Crushing College Hopes: How High-Stakes Testing Produced Inequities in College Access in Chile

IHELG Monograph

12-04

Paulina Pérez-Mejías
Researcher – Higher Education Research Center
Universidad de Santiago de Chile
paulina.perez@usach.cl
240-472-0503

Alberto F. Cabrera
Professor, College of Education
University of Maryland, College Park
cabrera@umd.edu
301-405-2875

Robert G. Croninger
University of Maryland, College Park
croninger@umd.edu
301-405-1818


© Paulina Pérez-Mejías, Alberto F. Cabrera and Robert G. Croninger 2012
University of Houston Law Center/Institute for Higher Education Law and Governance (IHELG)

The University of Houston Institute for Higher Education Law and Governance (IHELG) provides a unique service to colleges and universities worldwide. It has as its primary aim providing information and publications to colleges and universities related to the field of higher education law, and also has a broader mission to be a focal point for discussion and thoughtful analysis of higher education legal issues. IHELG provides information, research, and analysis for those involved in managing the higher education enterprise internationally through publications, conferences, and the maintenance of a database of individuals and institutions. IHELG is especially concerned with creating dialogue and cooperation among academic institutions in the United States, and also has interests in higher education in industrialized nations and those in the developing countries of the Third World.

The UHLC/IHELG works in a series of concentric circles. At the core of the enterprise is the analytic study of postsecondary institutions—with special emphasis on the legal issues that affect colleges and universities. The next ring of the circle is made up of affiliated scholars whose research is in law and higher education as a field of study. Many scholars from all over the world have either spent time in residence, or have participated in Institute activities. Finally, many others from governmental agencies and legislative staff concerned with higher education participate in the activities of the Center. All IHELG monographs are available to a wide audience, at low cost.

Programs and Resources

IHELG has as its purpose the stimulation of an international consciousness among higher education institutions concerning issues of higher education law and the provision of documentation and analysis relating to higher education development. The following activities form the core of the Institute’s activities:

Higher Education Law Library

Houston Roundtable on Higher Education Law

Houston Roundtable on Higher Education Finance

Publication series

Study opportunities

Conferences

Bibliographical and document service

Networking and commentary

Research projects funded internally or externally
Crushing College Hopes: How High-Stakes Testing Produced Inequities in College Access in Chile

Paulina Pérez-Mejías, Ph.D.
Visiting Scholar – Higher Education Program
University of Maryland, College Park
pperezm@umd.edu
240-472-0503
Professor – Universidad de Santiago de Chile
paulina.perez@usach.cl

Alberto F. Cabrera
Professor, College of Education
University of Maryland, College Park
Cabrera@umd.edu
301-405-2875

Robert G. Croninger, Ph.D.
Associate Professor, College of Education
University of Maryland, College Park
croninge@umd.edu
301-405-1818

Abstract

The Chilean education system has enjoyed a solid reputation in Latin America for decades. However, this reputation was severely shattered by a wave of student protests that brought Chilean schools and colleges to a halt for most of 2011. Among students’ numerous demands was a call to broaden access to higher education for traditionally underrepresented students, who persistently underperform in the national college admissions tests. Our purpose was to examine whether school and student characteristics account for the scoring gap. We found that student performance on admission tests is mostly influenced by the type of school attended rather than by individual characteristics. Our findings uncover a highly segregated school system that provides unequal quality education and learning opportunities. Far from leveling the field, the use of admissions tests to select applicants and to grant financial aid puts graduates of public schools at a great disadvantage, which can be even greater for women and low-income students.

Keywords

College admissions tests, admission policies, multilevel modeling, access to higher education
CRUSHING COLLEGE HOPES

Introduction

The Chilean education system has enjoyed a solid reputation in Latin America for several decades, and it has been often cited as an exemplary model of educational reform (UNESCO, 2009). Chile boasts one of the highest levels of literacy in Latin American (UNES, 2009); it is among the countries with the highest student achievement improvement rates in the last decade (Hanushek, Peterson & Woessmann, 2012; Moursched, Chijioke & Barber, 2007); and in 2009, Chile ranked first among Latin American countries in the Program for International Student Assessment test –PISA\(^1\) (OECD, 2010).

However, in 2011, this ideal picture of the Chilean education system was severely shattered by a wave of student protests. Known as the Chilean Winter (Villalobos-Ruminott, 2012), the student movement brought elementary, secondary and higher education to a halt for most of the academic year. Students demanded structural reforms of an education system they claim perpetuates and reproduces the stagnant inequality of the Chilean society (Al Jazeera, 2012). Given the Chilean achievements in education, the demands of the student movement, which were overwhelmingly supported by the public, came as a complete surprise to the international forum (Salinas & Fraser, 2012).

The student movement appears to be rooted in the inequitable structure of the Chilean educational system shaped by more than three decades of neoliberal educational reforms (Salinas & Fraser, 2012). These reforms were closely aligned to the policies sponsored by the World Bank, the International Monetary Fund, and the US Treasury Department, known as the Washington Consensus (Williamson, 2004), and widely supported by the “Chicago Boys”, an

---

\(^1\) The Program for International Student Assessment (PISA) is an international test that measures the performance of 15-year-olds in reading literacy, mathematics literacy, and science literacy every 3 years. To date over 70 countries and economies have participated in PISA.
influential group of Chilean economic advisors who studied under Milton Friedman at the University of Chicago (Espinoza, 2008; Pitton, 2006; Salinas y Frasier, 2012). Among those policies were the decentralization of public education, the creation of voucher schools, the advancement of cost-sharing as the main source of funding of higher education, and the *de facto* privatization of higher education (Espinoza, 2008; Pitton, 2006). As a result, Chile ended up with a highly privatized, market-oriented, and inequitable system of education (OECD & The World Bank, 2009).

**Access to Higher Education**

Access to colleges and universities substantially changed in this era of neoliberal reforms. Public expenditures to higher education decreased while most of the burden of financing postsecondary education was transferred to students and their families (Espinosa, 2008; Pitton, 2007). Additionally, the state created a loan system targeted at low-income students. However, in order to qualify for financial aid, students were required to score above a certain threshold on admissions tests. These college admission tests, modeled after the American SAT, had been used since the sixties as the main criterion to select applicants to public universities (Koljatic & Silva, 2006). Coupling financial aid to admissions test score requirements increased the stakes of performing well, especially for low-income students, who have historically scored lower in college admissions tests than their more affluent counterparts (Beyer & Le Foulon, 2002; Fontaine, 2002; Gil & Grez, 2002; Gil & Ureta, 2003; Le Foulon, 2002).

The testing model was overhauled in 2003 with the aim of better aligning the test with the contents and standards of the national high school curriculum. It was expected that this change of focus of evaluation, from aptitude to achievement, would reduce stagnant school, gender and income-based test score gaps (Koljatic & Silva, 2006; 2013). According to Koljatic and Silva (2006), these expectations were based on Atkison’s study (2001) on the predictive validity of
SAT I and SAT II tests of college performance in the University of California system. Atkinson concluded that achievement tests (SAT II), predicted performance in college success better than aptitude tests (SAT I); he also found that when controlling for the socioeconomic backgrounds of applicants, the predictive value of achievement tests remained intact, while that of aptitude tests virtually disappeared. Additionally Atkinson’s claimed that achievement tests were less vulnerable to charges of cultural or socioeconomic bias, and switching the focus to achievement tests would have a positive impact on high schools by setting clear curricular guidelines (Atkinson, 2001).

The purpose of this study was to examine the extent to which the revised college admission tests indeed fulfilled its purpose of decreasing gaps. That is, we sought to investigate test performance gaps among graduates of public schools, female and low-income students in relation to their better off counterparts after the testing procedures were changed in 2003. To accomplish this purpose, we used a multilevel approach that allowed us to simultaneously estimate test score gaps, at both the individual- and school-level. Therefore, the research questions guiding the study are: 1) What is the magnitude of the gender- and income-based scoring gaps within schools? 2) Do theses gaps vary across schools? 3) Do school characteristics explain such variation of within-school gaps? 4) What is the magnitude of gaps between schools of different sectors and do student and school characteristics explain such variation in the among-school gaps?

**Literature Review**

**Admissions Test Score Gaps**

The historically lower rates of college enrollment among certain groups of students traditionally excluded from higher education have been attributed often to their highly unequal results in college admissions tests, a trend that has been substantially documented (Beyer & Le
However, since the admissions test model was reformulated in 2003, similar studies are scarce. We were able to find only two published studies that have investigated the achievement gap related to admissions tests after 2003. Valdivieso, Antivilo, and Barrios (2006) found that test scores varied according to students’ family income, parents’ education, parents’ occupational status, and the type of high school from which students graduated. However, this study did not report the statistical significance of their results. An additional study conducted by Contreras, Corbalán, and Redondo (2007), using single-level regression techniques, identified income, prior achievement in elementary school, high school GPA, and gender as the main factors influencing test scores among Chilean high school graduates. Contreras and associates also found that parental education and the type of elementary and high school students attended had a smaller but still statistically significant effect on test performance. Although these studies have made a valuable contribution to the understanding of inequalities associated with the new admissions tests, they have ignored the nested nature of educational data and thus potentially underestimated the effect of the social and educational contexts of the schools on individual performance.

In the next section, we review the Chilean admission process to public universities as part of a broader conceptualization of pathways to college. The college choice process model developed by Hossler and Gallagher (1987) is used here as a frame of reference to structure the description of the Chilean admission process and students’ decisions involved in selecting an institution and major. Although this model was developed to study the college choice process of American students, given its simple three-stage structure – predisposition, search and choice – it is easily transferable to the Chilean context. Figure 1 depicts the three stages of the college choice process with a flowchart of sequential steps (rectangles) and decisions (rhombuses) involved in the admissions process at each stage.
Stages of the Chilean College Choice Process

**Predisposition.** The predisposition stage is the phase in which students’ intentions to continue education beyond high school emerges, and it may start as early as seventh grade for American students (Hossler & Gallagher, 1987; Cabrera & La Nasa, 2000). Similarly, in the Chilean context, this stage starts around eighth grade, when students must choose whether to enroll in a college-track or in a vocational high school. As noted by Cabrera and La Nasa (2000), in this stage students have already developed occupational and educational aspirations. However, it is not clear the role these aspirations play in Chilean students’ decision to attend a vocational or a college-track high school. Although one may assume that vocational students do not aspire to attend college, the high proportion of them taking admissions tests suggests otherwise. In fact, in 2010, 30% of test takers were graduates of vocational high schools (Kis & Feld, 2009). Also, because admissions tests were designed to assess students’ knowledge of college-track high school curriculum, vocational students are less likely than college-track students to obtain a score high enough to be admitted to college (Koljatic & Silva, 2010). Therefore, this early decision about the type of high school to attend shapes students’ future chances to access higher education. Another seemingly less important decision of this stage is the choice of which intensity track to follow in the eleventh grade (math, biology, or social sciences) (MINEDUC, 2011). Although following a particular intensity track is not a college admission requirement, it helps students to get prepared for the optional set of admissions tests, which assess more advanced and specific knowledge in math and sciences; higher scores in these areas are also required to apply for certain majors.

**Search.** In this stage, students gather information about universities and decide to which institutions they want to apply (Hossler & Gallagher, 1987; Cabrera & La Nasa, 2000). In Chile, there are many reliable sources available to support students’ search for information about
majors, admissions requirements, types of higher education institutions, accreditation status, financial aid, and other kinds of information (OECD & The World Bank, 2009). Most of these sources are official government websites and media publications (OECD & The World Bank, 2009). However, it is unknown if and how Chilean students use information from these or other sources in their search for financial aid, institutions, and programs.

According to Cabrera and La Nasa (2000), becoming academically qualified and obtaining a high school diploma are critical steps for American students to enroll in a postsecondary institution. Additionally, McClafferty, McDonough, and Nunez (2002) point out that preparing for and taking college admissions tests are also critical steps for American students on their path to college. Similarly, in the Chilean context, becoming ready for college requires graduating from high school and preparing for, registering for, and taking the admissions test, which are administered by the Department of Evaluation, Measurement and Educational Records (DEMRE). Although, in theory, becoming proficient at the contents of the national high school curriculum should be enough to perform well in admissions tests, in practice, this may not be the case depending on the quality of the school attended (OECD & The World Bank, 2009). Most high-income students rely on private coaching to get prepared for the tests, even if attending a high-quality school (Uribe & Salamanca, 2008). Most private preparatory courses for admissions tests are offered to eleventh and twelfth graders. Although there is not much evidence regarding the effects of test coaching in the Chilean context, it is widely believed that this preparation is absolutely necessary in order to earn competitive scores on the admissions tests (Williamson & Rodríguez, 2010).

Another step towards college for both vocational and college track high school graduates within the search stage is to register for the admissions tests (process 1 in Figure 1). Once applicants are registered, they can access a set of useful online resources at the DEMRE website.
For example, students can access free-of-charge practice test and use a simulation web tool that allows them to predict their chances of getting into a certain major at a specific institution. These are very useful resource to help students ponder their aspirations and to select a list of majors and institutions that maximize their chances to get admitted. Because Chilean universities admit students directly to majors, it is critical that upon high school graduation students have a clear idea of the major and the university to which they want to apply.

The final step in the search stage is taking the admissions tests (process 2 in Figure 1). College admissions tests consist of four standardized paper-based exams: 1) language, 2) mathematics, 3) science, and 4) history and social sciences. The language and mathematics sections of the test are mandatory, while the other sections are optional, although students are required to take at least one of the optional sections of the test. These tests attempt to measure applicants’ cognitive abilities, defined as their ability to recall information as it was learned; to conduct data analysis; to apply knowledge in problem solving; and to analyze, synthesize, and evaluate concepts, procedures, and problems. The questions are multiple choice with five possible answers each, of which only one is correct. In order to factor in applicants’ random answers, the final score is obtained by subtracting one quarter of all the wrong answers from the total number of correct answers. Then, the scores are standardized (z-scored) and adjusted so that the median is 500 score points and the standard deviation is 110 score points (the final score is $110 \times z$-score + 500). The test scale has a minimum of 150 score points and a maximum of 850 score points.

Admissions tests are offered just once annually, at the end of the academic year, in December. The tests are taken simultaneously in designated test locations across the country. The examination process takes three days; the first day students attend an orientation session, and during the next two days all registered applicants across the country simultaneously take the
different sections. If students do not complete the entire test-taking process, they are automatically eliminated from the admission process. Having actually taken the test becomes a critical step in the search process, because only after knowing their actual test scores can applicants ponder their real chances of being admitted to the programs and institutions to which they want to apply. Surprisingly, a significant number of students who register for the test ultimately do not actually take them. According to statistics from the DEMRE (2011), 13% of registered applicants did not take the test in 2010. The reasons behind this behavior have not been addressed yet in the literature.

In order to move to the next stage, the application process, high school graduates need to have scored above a minimum score. However, in the last three administrations of the test, one third of test takers did not score above that threshold. For these students, this is the end of the admission process. This is also the case for those applicants who earned the necessary score to move forward the application process but opted themselves out of it. The reasons why the unreported number of students makes such a decision are unknown.

**Choice.** In this stage, students must apply to a set of institutions and decide among alternative institutions to attend (Cabrera & La Nasa, 2000). The application to Chilean universities (process 3 in Figure 1) is done through the DEMRE website. Students must identify up to eight preferences of majors and institutions in order of importance. In 2005, only 43% of registered applicants completed the application process. The proportion of students who did so was lower than average for low-income students (34%) and graduates from public (35%) and vocational (21%) high schools (Valdivieso, Antivilo & Barrios, 2006).

Regarding the choice of institution and major, in the Chilean context it is the automatic system managed by DEMRE which makes the decision on behalf of the applicants (process 4 in Figure 1) based on their admissions test scores and stated preferences. DEMRE system admits
applicants into only one institution and major, although a significant proportion of students are not admitted to any of their preferred majors and institutions, either because they did not score above the threshold set by each university for a certain major, or because their application scores were not competitive enough to gain them admission to the majors listed in their preferences. The final step of this stage is actual enrollment and attendance at a particular institution (process 5 in Figure 1). In 2005, only 24% of registered applicants were able to enroll, and this proportion was lower for low-income (19%), graduates of public (20.1%) and vocational (11%) schools. In contrast, this proportion was higher for high-income students (35.2%) and graduates from private (34.1%) high schools (Valdivieso, Antivilo & Barrios, 2006).

**Theoretical Framework**

Given the scarcity of Chilean research on access to higher education, we relied mostly on the American literature to build a conceptual frame of reference. First, we looked at the literature on students’ performance on standardized tests, particularly on college admissions tests, to look for individual characteristics that may impact students’ performance in standardized tests. Most studies in this area use psychological theories to study how individual measures such as test anxiety, fear of failure, and stereotype threat, affect student performance on standardized tests (e.g. Cassady & Johnson, 2002; Cullen, Hardison & Sackett, 2004; McCarthy & Goffin, 2005; Sackett, Borneman & Connelly, 2008; Zwick, 2002).

Then, we looked at the research on college choice and access to higher education. Based on sociological theories, research identifies students’ social contexts, socioeconomic status, gender, and race as the main determinants of the likelihood of gaining access to higher education (e.g. Bourdeau & Passeron, 2006; Cabrera & La Nasa, 2000; Hossler & Gallagher, 1987; McDonough, 1997, 1998; Perna & Titus, 2004, 2005; Tierney & Colyar, 2005). College choice and access research based on economic perspectives, however, assume students are rational
individuals, thus identifying college costs and benefits (e.g. tuition costs, financial aid availability, expected future income) as the main determinants of the college choice process (e.g. DesJardins, Ahlburg, & McCall, 2006; St. John & Paulsen & Carter, 2005; Perna, 2006b).

Finally, there are researchers that integrate and combine both sociological and economic perspectives into a more complex set of conceptual models (Hossler & Galagher, 1987; Cabrera & La Nasa, 2000; Perna, 2006; McDonough and Fan, 2007).

We also reviewed the literature on school effectiveness seeking to single out organizational factors influencing student achievement. This body of research has been dominated by two main theoretical approaches. The structural or bureaucratic perspective focuses on the impact of school structures, resources, constraints, and contingencies on student outcomes perspective (Bolman & Deal, 2003; Lee, Bryk, & Smith, 1993; Ma, Ma, & Bradley, 2008; McDonough, 1998). Alternatively, the communitarian perspective (Lee, Bryk, & Smith, 1993; Phillips, 1997) has emphasized the role of school culture (McDonough, 1998). From this viewpoint, school culture is seen as a set of implicit rules, traditions, and norms that shape the way in which members of that community behave (Deal & Peterson, 2009). Based on these two perspectives, we identified a set of school variables influencing student achievement that can be grouped into three main categories: structure, composition, and practices.

We proposed a conceptual model that builds upon the literature on school effects on one side (Lee & Croninger, 1994; Lee, Bryk, & Smith, 1993; Ma, Ma, & Bradley, 2008) and concepts from integrated models of college choice and access (Perna, 2006; McDonough & Fann, 2007) on the other hand. By overlapping the factors identified in the literature and the variables available in the datasets used for this study, we built a nested model in which we hypothesized that: 1) at the school level, students are affected by the particular contexts of their
respective schools, not only by the structural school characteristics but also by the aggregated social and academic characteristics of their peers, as well as school practices and policies; and 2) at the individual level, socioeconomic status, academic achievement, and demographic characteristics impact student performance on admissions tests. We hypothesized that both individual agency and school characteristics would have an effect on student performance. Consequently, the model consists of a two-level hierarchical structure whereby students are nested within schools as depicted in Figure 2.

Methods

Data and Sample

This study draws from two main Chilean datasets: 1) the admissions test dataset containing information about approximately 249,000 Chilean high school graduates registered to take the college admissions tests in 2009, provided by the The Presidents’ Council of Chilean Universities (CTA-CRUCH, 2012); and, 2) the prior achievement dataset containing 2006 mathematics and language achievement tests for approximately 245,000 tenth graders, provided by the Department of Learning Outcomes Assessment of the Chilean Ministry of Education (SIMCE).

We focused our analysis on recent high school graduates of college-track schools. As such, graduates from vocational schools and test takers that graduated prior 2009 were excluded from the sample. We decided to exclude vocational high school graduates because their curriculum does not prepare them to take admissions tests. As such, they constitute a different population than college-track students. As for test takers who graduated prior to 2009, they likely

---

2 The Presidents’ Council of Chilean Universities and the Department of Learning Outcomes Assessment of the Chilean Ministry of Education authorized the use of admissions and prior achievement datasets, respectively, following the corresponding submission of a formal request for access to restricted files.
correspond to applicants who are retaking the test. Due to repetition, these applicants are more likely to obtain higher scores (Zwick, 2002). Moreover, applicants who graduated before 2009 may have spent a period in college, in a private coaching program, or in the labor market. These types of academic and work experiences have been found to increase students’ performance on college admissions tests (Briggs, 2004; Zwick, 2002).

After merging, cleaning, and imputing the data for these datasets, the final sample for the study was comprised by 106,414 students nested within 1,887 schools. This sample constitutes almost a full population sample of recent high school graduates who attended college-track schools in Chile. As such, it can be considered a random sample of exchangeable units of a hypothetical larger population (Snidjers, 2005).

**Data Preparation**

**Verification of assumptions.** Prior to model specification we judged the extent to which our data sample met linear-model assumptions. We checked normality by looking at histograms and quantile-quantile plots, frequency distribution, and skewness. We found only one variable (school size) violating this assumption and transformed it into a categorical variable. We also obtained the Mahalanobis distance to search for outliers, and verified that identified outliers corresponded to legitimate variability. There were no outliers in the student-level dataset, while in the school dataset we identified just a few cases as outliers, and most of them were no longer outliers after we categorized school size. These outliers were all kept in the sample.

**Multicollinearity.** Using Vector Inflation Factor (VIF) values, we assessed the degree of multicollinearity among continuous variables. The results of these analyses showed a high degree of multicollinearity among measures of prior achievement at 10th grade, so we created a

---

3 Applicants that were not admitted in past admission years, and college students who want to change major or institution; they were both required to retake the test and go through the admissions process all over again. Since 2011, a new regulation allowed scores to be valid for two admission periods.
composite out math, reading, and English achievement test scores. And yet, this composite measure ended being highly correlated with school sector, school SES, and with the outcome variable. For this reason, we decided to exclude this composite from the study.\textsuperscript{4} For categorical school variables, we conducted an ordinal factor analysis in LISREL 8.8 to obtain polychoric correlations and to determine the amount of shared variance among these variable. We found that school sector and school SES were highly correlated. However, we decided to keep these variables and reassess their exclusion when specifying the model.

**Missing Data.** We conducted a missing value analysis in SPSS 20 indicating the amount of missing data was not higher than 0.4 percent for any student- or school- variable in the dataset. Given the small amount of missing data, and the monotonic pattern of missing data, we relied on single imputation of variables with the fully conditional specification in SPSS 20 to handle missing cases. This method is appropriate to use when the pattern of missing data is random and uses iterative algorithms based on Markov Chain Monte Carlo techniques to estimate missing data (IBM SPSS, 2011). After imputing our dataset, our final analytical sample contained 106,414 students nested within 1,887 schools.

**Constructs and Measures**

The selection of variables to be included in our model was guided by theory, availability and measure quality of such variables in the datasets. The descriptive statistics of the final set of variables considered are displayed in Table 1, labeled as either student- or school-level variables.

**Student-Level Characteristics.** The *outcome variable* is students’ mean scores on the verbal and math sections of the admissions test, in year 2009. The admissions test scores range

\textsuperscript{4} Further analyses during the preliminary stages of the model specification resulted in unstable parameter estimation when the prior achievement composite was included at level 2. This confirmed our previous decision of excluding this composite from the model.
from a minimum of 201.5 points to a maximum of 846.0 points; the average test score is 528.7 points with a standard deviation (SD) of 102.4 points.

We relied only on *family income* as a proxy for student socioeconomic status at the individual level. Family income is originally included in the PSU dataset as a categorical variable of twelve income brackets, which we recoded into fewer categories (low-income, 60%; middle-income, 16%; and high-income, 24%) for the purpose of simplifying the analysis.

We used *high school GPA* as a measure of academic ability. In 2009, the average high school GPA was 5.65 with a standard deviation of 0.49 (see Table 1). Additionally, we included students’ *high school class ranking* (how a student’s GPA compares to the GPAs of other students in her class) as a dummy coded variable (top ten percent = 1, not in top ten percent = 0). We incorporated *gender* via a dummy variable whereby males were coded as 0 and females as 1. Our sample is predominately female (55%). We also included *age* as measured by a three-category variable (born in 1990 or before, 10.9%; born in 1991, 58.1%; born in 1992 or after, 31.0%).

**School-Level Characteristics.** As for structural characteristics of the schools we included *school sector* (private, 19.2%; subsidized, 55.5%; public, 25.3%), whether the school was a *free-of-charge school* (no, 27.4%; yes, 72.6%), and whether the school received additional *preferential subsidies* from the Chilean state (yes, 33.7%; no, 67.3%). We also included measures for *average class size* (Mean=29.1, SD = 7.8, min=1, max=45) and *school size*; the latter was transformed from a continuous measure into three categories: small (1-400 students, 26%), medium (400-1,000 students, 61%), and large (more than 1,000 students, 13%). Regarding school composition, we included a measure of the *socioeconomic status* of the school available in the prior achievement dataset, which classifies schools into five categories based on survey data about parental level of education, family income, and a school vulnerability index. We
recoded this variable into three categories to facilitate analysis (low-income, 30.9%; middle-income, 26.1%; high-income, 43.0%). Regarding compositional variables, we included the proportion of female students within schools, school prior achievement as measured by the average school test score on tenth grade achievement tests (math, reading, and English), and the average GPA of school seniors. As for school practices and policies, we incorporated school selectivity as a dummy-coded variable (0= open admission, 1= entrance examination).

Approximately, 30% of the participating high schools reported using entrance examinations to select students. We also included the proportion of students taking admissions tests (Mean 86.7, SD 18.3). This measure is a proxy of what McClafferty, McDonough and Nunez (2002) define as college-going culture within a school.

[Insert Table 1 about here.]

**Analytical Approach**

We analyzed the data using a multilevel approach and specified a two-level model, usually referred to as an intercepts- and slopes-as-outcomes model (Raudenbush & Bryk, 2001). The equations that define this model are:

\[
\text{Level 1 : } Y_{ij} = \beta_{0j} + \sum_{q=1}^{Q} \beta_{qj} X_{qij} + r_{ij}
\]

where \(Y_{ij}\) is the individual standardized admissions test score, \(i\) is an individual student, \(j\) is an individual school, \(Q\) is the number of student predictors, \(X_{qij}\) is a student variable, \(\beta_{0j}\) is the average students’ test score for school \(j\), \(\beta_{qj}\) is the effect of the \(q\)th student variable on students’ test scores in school \(j\), and \(r_{ij}\) is the error for student \(i\) in school \(j\).

\[
\text{Level 2 : } \beta_{qj} = \gamma_{q0} + \sum_{s=1}^{S} \gamma_{qs} W_{sj} + u_{qj}
\]
where $\beta_{qj}$ are level-1 coefficients, $\gamma_{q0}$ the mean of $Y_{ij}$ for schools at the 0 value of all school level predictors, $S$ is the number of school predictors, $\gamma_{qs}$ is the effect of the $s$th school predictor on the relationship between $Y_{ij}$ and the $q$th student predictor, $W_{sj}$ is a school predictor, and $u_{qj}$ is the variance of $\gamma_{qs}$.

**Model Specification**

We used HLM 7 to estimate our model. Given the large number of level-2 units in the sample (1,887 schools), we deemed appropriate to use full maximum likelihood (MLF) as the estimation method. We first fitted a fully unconditional model to determine whether there was enough variability to model between schools. Then, using a step-up approach as recommended by Raudenbush and Bryk (2001, p. 257), we started fitting a level-1 model to incorporate student-level predictors one by one and tested whether their effect varied across schools. To avoid overfitting the model by including too many random effects, we fixed the effect of the slopes that either had low reliabilities (less than .1) or that were not theoretically interesting to model as random effects. To detect and properly estimate the slope heterogeneity of random slopes, the variables with random effects were entered to the model as group-mean centered, while level-1 variables with fixed effects were grand-mean centered to adjust for possible mean differences between schools (Raudenbush & Bryk, 2001).

Once the level-1 model was specified, we included school-level variables to the random intercept in a stepwise manner. The aggregates of level-1 group-centered variables with random effects were included on the intercept to test for compositional effects and to avoid possibly omitting important level-2 predictors associated with level-1 random effects (Raudenbush & Bryk, 2001, p. 262). Finally, we entered level-2 predictors to the random slopes. We started this final stage by fitting a parallel model (same predictors on the intercept and varying slopes), and
then conducted hypothesis testing to eliminate predictors from the slopes that were not significantly contributing to the model to obtain a parsimonious final model.

**Limitations**

This study is limited in three ways. First, we relied on a single-year data sample. Therefore, these data provide just a snapshot in time of the way in which student- and school-level variables influence students’ performance on college admissions tests. Secondly, after fitting the model we tested the homogeneity of variance assumption, and we found evidence suggesting heterogeneity of variance present in our data. This indicates that unidentified sources of within-group variability may explain such heterogeneity (Bryk & Raudenbush, 1988). In fact, we were not able to include some level-1 and level-2 variables that are known to influence performance on standardized tests, which constitute the last limitation of our study. The omission or exclusion of such variables in the sample was due to multicollinearity issues (e.g. prior achievement was excluded because of its high correlation with school SES and school sector), or simply because they were not available in the datasets (e.g. test coaching). Omitted predictors that are indeed related to the outcome variable may result in confounded estimated effects (Raudenbush & Bryk), and may be as well the cause of heterogeneity of variance present in our data. This challenge, however, is not unique to our study.

**Results**

**Fully Unconditional Model**

The fully unconditional model (One-Way ANOVA with random effects), yields estimates of the population mean and the amount of variance in the outcome variable that lies within and

---

5 It is worth noting that the statistic used to test the assumption of homogeneity of variance is quite sensible to non-normal data (Raudenbush & Bryk, 2001, p. 265). And although the continuous variables in our model are normally distributed, some degree of non-normality is always present in real large datasets, especially in the presence of nominal and ordinal variables. We cannot discard this as a possible explanation of the heterogeneity of variance found in our data.
between schools. Table 2 provides estimates of the school grand mean ($\gamma_{00} = 514.22$), the variance of the school means around the grand mean ($u_0 = 5671.69$), and the within-school level variance ($r_{ij} = 5168.97$).

[Insert Table 2 about here.]

Because the estimate of the between-variance ($u_0$) is significantly different than zero ($\chi^2 = 115,449.2$, $df = 1,886$, $p$-value <0.001), we concluded that there is statistically significant variation of the admissions tests score means across schools. The span of this variability can be obtained with 95% confidence, i.e. the range of plausible values within which the school means fall:

$$[\gamma_{00} \pm 1.96 \sqrt{u_0}] = [514.22 \pm 1.96\sqrt{5671.69}] = [366.6, 661.6] \tag{3}$$

Additionally, using the estimates of the variance components obtained with the fully unconditional model, we estimated the intraclass correlation (ICC) coefficient as follows:

$$ICC = \frac{\hat{r}_{00}}{\hat{\sigma}^2 + \hat{r}_{00}} \times 100 = \frac{5671.7}{5671.7 + 5168.9} = 0.523 \tag{4}$$

The ICC reveals that more than 52% of the variability in admissions tests scores lies between schools, providing empirical evidence to use a multilevel approach. Therefore, a two-level random intercepts and slopes model was specified according to equations (1) and (2). The final estimation of the parameters for the model are summarized in Table 3, in which $\beta$ ’s represent student-level effects, and $\gamma$ ’s refer to school-level effects. These effects are presented in terms of changes in raw test scores, and in terms of changes in standard deviations (SD) units. The effects expressed in terms OF SD serve to estimate the size of the effects based on Cohen’s taxonomy of effect sizes (1988).

[Insert Table 3 about here.]
**Student-Level Effects**

The individual characteristics that had a statistically significant impact on these students’ performance on college admissions tests were age, gender, income, high school GPA, and class ranking. Students’ high school GPA had the largest effect on individual performance. One SD increase in student GPA resulted, on average, in an increase of 1.6 SD, and this effect significantly varied between schools with a 95% confidence interval of 1.1 and 2 SD. On average, students ranked in the top ten percent of their class scored over a third of a SD higher than lower achieving students, even after controlling for individual GPA. Female students scored 1 SD lower than did male students, which is considered also a large size effect. This effect varied significantly across schools with a 95% confidence interval of 0.5 to 1.5 SD. High-income and middle-income students scored 0.33 and 0.17 SD points higher than their low-income counterparts, respectively. Finally, high school senior students that were older than other students of the same cohort (born in 1990 or before) scored on average almost 0.30 SD lower than their younger classmates.

**School-level Effects**

**Structural School Characteristics.** School sector had the largest effect on school means of students’ performance on college admissions test scores. Students from private schools scored, on average, 1.7 SD higher than students in public schools, which is a considerable effect. The gap between subsidized private and public schools is also significant, but smaller. On average, students in subsidized private schools scored almost 0.30 SD higher than students in public schools, which is a more moderate effect size. School and class size also had an impact on student performance on college admissions tests. Students attending large schools (more than 1,000 students) scored, on average, 0.83 SD higher than those attending small schools (less than 400 students). The test score gap between students attending small and medium-sized schools
CRUSHING COLLEGE HOPES

(between 400 and 1,000 students) is more than a third of a SD, a moderate and statistically significant effect. Regarding average class size, for each SD increase in class size, students score 0.29 SD. Students attending tuition-free schools scored, on average, 0.37 SD lower than their peers who attended schools that charge tuition. Students attending schools that receive additional subsidies from the government scored 0.26 SD than those who attended more privileged schools that did not need this type of assistance.

**Compositional School Characteristics.** A school’s average GPA had a large effect on its students’ performance on admissions tests. One SD increase in a school’s average GPA resulted in an average 0.163 SD increase in admissions test scores. This effect holds true after controlling for individual GPA, thus revealing a contextual effect of individual GPA on student performance on college admissions tests. Also, schools serving high SES students scored 1.5 SD higher than schools serving less privileged students, even after controlling at level-1 for individual income. Such a school effect is substantial by any standard. As for the proportion of female students enrolled at schools, a one SD increase in the proportion of female students at a school produced an average decrease in that school’s mean score 0.15 SD. Although significant, this is a relatively small effect.

**School Practices and Policies.** Students attending schools that administer entrance examinations for admissions purposes scored, on average, 0.29 SD higher than students who attended schools with open admission. Also, the college-going culture of a school, as measured by the proportion of students who take admissions tests, had a moderate but significant effect on student scores on admissions tests. A SD increase in the proportion of students who take the tests resulted in an increase of 0.29 SD on the test.
Cross-level Interactions

The Female Slope. Only two school-level variables were found to have a significant effect on the female slope. The female slope was slightly moderated by class size and slightly exacerbated by the proportion of female students within the school. On average, the slope for female students decreased by 0.05 SD in schools with larger average class sizes (+1 SD). On the other hand, a greater proportion of female students in a school negatively affected the individual performance of female students. A 1 SD increase in the proportion of female students at the school widened the gap for female students by 0.07 SD.

The High School GPA Slope. The variability across schools associated with high school GPA was explained partially by school sector, school size, the proportion of students taking admissions tests, and the average GPA of the school. The performance-GPA slope was moderated at private and subsidized private schools by almost 0.1 SD, respectively. On the contrary, as school size increased, the performance-GPA slope became steeper by almost 0.1 SD in medium-sized (between 400 and 1,000 students) schools and by a bit more than 0.1 SD in large schools (more than 1,000 students). Also, the composition of the student body had an impact on the performance-GPA slope; namely, as the proportion of students who took admissions tests and the average GPA at the school increased, the performance-GPA slope became steeper by 0.13 and 0.25 SD, respectively.

Model Fit

As a final step, we obtained the proportion reduction in variance for the residuals at level-1, for the intercept, and for the slopes. The results of this calculation correspond to the percent reduction in each variance component obtained in each of the subsequent stages of the model specification: the null model, the within-school model, and the between-school model.

[Insert Table 4 about here.]
As illustrated in Table 4, the final model successfully explained 50% of the variance at the student level (within schools). Thus, there is 50% of variance within schools left to be explained by other factors. Also, the final model explained 81.7% of the variance in the intercept (between schools). This means that the variables included in the model successfully explained most of the variability of school mean test scores. Also, the model shows that school characteristics account for the variability in the high school GPA slope, although there is more than 57% of the variance left unexplained. Finally, the model was not able to explain the variability in the female slope. This is, the school characteristics that could moderate the gap in test scores of female and male students were not the ones included in the model.

Discussion

Our results show that the school characteristics explain to a large extent Chilean students’ performance on college admissions tests. The school attended explained more than fifty percent of the variability of students’ scores on the college admissions tests in 2009. This finding is consistent with the ICC obtained in prior achievement assessments in Chile. The OECD (2010) found that type of school attended accounted for slightly above 50% of the variability in math and reading scores in the Programme for International Student Assessment (PISA) test. This ICC for Chilean schools in performance is relatively higher than that of many other countries, even when compared to other Latin American developing countries. For example, the 2006 PISA study found that the ICC for the United States was approximately 25% and that for Norway was 11%, while for other Latin American countries, such as Colombia and Uruguay, the ICC was approximately 35% and 40%, respectively. This high ICC limits what students can do individually to improve their chances to access higher education. As such, performance in college admissions tests is more a reflection of schooling opportunities than individual efforts and capabilities.
Nonetheless, at the individual level, students can improve their chances through superior academic achievement in high school, as measured by their GPA during high school. This is the most important predictor of performance on admissions tests, after controlling for gender and individual income. The effect of academic achievement has such an impact on performance that high-achieving students (those in the top ten percent) outperform their classmates on the admission test, even after controlling for GPA and type of school attended. Across all schools female and low-income students obtain, on average, lower scores than their male and more privileged counterparts. This is consistent with previous research that had reported differences in mean scores favoring males and high-income students (Contreras, Corbalán, & Redondo, 2007; CTA-CRUCH, 2004; 2005; Koljatic & Silva, 2006; 2010; OECD & The World Bank, 2009).

Finally, at the individual level we also found that age is a significant predictor of performance on admissions tests. Older students scored significantly lower than their younger counterparts, although this difference was relatively small.

At the school level, most of the difference in average scores between schools is explained by school sector and socioeconomic status of students, an expected finding given the high level of socioeconomic segregation within school sectors in Chile (McEwan et al, 2008; Villalobos & Valenzuela, 2012). The gap between public and private schools is substantial, with public schools scoring 1.7 SD lower than private schools. In other words, students in a typical private school, on average, score 41 percentile points higher than students in an average public school. Although subsidized private schools also outperform public schools, the scoring gap between them is only one third of a SD. This means that students of a typical subsidized private school, on average, outperform students of a typical public school by only 8 percentile points. It is worth noting the fact that this private school advantage remains even after controlling for the compositional effect of socioeconomic status of the school and for school selectivity; this gap
suggests that there is something other than SES and selectivity that explains the superiority of private-school students’ performance on admissions tests. Perhaps, private schools are doing a better job preparing students for college admissions tests than are public and private subsidized schools.

DEMRE (2004, 2005), the institution that designs and administers the test in Chile, reported that the test score gap among private and public schools was approximately 1 SD, which correspond to a lower gap than the one we obtained (1.7 SD). Because our study is based in 2009 data, this difference in the magnitude estimated of the gap between private and public suggests that the gap may have increased between 2004, 2005 and 2009. Longitudinal analyses of schools’ average scores revealed that the gap between public and private schools has been relatively stable around 2 SD since the new set of test was implemented in 2003 (Pérez-Mejías & Croninger, 2012).

In relation to school size, our findings show that as the school size increases, the average performance of students attending the school increases, as well. This contradicts past research findings that smaller school sizes result in higher scores on standardized tests among elementary students (Arzola & Troncoso, 2011). These contradictory findings could be due to methodological differences, or because the effects of school size on performance in standardized tests may be spurious. In other words, it might be the case that the effect of school size is either mediated by other school characteristics, such as school resources, infrastructure, and principal’s leadership that this study was not able to control for.

The proportion of students taking admissions tests also influences the variability in school means. The higher the proportion of students who take the admissions test within a school, the higher is the school average admissions test score. The proportion of students who take the admissions test within a school may be reflecting what some researchers define as
college-going culture (Corwin & Tierney, 2007; McClafferty McDonough, & Nunez, 2002), which increases the chances of students following the necessary steps to get to college.

Regarding cross-level interactions, we found significant variation of the effect of GPA on test scores across schools; the weakest effects seem to be in small private schools, and the largest in large public schools. This finding suggests that GPA is a greater sorter in public schools than it is in private schools. As for gender, we found that its relationship with student performance was influenced by class size, with larger gaps in smaller classes; and with the proportion of female students within each school, with female students scoring even lower in schools in which the proportion of females in the schools is higher. However, the explanatory power of these two variables is very low, as most of the variability in the gender slope was left unexplained (89%).

In general, our results are consistent with Contreras et al. (2007), in relation to the direction of the effects. As for the size of the effects, although Contreras and colleagues did not report effect sizes for all variables in the study, they concluded that the impact of individual variables were much larger than school effects. Because they did not take into account the nested nature of the data, their results likely underestimated school effects.

Conclusions

The massive Chilean student movement of 2011 focused most of its discontent towards the inequities associated with the education system; a system which Villalobos and Valenzuela (2012) find to be highly stratified by income and socially polarized. This stratified system produces an uneven distribution of quality education across schools, which seem to explain the achievement gap in admissions tests between schools.

Based on a representative sample and an appropriate methodology to deal with nested data, we find that the type of school attended largely shapes a student’s performance on college admissions tests. The overwhelming role of the school severely constrains what students can do
individually to improve their chances to access higher education. Contrary to Atkinson’s arguments (2001) in favor of achievement-based tests, performance in college admissions tests in Chile is more a reflection of learning and schooling opportunities than individual efforts and capabilities. Most of the difference in average scores between schools is explained by school sector, which is highly correlated with available measures of school average socioeconomic status. Such correlation makes it difficult to tease apart the effects of socioeconomic status and school sector.

The awareness of such unequal schooling opportunities is probably crushing college hopes for many students attending low-quality public schools; this could explain what drove students to revolt while asking for a more equitable education system. Nonetheless, at the individual level, we conclude students can still improve their chances through superior academic achievement in high school, as measured by their GPA during high school. This is the most important predictor of performance on admissions tests, after controlling for gender and individual income.

At the beginning of this work, we used the college choice process model to identify critical decisions that students ought to make in their path to college, and the possible factors influencing those decisions. This frame of reference served our purpose of putting the college admissions tests into a much broader perspective, and to emphasize the fact that student performance on these tests is just the final outcome of a process that formally started at 8th grade. However, using college admissions tests as almost the only measure of potential college-performance means judging applicants by the final point of their educational course only, and disregarding the starting point and trajectories followed. Students able to overcome the drawbacks associated with underprivileged circumstances are judged only by their deficits,
leaving aside their gains. Far from leveling the field, admission decisions based solely on admissions tests “merely transforms privilege into merit” (Bourdieu & Passeron, 1979).

Implications of the Study

Implications for Policymaking. Our results suggest that the Chilean policy makers need to reconsider current financial aid and admissions policies to fulfill their goal of increasing access to postsecondary education on the part of underrepresented students. We found substantial gaps in test performance associated with school sector, gender and income. In a context given by an unequal distribution of quality education and learning opportunities, the use of admissions tests scores almost as the single criterion to select students goes against principles of test-fairness use (APA, 2012; see Willingham, 1999, p. 226). According to the Code of Fair Test Practices in Education of the American Psychological Association (2012), test use fairness urges that all test takers have equal opportunities to prepare for the test (APA, 2012) and to “avoid using a single test score as the sole determinant of decisions about test takers” (APA, 2012, C-5). It is recommended by APA to interpret test scores along with other information about individuals. Although high school GPA has always been part of the admission formula, the weight assigned to this variable ranges from 10% to 20%. The study findings’ reveal that high school GPA and student’s GPA ranking within each school is related to higher level of performance within schools. Including such variables as additional admission criteria would serve to compensate for large group differences and would widen the chances of graduates of low-performing schools to gain access to higher education. Recently, participating universities of the admission system managed by DEMRE agreed to include students’ class ranking in the admission formula starting with student admissions in 2013 (CRUCH, 2012). However, the weight assigned to these two criteria cannot exceed 20% of the total. Further studies should evaluate the effectiveness of the
inclusion of these new criteria to the admission formula and if the weight assigned is enough to broaden access to graduates of public schools and low-income and female applicants.

Additionally, using an arbitrary minimum threshold score to decide whether a student should receive financial aid is also an unfair use of the test (APA, 2012; Willingham, 1999, p. 232). The Code of Fair Test Practices in Education of the American Psychological Association (2012) explicitly states test users should “avoid using tests for purposes other than those recommended by the test developer, unless there is evidence to support the intended use or interpretation” (APA, 2012, C-3). We provided evidence that low-income students score, on average, below the minimum thresholds required to apply for financial aid. Therefore, their chances of complying with test score requirements are low. This evidence should encourage policymakers to consider rescinding current test score requirements used to determine students’ eligibility for state financial aid, at least when the student intention is to study at an open-admission institution.

Another financial policy that deserves to be reassessed is the one that allocates additional state appropriations to those institutions that enroll the 27,500 highest scoring students. Instead, we propose to devote those resources to incentive institutions to enroll students that may not score relatively high when compared with all applicants, but who outperform their peers and come from low-income backgrounds and public schools. These resources can serve to fund programs to adjust institutions practices to this new population of students.

Implications for Practice. Future work may develop appropriate and reliable measures of variables for which currently there are no available data in Chile. The need to have more and better information is critical to conduct high quality research that can inform policymaking. It would also be very beneficial for both researchers and policymakers to have a more integrated system of educational data, and more transparent procedures to be granted data access.
Educational data are dispersed over different Chilean government agencies; accessing such data is highly burdensome. Improvements in this area may result in more and better educational research.

**Recommendations for Future Research.** This study left some unresolved question worth answering by future research, such as the school characteristics that make some schools more equitable for women, whether the gaps found here are invariant across years, and whether other more complex modeling techniques, such as structural multilevel models could serve to better estimate individual- and school-level effects. Additionally, our review of the literature revealed there are many other critical issues related to the college choice process of Chilean students that deserve more attention. These include the influence of parents and teachers in students’ pathways to college, the role occupational and educational aspirations play in Chilean students’ decision to attend a vocational or a college track high school, if and how Chilean students use available sources information in their search for institutions and programs, the effect of test coaching on student performance, the role of financial aid in the decision to attend college, the reasons why an important proportion of students do not follow through the application process, the criteria used by students to state application preferences, and if and what type students defer their enrollment, just to name a few. Also, unknown is how different types of students navigate each of the stages of the college process and the role that socioeconomic status, gender, and schools play in such a process. Given the high inequalities of access to higher education in Chile, there is an urgent need to understand the ways in which students make decisions about college so policymakers and practitioners can implement appropriate policies and programs to enhance students’ educational opportunities.
References


Figures and Tables
Figure 1. The Chilean admission process in terms of the college choice process.
Figure 2. Conceptual model of student and school characteristics that influence student performance in college admissions tests.
Table 1. *Summary of descriptive statistics of student and school variables.*

<table>
<thead>
<tr>
<th>Student Variables <em>(n=106,415)</em></th>
<th>Mean (SD) / Percent</th>
<th>Range [min; max]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Admission Test Scores</td>
<td>528.73 (102.4)</td>
<td>[201.49; 846.0]</td>
</tr>
<tr>
<td>High School GPA</td>
<td>5.65 (0.49)</td>
<td>[4.0; 7.0]</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>45.0</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>55.0</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Born in 1990 or before</td>
<td>10.9</td>
<td></td>
</tr>
<tr>
<td>Born in 1991</td>
<td>58.1</td>
<td></td>
</tr>
<tr>
<td>Born in 1992 or after</td>
<td>31.0</td>
<td></td>
</tr>
<tr>
<td>Family Income</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low-income</td>
<td>60.0</td>
<td></td>
</tr>
<tr>
<td>Middle-income</td>
<td>16.0</td>
<td></td>
</tr>
<tr>
<td>High-income</td>
<td>24.0</td>
<td></td>
</tr>
<tr>
<td>Siblings in Higher Education</td>
<td>31.2</td>
<td></td>
</tr>
</tbody>
</table>

| School Variables *(N=1,887)*     |                     |                 |
| 10th Graders Math Average Test Score | 265.7 (44.9)      | [176; 365]      |
| 10th Graders Reading Average Test Score | 266.7 (31.6)      | [197; 337]      |
| 11th Graders English Average Test Score | 107.8 (23.3)     | [69; 178]       |
| College Admissions Average Test Score | 509.9 (78.8)     | [297; 723]      |
| Percent of students who took college admissions tests | 86.7 (18.3) | [6; 100] |
| Class Size                       | 29.1 (7.8)         | [1; 45]         |
| School Size                      | 739.3 (516.2)      | [5; 4,436]      |
| Small                            | 26.0                |                 |
| Medium                           | 61.0                |                 |
| Large                            | 13.0                |                 |
| School Sector                    |                     |                 |
| Private                          | 19.2                |                 |
| Subsidized private               | 55.5                |                 |
| Public                           | 25.3                |                 |
| School Socioeconomic Group       |                     |                 |
| Low-income                       | 30.9                |                 |
| Middle-income                    | 26.1                |                 |
| High-income                      | 43.0                |                 |
| Tuition-free                     | 27.4                |                 |
| Receives Preferential State Subsidy | 33.7            |                 |
| Selective (entrance examination) | 30.1                |                 |
Table 2. *Results from the Fully Unconditional Model*

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average school mean, $Y_{oo}$</td>
<td>514.22</td>
<td>1.76</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Variance component</th>
<th>p-value</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>School means (intercepts), $u_0$</td>
<td>5671.69</td>
<td>&lt;0.001</td>
<td>0.969</td>
</tr>
<tr>
<td>Student-level, $r_{ij}$</td>
<td>5168.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed Effects</td>
<td>Score points (se)</td>
<td>Effect in SD (se)</td>
<td>p-value</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------</td>
<td>------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Intercept, $\beta_{0j}$</td>
<td>465.1 (4.9)</td>
<td>1.72 (0.13)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Average school mean, $\gamma_{00}$</td>
<td>55.4 (4.2)</td>
<td>0.30 (0.10)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Private, $\gamma_{01}$</td>
<td>9.6 (3.3)</td>
<td>0.68 (0.10)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Subsidized private, $\gamma_{02}$</td>
<td>21.8 (3.1)</td>
<td>1.46 (0.12)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Middle-SES school, $\gamma_{03}$</td>
<td>47.0 (3.8)</td>
<td>0.33 (0.07)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>High-SES school, $\gamma_{04}$</td>
<td>10.4 (2.2)</td>
<td>0.83 (0.10)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Medium size, $\gamma_{05}$</td>
<td>26.7 (3.1)</td>
<td>0.83 (0.10)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Large size, $\gamma_{06}$</td>
<td>465.1 (4.9)</td>
<td>1.72 (0.13)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Average class size (standardized), $\gamma_{07}$</td>
<td>9.4 (1.4)</td>
<td>0.29 (0.04)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Prop. of students taking adm. tests, $\gamma_{08}$</td>
<td>9.4 (1.2)</td>
<td>0.29 (0.04)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Average school GPA, $\gamma_{09}$</td>
<td>9.4 (1.2)</td>
<td>0.29 (0.04)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Proportion of females, $\gamma_{10}$</td>
<td>9.4 (1.2)</td>
<td>0.29 (0.04)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Entrance examination, $\gamma_{11}$</td>
<td>9.4 (1.2)</td>
<td>0.29 (0.04)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Tuition-free, $\gamma_{12}$</td>
<td>9.4 (1.2)</td>
<td>0.29 (0.04)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Preferential State Subsidy, $\gamma_{13}$</td>
<td>9.4 (1.2)</td>
<td>0.29 (0.04)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

| Older student slope, $\beta_{1j}$ |  9.9 (0.7) | 0.30 (0.02) | <0.001 |
| Female slope, $\beta_{2j}$ | -30.3 (2.3) | -0.94 (0.07) | <0.001 |
| Intercept, $\gamma_{10}$ | -30.3 (2.3) | -0.94 (0.07) | <0.001 |
| Class size (standardized), $\gamma_{11}$ | 1.2 (0.5) | 0.05 (0.02) | 0.007 |
| Proportion of females, $\gamma_{12}$ | -2.1 (0.8) | -0.07 (0.13) | 0.009 |

| Middle-income slope, $\beta_{3j}$ | 5.4 (0.5) | 0.17 (0.02) | <0.001 |
| High-income slope, $\beta_{4j}$ | 10.6 (0.5) | 0.33 (0.02) | <0.001 |
| Top ten class ranking slope, $\beta_{5j}$ | 10.9 (0.7) | 0.34 (0.02) | <0.001 |

| High school GPA slope, $\beta_{6j}$ | 50.8 (1.2) | 1.58 (0.04) | <0.001 |
| Intercept, $\gamma_{10}$ | 50.8 (1.2) | 1.58 (0.04) | <0.001 |
| Private, $\gamma_{11}$ | -2.4 (1.0) | -0.07 (0.03) | 0.019 |
| Subsidized private, $\gamma_{12}$ | -2.7 (0.7) | -0.08 (0.02) | 0.006 |
| Medium size, $\gamma_{13}$ | 2.2 (0.8) | 0.07 (0.02) | <0.001 |
| Large size, $\gamma_{14}$ | 3.4 (0.9) | 0.10 (0.03) | <0.001 |
| Prop. of students taking adm. tests, $\gamma_{15}$ | 4.1 (0.4) | 0.13 (0.01) | <0.001 |
| Average school GPA, $\gamma_{16}$ | 8.3 (0.4) | 0.26 (0.01) | <0.001 |

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Variance component</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>School means (intercepts), $u_0$</td>
<td>1035.9</td>
<td>0.923</td>
</tr>
<tr>
<td>Female, $u_2$</td>
<td>55.9</td>
<td>0.183</td>
</tr>
<tr>
<td>High School GPA, $u_6$</td>
<td>51.8</td>
<td>0.369</td>
</tr>
<tr>
<td>Level-1, $r_{ij}$</td>
<td>2566.8</td>
<td></td>
</tr>
</tbody>
</table>
Table 4. *Percentage of variance explained and deviance statistic for the null, within-school, and final random intercepts and slopes model.*

<table>
<thead>
<tr>
<th>Variance Component</th>
<th>Null Model</th>
<th>Within Model</th>
<th>Final model</th>
<th>% Variance Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student-level, $r_{ij}$ sigma</td>
<td>5,169.0</td>
<td>2,565.8</td>
<td>2,566.8</td>
<td>50.3%</td>
</tr>
<tr>
<td>Intercept, $u_{0j}$ tau</td>
<td>5,671.7</td>
<td>5,266.6</td>
<td>1,035.9</td>
<td>81.7%</td>
</tr>
<tr>
<td>Female slope, $u_2$</td>
<td>-</td>
<td>55.8</td>
<td>55.8</td>
<td>0.0%</td>
</tr>
<tr>
<td>High school GPA slope, $u_6$</td>
<td>-</td>
<td>88.6</td>
<td>50.6</td>
<td>42.8%</td>
</tr>
</tbody>
</table>

**Model Fit**

<table>
<thead>
<tr>
<th>Deviance statistic</th>
<th>Null Model</th>
<th>Within Model</th>
<th>Final model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,219,062.2</td>
<td>1,146,547.3</td>
<td>1,144,162.7</td>
<td></td>
</tr>
<tr>
<td>Number of parameters</td>
<td>3</td>
<td>14</td>
<td>35</td>
</tr>
</tbody>
</table>