td>0.60</td>
<td>0.85</td>
<td>0.94</td>
<td>0.88</td>
</tr>
</tbody>
</table>

<sup>a</sup> Sample sizes are unweighted, whereas means are computed using the normalized student-level weight supplied by PISA. This convention applies to all tables in this study.

<sup>b</sup> Reading achievement is a standardized score across the 28 OECD countries in PISA to M=500, SD=100.

<sup>c</sup> Based on students’ responses to whether they attended remedial classes during the last three years.

<sup>d</sup> SES is a z-score (M=0, SD=1) across the 28 OECD countries and 4 non-OECD in the PISA-2000 assessment. It is a factor-weighted index of students’ families’ economic, social and cultural background.

<sup>e</sup> Data source: World Development Indicators 2003, World Bank. In secondary education, students who are enrolled in both lower secondary and upper secondary schools are counted. In PISA, only students aged between 15 and three months and 16 and two months, who were enrolled in the 7<sup>th</sup> grade or, above were included. Specifically, for Brazil and Mexico the percentage of 15 year-old students that participated in PISA represented only around 58% of the age group of students in these countries.
Table 2: Students’ Achievement, Retention, and Social Background in Brazil and Four Comparison Countries

<table>
<thead>
<tr>
<th></th>
<th>BRAZIL Not Retained</th>
<th>BRAZIL Retained</th>
<th>MEXICO Not Retained</th>
<th>MEXICO Retained</th>
<th>PORTUGAL Not Retained</th>
<th>PORTUGAL Retained</th>
<th>SPAIN Not Retained</th>
<th>SPAIN Retained</th>
<th>USA Not Retained</th>
<th>USA Retained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Proportion</td>
<td>0.58</td>
<td>0.42</td>
<td>0.85</td>
<td>0.15</td>
<td>0.80</td>
<td>0.20</td>
<td>0.98</td>
<td>0.02</td>
<td>0.96</td>
<td>0.04</td>
</tr>
<tr>
<td>Reading Achievement</td>
<td>436.20</td>
<td>359.87</td>
<td>439.93</td>
<td>342.53</td>
<td>500.47</td>
<td>363.68</td>
<td>496.73</td>
<td>373.09</td>
<td>508.95</td>
<td>357.12</td>
</tr>
<tr>
<td>SES</td>
<td>-0.98</td>
<td>-1.57</td>
<td>-1.06</td>
<td>-2.15</td>
<td>-0.25</td>
<td>-1.03</td>
<td>-0.23</td>
<td>-0.87</td>
<td>0.21</td>
<td>-0.97</td>
</tr>
<tr>
<td>Female</td>
<td>0.57</td>
<td>0.49</td>
<td>0.52</td>
<td>0.44</td>
<td>0.54</td>
<td>0.42</td>
<td>0.51</td>
<td>0.43</td>
<td>0.52</td>
<td>0.30</td>
</tr>
<tr>
<td>Remedial</td>
<td>0.37</td>
<td>0.48</td>
<td>0.42</td>
<td>0.49</td>
<td>0.43</td>
<td>0.72</td>
<td>0.21</td>
<td>0.35</td>
<td>0.27</td>
<td>0.30</td>
</tr>
<tr>
<td>Single Parent</td>
<td>0.17</td>
<td>0.19</td>
<td>0.17</td>
<td>0.21</td>
<td>0.11</td>
<td>0.12</td>
<td>0.17</td>
<td>0.31</td>
<td>0.21</td>
<td>0.29</td>
</tr>
</tbody>
</table>
Table 3: Characteristics of Schools Attended by 15-Year-Old Students in Brazil and Four Comparison Countries

<table>
<thead>
<tr>
<th></th>
<th>BRAZIL</th>
<th>MEXICO</th>
<th>PORTUGAL</th>
<th>SPAIN</th>
<th>USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>School Sample Size&lt;sup&gt;a&lt;/sup&gt;</td>
<td>324</td>
<td>183</td>
<td>149</td>
<td>185</td>
<td>153</td>
</tr>
<tr>
<td>Average SES</td>
<td>-1.38</td>
<td>-1.56</td>
<td>-0.62</td>
<td>-0.23</td>
<td>-0.10</td>
</tr>
<tr>
<td>Social Mix</td>
<td>-2.28</td>
<td>-2.51</td>
<td>0.15</td>
<td>-0.37</td>
<td>0.10</td>
</tr>
<tr>
<td>Private Schools&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.23</td>
<td>0.18</td>
<td>0.10</td>
<td>0.52</td>
<td>0.17</td>
</tr>
<tr>
<td>Collective Responsibility&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.55</td>
<td>0.12</td>
<td>0.49</td>
<td>0.21</td>
<td>0.41</td>
</tr>
<tr>
<td>Academic Press&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.60</td>
<td>-0.14</td>
<td>0.15</td>
<td>0.06</td>
<td>0.50</td>
</tr>
<tr>
<td>Students’ Sense of Belonging&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.22</td>
<td>0.03</td>
<td>-0.07</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td>Teacher Commitment&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-0.31</td>
<td>0.47</td>
<td>-0.51</td>
<td>-0.24</td>
<td>-0.18</td>
</tr>
<tr>
<td>School’s Physical Infrastructure&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-0.18</td>
<td>0.58</td>
<td>-0.34</td>
<td>-0.20</td>
<td>-0.14</td>
</tr>
<tr>
<td>Student Absenteeism&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.18</td>
<td>0.14</td>
<td>0.71</td>
<td>0.01</td>
<td>-0.25</td>
</tr>
<tr>
<td>Pedagogical Resources</td>
<td>-0.92</td>
<td>-0.12</td>
<td>-0.09</td>
<td>0.10</td>
<td>-0.03</td>
</tr>
</tbody>
</table>

<sup>a</sup> Sample sizes are unweighted, whereas means are produced using the normalized student-level weight supplied by PISA. This convention applies to all tables in this study. In the case of USA, total school sample size includes 32 principals that didn’t respond to the school questionnaire.

<sup>b</sup> Private schools include Catholic, non-Catholic religious, non-religions, and for-profit schools.

<sup>c</sup> The scores of these variables are standardized over the OECD sample to Mean=0, SD=1.
Table 4: Fully Unconditional HLM Models of Reading Achievement in Brazil and Four Comparison Countries

<table>
<thead>
<tr>
<th></th>
<th>BRAZIL</th>
<th>MEXICO</th>
<th>PORTUGAL</th>
<th>SPAIN</th>
<th>USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Sample Size/</td>
<td>15.1</td>
<td>25.1</td>
<td>30.8</td>
<td>33.6</td>
<td>25.1</td>
</tr>
<tr>
<td>School Sample Size (n)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within-School Variance</td>
<td>4030.93</td>
<td>3226.95</td>
<td>5466.70</td>
<td>5090.80</td>
<td>4273.98</td>
</tr>
<tr>
<td>(σ²)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between-School Variance</td>
<td>3097.19</td>
<td>3611.73</td>
<td>3134.49</td>
<td>1306.26</td>
<td>3014.11</td>
</tr>
<tr>
<td>(τ₀₀)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICCᵃ</td>
<td>0.43</td>
<td>0.52</td>
<td>0.35</td>
<td>0.20</td>
<td>0.41</td>
</tr>
<tr>
<td>Reliability (λ)</td>
<td>0.87</td>
<td>0.92</td>
<td>0.90</td>
<td>0.85</td>
<td>0.83</td>
</tr>
</tbody>
</table>

ᵃ ICC = τ₀₀ / (τ₀₀ + σ²)
### Table 5: HLM Within School Model of Reading Achievement in Brazil and Four Comparison Countries

<table>
<thead>
<tr>
<th>Fixed Effect $^{b,c}$</th>
<th>BRAZIL</th>
<th>MEXICO</th>
<th>PORTUGAL</th>
<th>SPAIN</th>
<th>USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept, $\gamma_{00}$</td>
<td>$389.3^{***}$</td>
<td>419.9***</td>
<td>474.3***</td>
<td>494.0***</td>
<td>493.6***</td>
</tr>
<tr>
<td>SES, $\gamma_{10}$</td>
<td>5.8***</td>
<td>6.0***</td>
<td>11.5***</td>
<td>19.2***</td>
<td>26.0***</td>
</tr>
<tr>
<td>Remedial class, $\gamma_{20}$</td>
<td>$-19.1^{***}$</td>
<td>$-16.4^{***}$</td>
<td>$-38.9^{***}$</td>
<td>$-37.0^{***}$</td>
<td>$-40.9^{***}$</td>
</tr>
<tr>
<td>One grade from HS, $\gamma_{30}$</td>
<td>$-48.1^{***}$</td>
<td>$-28.7^{***}$</td>
<td>$-63.8^{***}$</td>
<td>$-100.2^{***}$</td>
<td>$-90.4^{***}$</td>
</tr>
<tr>
<td>Two grades from HS, $\gamma_{40}^{d}$</td>
<td>$-80.8^{***}$</td>
<td>$-45.0^{***}$</td>
<td>$-111.4^{***}$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>Sex (1=female), $\gamma_{50}$</td>
<td>13.3***</td>
<td>7.8**</td>
<td>13.1***</td>
<td>21.3***</td>
<td>30.9***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effect $^{c}$</th>
<th>Variance Components $^{a}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept, $\mu_{0j}$</td>
<td>1845.93***</td>
</tr>
<tr>
<td>SES slope, $\mu_{1j}$</td>
<td>81.77***</td>
</tr>
<tr>
<td>Level-1 error, $\sigma_{ij}^2$</td>
<td>3615.20</td>
</tr>
</tbody>
</table>

---

a. ** p<.01 / *** p<.001

b. Between school variance is constrained to zero for all independent variables except SES. SES is centered around individual school means, whereas other variables are centered around the population mean for each country.

c. The statistical testing of fixed effects is based on T-test.

d. No students in Spain or USA reported to being more than one year behind their cohort.

e. The statistical testing of random effects is based on chi-square statistics.
Table 6: Final HLM Between School Model of Reading Achievement in Brazil and Four Comparison Countries

<table>
<thead>
<tr>
<th></th>
<th>BRAZIL</th>
<th>MEXICO</th>
<th>PORTUGAL</th>
<th>SPAIN</th>
<th>USA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average Reading Achievement</strong>&lt;sup&gt;b,c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{00}$</td>
<td>389.3 ***</td>
<td>419.5 ***</td>
<td>474.7 ***</td>
<td>494.1 ***</td>
<td>493.8 ***</td>
</tr>
<tr>
<td>Average SES, $\gamma_{01}$</td>
<td>35.9 ***</td>
<td>39.5 ***</td>
<td>21.5 ***</td>
<td>31.3 ***</td>
<td>50.3 ***</td>
</tr>
<tr>
<td>Collective Responsibility, $\gamma_{02}$</td>
<td>5.3 *</td>
<td>-</td>
<td>-</td>
<td>5.2 **</td>
<td>-</td>
</tr>
<tr>
<td>Academic Press, $\gamma_{03}$</td>
<td>5.1 *</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Students’ Sense of Belonging, $\gamma_{05}$</td>
<td>-</td>
<td>17.0 ***</td>
<td>10.2 ***</td>
<td>-</td>
<td>-11.4 *</td>
</tr>
<tr>
<td>Teacher Commitment, $\gamma_{06}$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3.8 ~</td>
<td>5.14 *</td>
</tr>
<tr>
<td>School’s Physical Infrastructure, $\gamma_{07}$</td>
<td>-</td>
<td>-</td>
<td>6.3 ***</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Student Absenteeism, $\gamma_{08}$</td>
<td>-4.7 *</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pedagogical Resources, $\gamma_{09}$</td>
<td>3.4 ~</td>
<td>-</td>
<td>8.9 ***</td>
<td>-</td>
<td>9.3 *</td>
</tr>
<tr>
<td>Private school, $\gamma_{011}$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>17.7 **</td>
</tr>
</tbody>
</table>

| **SES/ Achievement Slope**<sup>b,c</sup> |          |          |          |          |          |
| Intercept, $\gamma_{10}$ | 6.1 ***  | 4.6 **   | 11.2 *** | 19.5 *** | 27.0 *** |
| Academic Press, $\gamma_{11}$ | -        | -        | -2.5 ~   | -        | -        |
| Students’ Sense of Belonging, $\gamma_{12}$ | -        | 3.6 ~    | -        | -        | -        |
| Student Absenteeism, $\gamma_{13}$ | 2.9 *    | -        | -        | -        | -        |

| **Random Effect**<sup>d</sup> |          |          |          |          |          |
| Intercept, $\mu_{0j}$ | 631.7 *** | 710.0 *** | 320.4 *** | 439.3 *** | 481.0 *** |
| SES slope, $\mu_{1j}$ | 38.4 **  | 94.76 ** | 82.0 **  | 30.5 ~   | 220.6 *** |

<sup>a</sup> ~ p<.10;  * p<.05;  ** p<.01;  *** p<.001

<sup>b</sup> The dashes in each model represent variables dropped due to non-significance. The within school models are not reported here, as they are almost identical to the within-school models shown in table 5.

<sup>c</sup> The statistical testing of fixed effects is based on tests.

<sup>d</sup> The statistical testing of random effects is based on chi-square statistics.
Appendix 1: Description of the Variables Used in HLM Models

Here we provide considerable detail about the PISA variables that have been included in our HLM models. We believe this information is useful in two ways: (1) so that readers can understand exactly what constructs our measures are actually capturing; and (2) so that other researchers interested in using PISA data may want to replicate our analyses in other countries that participated in the study. To provide some logic for our model-building strategy, we also describe number of school-level variables we considered in our analyses but later eliminated them due to non-significant associations in the five countries we studied.

A. MEASURES INCLUDED IN OUR ANALYSES

Student-level Dependent Variable

Reading Achievement: A reading test that was administered as part of the PISA 2000 data collection to all sampled 15-year-old students in each country. The scores were estimated by Weighted Maximum Likelihood (PISA Technical Report 2001b). Proficiency in reading was standardized across the 28 OECD countries to a mean of 500 and a standard deviation (SD) of 100 [WLREAD]. We used original scoring of this variable for our separate-by-country analyses, without standardizing within each country.

Student-level Predictors

SES: Composite measure of socioeconomic status described in Annex A1 of the Pisa Technical Report (2001b). The measure was created from a cross-country principal components factor analysis including the 28 OECD countries and 4 non-OECD countries participating in PISA 2000. The SES composite includes the following PISA variables:

- ISEI. International socioeconomic index, derived from the classification of students' responses about father's and mother's occupation into the International Standard Classification of Occupations. If student provided responses for both father's and mother's occupation, the value of ISEI was the higher or the two;
- Parents' Education. Variables were recoded from their original codings into actual years of education. If student provided answers for both father's and mother's education, the greater value was taken [from FISCED and MISCED];
- PISA Wealth Index. A composite that measures family wealth [WEALTH];
- PISA Home Educational Resources Index. A composite [NEDRES]; and
- PISA Cultural Possessions Index. A composite variable that adds the numbers of possessions in the student's household that are related to "classical culture" [CULTPOSS];

SES was created as a standardized and z-scored composite (M = 0, SD = 1) across full PISA 2000 sample. For our separate-by-country analyses, we re-standardized to a z-score in each country.

Retained Once: Dummy-coded variable based on the grade the student was in [ST02Q01]. We coded this "1" if student was enrolled in the 8th grade (i.e., one year behind the beginning of high school), "0" otherwise.

Retained Twice: Dummy-coded variable also based on student's grade [ST02Q01]. We coded this "1" if student was enrolled in the 7th grade (i.e., two years behind the beginning of high school), "0" otherwise.

Remedial Class: Dummy-coded variable based on students' responses to the following questionnaire items:
During the last three years, have you attended any remedial courses in the test language at your school to improve your results? [ST23Q02];
During the last three years, have you attended any remedial courses in other subjects at your school to improve your results? [ST23Q03].

Students' response options were: No, never; Yes, sometimes; Yes, regularly. We coded "1" if students attended either type of remedial class at least "sometimes," "0" otherwise.

**Sex:** Dummy-coded variable, coded "1" for females, "0" for males [ST03Q01].

### School-level Predictors

These variables were created either as aggregates from student-level reports (PISA variables beginning with "ST") or from reports by the school principal (PISA variables beginning with "SC:). In our separate-by-country HLM analyses, all continuous variables were z-scores (M = 1, SD = 0) across the school samples in each country.

#### School Composition

**Average SES:** School-level aggregate of student SES, described above.

#### School Structure and Resources

**Private:** Dummy-coded variable, coded "1" for private schools, "0" for public schools [SC03Q01].

**Pedagogical Resources:** School-level aggregate of student-level PISA item, "At your school, how often do you use science laboratories?" [ST39Q05]. Response categories included 5 options: never or hardly ever, a few times a year, about once a month, several times a month, several times a week. We converted this measure to a z-scored, M = 0, SD = 1, within each country.

**School's Physical Infrastructure:** School-level PISA composite, taken from "Index of the Quality of Schools' Physical Infrastructure" (PISA Technical Report 2001b:249). The PISA composite was made up of measures asking principals about the extent to which learning by the 15-year-olds attending their school was hindered by:

- Poor condition of buildings [SC11Q01];
- Poor heating, cooling, and/or lighting systems [SC11Q02]; and
- Lack of instruction space (e.g., classrooms) [SC11Q03].

Response categories included 4 options: not at all, very little, to some extent, and a lot. Scaled scores were created with Weighted Maximum Likelihood estimates where positive values indicate that 15-year-olds' learning was not hindered by the school's infrastructure [SCMATBUI]. We converted this measure to a z-scored, M = 0, SD = 1, within each country.

#### School Academic Organization

**Academic Press:** School level aggregate of student-level PISA composite "Achievement Press" (PISA Technical Report, 2001b:226). The PISA composite was made up of students' responses on the frequency with which, in their <test language> lessons:

- The teacher wants students to work hard [ST26Q02];
The teacher tells students they can do better [ST26Q03];
- The teacher does not like it when students deliver careless work
  [ST26Q04]; and
- Students have to learn a lot [ST26Q15].

Response categories included 4 options: never, some lessons, most lessons, and every lesson. Scaled scores were created with Weighted Maximum Likelihood estimates where positive values indicate higher levels and negative values indicate lower levels of achievement press [ACHPRESS]. We converted this measure to a z-scored, \( M = 0, \ SD = 1 \), within each country.

**School Social Organization**

**Student Absenteeism:** School-level PISA composite, made up of measures asking principals about the extent to which learning by the 15-year-olds attending their school was hindered by:
- Student absenteeism [SC19Q02];
- Disruption of classes by students [SC19Q06]; and
- Students skipping classes [SC19Q09].

Response categories included 4 options: not at all, very little, to some extent, and a lot. The composite index was computed using non-parametric Item Response Theory methods, in particular Molenaar's generalization for polotomous items of Moken's scale for dichotomous items (Molenaar 1997). Positive values indicate that the learning of 15-year-olds was not hindered by students' absenteeism. We converted this measure to a z-scored, \( M = 0, \ SD = 1 \), within each country.

**Collective Responsibility:** School-level aggregate of student-level PISA composite "Teacher Support" (PISA Technical Report 2001b: 226). The PISA composite was made up of students' responses on the frequency with which their teachers:
- Show an interest in every student's learning [ST26Q05];
- Gives students an opportunity to express opinions [ST26Q06];
- Helps students with their work [ST26Q07];
- Continues teaching until the students understand [ST26Q08];
- Does a lot to help students [ST26Q09]; and
- Helps students with their learning [ST26Q10].

Response categories included 4 options: never, some lessons, most lessons, and every lesson. Scaled scores were created with Weighted Maximum Likelihood estimates where positive values indicate higher levels and negative values indicate lower levels of teacher support [TEACHSUP]. We converted this measure to a z-scored, \( M = 0, \ SD = 1 \), within each country.

**Students' Sense of Belonging:**
- Feel like an outsider [ST31Q01, reversed];
- Make friends easily [ST31Q02];
- Feel like they belong [ST31Q03];
- Feel awkward and out of place [ST31Q04, reversed];
- Other students seem to like them [ST3105]; and
- Feel lonely [ST31Q06, reversed].

Response categories included 4 options: strongly disagree, disagree, agree, and strongly agree. Scaled scores were created with Weighted Maximum Likelihood estimates where positive values indicate higher levels and
negative values indicate lower levels of sense of belonging [BELONG]. We converted this measure to a z-scored, $M = 0$, $SD = 1$, within each country.

**Teachers' Commitment:** School-level PISA composite, taken from "School Principals' Perceptions of Teachers' Morale and Commitment" (PISA Technical Report, 2001b, p.247). The PISA composite was made up of measures asking the degree to which principals agreed with statements about teachers working in their schools:

- The moral of teachers in this school is high [SC20Q01];
- Teachers work with enthusiasm [SC20Q02];
- Teachers take pride in this school [SC20Q03]; and
- Teachers value academic achievement [SC20Q04].

Response categories included 4 options: strongly disagree, disagree, agree, and strongly agree. Scaled scores were created with Weighted Maximum Likelihood estimates where positive values indicate a higher perception of teacher morale and commitment, and negative values indicate lower levels [TCMORAL]. We converted this measure to a z-scored, $M = 0$, $SD = 1$, within each country.

**II. MEASURES CONSIDERED BY ULTIMATELY EXCLUDED**

All of the variables we list here we considered at the school level, but ultimately we eliminated these due to non-significant associations. More detail of our model-building strategy is available in the text. Some are made as school-level aggregates from students' responses. Others are composites created from principals' responses. All are composite variables, which include several individual variables. For researchers interested in pursuing analyses with these measures, more detail about their construction is available from the Pisa Technical Manual (2001b).

- Index of Disciplinary Climate;
- Index of Teacher-Student Relations;
- Index of Computer Usage;
- Index of Interest in Computers;
- Index of Comfort With and Perceived Ability to Use Computers;
- Index of School Autonomy;
- Index of Teacher Autonomy;
- Index of Principals' Perceptions of Teacher-Related Factors Affecting School Climate;
- Index of Principals' Perceptions of Student-Related Factors Affecting School Climate;
- Index of the Quality of The School's Education Resources;
- Index of Teacher Shortage;
- Proportion of Teachers Fully Certified; and
- Proportion of Language Teachers with ISCED 5 A-Level in Pedagogy.
### Appendix 2: Saturated HLM Between School Model of Reading Achievement in Brazil and Four Comparison Countries

<table>
<thead>
<tr>
<th></th>
<th>BRAZIL</th>
<th>MEXICO</th>
<th>PORTUGAL</th>
<th>SPAIN</th>
<th>USA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average Reading Achievement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{00}$</td>
<td>389.2***</td>
<td>419.6***</td>
<td>475.3***</td>
<td>494.1***</td>
<td>493.5***</td>
</tr>
<tr>
<td>Average SES, $\gamma_{01}$</td>
<td>33.5***</td>
<td>42.8***</td>
<td>22.7***</td>
<td>29.8***</td>
<td>50.8***</td>
</tr>
<tr>
<td>Collective Responsibility, $\gamma_{02}$</td>
<td>4.4~</td>
<td>3.9</td>
<td>1.7</td>
<td>4.3*</td>
<td>-3.1</td>
</tr>
<tr>
<td>Academic Press, $\gamma_{03}$</td>
<td>5.1*</td>
<td>-4.0</td>
<td>0.02</td>
<td>1.4</td>
<td>3.1</td>
</tr>
<tr>
<td>Students’ Sense of Belonging, $\gamma_{05}$</td>
<td>1.2</td>
<td>15.5***</td>
<td>9.2**</td>
<td>1.0</td>
<td>-10.0~</td>
</tr>
<tr>
<td>Teacher Commitment, $\gamma_{06}$</td>
<td>0.12</td>
<td>2.0</td>
<td>-0.02</td>
<td>3.0</td>
<td>4.3~</td>
</tr>
<tr>
<td>School’s Physical Infrastructure, $\gamma_{07}$</td>
<td>1.6</td>
<td>-3.1</td>
<td>5.7***</td>
<td>-0.77</td>
<td>-1.4</td>
</tr>
<tr>
<td>Student Absenteeism, $\gamma_{08}$</td>
<td>-4.2*</td>
<td>-1.9</td>
<td>-2.8</td>
<td>0.90</td>
<td>-1.1</td>
</tr>
<tr>
<td>Pedagogical Resources, $\gamma_{09}$</td>
<td>3.3~</td>
<td>-0.86</td>
<td>9.52***</td>
<td>-0.66</td>
<td>9.7*</td>
</tr>
<tr>
<td>Private school, $\gamma_{011}$</td>
<td>9.7</td>
<td>-11.6</td>
<td>-12.8</td>
<td>6.9</td>
<td>15.3~</td>
</tr>
</tbody>
</table>

| **SES/ Achievement Slope** |         |         |          |         |         |
| Intercept, $\gamma_{10}$ | 6.0*** | 4.6** | 11.6*** | 19.4*** | 26.8*** |
| Academic Press, $\gamma_{11}$ | 0.43 | -0.04 | -2.0 | 0.0 | -4.4 |
| Students’ Sense of Belonging $\gamma_{12}$ | 0.71 | 3.8~ | 2.1 | -1.6 | -4.5 |
| Student Absenteeism, $\gamma_{13}$ | 3.0* | 1.4 | -0.09 | 2.6 | -0.20 |

| **Random Effect** | **Variance Component** |         |          |         |         |
| Intercept, $\mu_{0j}$ | 643.1*** | 725.9*** | 320.6*** | 447.8*** | 496.5*** |
| SES slope, $\mu_{1j}$ | 40.1** | 100.5*** | 86.9* | 30.0 | 214.5*** |

---

a. The within school models are not reported here, as they are almost identical to the within-school models shown in table 5.

b. ~ p<.10; * p<.05; ** p<.01;