

Kiel Institute for World Economics
Duesternbrooker Weg 120
24105 Kiel (Germany)

Kiel Working Paper No. 1145

**Which School Systems Sort
Weaker Students into Smaller Classes?
International Evidence**

by

Martin R. West
Ludger Wößmann

January 2003

The responsibility for the contents of the working papers rests with the author, not the Institute. Since working papers are of a preliminary nature, it may be useful to contact the author of a particular working paper about results or caveats before referring to, or quoting, a paper. Any comments on working papers should be sent directly to the author.

Which School Systems Sort Weaker Students into Smaller Classes? International Evidence*

Abstract:

We examine whether the sorting of differently achieving students into differently sized classes results in a regressive or compensatory pattern of class sizes for a sample of national school systems. Sorting effects are identified by subtracting the causal effect of class size on performance from their total correlation. Our empirical results indicate substantial compensatory sorting within and especially between schools in many countries. Only the United States, a country with decentralized education finance and considerable residential mobility, exhibits regressive between-school sorting. Between-school sorting is more compensatory in systems with ability tracking. Within-school sorting is more compensatory when administrators rather than teachers assign students to classrooms.

Keywords: Student sorting, class size, educational achievement

JEL Classification: I28, H52, D73

Martin R. West

Harvard University
John F. Kennedy School of Government
Program on Education Policy and Governance
79 J. F. Kennedy Street, Taubman 306
Cambridge, MA 02138
United States
Phone: (+1) 617 496-5488
E-mail: mrwest@fas.harvard.edu

Ludger Wößmann

Kiel Institute for World Economics
24100 Kiel
Germany
Phone: (+49) 431 8814-497
Fax: (+49) 431 8814-500
E-mail: woessmann@ifw.uni-kiel.de
Internet: [www.uni-kiel.de/ifw/
staff/woessman.htm](http://www.uni-kiel.de/ifw/staff/woessman.htm)

* We would like to thank Claudia Buch, David Ellwood, Erich Gundlach, Eric Hanushek, Caroline Hoxby, Christopher Jencks, Paul Peterson, and seminar participants at Harvard and Kiel Universities for helpful comments. Wößmann would like to thank the Program on Education Policy and Governance (PEPG) at the Kennedy School of Government, Harvard University, for its hospitality during two visits which allowed much of this work to be completed. West is equally grateful for the support and hospitality provided by the Kiel Institute for World Economics.

I. Introduction

The sorting of students into classes of different sizes may reflect a variety of motives. While many policies influencing class sizes are explicitly compensatory, with low-performing students placed in smaller classes, other forces at work in many education systems are likely to have a regressive impact.¹ Of course, the question of whether allocating additional class-size resources to low-performing students is an effective and cost-efficient strategy for improving their performance is a matter of considerable ongoing debate (cf. Hanushek 2002; Krueger 2002). Yet the interest parents, teachers, and administrators express in reducing class sizes suggests that the distributional outcomes of this sorting process are of interest from a political economy perspective.² First, they can serve as a test of theories about the relative influence of various groups and individual actors in education systems with different institutional characteristics, and second as an imperfect but informative indicator of the extent to which these school systems reinforce or counterbalance existing academic and social inequities.

The mechanisms through which students of differing abilities might be sorted in smaller or larger classes are countless, and stem from sources as diverse as parental choices about where to live and which school their child will attend, the placement of students into classrooms within schools, and school-level placement policies of the school system as a whole. Although endogeneity in the relationship between class size and student achievement is widely recognized as a potential source of bias in estimates of resource effects (cf. Card and Krueger 1996), most previous research has concerned itself with the placement policies responsible only insofar as they mask the true causal impact of class size on achievement, and thus motivate the development of experimental or quasi-experimental research designs (Krueger 1999; Angrist and Lavy 1999; Case

¹ Although public finance scholars typically distinguish regressive policies from *progressive* policies, we use the term *compensatory* in order to emphasize that we are concerned here with the allocation of class size resources with respect to achievement rather than income and to avoid the connotations of the term ‘progressive education’ as it is most commonly used in the field of education.

² Viadero (1998) reports the consistent popularity of class-size reductions with U.S. politicians, and a recent poll shows that reducing class size is second only to early-childhood education as a priority for education spending, far ahead of items such as increasing teacher pay, putting computers in the classroom, and school construction and modernization (Jacobson 2002).

and Deaton 1999; Hoxby 2000).³ As yet, the relationship between students' academic ability and the relative size of their classes between and within schools has not been estimated, and the causes of different patterns of resource allocation remain unclear.

In this paper, we estimate the extent to which students of different ability levels are sorted into differently sized classes both between and within schools in 18 education systems around the world.⁴ To account for the possibility that the size of the class to which a student is currently assigned may affect her academic performance, we use a combination of school-fixed-effects and instrumental-variables identification strategies to decompose the simple correlation between class size and student achievement in each country into three parts (Section II): (i) the effect of students being sorted into schools with different average class sizes (the between-school sorting effect); (ii) the effect of students being sorted within schools into smaller or larger classes (the within-school sorting effect); and (iii) the causal effect of class size on student achievement.

We implement this identification strategy using the international database of the Third International Mathematics and Science Study (TIMSS), which provides data rich enough to support the estimations for a representative sample of middle-school students in 18 countries (Section III). To provide unbiased estimates of both between- and within-school sorting effects in each of these school systems, our identification strategy takes advantage of unique characteristics of the TIMSS database, namely its information on the performance and the actual and grade-average class size of students in two adjacent grades within the same school on a single achievement test.

Our empirical estimates reveal whether compensatory or regressive effects dominate the sorting of students into differently sized classes in our sample of school systems (Section IV). For the majority of countries in our sample, we find a statistically significant compensatory pattern of between-school sorting. The United States is the

³ A few studies of class-size effects have attempted to characterize in general terms the allocation of class-size resources in the specific contexts they examine. Akerhielm (1995) uses teachers' subjective ratings of the quality of the students in their classrooms to show that in the United States, teachers of relatively small classes were more likely to describe their students' abilities as below average. Similarly, a survey of a random sample of 500 teachers in New Jersey conducted by Boozer and Rouse (2001: 166) revealed an "overwhelming tendency of schools ... to allocate resources in a compensatory fashion."

⁴ While many kinds of sorting are possible, for example based on social class, race, or other family background features, this paper examines only sorting with respect to students' academic abilities. Likewise, it restricts itself to the allocation of class size, leaving the allocation of other resources for future research.

only country that exhibits statistically significant regressive between-school sorting. We also find evidence of sorting at the within-school level, with the allocation being compensatory in some countries and regressive in others.

In Section V, we develop hypotheses concerning the determinants of different patterns of resource allocation between and within schools. At the between-school level, we argue that the political economy of ability tracking results in compensatory resource allocations, while the residential choices of parents, particularly in combination with decentralized educational finance, further regressive resource allocations. At the within-school level, we argue that sorting tends to be regressive where teachers control student placements and compensatory where administrators control the placement of students and teachers into classrooms and where external exams make efficiency considerations more salient. Using the international variation in our estimates of sorting effects in combination with institutional background data on the school systems included in TIMSS, we present evidence consistent with each of these hypotheses.

II. Identifying Sorting Effects

A. Definition

We define sorting effects as *the observed relationship between the academic achievement of students (as measured by performance on a standardized assessment) and the size of the class in which they are taught, exclusive of any causal effect of class size on achievement*. The most obvious sorting effects occur when students are placed into specific schools and classrooms explicitly according to their prior academic performance. However, school systems may also sort students according to a variety of other characteristics, such as race, sex, disruptiveness, or socioeconomic status. To the extent that these characteristics are correlated with academic achievement, these alternative forms of sorting may play a substantial indirect role in generating the overall patterns presented in this paper.

Sorting effects so defined are conceptually equivalent to omitted-variable biases in an estimation of the effect of class size on student performance using observational data. Let O be a vector of variables “omitted” from a specification that tries to estimate the causal effect γ of class size S on test-score performance T :

$$(1) \quad T = \gamma S + O\theta + \varepsilon \quad .$$

Estimating this equation without the omitted variables O yields a biased estimate of the class-size coefficient which we denote α , and the standard formula for omitted-variable bias (cf., e.g., Greene 2000: 334) tells us that the expected value of this estimate is

$$(2) \quad E(\alpha) = \gamma + [(S'S)^{-1} S'O] \theta \quad .$$

Given an estimate γ of the causal effect of class size on performance, we can therefore measure the bias β introduced by sorting effects as follows:

$$(3) \quad \beta = E(\alpha) - \gamma = [(S'S)^{-1} S'O] \theta \quad .$$

Assuming for the moment only one omitted variable, we can also write

$$(4) \quad \beta = \frac{\text{cov}(S, O)}{\text{var}(S)} \theta \quad .$$

That is, the larger the covariance between the omitted variable and class size and the larger the effect of the omitted variable on performance, the larger the sorting effect (everything in absolute terms). The estimate β can essentially be interpreted as the effect of the “omitted variable” on test-score performance insofar as it is correlated with class size. We generally assume below that the size of the effect of specific omitted variables on performance is the same across school systems, so that any variation in aggregate sorting effects across school systems reflects differences in the covariance between class size and the complete set of “omitted” variables that influence performance and are at the same time related to class size (relative to the variance of class size in each system).

B. A Thought Experiment

A simple hypothetical example may help to clarify this definition. Imagine two school systems, MUCHSORTING and LITTLESORTING, each with the same number of students to be placed into classes of only two sizes, 10 students and 20 students per class. For simplicity, assume that there is no causal effect of class size on student performance in either system within this range of variation. Also assume that there is again only one omitted variable influencing both performance and class size:

disruptiveness. To simplify even further, assume that there are only two types of students – disruptive ones and non-disruptive ones – and that there is an equal number of each in both systems. Assume that more disruptive students are also more likely to be low performers – say, disruptive students perform 100 test-score points worse than non-disruptive students – and that administrators in both systems attempt to equalize the level of disruption across classrooms by placing more disruptive students in smaller classes. Each system is therefore characterized by a compensatory pattern of sorting, with low performers more likely to be taught in relatively small classes.

The two systems differ only in the proportion of disruptive students placed into the smaller classes. Assume that MUCHSORTING places all its disruptive students and none of its non-disruptive students into smaller classes, while LITTLESORTING places only 60 percent of its disruptive students into smaller classes – implying that 40 percent of its non-disruptive students also end up in smaller classes. (Perhaps teachers of small classes in LITTLESORTING object to having to instruct only disruptive students, forcing administrators to distribute non-disruptive students more evenly.) Given these assumptions, the average test score will be 100 points lower in the small classes than in the large classes in MUCHSORTING, but only 20 points lower in LITTLESORTING. Estimates of sorting effects β for the two school systems would indicate that for every 1 additional student in a class, average performance is 10 test-score points higher in MUCHSORTING, while it is only 2 points higher in LITTLESORTING.

In reality, of course, the situation is more complex. There are numerous examples of potential “omitted variables” other than disruptiveness that influence performance and may also be correlated with class size. The prior performance of students affects their current performance, and it will be correlated with class size in systems with ability tracking and different class sizes in different tracks. Family income would be another example of such an “omitted variable” in as much as it affects both students’ performance and the size of the classes in which they are taught, e.g. through residential choices of parents combined with local school financing. Sorting effects as we define them reflect the joint impact of decisions made by parents, educators, policymakers, and anyone else who influences the placement of students into classrooms on the basis of any such characteristics. It is this diverse and highly decentralized nature of placement

decisions that makes an empirical assessment of aggregate sorting effects essential to be able to characterize the allocation of class-size resources in different school systems.⁵

C. The Identification Strategy

As equation (3) indicates, we can identify sorting effects by decomposing the correlation between class size and achievement into the causal effect of class size on achievement and the bias introduced by sorting. Sorting effects can be usefully divided into two broad categories: sorting taking place *between schools* with different average class sizes, due to factors such as residential choice or tracking by school; and sorting taking place *within schools*, for example as a result of the policies schools use to assign students to different classes. The identification strategy advanced here generates separate estimates of the sorting effects arising at each of these two levels.

We begin with a standard least-squares (LS) regression of test scores on class size. Using test-score data pooled from two subjects and grades, the following education production function is estimated:

$$(5) \quad T_{icgs} = \alpha_{LS} S_{cgs} + \delta_1 M_{icgs} + \eta_1 G_{gs} + \nu_{1,cgs} + \varepsilon_{1,icgs} \quad ,$$

where T_{icgs} is the test score of student i in class c at grade level g in school s , S is the class size, M is a subject dummy (indicating test scores in mathematics as opposed to science), and G is a grade level dummy (indicating test scores from eighth as opposed to seventh grade). Additional variables to control for student and family background characteristics are intentionally omitted from this initial specification so that the sorting of students according to these characteristics will be included in our estimates of aggregate sorting effects. The coefficients α_{LS} , δ_1 , and η_1 are parameters to be estimated, ν is a class-specific component of the error term, and ε is a student-specific

⁵ Because of the largely unobservable character of various decentralized placement decisions, it would be hard to identify a system of two simultaneous equations, which could in principle be used as an alternative strategy to estimate sorting effects. In this simultaneous equations system, the first equation would specify performance as a function of class size (the causal class-size effect) and some identifying variables, while the second equation would specify class size as a function of performance (the sorting effect) and some identifying variables. Another alternative estimation strategy would be to simply assume that there is no causal class-size effect, in which case it would be possible to estimate the sorting effect by just regressing class size on performance. However, our results show that there *are* sizeable class-size effects in some countries, so that this alternative estimation strategy would yield substantially biased estimates of sorting effects. Data on performance and class size at the beginning of a school year would mitigate this problem, but no such data exist for a large cross-section of countries.

component of the error term. The following subscripts are applied throughout: i for student, c for class, g for grade level, and s for school.

Although often interpreted as the causal effect of class size on student performance (cf. Hanushek 2002; Krueger 2002), the estimated parameter α_{LS} also reflects the consequences of any sorting taking place at either the between- or the within-school level. Substituting the estimate α_{LS} into equation (3), we write the decomposition of α_{LS} as follows:

$$(6) \quad \alpha_{LS} = \gamma + \beta_B + \beta_W \quad ,$$

where γ again represents the true causal effect of class size on student achievement, β_B is a measure of between-school sorting, and β_W is a measure of within-school sorting.

We eliminate the effects of between-school sorting from the coefficient on class size estimated in equation (5) by controlling for school fixed effects (FE):

$$(7) \quad T_{icgs} = \alpha_{FE} S_{cgs} + D_s \lambda_1 + \delta_2 M_{icgs} + \eta_2 G_{gs} + \nu_{2,cgs} + \varepsilon_{2,icgs} \quad ,$$

where D is a vector of school dummies. This specification essentially relates differences in the relative performance of students in adjacent grades in the same school to differences in class size between the two grades. Any systematic between-school variation in student performance, regardless of its source, is accordingly excluded from the coefficient on class size. The estimate α_{FE} therefore includes only the causal effect of class size on achievement and the within-school sorting effect:

$$(8) \quad \alpha_{FE} = \gamma + \beta_W \quad .$$

To disentangle the true class-size effect γ from the within-school sorting effect β_W , we introduce into equation (7) a vector of controls for student and family-background characteristics C and apply an instrumental variable (IV) strategy to ensure that the variation in class size is exogenous to student achievement. The variable we use as an instrument for class size is the average class size at the respective grade level in each school. To be used as an instrument, a variable needs to be correlated with the endogenous right-hand-side variable (class size) but must be unrelated to the dependent variable (achievement) apart from the indirect effect resulting from its relationship with the endogenous right-hand-side variable. We demonstrate below that schools' average

class size in each grade is in fact highly correlated with the size of the class actually tested in that grade. Bound by staffing rigidities and administrative rules that determine the number of classes in a grade on the basis of cohort size, schools generally do not have the flexibility needed to allocate class-size resources across grades in response to differences in the performance level of adjacent cohorts. Thus, differences in average class size between grades within a school should be unrelated to between-grade differences in student performance. And apart from the effect of grade-average class size on actual class size, there is no reason to expect average class sizes to affect the performance of students in a specific class. The second stage of the two-stage least-squares (2SLS) estimation is then:

$$(9) \quad T_{icgs} = \alpha_{IV} \hat{S}_{cgs} + D_s \lambda_2 + C_{icgs} \chi_1 + \delta_3 M_{icgs} + \eta_3 G_{gs} + \nu_{3,cgs} + \varepsilon_{3,icgs} \quad ,$$

where \hat{S}_c is the predicted value of the first-stage regression of actual class size S_c on the average class size of the grade level in the school A_c :

$$(10) \quad S_{cgs} = \phi A_{gs} + D_s \lambda_3 + C_{icgs} \chi_2 + \delta_4 M_{icgs} + \eta_4 G_{gs} + \nu_{4,cgs} + \varepsilon_{4,icgs} \quad .$$

With the average difference in performance between the two adjacent grades absorbed by the grade-level dummy G , the remaining difference in performance between the classes from the two grades is unique to each school. This idiosyncratic performance variation is now related to only that part of the between-grade difference in actual class sizes that is caused by between-grade differences in average class sizes. In effect, this identification strategy asks whether students in a particular grade of a school performed better than students in the adjacent grade at the same school (both relative to the national averages for their respective grades) when their classes were on average smaller than those of students in the adjacent grade at the same school. Arguably, this remaining class-size variation is a consequence of random fluctuations in the cohort size between two adjacent grades of a school. Given adequate control variables and exogenous variation in class size, the coefficient α_{IV} estimated in equation (9) is uncontaminated by any sorting effects and reflects only the causal effect of class size on achievement (cf. Wößmann and West 2002):

$$(11) \quad \alpha_{IV} = \gamma \quad .$$

Equations (6), (8), and (11) along with the estimates of α_{LS} , α_{FE} , and α_{IV} enable us to disentangle within- and between-school sorting effects from causal class-size effects as follows:

$$(12) \quad \beta_W = \alpha_{FE} - \alpha_{IV}$$

$$(13) \quad \beta_B = \alpha_{LS} - \alpha_{FE}$$

$$(14) \quad \beta_T \equiv \beta_B + \beta_W = \alpha_{LS} - \alpha_{IV} .$$

The estimate β_T , defined as the sum of the between-school sorting effect β_B and the within-school sorting effect β_W , is an aggregate measure of the extent to which class-size resources in a given national education system are allocated in a compensatory or regressive manner with respect to student achievement.

Positive estimates of sorting effects indicate that class-size resources are allocated in a compensatory manner, with low-performing students placed in relatively small classes. The size of the estimates specifies how many fewer test-score points students placed in a class that is one student smaller tend to score, taking into account any causal effect of class size on achievement. Conversely, negative estimates indicate that additional class-size resources are targeted at more advanced students.

The statistical significance of the estimates of each of these sorting effects, β_B , β_W , and β_T , can be calculated using the specification test advanced by Hausman (1978) to compare alternative estimators of the same parameter. The null hypothesis is that the difference in the estimates between the two specifications is not systematic. Intuitively, the test assesses whether the bias affecting a parameter in a given specification – in our case, the bias resulting from sorting effects – is statistically significant. Given this approach, the statistical significance of our estimates of sorting effects for each country depends on the precision of the coefficients on class size estimated in equations (5), (7), and (9) and on the number of independent observations of class size in the data.

III. The TIMSS Database

The identification strategy developed above to estimate between- and within-school sorting effects was designed to take advantage of certain unique features of the data

collected as part of the Third International Mathematics and Science Study (TIMSS). The use of school-level fixed effects is made possible by the fact that the study sample included more than one class from each school. Using each school's average class size in each grade as an instrument imposes the additional requirement of data on achievement, actual class size, and grade-average class size for adjacent grades taking part in the same achievement test. Among large-scale international studies of student achievement, TIMSS is the only dataset with this particular set of characteristics.

Conducted in 1994/95, TIMSS tested nationally representative samples of middle school students in each participating country. The target population was defined as all students enrolled in the two adjacent grades that contained the largest proportion of 13-year-old students at the time of testing (grades seven and eight in most countries). In addition to testing students' educational performance in mathematics and science, extensive questionnaires were administered to students, teachers, and school principals in order to gather background information on the students themselves and on their institutional environments. Datasets for the middle school years were ultimately made available for 39 school systems around the world.⁶ Schools were sampled randomly within each country, and as a general rule, one class per grade was selected at random within each sampled school. Schools serving only students with special needs were excluded from the target population, implying that our estimates of sorting effects will be unaffected by cross-country differences in the treatment of these students. Within sampled classes, however, all students were generally required to participate.

Our reliance on within-school variation in performance and class size to identify sorting effects required that we restrict the sample to those schools in which both a seventh-grade and an eighth-grade class were actually tested. Furthermore, we only included in our analysis schools in which data on the actual class size and data on the grade-average class size were available for both the seventh-grade and the eighth-grade class. We ultimately conducted our analysis on the 18 school systems for which data from at least 50 schools in both mathematics and science remained after applying these

⁶ Separate datasets were collected for the French Belgian and Flemish Belgian school systems; both are included in the sample of school systems we examine in this paper.

criteria.⁷ Despite its reduced size, the sample of countries with which we are left includes systems from four different continents and with a wide range of distinctive institutional configurations.

Table 1 presents descriptive statistics on sample size, student performance, and class size for each of the countries included in our sample. The first five columns provide information on the number of students, classes, and schools. Each student in sampled classrooms was tested separately in mathematics and science. For our analysis of sorting effects we combine these observations into a single database. The number of observations (column (1)) is therefore roughly twice the number of individual students actually tested (column (2)).⁸ The number of observations of student performance in our sample ranges from 3,120 in Iceland to 20,209 in Japan. The number of classrooms sampled per country (column (3)) represents the number of independent observations of class size in our models, ranging from 134 in Hong Kong to 433 in Canada.

In TIMSS, student performance in mathematics and science were measured separately using international achievement scores with an international mean of 500 and an international standard deviation of 100. The mean performance for each country among the students included in our sample is presented in column (5). Portugal exhibits the lowest average test score (446) and Singapore the highest (600).

Data on the actual class size (column (6)) of each mathematics and science class come from the background questionnaires completed by each teacher, while data on the school-level average class size in grades seven and eight (column (7)) are from the school-principal background questionnaires. The minimum country-average class size of 20.3 students per class is found in Iceland, followed closely by the two Belgian school systems. With an average of 52.9 students per class, Korea has the largest classes by far. The other three East Asian countries in our sample also feature average class sizes of more than 30 students. The country averages for grade-average class size in a school are generally quite similar to the averages for actual class sizes. The variance in grade-average class sizes is somewhat smaller than the variance in actual class sizes –

⁷ This follows Wößmann and West (2002), who also report the specific reasons for the exclusion of the other TIMSS participants from the database.

⁸ The precise number is less than half in each country due to missing data on actual class size in one of the two subjects for some students.

as would be expected, since relatively small or large classes included in the sample should be balanced out by other classes within the same grade.

Column (8) reports the difference in the grade-average class size between seventh and eighth grade in the sampled schools. There are generally no sizable differences in grade-average class size between seventh and eighth grade, suggesting that the rules governing class size are the same between seventh and eighth grade in the included countries. The sole exception is Singapore, where grade-average class size appears to be nine students larger in the eighth grade.

A comparison of the standard deviations reported in parentheses in columns (6) to (8) demonstrates that the dispersion in the grade difference in class size is by and large comparable to the dispersion in actual class sizes in each country. The standard deviation in the between-grade difference in average class size ranges from 1.1 in Hong Kong to 6 in Spain and Singapore, with an average over the 18 countries in our sample of 3.5. It is less than 2 in only two countries: Hong Kong and Scotland. In these two countries, little variation in class size remains after having eliminated both between-school variations and within-grade variations in individual schools, leaving us with little variation on which to base our instrumental-variables estimation of causal class-size effects. In Hong Kong, for example, the largest positive class-size difference between eighth- and seventh-grade classes in either mathematics or science in any school is only 2, and the largest negative difference between eighth- and seventh-grade classes is only 3. With these two likely exceptions, however, there seems to be enough of the specific variation necessary to implement our instrumental-variables identification strategy.

While data on actual class size is taken from the teacher questionnaires, data on grade-average class size is collected from school principals. In column (9), coefficient estimates of a simple regression of actual class size on grade-average class size without a constant are reported for each country. The estimates are very close to 1 in all countries.⁹ This indicates that the data from the different background questionnaires are consistent. The estimates also confirm that the sampled classes are, on average, of the same size as the typical class sizes of the grades of the sampled schools; neither large classes nor small classes are over-represented.

⁹ Wald tests confirm that even though these coefficients are very precisely estimated, they are statistically indistinguishable from 1 in 11 of the 18 countries.

The following data on student and family-background characteristics, compiled from student background questionnaires, are included as control variables when estimating equations (9) and (10): the student’s sex, age, and country of birth, whether the student is living with both parents, parental education, and the number of books in the student’s home. The latter two are categorical variables with five categories each, so four dummy variables each are separately included in the regressions. Descriptive statistics on this data for each of the countries in our sample are presented in Appendix Table A1.¹⁰

IV. Estimation Results

A. Base Regression Results

Table 2 presents the results of the three base regressions – equations (5), (7), and (9) in Section II – used to identify sorting effects. The dependent variable in each regression is the TIMSS test score, pooling mathematics and science data to perform a single set of regressions for each country.¹¹ To facilitate international comparisons, we do not use scores that have been standardized for individual countries, but instead rely on the international achievement scores described in Section III. The regressions include a control variable indicating the subject to absorb any consistent within-country differences in performance level between the two subjects, as well as a control for the student’s grade level.¹²

¹⁰ Wößmann and West (2002: Appendix 1) compare descriptive statistics for the sample of students included in our study to the full sample of students tested by TIMSS for each country. Recall that the only students excluded were those attending schools in which students from only a single grade were tested, and those attending schools for which data on either actual or grade-average class size for one of the tested grades was missing. Apart from a few minor exceptions, the sample of students that we include in our study has very similar background and performance characteristics to the full sample of students tested in each country.

¹¹ When performing the analysis separately for each subject, the resulting estimates of sorting effects are highly correlated. Given this correlation and the fact that the determinants of sorting we consider below do not vary between subjects, we present only the more precise pooled results.

¹² The method used to calculate standard errors for these estimates takes into account the hierarchical structure of the TIMSS database. While student achievement is measured at the level of the individual student, class size is measured at the classroom level. Moreover, individual students in the same class may have various characteristics in common that are not fully captured by the included control variables. Regression analysis using hierarchically structured data requires the addition of a higher-level error term in order to avoid spurious results (Moulton 1986). Equations (5), (7), and (9) accordingly include a class-specific error component ν_{cgs} in addition to the student-specific error component ε_{icgs} . This technique, called clustering-robust linear regression (CRLR), delivers consistent estimates of standard

Column (2) reports the coefficient on class size from a weighted least-squares (LS) regression of achievement on class size as in equation (5).¹³ This coefficient reflects the causal effect of class size on student achievement as well as the effects of any sorting of students between- and within-schools into differently sized classes according to their performance. Eleven of the 18 LS estimates are positive in sign and statistically significant, indicating that students in larger classes perform better than students in smaller classes. In Hong Kong, where this pattern is most pronounced, students score 5.6 additional points (or 5.6 percent of an international standard deviation) on the TIMSS exam for each additional student in their class. In no country in our sample is the coefficient on class size statistically significantly less than zero, as would be the case if students in smaller classes outperformed students in larger classes.

Results of the specification that includes school fixed effects (FE) to control for any systematic between-school variation in student ability or the quality of education (equation (7)) are presented in column (4). The number of statistically significant positive estimates decreases to 4 from the initial total of 11 generated by the LS method, and there are 2 statistically significant negative estimates. The increased prevalence of statistically insignificant results cannot be attributed to a reduction in the precision in our estimates, which is modestly higher than in the LS specification. Rather, it appears that taking into account the effects of the sorting of students between schools yields a substantially different picture of the relationship between class size and student achievement.

The identification strategy we use to eliminate any effects of between- and within-school sorting combines school fixed effects with an instrumental-variables approach (FE-IV), as in equation (9). Also included in this specification is the full set of student-background control variables described in Section III. Column (6) presents the estimates produced by this specification, which can be interpreted as unbiased estimates of the

errors in the presence of hierarchically structured data (cf. Deaton 1997). It also takes account of any interdependence of the mathematics and science test scores of students from the same class.

¹³ The regressions weight each observation by the inverse of the probability of being sampled to ensure that the contribution to the parameter estimates made by students from each stratum in the stratified TIMSS sample reflects its proportional size in the general population (Wooldridge 2001). The use of weighted estimators is important in the case of models in which an important predictive variable, e.g. innate ability in this case, is unobserved (cf. DuMouchel and Duncan 1983).

causal effect of class size on student achievement.¹⁴ Once again, the pattern of results changes considerably as within-school sorting effects are excluded from the analysis. In no school system does the coefficient on class size remain statistically significant in the counter-intuitive, positive direction after the effects of both between- and within-school sorting have been excluded. On the other hand, the FE-IV estimates are statistically significant and negative in two countries: Greece and Iceland. In these two countries, smaller classes have an observable beneficial effect on student performance. The effects are substantial in magnitude, indicating students score just over two points (or 2 percent of an international standard deviation of test scores) higher for every one student less in their class. The 16 statistically insignificant estimates are rather evenly distributed around zero, with nine positive and seven negative.

B. Estimates of Sorting Effects

Table 3 presents our estimates of between-school, within-school, and total sorting effects for each of the 18 countries in our sample. Estimates of the between-school sorting effect (β_B) are reported in column (2). The standard errors for these estimates (column (3)), based on Hausman tests (see Section II.C), indicate that our identification method generates reasonably precise estimates of between-school sorting effects for each of the school systems we examine. Only in a single case – the Czech Republic – is the standard error slightly greater than 1.0, and the estimated sorting effect is nevertheless statistically significant at the one percent level.

The most notable feature of these results is the prevalence of statistically significant estimates, which indicates that students performing at different levels are indeed sorted into schools with differently sized classes in the majority of the countries in our sample. The predominance of positive estimates suggests that most of this between-school sorting is compensatory in character, with lower performing students placed in smaller classes. We find statistically significant positive estimates of β_B in 12 of our 18 countries. Only in the United States is β_B statistically significant and less than zero, indicating that students who perform better are typically placed in smaller classes.

Many of the estimated between-school sorting effects are also substantial in magnitude. In the Czech Republic, where the estimate of the between-school sorting

¹⁴ See Wößmann and West (2002) for detailed discussion of these estimates.

effect is the largest, students in classes that are one student larger as a result of the school they attend generally score nearly four test-score points (or four percent of an international standard deviation of student performance) higher on the relevant TIMSS exam, exclusive of any causal effect of class size on performance. The statistically significant negative estimate of β_B for the U.S., on the other hand, is quite small, indicating that students generally score only about 0.14 test-score points lower for each additional student in their class (once again excluding any causal effect of class size on performance).

The estimates of within-school sorting effects (β_W) reported in column (4) exhibit a different pattern. Fewer of these estimates achieve statistical significance, and those that do are more evenly divided between positive and negative results (although the majority is once again indicative of a compensatory pattern of resource allocation). The four countries with statistically significant positive estimates of β_W include France, Greece, Iceland, and Singapore. Statistically significant negative estimates are found for French Belgium and the Czech Republic.

The precision of our statistically insignificant estimates of β_W varies widely (column (5)). It is therefore useful to distinguish those cases in which we cannot reach any useful conclusions on the basis of our estimates from others in which we can with a reasonable degree of confidence rule out the existence of substantial sorting effects. Four countries in particular have extremely imprecise estimates: Australia, Hong Kong, Scotland, and the United States. A review of the results presented in Table 2 suggests that the high standard errors for these school systems result from imprecision in our estimates of α_{IV} , itself attributable to data insufficient to implement the demanding FE-IV estimation method.¹⁵ It is clear that our estimates of within-school sorting effects in these four cases are insufficiently precise to serve as the basis for any confident conclusions regarding the allocation of class-size resources. Elsewhere, however, the standard errors of these estimates are only modestly larger than the standard errors of our estimates of

¹⁵ In the cases of Hong Kong and Scotland, this imprecision reflects the lack of sufficient variation in the grade-average class sizes of adjacent grades in the same school, as discussed in Section III. The imprecision of the FE-IV estimates in Australia and the United States is due to the poor quality of the instrument, as demonstrated by its performance in the first-stage regression (cf. Wößmann and West 2002).

between-school sorting effects. In short, our strategy for identifying within-school sorting effects appears to have sufficient power to identify any substantial effects.

The results of the estimation of the joint effects of sorting at both the between- and within-school levels are presented in column (6) of Table 3. Each of the 7 statistically significant estimates of β_T is positive, indicating that lower performing students are placed in smaller classes. Singapore exhibits the most compensatory pattern of sorting, with students in classes that are one student smaller scoring more than five test-score points higher due to aggregate sorting effects. Among the statistically insignificant estimates of β_T , all but two are also greater than zero.¹⁶

In summary, we find evidence of statistically significant sorting effects at either the between- or within-school level in 15 of the 18 countries in our sample.¹⁷ These results confirm that students in each of these school systems are not assigned to differently sized classes randomly, but rather in a way that systematically reflects their ability. Variations in class size are indeed as much a consequence as a cause of differences in achievement.

These findings differ markedly from those of Hanushek and Luque (2002), who also examine the TIMSS database for evidence of non-random assignment into large and small classes. Using classroom-level performance data, they apply two alternative strategies to ensure that the results from standard LS regressions used to estimate class-size effects are not biased by what we refer to as sorting effects. They first restrict their sample to schools in rural areas, which they contend are unlikely to have more than one class per grade. Second, they include in an analysis of the full sample an indicator variable identifying classes smaller than the grade-level average in the same school. These two strategies produce few substantial changes in either size or statistical significance of their estimates of class-size effects, leading them to conclude that their “overall results are not heavily influenced by such selection effects” (Hanushek and Luque 2002: 19).

¹⁶ In terms of precision of the β_T estimate, which also relies on results of the FE-IV specification, the same conclusion applies as for the β_W estimate.

¹⁷ The three countries without statistically significant sorting effects are Canada, Korea, and Scotland, although our estimates of β_W for the Scottish school system are too imprecise for us to reach any firm conclusions regarding within-school sorting.

Their approach suffers from several potential flaws. In the absence of information on total school size, the fact that a school is located in a rural area is likely to be a poor indicator of whether or not it has multiple classes per grade, while a variable indicating simply that a class is smaller than average fails to differentiate among classes according to how much smaller they are than normal. Perhaps most importantly, neither of the strategies implemented by Hanushek and Luque (2002) addresses the possibility of the sorting of students *between* schools. In contrast, by considering sorting occurring at both the between- and within-school levels, we find evidence of between-school sorting effects in 13 of the 18 countries in our sample.

The relative importance of between-school sorting effects for the allocation of class-size resources is apparent in Figure 1, which displays the distribution of our estimates of β_B , β_W , and β_T into categories according to sign and statistical significance. The lower prevalence of within-school sorting may reflect the fact that many schools have only one class in each subject per grade, precluding the sorting of student into different classes. Figure 1 also draws attention to the fact that sorting effects in most of the school systems we examine tend to be compensatory, resulting in low-performing students being placed in smaller classes. However, exceptions with regressive sorting effects are found at both the between- and within-school levels, and the magnitude of both compensatory and regressive sorting effects varies considerably. In what follows, we discuss institutional characteristics of the school systems that may be responsible for these divergent patterns.

V. Determinants of Sorting

A. How Class-Size Resources Are Allocated: Towards a Theory of Student Sorting

The sorting effects estimated above reflect the aggregate impact of decisions made by a diverse set of actors, including parents, teachers, administrators, and central policymakers. Understanding the potential determinants of between- and within-school sorting effects requires that we narrow our focus to those stakeholders most likely to be able to influence the allocation of class-size resources at specific points. We postulate two kinds of actors as particularly important for sorting occurring at each of the two levels, depicted in Table 4.

Between-school sorting effects reflect decisions regarding which schools students attend and the resources available to them, decisions which should be made primarily by parents and central policymakers. In an education system with decentralized finance, Tiebout (1956) choice allows parents who demand more education to live in districts with a higher implicit school tax rate and thus more resources and the potential for smaller classes. Moreover, if the level of resources available in local schools is positively correlated with the price of housing, the children of wealthy parents, who will tend to be high performers due to their families' backing, will also have access to superior resources. Provided the education system allows parents to "purchase" additional resources for their children by engaging in sorting at the between-school level (i.e. resources are not fully equalized across districts and schools), we would thus expect that sorting effects resulting from parental residential choices will be regressive.¹⁸

The influence of policymakers on between-school sorting effects is likely to reflect political considerations. Given popular norms concerning equal educational opportunity, these considerations may be expected to lead them to allocate more resources to schools with disadvantaged students. Political pressure on policymakers should be particularly acute in systems that place students into distinct types of schools according to their ability, as parents and other advocates of the interests of poor performers are unlikely to consent to a tracked system unless students placed in the tracks for lower performers are compensated with tangible additional resources. Provided that such advocates have at least some influence on resource allocation, a simple political economy model of tracking suggests that tracked systems which assign students to schools according to their ability will tend to allocate class-size resources in a more compensatory manner.

Whereas between-school sorting effects are likely to be controlled by parents and central policymakers, patterns of within-school sorting should be determined primarily by the teachers and administrators of individual schools.¹⁹ These two sets of actors have

¹⁸ It might also be the case that parents of struggling students select school districts or individual schools with small classes to provide their children with extra attention, a consideration we assume here is of secondary importance.

¹⁹ To some extent, parents and policymakers may certainly exert an indirect impact on within-school sorting by influencing the actors with direct control over placements within schools.

divergent interests. The process of assigning students to classrooms has a direct influence on the conditions under which individual teachers will perform their job, giving them a clear personal stake in its outcome. All else equal, most would presumably prefer to teach gifted students in small classes, particularly if they feel that their own performance will be evaluated on the basis of that of their students. To see how this might affect patterns of within-school sorting, assume that a certain individual teachers – e.g. the most senior one – is allowed to govern the placement of students in classes and assign teachers to classrooms. He will then choose to place the best students into small classes and to teach these classes himself, leaving more junior teachers to deal with relatively large classes of relatively poor performers. If decision-making power rests with teachers collectively rather than with a single teacher, a similar pattern of regressive within-school sorting will emerge as long as some teachers exert more power in the collective decision-making process than do others.

Unlike teachers, the administrators of individual schools do not have a personal interest in teaching conditions in specific classrooms. Insofar as they are subject to political pressures in their within-school placements, these should lead them to engage in compensatory sorting as with policymakers in the between-school case. Additionally, assuming that school principals have limited resources and are evaluated on the basis of the overall performance of the student body as a whole, they will also be motivated to place students in classrooms in the manner most efficient for increasing overall student achievement. In practice, principals may transfer these incentives onto subordinate administrators, such as department heads who should possess more accurate information regarding the abilities of students and teachers in the specific domain under their control. There is scant empirical evidence available on efficiency in the assignment of students to small and large classes. However, Lazear (2001) argues that classroom teaching is a public good with congestion effects, so that it is optimal to place disruptive students in relatively small classes and well-behaved students in larger classes. Assuming disruptive students are also more likely to be low achievers, an efficient school would be characterized by compensatory within-school sorting effects. The incentives for principals to pursue efficient placement policies should be accentuated in education systems in which they are held accountable for student performance by a strong system of external exams (Bishop 1997). External exams also serve to draw

attention to the performance of low-performers, and may therefore lead to more compensatory placement policies independent of broader concerns for efficiency.

This brief discussion suggests several testable hypotheses regarding the likely determinants of sorting effects, which are summarized in Table 4. While extensive parental choice among schools and decentralized school finance should result in regressive between-school sorting effects, tracking policies may lead to more compensatory patterns. Meanwhile, within-school sorting should be more regressive where subgroups of teachers can influence the assignment of students and teachers to classrooms, and more compensatory where school administrators control this process. Finally, within-school sorting should be more compensatory where external exams hold schools accountable for student performance.

B. Evidence on the Determinants of Between-School Sorting

We test these theoretical hypotheses on the determinants of between- and within-school sorting by regressing our estimates of sorting effects on proxies for the various characteristics of school systems our theory suggests should be important. Based on only a few observations and relatively weak proxies, our empirical analysis of the determinants of sorting effects is necessarily speculative. Nevertheless, the results are informative.

As our dependent variable in this analysis is estimated rather than observed, the error term from an ordinary least squares regression is heteroskedastic with mean zero and variance equal to the sum of the variance of the actual error term and the variance of the estimated sorting effect. We account for this heteroskedasticity with the weighted estimation strategy advanced by Anderson (1993; cf. Slaughter 2001): First, we regress the estimated sorting effect on the relevant explanatory variable(s) using ordinary least squares. We then regress the squared residuals from this first regression on the variance of the estimated sorting effect and this variance squared and cubed. Finally, we generate the predicted values from the latter regression. These predicted values indicate the extent to which the original residuals can be explained by the variance of the sorting effects. The inverses of these predicted values are then used as the weights in a weighted least squares estimation of the initial model, thereby assigning less weight to those observations that are relatively imprecisely estimated.

Table 5 presents the results of regressions of our estimate of the between-school sorting effect β_B on a series of variables compiled from the TIMSS background questionnaires administered to the principal of each school and from supplementary sources.²⁰ We use two proxies for ability tracking: the percentage of TIMSS schools in each school system reporting that they use academic performance as a criteria for admission (*performance-based admissions*) and a dummy variable indicating whether students are placed in different tracks prior to the completion of secondary education (*tracking*). We constructed the latter variable on the basis of information contained in the *International Encyclopedia of National Systems of Education* (Postlethwaite 1996). The impact of parental choices on between-school sorting effects depends on two factors: families' ability to move between schools on the basis of class size and decentralized responsibility for education funding. Our proxy for mobility among the population with school-aged children is the mean percentage of current students in the schools sampled by TIMSS that transferred into the school after the start of the academic year (*transfer rate*).²¹ As a measure of fiscal decentralization, we take the percentage of final funds for primary and secondary education in 1995 allocated by local (as opposed to regional or central) governments (*local finance*; OECD 1998: 137, Table B6.1a). Descriptive statistics for each of these variables are presented in Appendix Table A2.

Columns (1) to (4) of Table 5 present the simple bivariate relationships between these four variables and our estimates of β_B . A positive coefficient indicates that the variable is associated with a more compensatory pattern of resource allocation, while a negative estimate suggests the relevant variable is associated with more regressive sorting. Both measures of tracking are in fact associated with compensatory between-school sorting effects. A high level of student mobility between schools is weakly associated with more regressive between-school sorting effects, as is the percentage of funds for education allocated by local governments.

²⁰ The value assigned to each country for the TIMSS variables represents the simple mean of the figures reported separately in the data almanacs for the seventh- and eighth-grade samples.

²¹ Using the mean percentage of students of current students in the schools sampled by TIMSS that did not finish the academic year in the same school as an alternative proxy for mobility produces equivalent results.

Each of these relationships becomes stronger in the multivariate specifications reported in columns (5) to (7).²² Although the small number of observations urges extreme caution in generalizing these results, the basic patterns among the countries with available data are clear. The use of academic admissions criteria by a large number of schools could in principle be used to foster any pattern of resource allocation. Its strong association with more compensatory sorting effects across specifications therefore serves to validate our hypothesis that political constraints ensure that highly selective education systems are characterized by compensatory patterns of resource allocation.

While our empirical results indicate that higher levels of student mobility and fiscal decentralization are independently associated with sorting effects, it is actually the combination of these two factors that we suggest is necessary for parental choices to lead to regressive between-school sorting. Data on both is available for only six of the countries in our sample. Nevertheless, the interaction term between transfer rate and local finance included in the column (7) specification is in fact statistically significantly associated with regressive sorting between schools.²³

The explanatory power of this interaction term reflects the fact that it essentially functions as an indicator variable for the United States, the only country in our sample for which we find statistically significant evidence of regressive between-school sorting. The United States is of course a country with high residential mobility and a system of education finance dominated by local property taxes. The combination of these two factors appears to allow families with high-performing students to live in school districts with relatively small classes. Programs intended to allocate additional resources to districts with a large percentage of students living in poverty are apparently insufficient to compensate fully for the decentralized decisions of parents regarding the schools their children will attend. This same logic may also explain the absence of compensatory between-school sorting effects in Canada, the country whose institutional

²² Qualitatively the same multivariate results emerge when using *tracking* rather than *performance-based admissions* as our proxy for tracked systems, although the statistical significance of the results is generally lower. This may reflect the fact that the former is a dichotomous variable, while the latter is continuous.

²³ Due to the small number of degrees of freedom, we did not include the main effects of transfer rate and local finance in this specification.

arrangements for financing education most closely parallel those of the United States (cf. Hoxby 2003).²⁴

C. Evidence on the Determinants of Within-School Sorting

Table 6 presents our analysis of the determinants of within-school sorting patterns, using our estimates of within-school sorting effects β_W . The TIMSS school background questionnaires included a battery of questions concerning who exercised primary responsibility for assigning students and teachers to classrooms. We use the percentage of schools reporting that teachers are responsible for placing students as a proxy for the ability of certain teachers to pursue their interests in within-school sorting. Conversely, department heads are likely to share their administrative superiors' joint goal of compensatory and efficient placements, and their greater familiarity with the students and teachers under their control may actually make them most effective in pursuing that goal. The percentages of schools in which department heads have primary responsibility for placing students and for assigning teachers in classes are therefore used as proxies for the dominance of compensatory considerations.²⁵

Columns (1) to (3) present the bivariate regressions of within-school sorting effects β_W on these proxies for within-school placement responsibilities. Although one of the observed relationships falls (marginally) short of conventional levels of statistical significance, each is in the expected direction. Greater teacher influence on student placements is associated with more regressive within-school sorting effects, while control by department heads is associated with more compensatory patterns. These relationships remain on the fringes of statistical significance when the proxies for responsibility of teachers and department heads are entered jointly in the multivariate specifications reported in columns (4) and (5). However, all continue to be in the expected direction.

Finally, column (6) presents the relationship between β_W and the percentage of principals reporting that external exams have a large influence on the curriculum in their schools, which we use as a proxy for the influence of external exams in general. Although this question was only asked in 15 countries, the variable has the statistically

²⁴ The estimate of β_B for Canada is extremely precise but statistically indistinguishable from zero.

²⁵ Descriptive statistics for these variables are also presented in Appendix Table A2.

significant association with more compensatory placement policies our theoretical hypothesis suggested.²⁶

VI. Conclusions

Do countries place less skilled students in classes smaller than those of their more skilled counterparts? While the relative incidence of compensatory sorting effects among our 18 sample countries implies they do, exceptions confirm that the internal allocation of class-size resources between low- and high-performing students defies easy generalization. Nevertheless, our analysis of sorting effects and their potential determinants suggest several broad conclusions.

First of all, the prevalence of sorting effects across a diverse sample of countries – a finding that contrasts with previous research using international data (cf. Hanushek and Luque 2002) – has important implications for estimating the causal effects of educational resources on student achievement. To the extent that school systems are compensatory in their allocation of particular resources, conventional estimates of resource effects will be biased against finding that those resources increase student achievement. Yet international variation in the direction and magnitude of sorting effects indicates that this bias must be assessed individually for specific school systems. Progress in the estimation of resource effects in international education production functions will therefore depend on the development and application of identification strategies designed to address the problem of resource endogeneity independently in each school system examined. Furthermore, such strategies need to be robust enough to take into account the possibility of endogeneity resulting from non-random placements both at the between- and the within-school levels.

The relationship between the specific sorting effects estimated in this paper and educational equity broadly conceived is not straightforward. The necessity of this qualification is underscored by the fact that a statistically significant causal effect of class size on achievement is found for only 2 of the 18 countries of our sample, while

²⁶ An interaction term between external exams and each of the variables on responsibility for placing students and teachers enters the specifications of columns (1) to (4) with a positive coefficient in each case. Although the estimates on the interaction terms never reach statistical significance at conventional levels, the direction of these relationships is consistent with the claim that central exams induce more compensatory placement policies.

the possibility of a substantial causal class-size effect can be rejected in 11 countries (cf. Wößmann and West 2002). In view of this finding, it is important to know whether the allocation of other educational resources accords with the findings reported here. Do low-performing students in countries with compensatory sorting effects also receive a disproportionate share of other resources, or is class size instead used to compensate for other inequities? Given budget constraints, investments in teacher quantity of the type necessary to reduce class size for low-achieving students might themselves serve to lower the quality of these students' teachers. If the quality of teachers represents the more important input in educational production (cf. Hanushek et al. 1998; Tamura 2001), compensatory class-size policies may even serve to decrease educational opportunities for low-performers.

Our analysis of the potential determinants of sorting effects enhances our understanding of the allocation of educational resources in the political economy of education, offering a solid foundation for future research in this area. Both our theory and our evidence suggest that aggregate patterns of sorting at the between-school level reflect the decisions of parents and central policymakers. While fiscal decentralization and the ability to move between schools allow wealthy parents to purchase additional resources for high-performing students, political considerations lead policymakers in selective systems with multiple tracks to direct additional class-size resources towards schools with low-performing students. By contrast, our model of within-school sorting emphasizes the role played by actors within specific schools. While individual teachers have an interest in regressive placement policies (provided they are permitted to teach the smaller classes), efficiency concerns may lead principals and other administrators to implement compensatory within-school placements schemes – especially in systems in which they are held accountable for student performance by external exams. All these hypotheses are consistent with our empirical evidence.

More generally, our findings regarding the determinants of sorting effects cast doubt on the empirical relevance of models of education production that treat schools as unitary actors maximizing educational productivity, thus ignoring both the institutional context in which they are situated and the interests of the individuals working within them. Lazear (2001), for example, assumes that schools always attempt to maximize educational productivity, and thereby pursue an optimal allocation of students to

classrooms. While his model's implication that it would be *optimal* to sort more disruptive students into smaller classes may well hold, the interests of some teachers may run in a different direction. Where teachers or other parties with personal interests at stake are allowed to influence student placements, educational considerations may not be paramount and school systems may *not* operate optimally.

References

- Akerhielm, Karen (1995). Does Class Size Matter? *Economics of Education Review* 14 (3): 229-241.
- Anderson, Patricia M. (1993). Linear Adjustment Costs and Seasonal Labor Demand: Evidence from Retail Trade Firms. *Quarterly Journal of Economics* 108 (4): 1015-1042.
- Angrist, Joshua D., Victor Lavy (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *Quarterly Journal of Economics* 114 (2): 533-575.
- Bishop, John H. (1997). The Effect of National Standards and Curriculum-Based Exams on Achievement. *American Economic Review* 87 (2): 260-264.
- Boozer, Michael, Cecilia Rouse (2001). Intraschool Variation in Class Size: Patterns and Implications. *Journal of Urban Economics* 50 (1): 163-189.
- Case, Anne, Angus Deaton (1999). School Inputs and Educational Outcomes in South Africa. *Quarterly Journal of Economics* 114 (3): 1047-1084.
- Card, David, Alan B. Krueger (1996). School Resources and Student Outcomes: An Overview of the Literature and New Evidence from North and South Carolina. *Journal of Economic Perspectives* 10 (4): 31-50.
- Deaton, Angus (1997). *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*. Baltimore, MD: The Johns Hopkins University Press.
- DuMouchel, William H., Greg J. Duncan (1983). Using Sample Survey Weights in Multiple Regression Analyses of Stratified Samples. *Journal of the American Statistical Association* 78 (383): 535-543.
- Greene, William H. (2000). *Econometric Analysis*. 4th ed. Upper Saddle River, NJ: Prentice Hall.
- Hanushek, Eric A. (2002). The Failure of Input-Based Schooling Policies. NBER Working Paper 9040, Cambridge, MA: National Bureau of Economic Research (forthcoming: *Economic Journal*).
- Hanushek, Eric A., John F. Kain, Steven G. Rivkin (1998). Teachers, Schools, and Academic Achievement. NBER Working Paper 6691, Cambridge, MA: National Bureau of Economic Research.
- Hanushek, Eric A., Javier A. Luque (2002). Efficiency and Equity in Schools around the World. NBER Working Paper 8949, Cambridge, MA: National Bureau of Economic Research (forthcoming: *Economics of Education Review*).
- Hausman, Jerry A. (1978). Specification Tests in Econometrics. *Econometrica* 46 (6): 1251-1271.
- Hoxby, Caroline M. (2000). The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *Quarterly Journal of Economics* 115 (4): 1239-1285.

- Hoxby, Caroline M. (2003). Introduction to *The Economic Analysis of School Choice*. Chicago, IL: University of Chicago Press (forthcoming).
- Jacobson, Linda. (2002). Poll: Public Sees Schools as a Priority. *Education Week* 21 (32): 1,22 (April 24).
- Krueger, Alan B. (1999). Experimental Estimates of Education Production Functions. *Quarterly Journal of Economics* 114 (2): 497-532.
- Krueger, Alan B. (2002). Economic Considerations and Class Size. NBER Working Paper 8875, Cambridge, MA: National Bureau of Economic Research (forthcoming: *Economic Journal*).
- Lazear, Edward P. (2001). Educational Production. *Quarterly Journal of Economics* 116 (3): 777-803.
- Moulton, Brent R. (1986). Random Group Effects and the Precision of Regression Estimates. *Journal of Econometrics* 32 (3): 385-397.
- Organisation for Economic Co-Operation and Development (OECD) (1998). *Education at a Glance: OECD Indicators 1998*. Paris.
- Postlethwaite, T. Neville (ed.) (1996). *International Encyclopedia of National Education Systems*. 2nd ed. Oxford: Pergamon.
- Slaughter, Matthew J. (2001). International Trade and Labor-Demand Elasticities. *Journal of International Economics* 54 (1): 27-56.
- Tamura, Robert (2001). Teachers, Growth, and Convergence. *Journal of Political Economy* 109 (5): 1021-1059.
- Tiebout, Charles M. (1956). A Pure Theory of Local Expenditures. *Journal of Political Economy* 64 (5): 416-424.
- Viadero, Debra (1998). Small Classes: Popular, But Still Unproven. *Education Week* 17 (February 18).
- Wooldridge, Jeffrey M. (2001). Asymptotic Properties of Weighted *M*-Estimators for Standard Stratified Samples. *Econometric Theory* 17 (2): 451-470.
- Wößmann Ludger, Martin R. West (2002). Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS. Harvard University, Program on Education Policy and Governance Research Paper PEPG/02-02.

Table 1: Descriptive Statistics: Sample Size, Student Performance, and Class Size

(1)-(4): Absolute numbers. – (5)-(8): Weighted means; standard deviations in parentheses. – (9): Coefficient of a regression of actual on grade-average class size without a constant, including a subject dummy; robust standard errors in parentheses.

	(1) Obs.	(2) Sample Size Students	(3) Classes	(4) Schools	(5) Test Score	(6) Actual Class Size	(7) Grade-Average Class Size	(8) Between-Grade Diff. in Avg. C.S.	(9) Actual on Average C. S., No Constant
Australia	16824	9963	426	123	519.566 (100.484)	27.230 (4.373)	27.005 (3.074)	-0.024 (2.359)	1.021* (0.006)
Belgium (Fl)	6772	4483	232	111	556.731 (82.059)	20.821 (4.316)	20.333 (3.727)	-0.550 (2.603)	0.973* (0.025)
Belgium (Fr)	5856	3500	184	87	493.025 (87.620)	21.088 (3.742)	20.533 (2.785)	0.540 (2.486)	1.015* (0.016)
Canada	12179	9754	433	204	518.372 (86.967)	27.885 (9.612)	27.735 (3.660)	0.230 (4.375)	1.019* (0.052)
Czech Rep.	7196	5234	230	115	549.875 (90.560)	25.782 (3.650)	25.782 (3.650)	0.191 (4.163)	1.002* (0.011)
France	7288	4463	186	93	498.847 (81.844)	25.316 (3.581)	25.660 (2.542)	-0.667 (2.833)	0.973* (0.011)
Greece	11526	6676	262	131	465.942 (89.392)	27.974 (8.109)	28.473 (8.176)	-0.076 (4.174)	0.891* (0.062)
Hong Kong	8360	5323	134	67	545.625 (100.583)	39.808 (4.3630)	40.407 (3.129)	-0.716 (1.139)	0.999* (0.006)
Iceland	3120	1948	149	74	467.746 (74.611)	20.336 (5.903)	20.215 (5.909)	-0.149 (4.791)	0.986* (0.019)
Japan	20209	10142	298	149	570.349 (97.421)	36.557 (4.036)	36.322 (4.596)	1.315 (4.776)	0.996* (0.005)
Korea	9731	5634	290	145	572.894 (103.726)	52.949 (20.684)	50.260 (4.717)	0.179 (2.426)	0.968* (0.016)
Portugal	10961	5989	252	126	446.210 (71.280)	25.147 (4.513)	25.735 (3.901)	-0.413 (4.146)	0.962* (0.008)
Romania	7270	4793	182	91	475.180 (95.685)	28.129 (5.861)	27.400 (5.519)	-0.780 (4.085)	0.989* (0.016)
Scotland	6766	4112	182	90	485.308 (92.311)	23.961 (5.250)	26.444 (3.165)	-0.011 (1.857)	0.817* (0.010)
Singapore	15931	8223	272	136	600.205 (100.395)	33.321 (6.966)	32.489 (6.218)	9.007 (5.999)	1.016* (0.007)
Slovenia	7667	4350	190	95	530.877 (89.451)	24.455 (3.9423)	24.272 (4.189)	-1.579 (4.073)	0.984* (0.0130)
Spain	8528	5060	199	98	482.795 (78.785)	29.390 (9.068)	28.519 (6.853)	0.959 (6.156)	1.016* (0.020)
United States	11018	6945	332	113	507.131 (100.312)	27.808 (15.766)	25.880 (4.304)	-0.487 (3.412)	1.048* (0.034)

* 1 percent significance level (based on clustering-robust standard errors).

Table 2: The Coefficient on Class Size in the Three Base Regressions

Estimates of the coefficient on class size. Dependent variable: TIMSS test score in mathematics or science. – (LS): Controlling for grade and subject. – (FE): Controlling for grade, subject, and school fixed effects. – (FE-IV): Controlling for grade, subject, school fixed effects, and 12 student and family-background variables, with class size instrumented. – Clustering-robust standard errors in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Observations	LS ^a		FE ^a		FE-IV ^a	
		Coefficient (α_{LS})	Standard Error	Coefficient (α_{FE})	Standard Error	Coefficient (α_{IV})	Standard Error
Australia	16824	4.927*	(0.730)	3.098*	(0.559)	-1.741	(4.867)
Belgium (Fl)	6772	2.422*	(1.372)	0.633	(1.036)	3.924	(2.611)
Belgium (Fr)	5856	0.143	(1.186)	-1.058†	(0.860)	0.156	(0.859)
Canada	12179	0.297	(0.249)	0.247	(0.227)	0.052	(0.616)
Czech Rep.	7196	2.222†	(0.983)	-1.545	(1.407)	0.512	(1.274)
France	7288	2.532*	(0.665)	0.659	(0.530)	-1.380	(1.028)
Greece	11526	0.468†	(0.210)	-0.013	(0.121)	-2.011†	(0.943)
Hong Kong	8360	5.587*	(1.028)	2.982*	(0.543)	-8.043	(8.178)
Iceland	3120	-0.306	(0.369)	-0.169	(0.501)	-2.197†	(0.877)
Japan	20209	3.264*	(0.762)	-0.356	(0.294)	-0.101	(0.394)
Korea	9731	-0.057	(0.075)	-0.137*	(0.046)	-0.796	(0.570)
Portugal	10961	1.082*	(0.296)	0.931*	(0.298)	0.447	(0.556)
Romania	7270	2.854*	(0.654)	0.294	(0.587)	1.391	(1.731)
Scotland	6766	0.333	(0.435)	0.073	(0.285)	-22.741	(34.651)
Singapore	15931	5.499*	(0.511)	3.079*	(0.420)	0.492	(0.489)
Slovenia	7667	0.963	(0.671)	0.113	(0.572)	0.769	(1.051)
Spain	8528	0.450*	(0.174)	0.036	(0.144)	-0.533	(0.427)
United States	11018	-0.140	(0.125)	-0.000	(0.092)	3.345	(5.515)

^a See text for a complete explanation of the different estimation methods.

Significance levels (based on clustering-robust standard errors): * 1 percent. — † 5 percent. — ‡ 10 percent.

Table 3: Sorting Effects in 18 Countries

Standard errors from Hausman specification tests in parentheses.

	(1)	(3)		(5)		(7)	
	Observations	Between-school sorting ^a		Within-school sorting ^a		Total sorting ^a	
		Estimate (β_b)	Standard Error	Estimate (β_w)	Standard Error	Estimate (β_t)	Standard Error
Australia	16824	1.829*	(0.469)	4.839	(4.835)	6.668	(4.812)
Belgium (Fl)	6772	1.789 [†]	(0.899)	-3.291	(2.397)	-1.502	(2.222)
Belgium (Fr)	5856	1.201	(0.817)	-1.214*	(0.040)	-0.013	(0.818)
Canada	12179	0.051	(0.103)	0.195	(0.572)	0.246	(0.563)
Czech Rep.	7196	3.767*	(1.007)	-2.056*	(0.598)	1.711 [†]	(0.810)
France	7288	1.873*	(0.401)	2.039 [†]	(0.880)	3.912*	(0.784)
Greece	11526	0.481*	(0.172)	1.998 [†]	(0.936)	2.479*	(0.920)
Hong Kong	8360	2.604*	(0.873)	11.025	(8.160)	13.629	(8.113)
Iceland	3120	-0.137	(0.340)	2.029*	(0.719)	1.892 [†]	(0.795)
Japan	20209	3.620*	(0.703)	-0.255	(0.263)	3.365*	(0.652)
Korea	9731	0.080	(0.059)	0.659	(0.568)	0.739	(0.565)
Portugal	10961	0.151*	(0.032)	0.485	(0.469)	0.636	(0.470)
Romania	7270	2.560*	(0.290)	-1.097	(1.629)	1.463	(1.603)
Scotland	6766	0.406	(0.328)	22.667	(34.650)	23.073	(34.649)
Singapore	15931	2.420*	(0.290)	2.587*	(0.250)	5.007*	(0.147)
Slovenia	7667	0.849 [†]	(0.351)	-0.656	(0.881)	0.193	(0.808)
Spain	8528	0.414*	(0.097)	0.569	(0.402)	0.983 [†]	(0.390)
United States	11018	-0.140 [‡]	(0.084)	-3.345	(5.514)	-3.484	(5.514)

^a See text for an explanation of the methods used to calculate the sorting effects.Significance levels: * 1 percent. —[†] 5 percent. —[‡] 10 percent.

Table 4: Hypotheses on Main Determinants of Sorting

	Regressive	Compensatory
Between-school sorting	Parental residential choice (combined with local finance)	Policymakers influenced by the political economy of tracking
Within-school sorting	Teachers govern student placement	Administrators govern student placement (especially when held accountable for efficiency and equity)

Table 5: Determinants of Between-School Sorting Effects

Least-squares regressions. – Standard errors in parentheses. – Dependent variable: Between-school sorting effect (β_b – see Table 3). – All observations are weighted to account for estimated dependent variable. – See text for specification details.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Performance-based admissions	0.099 [‡] (0.046)				0.184* (0.018)	0.217 [†] (0.050)	0.214* (0.033)
Tracking		1.180 [‡] (0.600)					
Transfer rate			-0.178 [‡] (0.096)		-0.399* (0.089)		
Local finance				-0.012 (0.009)		-0.037 [†] (0.009)	
Transfer rate*local finance							-0.003* (0.001)
Constant	0.291 (0.444)	0.391 (0.501)	1.625* (0.398)	1.331 [†] (0.459)	0.793* (0.210)	-0.374 (0.225)	-0.383 [°] (0.174)
Observations	10	18	18	11	10	6	6
F	4.6	3.9	3.5	1.8	51.9	10.0	22.3
Probability > F	(0.064)	(0.067)	(0.082)	(0.211)	(0.000)	(0.047)	(0.016)
R ²	0.367	0.195	0.177	0.168	0.937	0.870	0.937
Adjusted R ²	0.288	0.145	0.126	0.076	0.919	0.783	0.895

Significance levels: * 1 percent. —[†] 5 percent. —[‡] 10 percent.

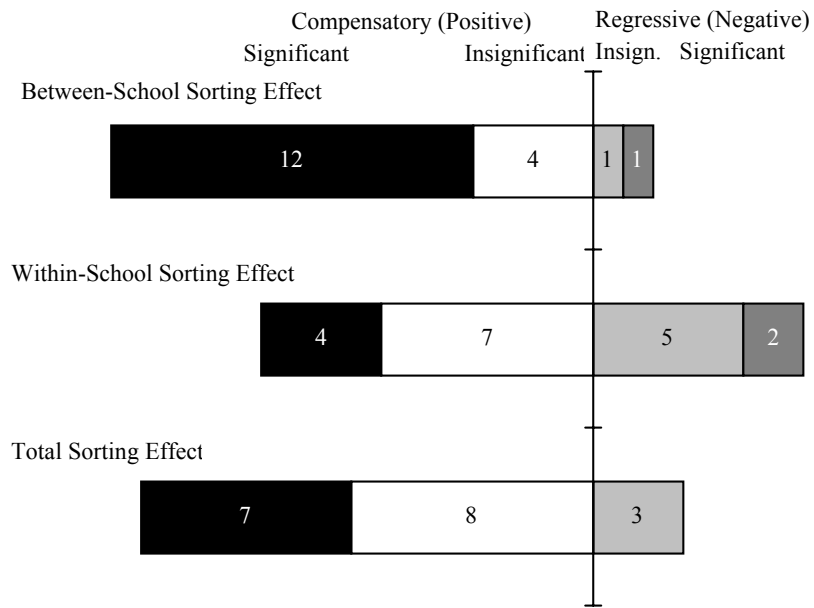
Table 6: Determinants of Within-School Sorting Effects

Least-squares regressions. – Standard errors in parentheses. – Dependent variable: Within-school sorting effect (β_s – see Table 3). – All observations are weighted to account for estimated dependent variable. – See text for specification details.

	(1)	(2)	(3)	(4)	(5)	(6)
Teachers place students	-0.053 [°] (0.033)			-0.053 [°] (0.033)	-0.030 (0.030)	
Department heads place students		0.040 [†] (0.018)		0.035 [†] (0.018)		
Department heads assign teachers			0.115 [†] (0.039)		0.101 [†] (0.041)	
External exams						0.051 [†] (0.022)
Constant	1.348 [‡] (0.708)	-0.275 (0.502)	-0.290 (0.405)	0.722 (0.793)	0.338 (0.748)	-0.62 (0.422)
Observations	17	17	17	17	17	15
F	2.7	4.8	8.6	3.94	4.9	5.6
Probability > F	(0.124)	(0.044)	(0.010)	(0.044)	(0.025)	(0.034)
R ²	0.151	0.243	0.365	0.360	0.410	0.302
Adjusted R ²	0.094	0.193	0.323	0.269	0.326	0.248

Significance levels: * 1 percent. — † 5 percent. — ‡ 10 percent. — ° 15 percent.

Figure 1: Sorting Effects across 18 School Systems^a



^a Number of systems showing a statistically significant positive (black), a statistically insignificant positive (white), a statistically insignificant negative (light gray), and a statistically significant negative (dark gray) estimate, respectively. — See text for details on the identification methods.

Table A1: Descriptive Statistics: Student Background Data

(1): Absolute number. – (2)-(14): Weighted means; standard deviations in parentheses.

	(1) Observations	(2) Upper Grade	(3) Female	(4) Age	(5) Born in Country	(6) Living with both Parents	(7) Some Secondary	(8) Parents' Education Finished Secondary	(9) Some after Sec.	(10) Finished University	(11) 11-25	(12) Books at Home 26-100	(13) 101-200	(14) More than 200
Australia	16824	0.487 (0.500)	0.514 (0.500)	13.725 (0.692)	0.891 (0.312)	0.834 (0.372)	0.209 (0.407)	0.227 (0.419)	0.267 (0.443)	0.284 (0.451)	0.060 (0.237)	0.247 (0.431)	0.263 (0.440)	0.408 (0.491)
Belgium (Fl)	6772	0.528 (0.499)	0.525 (0.499)	13.592 (0.721)	0.971 (0.167)	0.921 (0.270)	0.138 (0.345)	0.366 (0.482)	0.235 (0.424)	0.212 (0.409)	0.158 (0.365)	0.355 (0.479)	0.191 (0.393)	0.216 (0.412)
Belgium (Fr)	5856	0.528 (0.499)	0.570 (0.495)	13.710 (0.881)	0.914 (0.281)	0.858 (0.349)	0.066 (0.248)	0.182 (0.386)	0.392 (0.488)	0.318 (0.466)	0.100 (0.300)	0.274 (0.446)	0.218 (0.413)	0.350 (0.477)
Canada	12179	0.495 (0.500)	0.497 (0.500)	13.653 (0.763)	0.919 (0.272)	0.817 (0.387)	0.089 (0.285)	0.188 (0.391)	0.339 (0.473)	0.341 (0.474)	0.104 (0.306)	0.279 (0.449)	0.253 (0.435)	0.323 (0.467)
Czech Rep.	7196	0.500 (0.500)	0.498 (0.500)	13.886 (0.650)	0.988 (0.111)	0.913 (0.281)	0.202 (0.401)	0.355 (0.479)	0.207 (0.405)	0.211 (0.408)	0.042 (0.201)	0.321 (0.467)	0.301 (0.459)	0.331 (0.471)
France	7288	0.484 (0.500)	0.493 (0.500)	13.747 (0.837)	- (-)	0.873 (0.333)	0.165 (0.371)	0.414 (0.493)	0.259 (0.438)	0.114 (0.318)	0.169 (0.375)	0.377 (0.485)	0.201 (0.401)	0.209 (0.406)
Greece	11526	0.483 (0.500)	0.482 (0.500)	13.111 (0.734)	0.940 (0.237)	0.875 (0.331)	0.202 (0.401)	0.213 (0.410)	0.226 (0.418)	0.154 (0.361)	0.227 (0.419)	0.429 (0.495)	0.175 (0.380)	0.116 (0.320)
Hong Kong	8360	0.506 (0.500)	0.452 (0.498)	13.686 (0.873)	0.880 (0.325)	0.913 (0.282)	0.424 (0.494)	0.285 (0.451)	0.056 (0.230)	0.071 (0.257)	0.288 (0.453)	0.302 (0.459)	0.101 (0.302)	0.106 (0.308)
Iceland	3120	0.504 (0.500)	0.483 (0.500)	13.145 (0.587)	0.952 (0.213)	0.895 (0.307)	0.076 (0.265)	0.125 (0.331)	0.527 (0.499)	0.196 (0.397)	0.062 (0.240)	0.329 (0.470)	0.280 (0.449)	0.322 (0.467)
Japan	20209	0.512 (0.500)	0.483 (0.500)	13.902 (0.575)	- (-)	- (-)	- (-)	- (-)	- (-)	- (-)	- (-)	- (-)	- (-)	- (-)
Korea	9731	0.504 (0.500)	0.431 (0.495)	13.712 (0.607)	0.991 (0.094)	0.883 (0.321)	0.172 (0.377)	0.418 (0.493)	0.111 (0.314)	0.230 (0.421)	0.107 (0.310)	0.333 (0.471)	0.243 (0.429)	0.231 (0.422)
Portugal	10961	0.483 (0.500)	0.509 (0.500)	13.986 (1.120)	0.916 (0.277)	0.895 (0.306)	0.241 (0.428)	0.096 (0.295)	0.058 (0.233)	0.073 (0.260)	0.268 (0.443)	0.319 (0.466)	0.135 (0.342)	0.159 (0.365)
Romania	7270	0.508 (0.500)	0.499 (0.500)	14.157 (0.704)	0.955 (0.207)	0.755 (0.430)	0.248 (0.432)	0.288 (0.453)	0.324 (0.468)	0.096 (0.295)	0.208 (0.406)	0.212 (0.409)	0.129 (0.335)	0.259 (0.438)
Scotland	6766	0.513 (0.500)	0.494 (0.500)	13.216 (0.600)	0.923 (0.267)	0.842 (0.365)	0.148 (0.355)	0.383 (0.486)	0.352 (0.478)	0.117 (0.322)	0.165 (0.372)	0.308 (0.462)	0.193 (0.395)	0.222 (0.416)
Singapore	15931	0.503 (0.500)	0.489 (0.500)	13.938 (0.832)	0.922 (0.269)	1.000 (0.000)	0.002 (0.044)	0.563 (0.496)	0.135 (0.342)	0.073 (0.260)	0.219 (0.413)	0.409 (0.492)	0.145 (0.352)	0.121 (0.326)
Slovenia	7667	0.481 (0.500)	0.514 (0.500)	14.276 (0.636)	0.966 (0.182)	0.912 (0.283)	0.076 (0.265)	0.339 (0.473)	0.306 (0.461)	0.170 (0.376)	0.177 (0.382)	0.388 (0.487)	0.206 (0.405)	0.201 (0.400)
Spain	8528	0.499 (0.500)	0.495 (0.500)	13.737 (0.853)	0.973 (0.161)	0.907 (0.291)	0.221 (0.415)	0.124 (0.330)	0.125 (0.330)	0.160 (0.367)	0.179 (0.383)	0.330 (0.470)	0.196 (0.397)	0.243 (0.429)
United States	11018	0.497 (0.500)	0.508 (0.500)	13.727 (0.718)	0.932 (0.252)	0.805 (0.396)	0.055 (0.227)	0.178 (0.383)	0.409 (0.492)	0.346 (0.476)	0.120 (0.325)	0.284 (0.451)	0.213 (0.409)	0.305 (0.460)

Table A2: Descriptive Statistics: Institutional Background Data

Country-level data. See text for variable definitions and information on sources.

	(1) Observations	(2) Mean	(3) Standard Deviation
Performance-based admissions	10	8.56	(6.66)
Tracking	18	0.72	(0.46)
Transfer rate	18	2.84	(2.56)
Local finance	11	30.27	(38.92)
Teachers place students	17	17.80	(12.51)
Department heads place students	17	17.04	(25.52)
Department heads assign teachers	17	7.71	(13.82)
External exams	15	17.04	(24.83)