DISCUSSION PAPER SERIES

IZA DP No. 6161

The Measurement of Educational Inequality: Achievement and Opportunity

Francisco H. G. Ferreira
Jérémie Gignoux

November 2011

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

The Measurement of Educational Inequality: Achievement and Opportunity

Francisco H. G. Ferreira
World Bank and IZA

Jérémie Gignoux
Paris School of Economics

Discussion Paper No. 6161
November 2011

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 6161
November 2011

ABSTRACT

The Measurement of Educational Inequality: Achievement and Opportunity*

This paper proposes two related measures of educational inequality: one for educational achievement and another for educational opportunity. The former is the simple variance (or standard deviation) of test scores. Its selection is informed by consideration of two measurement issues that have typically been overlooked in the literature: the implications of the standardization of test scores for inequality indices, and the possible sample selection biases arising from the Program of International Student Assessment (PISA) sampling frame. The measure of inequality of educational opportunity is given by the share of the variance in test scores that is explained by pre-determined circumstances. Both measures are computed for the 57 countries in which PISA surveys were conducted in 2006. Inequality of opportunity accounts for up to 35 percent of all disparities in educational achievement. It is greater in (most of) continental Europe and Latin America than in Asia, Scandinavia, and North America. It is uncorrelated with average educational achievement and only weakly negatively correlated with per capita gross domestic product. It correlates negatively with the share of spending in primary schooling, and positively with tracking in secondary schools.

JEL Classification: D39, D63, I29, O54

Keywords: educational inequality, educational achievement, inequality of opportunity

Corresponding author:

Francisco H. G. Ferreira
The World Bank
1818 H Street NW
Washington, DC 20433
USA
E-mail: fferreira@worldbank.org

* We are grateful to Gordon Anderson, Markus Jäntti, Maria Ana Lugo, John Micklewright, Alain Trannoy and participants at conferences and seminars in Barcelona, Buenos Aires, Oxford and St. Gallen for helpful comments on earlier drafts. We are solely responsible for any remaining errors. The views expressed in this paper are those of the authors, and should not be attributed to the World Bank, its Executive Directors, or the countries they represent.

1. Introduction

Educational inequalities have long been a matter of significant policy concern, in both developed and developing countries. Some view educational achievement as a dimension of well-being in its own right, or at least as a fundamental input into a person's functionings and capacity to flourish (Sen, 1985). Education is also a powerful predictor of earnings, as we have known since the early days of work on human capital. More recent research has also found that inequality in educational achievement and earnings inequality are correlated, both over time within the United States and across countries (see, e.g., Blau and Kahn, 2005; and Bedard and Ferrall, 2003). Education is also correlated with health status, and in some cases with political participation in the democratic process, so that inequalities in the former may translate into undesirable gaps and gradients in other dimensions as well.

For all of these reasons, people care about the distribution of education. Those concerned about fairness and social justice care also about the distribution of opportunities for acquiring a good education and, in particular, about the degree to which family background and other pre-determined personal characteristics determine a person's educational outcomes. Nevertheless, there is much less agreement on how those concepts – inequality in educational outcomes, and inequality of opportunity to a good education – should be measured. Constrained by data availability, early work comparing inequality in education across countries focused on educational attainment: the number of years of schooling a person had completed or, in some cases, broader 'levels' of education, such as primary, secondary, or higher. Thomas, Wang and Fan (2001) compiled a set of Gini coefficients for years of schooling for 85 countries, over the period from 1960 to 1990. Castelló and Doménech (2002) and Morrisson and Murtin (2007) also examine inequality in years of schooling across a large number of countries.

Interesting though those comparisons were, there is widespread agreement that a year of schooling is a problematic unit with which to measure "education". Does a student learn the same amount in 6th grade in Zambia as in Finland? Is the value of one year of schooling the same even across different schools in a single country or city? The growing availability of data on student performance in comparable tests has confirmed what one already suspected: that the answer to these questions is generally 'no'. The quality – and hence the ultimate value – of education varies considerably, both within and across countries.

Over the last decade, different projects have compiled school-based surveys that administer identical cognitive achievement tests to samples of students across a number of countries, as well as collecting (reasonably) comparable information about the students' families and the schools they attend. The OECD's *Program of International Student Assessment* (PISA) and the International Association for the Evaluation of Educational Achievement's *Trends in International Mathematics and Science Study* (TIMSS) are perhaps the best known, but the

Progress in International Reading Literacy Study (PIRLS), which is applied to younger students, shares a number of common features.[1]

As anyone who has been to school may recall, performance in a test, while probably preferable to a simple indicator of enrollment or attendance, is not a perfect measure of learning either. For one thing, tests and test items (i.e. questions) vary in difficulty. The final result is known to measure scholastic ability or learning achievement only imperfectly. For this reason, all of the aforementioned surveys present scores constructed from the raw results by means of Item Response Theory (IRT) models, which attempt to account for "test parameters", so as to better infer true learning. This process generates an arbitrary metric for test scores, which are then typically standardized to some arbitrary mean and standard deviation.

Using these standardized test scores, a number of studies have attempted to provide international comparisons of educational inequality on the basis of *achievement*, rather than *attainment*. Micklewright and Schnepf (2007) and Brown et al. (2007) examine the robustness of measures of central tendency and dispersion in the distribution of student achievement obtained using different surveys, by comparing the measures and country rankings across them. They find broad agreement across surveys, but also some evidence that the specific statistical models used to estimate IRT adjustments do affect results, in particular for less developed countries. Marks (2005), Schultz, Ursprung and Wossmann (2008), and Macdonald et al. (2010) examine the question of intergenerational persistence in educational achievement, which is closely related to that of inequality of opportunity, and present cross-country comparisons of measures of the association between student achievement and certain family characteristics.

This paper seeks to contribute to that literature by proposing two simple and closely-related measures of inequality - one for educational achievement and another for opportunity to education – and reporting them for all countries that participated in the 2006 wave of PISA surveys. To measure inequality in achievement, we propose simply using the variance or the standard deviation of test scores. But we arrive at this simple proposal by considering the implications of two issues specific to the distribution of test scores for the measurement of inequality. These two issues are: (i) the fact that many common inequality indices are not ordinally invariant in the standardization to which IRT-adjusted test scores are generally subjected; and (ii) the fact that PISA student samples are likely to suffer from non-trivial selection biases in a number of countries. The choice of the variance (or the standard deviation) addresses the first issue. We also propose two alternative two-sample non-parametric procedures to assess the robustness of the inequality measure to the sample selection biases, and implement them in the four countries for which PISA sample coverage (as a share of the total population of 15 year-olds) is smallest.

The proposed measure of inequality of educational opportunity draws on the recent literature on inequality of opportunity in the income space, but is also adapted to the specificities of

[1] There is also an International Adult Literacy Survey (IALS), which is applied to adults long after they have left school.

educational data and the resulting choice of measure for inequality in achievement. It also utilizes information on student background more comprehensively than all previous studies we are aware of, and is additively decomposable both across circumstances and population subgroups. The measure is also isomorphic to (inverse) measures of educational mobility.

We report our measures of inequality in educational achievement and opportunity for the 57 countries that took part in the PISA 2006 exercise. Each measure was computed separately for each of the three tests applied by PISA: mathematics, reading and science. But there was a good measure of agreement between their rankings, and we often refer only to the math results in the text.[2] We find considerable variation in the standard deviation of test scores, from lows of around 80 (for Indonesia, Estonia and Finland) to highs near 110 (in Belgium and Israel).[3] Similarly stark variation exists in our measure of inequality of opportunity, from 0.10 – 0.15 for Macau (China), Australia, and Hong Kong SAR, China, up to 0.33 – 0.35 in Bulgaria, France and Germany. Inequality of opportunity is uncorrelated with mean achievement and only weakly (negatively) correlated with GDP per capita. Broadly speaking, it is higher in continental Europe (except for Italy) and Latin America than in Asia and Scandinavia, with the US and the UK in intermediate positions. It is negatively correlated with the share of public educational spending that accrues to primary schools, and positively correlated with the proportion of technical and vocational enrollment at the secondary level (a measure of "educational tracking").

The paper is organized as follows. Section 2 describes the data sets we use. Section 3 considers the implications of test score standardization and of the PISA sampling frame for the measurement of inequality in educational achievement, and reports the standard deviation in test scores for our sample of countries. Section 4 proposes our measure of inequality of educational opportunity (IOp), discusses some of its properties, and presents results. Section 5 applies the proposed measures by examining how they correlate with two educational policy indicators across countries. Section 6 concludes.

2. Data

Two broad kinds of data are used for the analysis in this paper. The first is the complete set of PISA surveys, for all 57 countries that participated in the 2006 round. The second is a group of four household surveys, for Brazil, Indonesia, Mexico and Turkey, which are used as ancillary surveys in the two-sample non-parametric sample selection correction procedures described in Section 3. We briefly describe each of these in turn.

[2] See Micklewright and Schnepf (2007) for a careful comparison of rankings from each of the PISA tests, as well as from TIMSS and PIRLS.

[3] But the low variance for Indonesia is a good example of the sensitivity of these measures to assumptions made about the nature of selection into the test-taking sample. Under our scenario of "extreme" selection on unobservables, the variance of math scores for Indonesia triples. See below.

The PISA 2006 data sets

The third round of the Program of International Student Assessment surveys was conducted in 57 countries between March and November, 2006. Two earlier rounds were collected in 2000/2002 (in 43 countries), and in 2003 (in 41 countries). A fourth round has since been collected in 2009. Most OECD countries were surveyed, as were a number of developing countries in Asia, Latin America, North Africa and the Middle East. Sample sizes range from 339 in Liechtenstein to 30,971 in Mexico. Table 1 lists all participating countries in the 2006 round, as well as their sample sizes.

In each country, fifteen year-olds enrolled in any educational institution, and attending grade 7 or higher, were sampled. All children surveyed took three tests: in reading, mathematics, and science.[4] Their performance in these tests forms the basis for the assessment of their learning or cognitive achievement. Yet, educationalists seem agreed that raw, unadjusted test scores are of little value. Test questions (or 'items') vary in their degree of difficulty, and simply adding up correct answers, or weighing them arbitrarily, does not correctly measure the latent variable of interest – cognitive achievement. Instead, the educational community in charge of international tests such as PISA, TIMSS, PIRLS and IALS processes raw scores through statistical techniques known as Item Response Theory (IRT). See Baker (2001) for a general introduction, and OECD (2006) for a description of how the method is applied to PISA surveys. In essence, an item response model consists of an equation of the form:

$$p(s|\theta,\alpha) \tag{1}$$

Equation (1) gives the probability of scoring s in a given test, conditional on individual latent cognitive ability θ and test item parameters α (such as their difficulty). Given an additional assumption about the distribution of latent ability in the population (usually a normal law such as $\theta \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$) and an observed distribution of raw scores, *F(s),* the IRT model can be used to back up a distribution of the latent variable θ.[5]

This process involves a number of functional form assumptions which are not innocuous. Brown et al. (2007) have shown, for instance, that the final distribution of test scores can be sensitive to differences in the specification of the model used to estimate equation (1). Here, however, we are concerned with the standardization that happens *after* the IRT adjustment. Once that procedure is complete, and a new distribution of 'adjusted' test scores (which we denote by *x*) has been generated, this latter variable is standardized, according to a simple formula such as:

$$y_{ij} = \hat{\mu} + \frac{\hat{\sigma}}{\sigma}\left(x_{ij} - \mu\right) \tag{2}$$

In equation (2), x_{ij} denotes the (post-IRT, pre-standardized) test score for individual *i* in country *j*. μ and σ denote their original mean and standard deviation across all countries in the sample

[4] The data for achievements in Reading for the United States were not issued after a problem occurred during the field operations in that country.

[5] See Mislevy (1991) and Mislevy et al. (1992) for a more detailed discussion.

(the world, or the OECD, for example). $\hat{\mu}$ ($\hat{\sigma}$) is the new arbitrary mean (standard deviation) for the standardized distribution. In the PISA procedure, it has a value of 500 (100). It is the distributions of y_{ij} that are used in computing means and inequality indicators for each country j in the PISA data set. As we will see in the next section, the operation described by equation (2), even if the IRT procedure that precedes it is taken as given, poses serious issues for inequality measurement.

In addition to standardized test scores, the PISA data set contains information on a number of individual, family and school characteristics for each test-taker. The presence of these covariates accounts for a large part of the interest of the research community on the PISA data. For the analysis of inequality of opportunity in education, we focus on a subset of these covariates that are informative of the family background and other inherited circumstances of the child. Ten such variables are used: gender, father's and mother's education, father's occupation, language spoken at home, migration status, access to books at home, durables owned by the households, cultural items owned, and the location of the school attended (used as an indicator or a rural or urban upbringing).[6]

Parental education is measured by the highest level completed and is coded using ISCED codes into four categories: a) no education or unknown level; b) primary education (ISCED level 1); c) lower secondary education (ISCED level 2), upper secondary (ISCED level 3), or post-secondary non-tertiary education (ISCED level 4); and d) college education (ISCED level 5)). Father's occupation is classified using ISCO codes. We aggregate occupations into three broad categories: a) legislators, senior officials and professionals, technicians and clerks; b) service workers, craft and related trades workers, plant or machine operators and assemblers, and unoccupied individuals; and c) skilled agricultural and fishery workers, elementary occupations or unknown occupation. The variable for language spoken at home is a dummy identifying a language other than the language of the test. The migration status variable is a dummy identifying a first or second generation migrant as an individual who was, or whose parents were, born in a foreign country.

The number of books at home variable, an indicator of parental human capital, is a categorical variable coded into four categories: a) 0 to 10 books; b) 11 to 25 books; c) 26 to 100 books; and d) more than 100 books. Ownership of durables, an indicator of family wealth, is captured by six dummy variables indicating the ownership of a) a dishwasher; b) a DVD or a VCR player; c) a cell phone; d) a television; e) a computer; f) a car. Ownership of cultural possessions is captured by three dummy variables indicating the ownership of a) books of literature; b) books of poetry; and c) works of arts (paintings are mentioned as an example of such works in the formulation of the question). School location is a proxy for the person's inherited spatial endowment and we recode it using three categories: a) villages or small towns (less than 15,000 inhabitants); b) towns (between 15,000 and 100,000 inhabitants); and c) cities (larger than 100,000 inhabitants).

[6] School-level variables are not used in this analysis deliberately, for reasons which should become clear in Section 4.

School location information was not collected in France; Hong Kong SAR, China; and Liechtenstein.

A final data issue worth highlighting is that of sample coverage and representativeness. PISA samples were designed to be representative of the population of 15 year-olds who are enrolled in grade 7 or higher in any educational institution. The samples are *not,* therefore, representative of the total population of 15 year-olds in each country: children who dropped out of school before they turned fifteen, as well as those who are so delayed that they are in grade 6 or lower at age fifteen, are purposively excluded. In addition, sampling flaws induce an additional under-coverage of enrolled 15 year olds. PISA documentation suggests that this arises from the fact that their sampling frame (a listing of schools and sampling weights) is established in the year preceding the surveys, on the basis of current school enrollment on that year. But some schools close down between the two years, and new ones are not included in the sample. Changes in the enrollment of 15 year-olds arising from this process are not taken into account.

The PISA sample coverage rate, defined as the ratio of the covered student population (using PISA expansion factors) to the total population of 15 year-olds, varies considerably across countries, and is reported in column 2 of Table 1. Although coverage is typically high in OECD countries, it is low in many developing ones: coverage rates are as low as 47% for Turkey, 53% for Indonesia, 54% for Mexico, and 55% for Brazil. Overall, coverage is less than 80% of the total population of 15 years-olds in fifteen countries. Table 2 provides a sense of the sources of exclusion for the four countries in our dataset with the lowest coverage rates, by decomposing those selected out of the sample into children no longer in school, children with excessive delays, and those missed due to PISA sampling issues. It should be obvious from these magnitudes that any international comparison of countries with vastly different coverage rates must seek to address the problem in some way, and we suggest two alternatives in Section 3.

Ancillary household survey data sets

Our proposed procedure to examine the sensitivity of inequality measures to sample selection, which is described below, relies on using information on fifteen year-olds from general-purpose household surveys. While these surveys may have their own sampling issues, these are not dictated by school enrollment or delay status, or by school closures, openings and reforms. We obtained such household surveys for the four countries with the lowest coverage rates in the 2006 PISA sample: those reported in Table 2. For Brazil, we used the Pesquisa Nacional por Amostra de Domicílios (PNAD) 2006. For Indonesia, we used the SUSENAS 2005. For Mexico, the Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) for 2006 was used. For Turkey, the Household Budget Survey (HBS) 2006 was used.

All four are large-sample household surveys with national coverage and representative down to the regional level, which are fielded on an annual basis by each country's national statistical authority. The PNAD 2006 collected information from a sample of about 119,000 households and 410,000 individuals; SUSENAS 2005 from 257,900 households and 1,052,100 individuals; the

ENIGH 2006 from 20,900 households and 83,600 individuals; and the HBS 2006 from 8,600 households and 34,900 individuals. We restrict the samples to children aged 15, for which we have 7,626 observations in the PNAD 2006; 22,600 in the SUSENAS 2005; 1,921 in the ENIGH 2006; and 683 in the HBS 2006. Although some children in boarding schools and other institutions are likely to be out of the sample frame, those samples should otherwise be representative for the total population of 15 year-olds.

In these four countries, these are the staple surveys for assessing the distribution of household income and, in some cases, consumption expenditures. But they also collect information on other topics, including labor supply, education and migration. We use information on parents' characteristics for estimating the total population of 15 year-olds in groups defined by similar gender, mother's education and father's occupation. The classification of the family background variable can be made comparable with the ones in the PISA by appropriate aggregation of coding categories. Parental characteristics are missing for orphans, children who do not live with their parents, or whose parents did not report their education. For instance, the information on mother's education is missing for about 15.0% of 15 year-olds in the PNAD 2006, 8.7% in the SUSENAS 2005, 11.9% in the ENIGH 2006, and 3.8% in the HBS 2006. When comparing the two surveyed populations, children with missing parental background information in the household surveys are not dropped, but associated with those with the same information missing in the PISA survey.

3. Measuring Inequality in Educational Achievement

Measures of inequality in educational achievement are based on distributions of standardized test scores (y_{ij}), constructed from the IRT-adjusted scores (x_{ij}) by means of a transformation such as equation (2). In the case of PISA, the transformation is given by (2) exactly, with $\hat{\mu} = 500$, and $\hat{\sigma} = 100$. That operation involves both a translation of the original distribution (by the difference between the new arbitrary mean and the original mean, re-scaled) and a rescaling (by the ratio of the new to the original standard deviations).

In the field of inequality measurement it is usual to impose axioms, or desirable properties, that individual indices should respect. Three common such axioms are:

(i) *symmetry*: which requires that the measure be insensitive to any permutation of the *y* vector;

(ii) *continuity* in any individual income;

(iii) and the *transfer principle*: which requires that the measure should rise (strong axiom) or at least not fall (weak axiom) as a result of any sequence of mean-preserving spreads.

In addition, inequality indices often satisfy *either* one of two invariance axioms:

(iv-a): *scale invariance*: which requires that the index be insensitive to any re-scaling of the *y* vector: $I(y) = I(\lambda y), \lambda > 0$, where *y* is the vector of interest, and λ is a positive scalar.

(iv-b): *translation invariance*: which requires that the index be insensitive to a translation of the y vector: $I(y) = I(y + a), a \neq 0$, where *a* is a non-zero constant vector of the same dimension as *y*.

An important result, due to Zheng (1994), is that no inequality index that satisfies axioms (i)-(iii) – known as "meaningful" inequality measures - satisfies *both* (iv-a) and (iv-b). This impossibility result, in other words, states that no meaningful inequality index can be both scale- and translation invariant. A direct implication of Zheng's result for the measurement of inequality of educational achievement using standardized data is stated below as our Remark 1:

Remark 1: No meaningful inequality index yields a cardinally identical measure for the pre- and post-standardization distributions of the same test scores.

Note that the remark derives from the standardization procedure (equation 2), rather than from the much more complex item response theory adjustments. It refers, therefore, to the measurement of inequality *in* IRT-adjusted test scores, and not to a comparison between adjusted and unadjusted scores. For the same reason, it is additional to and unconnected with any concerns about the sensitivity of summary statistics to changes in the IRT model specification, such as those discussed by Brown et al. (2007) with respect to the number of parameters used to estimate equation (1).

How important is Remark 1? Clearly this depends on whether or not inequality indices applied to pre- and post-standardization distributions are ordinally equivalent – that is to say, whether they *rank* distributions in precisely the same way, regardless of cardinal differences in value. After all, standardization is just a change in metric. The (post-standardization) mean score in each country *j*, for example is simply:

$$\mu_j^y = \hat{\mu} + \frac{\hat{\sigma}}{\sigma}\left(\mu_j^x - \mu\right) \qquad (3)$$

Where μ_j^x is the pre-standardization mean in country j, and other notation is as in equation (2). Since every other term in (3) is a constant, μ_j^y and μ_j^x are ordinally equivalent. One is a monotonic (and in this case, affine) transformation of the other. Country ranks based on either would be identical. The only effect of standardization on country mean scores is a change in metric. Since this was the point of the process in the first place, there seems to be no cause for concern.

The same is true for percentile-based measures of dispersion, such as the inter-quartile ratio, or the absolute difference P95-P5 used by Micklewright and Schnepf (2007) to compare dispersion across 21 countries and three different surveys. Equation (2) is itself a monotonic, and therefore rank-preserving, transformation. Since each score y_i occupies precisely the same rank in its distribution as the original score x_i did in its distribution, rank- or percentile-based measures – be they ratios or differences, will be cardinally different, but ordinally equivalent.

Yet this is not true of inequality measures in general. The post-standardization Gini coefficient in country *j* (G_j^y) for example, can be straight-forwardly shown to relate to the pre-standardization Gini (G_j^x) as follows:

$$G_j^y = \frac{\mu_j^x \hat{\sigma}}{\mu_j^y \sigma} G_j^x \qquad (4)$$

Unlike in equation (3), the terms multiplying G_j^x are not all constants. In particular, the post-standardization Gini is a function of the ratio of pre- to post- standardization means, which is an increasing function of μ_j^x (see equation 3). The existence of a second argument in (4) implies that the post-standardization Gini coefficient is *not* ordinally equivalent to its pre-standardization analogue.

Most other common meaningful inequality measures do not share the linearity of the Gini, so their post- and pre-standardization formulae cannot be related as straightforwardly. Nevertheless, substitution of equations (2) and (3) into the formulae for the Generalized Entropy or the Kolm-Atkinson classes of inequality measures yield expressions that are functions of both the central distance indicators of the measure in question, and of the ratio of pre- to post-standardization means ($\frac{\mu_j^x}{\mu_j^y}$). For the Generalized Entropy (GE) class, for example:

$$GE_j^y = \frac{1}{\alpha^2 - \alpha}\left[\frac{1}{n_j}\sum_{i \epsilon j}\left(\frac{\hat{\mu} + \hat{\sigma}/\sigma(x_{ij} - \mu)}{\mu_j^x}\right)^{\alpha}\left(\frac{\mu_j^x}{\mu_j^y}\right)^{\alpha} - 1\right] \qquad (5)$$

These results give rise to our second remark:

Remark 2: A number of well-known inequality indices are not even ordinally equivalent when applied to pre- and post-standardization distributions.

Ordinal equivalence with respect to standardization is clearly a desirable property for an index used for measuring inequality in educational achievement. The standardization operation given by (2) is meant merely to adjust an arbitrary metric. It is not intended to fundamentally alter our judgment of how countries compare with one another in substantive terms. Yet, when indices such as the Gini or Theil index are applied to these standardized distributions, we cannot be confident that the original rank in post-IRT adjusted inequality is preserved.[7]

What then are the options for those interested in the distribution of educational achievement? One could, of course, rely on rank-based measures such as the inter-quartile range or percentile differences which, as noted above, are ordinally equivalent. However, these measures do not satisfy the transfer principle: a progressive transfer (from above) to the income recipient on the 95th percentile will, for example, cause the p95-p05 measure to indicate an increase in inequality. And of course, because such indices are insensitive by construction to any chances in incomes that do not affect those on the percentiles of reference, they also violate continuity.

[7] Gamboa and Waltenberg (2011), for example, report Theil-L indices of post-standardized PISA test scores.

A possible alternative would be to use an absolute measure of inequality – such as the variance, or the absolute Gini coefficient[8] - which are ordinally invariant in the standardization. The variance of a post-standardized distribution (V_j^y), for example, is a monotonic (linear) function of the pre-standardization variance (V_j^x), and does not depend on any other moment of the pre-standardization distribution:

$$V_j^y = \left(\frac{\hat{\sigma}}{\sigma}\right)^2 V_j^x \qquad (6)$$

The variance is seldom used as an inequality measure because it is scale-dependent: it increases with the mean. It also fails the transfer sensitivity axiom, by placing greater weight on transfers higher up the distribution than to those lower down. While these are not trivial concerns, it appears to us that in the context of distributions of educational achievement, they are less severe than violating either the transfer principle itself (like the percentile based measures) or ordinal invariance in the standardization, which allows an apparently innocuous operation to fundamentally alter distributional rankings. The variance (and the standard deviation, of course) is a meaningful measure of inequality in the precise sense that it satisfies axioms (i)-(iii) above. The variance is also additively decomposable, and shares of the variance obtained from some such decompositions can be shown to be cardinally invariant to standardization, as discussed in the next section. These properties will prove instrumental in adapting an intuitive measure of inequality of opportunity to the context of education.

For these reasons, we adopt the variance and the standard deviation as our basic measures of inequality of educational achievement. Because users of this kind of data are generally more comfortable with the standard deviation than its square, this is the variable we report. Columns 3-11 in Table 1 present the mean and standard deviation (S.D.) of the standardized test scores in reading, math and science, in that order, for all 57 countries in the 2006 PISA surveys. The column immediately to the right of each S.D. column reports its bootstrapped standard error. Among the countries with higher inequality in math scores are Western European countries such as Austria, Belgium, France, Germany, and Italy; East European ones such as Czech Republic and Bulgaria, Latin American countries such as Argentina and Uruguay, but also Israel and Taiwan, China. Among the ones with lower inequality in achievements are other European countries such as Croatia, Denmark, Estonia, Finland, Ireland, and Latvia, but also Asian countries such as Indonesia, Thailand and Jordan. Countries such as the UK, Japan, and the United States take intermediate rankings.[9] Figure 1 portrays the S.D. (and its confidence interval) in the mathematics test scores for all countries in the sample.

[8] The absolute Gini coefficient, of course, is the standard (relative) Gini index scaled up by the mean.

[9] The inequality measures obtained for Azerbaijan seem particularly small and place the country as an outlier in all the analyses. It is unclear how much of this is due to the data collection procedures in this country, but such a different pattern is not likely due to real differences only.

Sample selection issues

Although we have established that the country ranking that can be derived from Table 1 is ordinally equivalent to the pre-standardization ranking, the issue of PISA sample selection remains a potential problem. As noted in Section 2, coverage rates range from a low of 0.47 in Turkey, to 1.02 in Switzerland.[10] Selection would not be a problem if one were interested *exclusively* in the performance of 15 year-olds that are in school, and within a reasonable range of their expected grade of attendance. But this is likely to be an excessively narrow prism through which to assess a country's educational system and – even more so – to make international comparisons. Consider the example of two hypothetical "educational strategies", illustrated by countries A and B, which have identical distributions of school and family characteristics, as well as of underlying ability in the population of 15 year-olds. Country A seeks to be inclusive, and allocates resources towards retaining as many students as possible in school, and towards promoting learning by those with the lowest demonstrated achievement. Country B, on the other hand, actively discourages enrollment by those with lower ability, and seeks to retain only the top half of performers in school by age 15. Looking only at the test scores for the samples of enrolled fifteen year-olds will naturally suggest that Country B has both a higher mean and a lower variance than country A, and thus a superior educational system altogether.

This is not to suggest, of course, that Brazil, Indonesia, Mexico, Turkey, or any of the other countries with low coverage rates in Table 1 actively pursue an exclusionary strategy like that of hypothetical country B. But dropping out and lagging behind are, nevertheless, extremely likely to be selective processes, in the sense that they are correlated with family and student characteristics that also affect test scores. If one is interested in comparing the educational achievement of the *population* of fifteen year-olds across countries, therefore, the PISA samples suffer from selection bias.

Correcting for such biases is never simple, and even less so when non-participants are not observed at all in the sample (unlike, say, when seeking to correct for labor force participation on the basis of surveys that contain information on both earners and non-participants). While we do not offer a sample selection bias correction procedure for all countries in the PISA sample in this paper, we propose a simple two-sample non-parametric mechanism for assessing the sensitivity of our inequality measures to alternative assumptions about the sample selection process.

Denote the (density of the) distribution of test scores *y* in a particular country *j* by $f_j(y)$. Consider a vector of covariates *X* that is observed both in the PISA sample and in an ancillary household survey, which is representative of the full population of 15 year-olds. Note that the density of test scores in the PISA sample can be written as:

[10] One presumes that coverage rates in excess of 1.00 must be due either to statistical discrepancies in the estimates of 15 year-olds in the total population, or to errors of inclusion in the sample of test-takers.

$$f_j(y) = \iiint \Phi_j(y, X)dX = \iiint g_j(y|X)\phi_j(X)dX \quad (7)$$

In (7), Φ denotes the joint distribution of *y* and *X*, *g* denotes the conditional distribution of *y* on *X*, and ϕ denotes the joint density of the covariates in the vector *X*.[11] If the joint density of the observable covariates *X* in a particular survey for country *j* is written $\phi_j\ (X|s = survey)$, then our first proposed estimate for a test-score distribution (density) corrected for sample selection on observables is given by:

$$f_j^{SO}(y) = \iiint g_j(y|X)\psi_j(X)\phi_j(X)dX \quad (8)$$

Where $$\psi_j(X) = \frac{\phi_j(X|s=HH)}{\phi_j(X|s=PISA)} \quad (9)$$

Equation (9) is simply the ratio of the density of fifteen year-olds whose observed characteristics *X* take certain values, in the ancillary household survey (HH), to the density of fifteen year-olds with the exact same observed characteristics in the PISA survey. $\psi_j(X)$ is a re-weighting function exactly analogous to that used by DiNardo, Fortin and Lemieux (1996) to construct counterfactual income densities in their study of inequality in the US. Whereas DiNardo et al. use the ratio of densities across different years (of the same survey), we use the ratio of densities across different surveys (for the same year). To the extent that test-taking (i.e. being in the PISA sample) is correlated with observed covariates in X, the counterfactual distribution in (8) should correct for the corresponding selection bias.[12] In practice, this procedure was implemented by partitioning both the PISA and the ancillary household survey into cells with identical values for three observed covariates: gender, mother's education, and father's occupation, with the latter two variables classified as in Section 2.[13] The ratios of densities in each cell in these partitions were used to construct the reweighting function (Equation 9), and both the S.D. and the IOp measures were computed over the counterfactual density of scores given by (8).

This procedure assumes that selection into the PISA sample is fully explained by observable variables, such as gender and family background. While such variables are likely to play a role in selection, it is also likely that other, unobserved variables do too. Within the set of girls, with mothers with no formal education and fathers who work in agriculture, for example, it is possible that a higher proportion of high-ability students than low-ability students stay in school long enough to enter the PISA sample. This kind of selection would imply that equation (8) may overstate the achievement of those students who are counterfactually "brought back into" the sample: simple re-weighting effectively assigns all those out-of-sample students the same scores

[11] The triple integral notation is short-hand for integrating out every element of *X*, so that there are as many integrals as there are elements in the vector of covariates common to both surveys. As it happens, in our application that dimension is three.

[12] The superscript SO stands for selection on observables.

[13] Surveys were thus partitioned into 24 cells. Given the sample sizes reported earlier, particularly for Turkey's HBS and, to a lesser extent, Mexico's ENIGH, it was not possible to further refine the partition by using additional covariates.

obtained by students similar to them (in terms of the variables in *X*). If they are, in fact, likely to perform somewhat less well because of unobserved differences, the procedure overstates their true performance.

By its very nature, of course, selection on unobservables is harder to account for. The ancillary household surveys used to construct the reweighting function do not contain information on test scores. To provide another sensitivity test for the possible magnitude of sample selection bias driven by unobservables, we consider the (rather extreme) assumption that all those students who are counterfactually "re-introduced" into the PISA sample by the above procedure – a proportion given by $\psi_j(X) - 1$, for each *X* – do no better than those who are actually in the sample. In practice, we ascribe to them the lowest observed score for their cell in the partition. As an illustration of the effects of these two re-weighting procedures on the distribution, Figure 3 shows the histograms and kernel density estimates of the distribution of mathematics test scores in Turkey, under each alternative sample selection correction scenario: no correction, correction for selection on observables, and correction for selection on observables and unobservables, under the assumption of no common support.

In order to provide a sense of how sensitive our estimates of educational inequality (reported in Table 1) might be to sample selection, Table 3 reports the results of both of the above scenarios for the four countries with the lowest PISA coverage ratios in Table 1.[14] To economize on space, Table 3 reports the effects of these 'selection correction' procedures both on the standard deviation of test scores and on our measure of inequality of educational opportunity, which is introduced in the next section. The first three columns report these measures (and standard errors) for the uncorrected, original PISA sample, for reading, math and science respectively. The next three report estimates for the correction that assumes selection on observables only (equation 8), and the final three for the correction that assumes selection on unobservables (with no common support).

The results in Table 3 provide a mixed message. Somewhat surprisingly, both inequality of achievement (measured by the standard deviation) and inequality of opportunity seem to be quite robust to selection on observables, despite very low coverage rates (of approximately 50% in these four countries). While this is encouraging, the same cannot be said for the estimates for selection on unobservables. Under these (admittedly extreme) assumptions, inequality in achievement increases by between 44% in Turkey and 92% in Mexico. Inequality of educational opportunity also rises in all countries, except Mexico.

It is possible to interpret these results as comforting, if one chooses to focus on the relative robustness of the measures to selection on observables, even in countries where PISA coverage is lowest. It seems most likely that, if these observed variables account for most of the sample selection process, the estimates of educational inequality in Table 1 are robust for all countries. The fact that those estimates are sensitive to selection on unobservables can be minimized by

[14] Coverage in these four countries – Brazil, Indonesia, Mexico and Turkey – was described in some detail in Section 2 and Table 2 above.

the strength of the "no common support" assumption that assigns the very lowest grade in each cell to all those students counterfactually added to the sample.

Yet, it would probably be wiser to interpret the results from Table 3 as providing grounds for caution. We simply do not know how much selection into the PISA sample takes place on the basis of variables other than gender, mother's education and father's occupation. Until more is known about the composition of the group of fifteen year-olds that is excluded from the PISA sample, the possibility remains that inequality in countries with low coverage is underestimated. Investigation of that group of teenagers would seem like an important – but so far neglected – area of study for those interested in the distribution of educational achievement, particularly in developing countries.

4. A Measure of Inequality of Educational Opportunity

At least as important as the total level of inequality in educational achievement is the question of how much of that inequality is explained by pre-determined circumstances, which individuals simply inherit, rather than controlling. While many may find some inequality in achievement – that might reflect differences in effort, or perhaps even differences in innate ability – quite acceptable, it is common to come across arguments against unequal opportunities among students. These are differences in achievement that do not reflect the choices or actions of today's students, but only inherited circumstances beyond their control. That such inequalities are morally objectionable is today a dominant view among social justice theorists. See, for example, Cohen (1989), Dworkin (1981), Roemer (1998) and Fleurbaey (2008) for some of the classic references. There is also a positive argument against the inheritance of educational inequality, namely that if scarce opportunities for educational investment are allocated on some basis other than talent – such as inherited wealth, for example – this will lead to an inefficient allocation of resources.[15]

The applied literature on the measurement of inequality of opportunity has focused primarily on opportunities for the acquisition of income, but there is no reason it cannot be adapted to the space of educational achievement.[16] Two main approaches characterize that empirical literature. Both approaches begin by seeking agreement on a set of individual characteristics which are beyond the individual's control, and for which he or she cannot be held responsible. These variables are known as '*circumstances*'. Once a vector *C* of circumstances has been agreed upon, society can be partitioned into groups with identical circumstances. Formally, such a partition is given by a set of *types*: $\Pi = \{T_1, T_2, ..., T_K\}$, such that $T_1 \cup T_2 \cup ... \cup T_K = \{1, ..., N\}$, $T_l \cap T_k = \varnothing, \forall l, k$, and the vectors $C_i = C_j, \forall i, j | i \in T_k, j \in T_k, \forall k$.

Given such a partition, the two approaches differ in how they define the benchmark of equality of opportunity. In the *ex-ante* approach, associated with van de Gaer (1993), the opportunity set

[15] See, e.g. Fernández and Galí (1999).

[16] Indeed Checchi and Peragine (2005), the working paper version of their 2010 paper, do apply the concept to educational achievement measures. See also Gamboa and Waltenberg (2011) for a more recent treatment.

faced by each type is evaluated, and equality of opportunity is attained when there is perfect equality in those values across all types. In practice, researchers have often used the mean income (or achievement) of the type as an estimate of the value of the opportunity set they face. Since equality of opportunity would imply equality in means across types, inequality of opportunity is then naturally seen as some measure of between-type inequality.

In the *ex-post approach*, associated with Roemer (1998), equality of opportunity obtains only when individuals exerting the same degree of effort, regardless of their circumstances, receive the same reward. Under certain assumptions, this amounts to requiring equality in the full conditional outcome distributions across all types. Inequality of opportunity would, in this case, best be captured by the (appropriately weighted) sum of inequality within groups characterized by the same degree of effort.[17] The two approaches are closely related but, for any society with a given joint distribution of achievement and circumstance variables, they yield different answers to the question "How much inequality of opportunity is there?" See Fleurbaey and Peragine (forthcoming) for a formal discussion of the relationship between the two approaches.

In what follows, we adapt the *ex-ante* approach employed by Ferreira and Gignoux (forthcoming) to the distributions of test scores described earlier.[18] These authors propose to measure inequality of opportunity (IOp) by between-type inequality. Specifically:

$$\theta_{IOp} = \frac{I(\{\mu_i^k\})}{I(y)} \tag{10}$$

where $\{\mu_i^k\}$ is the smoothed distribution corresponding to the distribution *y* and the partition *Π*.[19]

Naturally, θ_{IOp} can be computed non-parametrically by means of a standard between-group inequality decomposition (provided the chosen inequality index *I()* is properly decomposable). However, this procedure is data-intensive when the vector *C* is large. As the partition becomes finer, cells become small and sparsely populated, and the precision of the estimates of cell means declines, giving rise to an upwards bias in the estimation of θ_{IOp}. Following Bourguignon et al. (2007), Ferreira and Gignoux (forthcoming) then propose a parametric alternative for θ_{IOp}, based on an OLS regression of *y* on *C*:

$$\hat{\theta}_{IOp} = \frac{I(C_i'\hat{\beta})}{I(y)} \tag{11}$$

$\hat{\beta}$ in (11) is the OLS estimate of the regression coefficients in a simple regression of *y* on *C*:

$$y_i = C_i'\beta + \eta_i \tag{12}$$

[17] Under the standard Roemerian assumptions, these groups are Checchi and Peragine's (2010) *'tranches'*.
[18] Ferreira and Gignoux (forthcoming), in turn, build on Bourguignon et al. (2007) and Checchi and Peragine (2010).
[19] A smoothed distribution is obtained from a vector *y* and a partition *Π* by replacing each element of *y* in a given cell T_k with the mean value of *y* in its cell, μ^k. See Foster and Shneyerov (2000).

In (11), $C_i'\hat{\beta}$ denotes the vector of predicted incomes from regression (12). Under the maintained assumption of a linear relationship between achievement and circumstances, this vector is equivalent to the smoothed distribution, since all individuals with identical circumstances are assigned their conditional mean incomes.

Because of its unique path-independent decomposability properties, Checchi and Peragine (2010) and Ferreira and Gignoux (forthcoming) both use the mean logarithmic deviation as the inequality index *I().* However, as shown above, the mean log deviation is not ordinally invariant in the standardization to which test scores are submitted, and it is therefore unsuitable for use in the present context. Following the discussion in Section 3, we use the simple variance as our inequality index *I().* This choice yields our proposed measure of inequality of educational opportunity, as a special case of (11):

$$\hat{\theta}_{IOp} = \frac{Var(C_i'\hat{\beta})}{Var(y_i)} \tag{13}$$

This index has a number of attractive features. First, it is extremely simple to calculate: It is simply the R^2 of an OLS regression of the child's test score on a vector *C* of individual circumstances. In our application to the PISA data sets, *C* includes the following ten variables: gender, father's and mother's education, father's occupation, language spoken at home, migration status, access to books at home, durables owned by the households, cultural items owned, and the location of the school attended.

Second, despite its simplicity, it is a very meaningful summary statistic. It is a parametric approximation to the lower bound on the share of overall inequality in educational achievement that is causally explained by pre-determined circumstances. A formal proof is provided by Ferreira and Gignoux (forthcoming). But the basic intuition is to note that (12) can be seen as the reduced form of a (linearized version of a) model such as:

$$y = f(C, E, u) \tag{14}$$

$$E = g(C, v) \tag{15}$$

In (14) and (15), *y* denotes achievement, and *C* denotes the vector of circumstances, as before. *E* denotes a vector of efforts: all variables that affect achievement and over which individuals do have some measure of control. *u* and *v* denote random shocks. Because 15 year-olds may conceivably affect the choice of school they attend, the class they are assigned to, and thus the teachers they interact with, all school characteristic variables, for example, are included in *E*. So are any direct measures of the student's own efforts in preparing for exams, for instance. Of course, efforts *E* can be influenced by circumstances *C*, but the reverse cannot happen. Variables can only be treated as circumstances if they are pre-determined and entirely exogenous to the individual.

Now return to (12) as a linearized reduced form of (14)-(15). We know that circumstances *C* are economically exogenous to *y*. We also know that all effort (*E*) variables (whether or not one

could observe them in the data) are omitted deliberately: β is intended to capture the reduced-form effect of circumstances – both directly and through efforts. Since all relevant factors are classified into either circumstances or efforts, the only sources of bias to the estimates of β are omitted, unobserved circumstance variables. Although the observed vector *C* is economically exogenous, it may not be exogenous in the (econometric) sense that its components may be correlated with other (unobserved and thus omitted) circumstance variables. Individual elements of the vector $\hat{\beta}$ suffer from these omitted variable biases, and *cannot* be interpreted as causal estimates of the individual impact of a particular circumstance on test scores.

If one is interested, however, on the total *joint effect* of *all circumstances* on achievement and, more specifically, on the share of variation in y that is causally explained by the *overall* effect of circumstances (operating both directly and through efforts), then the R^2 of (12) - our $\hat{\theta}_{IOp}$- yields a valid lower bound for the object of interest. By construction, the only missing variables in (12) are other circumstances. If any were added, $\hat{\theta}_{IOp}$ might rise, but it cannot fall. While individual coefficients in $\hat{\beta}$ may be biased, $\hat{\theta}_{IOp}$ is a lower bound estimate of the joint causal effect of all circumstances on achievement, and thus an appropriate measure of inequality of opportunity. A formal proof is provided by Ferreira and Gignoux (forthcoming), for the perfectly analogous case of incomes.

A third attractive feature of (13) is that it allows for the use of more information on circumstances than previous studies, which typically rely on a smaller set of background variables, and thus capture a more limited share of heterogeneity in family resources. Schultz, Ursprung and Wossmann (2008), for example, focus on the number of books at home. Macdonald et al. (2010) look at the effect of gender and an index of household wealth but ignore, for example, information on parental education and occupation. Gamboa and Waltenberg (2011) see inequality of opportunity as determined by gender, parental education, and school type (public or private), which they treat as a circumstance. We consider the joint effect of all of these circumstances, and more.

A fourth attractive feature of $\hat{\theta}_{IOp}$ as a measure of inequality of educational opportunity is that, unlike any measure of the level of inequality (see Remark 1 above), it is a parametric estimator of a ratio (equation 10) that is cardinally invariant in the standardization of test scores. To see this, note that any sub-group mean is affected by standardization in a manner analogous to equation (3), so that:

$$Var\{\mu_i^k(y)\} = \left(\frac{\hat{\sigma}}{\sigma}\right)^2 Var\{\mu_i^k(x)\} \qquad (16)$$

Given (16) and equation (6), it follows that $\theta_{IOp} = \frac{Var(\{\mu_i^k(y)\})}{Var(y)} = \frac{Var(\{\mu_i^k(x)\})}{Var(x)}$.

A fifth attractive feature of this IOp measure is that it is neatly decomposable into components for each individual variable in the vector *C*. Equation (13) can be rewritten as:

$$\hat{\theta}_{IOp} = \left(\text{var } y\right)^{-1}\left[\sum_j \beta_j^2 \,\text{var}\, C_j + \frac{1}{2}\sum_k \sum_j \beta_k \beta_j \,\text{cov}\left(C_k, C_j\right)\right] \quad (17)$$

This in turn can be written as the sum over all elements (denoted by *j*) of the *C* vector:

$$\hat{\theta}_{IOp} = \sum_j \hat{\theta}^j = \sum_j \left(\text{var } y\right)^{-1}\left[\beta_j^2 \,\text{var}\, C_j + \frac{1}{2}\sum_k \beta_k \beta_j \,\text{cov}\left(C_k, C_j\right)\right] \quad (18)$$

This decomposition is an example of a Shapley-Shorrocks decomposition: it corresponds to the average between two alternative paths for estimating the contribution of a particular circumstance C_J to the overall variance. In the first (direct) path, all C_j, $j \neq J$ are held constant. In the second (residual) path, C_J is itself held constant, and its contribution is taken as the difference between the total variance and the ensuing variance. Either path is conceptually valid, and the Shapley-Shorrocks averaging procedure yields (18) as the path-independent additive decomposition.[20]

Finally, $\hat{\theta}_{IOp}$ can be seen as isomorphic to a measure of intergenerational persistence of inequality, itself the converse of a measure of educational mobility.[21] In the canonical Galton regression of a child's outcome (y_{it}) on the parent's outcome ($y_{i,t-1}$):

$$y_{it} = \beta y_{i,t-1} + \varepsilon_{it} \quad (19)$$

the coefficient β is sometimes used as measure of persistence, and 1-β as a measure of mobility. An alternative that gives equal weight to the variance in both father's and son's distributions is the R^2 of (19) which is, of course, also the square of the correlation coefficient between the two outcomes in the population. If one were to replace the parent's outcome $y_{i,t-1}$ with a vector of parental or family background variables, (19) would transform into something very close to (12), and the R^2 measure of immobility into our measure of inequality of opportunity, $\hat{\theta}_{IOp}$. Indeed, the only pre-determined circumstance among the ten variables previously listed which is not a family background variable is the child's own gender. Apart from the child's own gender, one could see $\hat{\theta}_{IOp}$ as a measure of intergenerational persistence, or immobility, in which the missing value for the parent's own test scores, $y_{i,t-1}$, is replaced with a proxy vector of family background circumstances, C_i.

[20] See Shorrocks (1999) for the original application of the Shapley value to distributional decompositions. Ferreira et al. (2011) provide a formal proof that (18) is the Shapley-Shorrocks decomposition of the variance into the effects of individual circumstances.

[21] Mobility is a multifaceted concept, and there are many distinct measures of it, often attempting to capture different aspects of "movement" across distributions. See Fields and Ok (1996) for a discussion. In the present context, we adopt a view of mobility as time- or origin-independence. See also Shorrocks (1978). Persistence would therefore correspond to the concept of origin-dependence, which is closely related to the notions of inequality of opportunity in both van de Gaer (1993) and Roemer (1998).

Having separately regressed test scores for each subject (in each country) on the vector *C* (equation 12), and computed the R^2 of each regression to obtain $\widehat{\theta}_{IOp}$, we report them on Table 4. These are our estimates of the inequality of educational opportunity (IOp) given by equation (13). They range between 0 and 1, and can be interpreted straight-forwardly as a lower-bound on the share of the total variance in educational achievement that is accounted for by pre-determined circumstances (gender and family background) in each country. Bootstrapped standard errors are reported next to each IOp measure. The IOp estimates range between 12.7% and 38.8% of the total variance of test scores in reading; between 4.4% (10.2% excluding the outlier Azerbaijan) and 35.1% of the variance of test scores in math; and between 11.1% and 37.9% in Science.[22]

Figure 2 provides the same results graphically for achievements in mathematics, after ranking the countries by the IOp measure. 95% confidence intervals are presented using the bootstrapped standard errors and assuming normal distributions of the estimates. No clear regional pattern emerges from the estimates presented in Table 4 and Figure 2. Among the countries with the highest levels of inequality of opportunity, with shares above 30%, are Western European countries (such as Belgium, France, and Germany) but also Eastern European countries (such as Bulgaria and Hungary), and Latin American countries (such as Argentina, Brazil and Chile). Among the countries with the lowest IOp, with shares below 20%, are Asian countries (such as Azerbaijan, Macao (China), and Hong Kong SAR, China), Nordic countries (such as Finland, Iceland, and Norway), Russia, Australia and Italy. The United States, the UK, and Spain lie in an intermediate range, with shares close to 25%.

One can use these results to make specific comparisons. For example, the degree of inequality of educational opportunity seems to be significantly higher in a few large European countries, such as France and Germany, than in the United States. However these inequalities are significantly lower in Nordic countries, such as Finland and Norway, or in Japan and Korea. Regarding developing economies, countries in Latin America tend to rank in the upper half of the distribution, while Asian countries, such as Indonesia and Thailand, rank in the lower half. Although the estimates are very imprecise for Indonesia, Thailand exhibits significantly lower inequalities than Latin American countries such as Brazil. The results for reading and science are not discussed in detail here, but IOp measures for the three subjects are highly correlated: the Spearman rank correlation coefficients for shares in Reading, Math and Science range from 0.75 to 0.92.

The absence of a clear geographical pattern in the cross-country distribution of inequality of educational opportunity is mirrored in the absence of a correlation between IOp and either the level of educational achievement, as measured by mean test scores, or the level of economic

[22] If one were interpreting these shares as proxies for the persistence measure given by the R^2 of (19), one should note that the numbers correspond to *squares* of the correlation coefficient. The square root of IOp for mathematics scores, for example, ranges from 0.21 to 0.59.

development, as measured by GDP per capita.[23] Figure 4 plots the relationship between IOp and mean achievement in mathematics. The regression line and a 95% confidence interval are shown on the graphs. The regression coefficient is statistically insignificantly different from zero at the 10% level. Figure 5 plots IOp in mathematics against GDP per capita, again showing the regression line and a 95% confidence interval. No statistically significant relationship is found. In order to test whether outliers such as Azerbaijan or Macao-China drive the statistical relationship, the procedure proposed by Besley, Kuh and Welsch (1980) is implemented to identify outliers and the test of a linear relationship is performed again after the exclusion of the corresponding observations. In this case, the negative regression coefficient is significant at 10% for mathematics, but remains insignificant for reading and science (not shown in figure).

The exact decomposition of inequality of opportunity into partial shares by individual circumstance, described in equation (18), is presented in Table 5 for mathematics scores. The shares of the ten circumstances add up to the total IOp given in the first column. As may be seen from inspection of equation (18), these partial shares are functions of individual regression coefficients from (12). As noted earlier, these individual coefficient estimates are likely to be biased, have not been presented here, and are not the focus of the paper. These partial shares reflect them, and should *not* be interpreted causally in any way. They are useful only as a description of the variables underpinning the overall (lower-bound) measure of inequality of opportunity.

With that caveat in mind, Table 5 suggests that family educational and cultural resources seem to be associated with the largest share of inequality of learning achievement. Mother's and father's education combined account for a mean of 3.7 and a maximum of 9.2 (in Hungary) percentage points of the overall shares of explained inequality in the set of 57 countries, which take the mean of 24.7. The number of books at home accounts for a mean of 7.2 and a maximum of 14.4 percentage points (in Austria). Add parental education, language at home, numbers of books, and cultural possessions, and this set of "educational and cultural variables" add up to a mean of 15.0 points. Family economic resources also appear as an important source of learning inequalities. Father's occupation and the "durable assets" indicator account for means of 3.6 and 3.8, respectively. With immigration status, the set of "economic variables" explains a mean of 7.8 points. Finally, the type of area where schools are located accounts for a mean of 1.6 and a maximum of 10.7 (in Kyrgyzstan) points of the overall shares, whereas the student's gender accounts for a rather limited mean of 0.6 and a maximum of 2.1 (in Chile) points of the overall shares. There are also interesting regional variations in these partial shares of learning inequality. For instance, the partial share associated with educational and cultural resources has a higher mean in Western and Eastern European countries than in other regions, whereas the share associated with economic resources has a higher mean in Latin America.

[23] GDP per capita is measured at purchasing power parity exchange rates, in 2006 US prices; the data are from the World Development Indicators (WDI) database.

5. A Descriptive Application: Correlations between IOp and Education Policies

As an illustration of potential applications, we now briefly investigate the cross-country correlation between the measure of inequality of educational opportunity presented in the previous section and two specific educational policy variables: the distribution of public spending across different levels of the education system, and the extent of early tracking of pupils between general and vocational schools or classes.

The incidence of public spending in education and the allocation of financial resources among the different segments of the education system have been examined by various studies (e.g. Birdsall, 1996; Castro-Leal et al., 1999; and Van de Walle and Nead, 1995). Given that children with disadvantaged backgrounds tend to drop out from school earlier than others, the allocation of resources to the primary level of schooling is generally thought more likely to be progressive.

The impacts of tracking policies on the efficiency and equity of educational systems are another example of education policies that have received considerable attention in recent studies (Ariga et al., 2006; Brunello and Checchi, 2007; Brunello et al., 2006; Hanushek and Woessman, 2006; Manning and Piskhe, 2006). Theory does not provide clear-cut predictions for the effect of early tracking on educational achievements. On the one hand homogenous classrooms, and the associated specialization of teaching and curricula to the needs and abilities of specific students, could lead to efficiency gains. But on the other hand, disadvantaged groups might be harmed by unfavorable allocations of resources, including less well endowed schools, teacher sorting, peer effects, or differences in curricula[24]. Moreover, since much of the early inequality in achievement – and thus the track placements themselves – are driven by differences in parental resources, a frequent concern has been that tracking might reinforce the effects of family background on educational achievements. I.e. that it might reduce intergenerational mobility, and exacerbate inequality of educational opportunity.

We briefly examine the correlation between our measure of IOp and these two policies, using data on the policy indicators from the UNESCO Institute for Statistics (UIS).[25] Our indicator of the distribution of educational expenditures is the share of spending in primary schools - defined as the first ISCED level, corresponding to grades 1 to 6 - in total public educational expenditure. The indicator of tracking is the share of technical or vocational enrollment at the secondary level (including lower and upper secondary or the second and third ISCED levels, usually corresponding to grades 7 to 12) in total enrollment at that level. The information on the distribution of education expenditure across levels is missing for six countries (Canada, Montenegro, Qatar, Russia, Serbia and Taiwan, China) and the information on the share of technical and vocational enrollment at the secondary level is missing in five countries (Latvia;

[24] Early tracking may also be costly in terms of the misallocation of students to tracks, and in terms of forgone versatility in the production of skills (Brunello and Checchi, 2007).

[25] The data for 2006 correspond to the school year 2005-06 for countries where the school year laps over two calendar years.

Montenegro; Serbia; Taiwan, China; and the United States). Two other countries are excluded from the analysis: Liechtenstein and Luxembourg. The number of observations for Liechtenstein (339 examinees) makes the estimates of learning inequalities unreliable and Luxemburg is too much of an outlier in terms of GDP per capita in 2006 (at about 69.000 US dollars, with the US in second place at 44.000 US dollars).

There is considerable variation in the share of expenditures allocated to the primary level of education in the remaining country sample. While the mean share is 27.0%, the lowest share is observed in Romania at 13.8% and the highest in Jordan at 41.7% (the first quartile is at 20.2% and the third quartile at 34.0%). Figure 6 provides an illustration of the relationship between the primary share of expenditures and IOp. Once again the regression line and a 95% confidence interval for the mean are shown. Table 6 gives the tests of significance of this relationship both without any controls (first panel) and controlling for per capita GDP and public education expenditure per pupil (second panel). Once outliers are excluded, significant negative correlations exist both for reading and science, with or without controls. For math, the negative correlation is only significant with controls. The coefficients lie between -0.001 and -0.003, indicating that an increase of 10 points in the share of resources allocated to primary schooling is associated with decreases of 1 to 3 points in inequality of educational opportunity.

There is also considerable heterogeneity in tracking in our country sample. The mean share is 20.8 percent and values range from 0.9% in Qatar to 51.4% percent in the Netherlands (the first quartile is at 12.9 and the third at 31.2). As before, Figure 7 provides a scatter plot of the relationship between tracking and IOp in this sample, while Table 7 lists coefficients and standard errors, both without any controls (upper panel) and controlling for per capita GDP and public education expenditure per pupil (bottom panel). There is a clear pattern of significant positive relationships across all three subjects and both regression specifications, with the statistical significance being stronger in the specification with controls. Higher inequality of opportunity tends to be associated with higher shares of technical and vocational enrollment. The regression coefficients lie between 0.001 and 0.002, indicating that an increase of 10 points of the share of technical or vocational enrollments is associated with an increase of 1 to 2 points in inequality of opportunity.

These correlations suggest that our measure of inequality of opportunity is negatively associated with the share of public spending on primary education, and positively associated with tracking into general or technical/vocational schooling at the secondary level. These associations allow for absolutely no inference of causality, of course, but the results seem in line with and extend those of studies devoted to these relationships. For instance, while Hanushek and Woessman (2006) find tracking to be associated with higher levels of overall inequality in test scores, our results suggest it also tends to come with higher levels of inequality of learning opportunities.[26] This analysis remains descriptive in nature, and does not control for the

[26] However, the long term effects of early tracking remain a matter of debate. For instance, Brunello and Checchi (2007) find that although it tends to increase the link between family background and educational attainments by

heterogeneity in education systems or pupil populations. They are only meant to illustrate the potential use of indicators of inequality of opportunity for future studies of the distributive impacts of education policies. Future extensions – notably involving the use of panel data - might allow for causal analysis of these relationships.

6. Conclusions

Internationally comparable information on learning outcomes, such as the standardized test scores collected by PISA surveys, represents a revolution in the quality of data available for research on education. It allows for potentially much greater insight into the determinants of educational achievement, and might therefore contribute to the design of policies that raise average learning levels, or that reduce educational disparities.

The measurement of educational disparities using this kind of data is not, however, a trivial extension of inequality measurement in years of schooling, or in other variables like income. This paper has highlighted two issues that require special attention in the measurement of inequality in educational achievement, and which appear to have been overlooked so far. The first is the standardization of test scores, to which all meaningful measures of inequality are cardinally sensitive. More importantly, many common measures of inequality, including the Gini coefficient and the Theil indices, are not event ordinally invariant to standardization, invalidating country rankings that are based on them.

We show that the simple variance (or the standard deviation) of test scores is ordinally invariant to standardization, and present estimates for all 57 countries that took part in the 2006 round of PISA surveys, in all three subjects for which tests are carried out: reading, mathematics and science. There is considerable international variation in educational inequality thus measured. The standard deviation in Math scores ranges from around 80 in Indonesia, Estonia and Finland, to nearly 110 in Belgium and Israel.

The second measurement issue that may compromise international inequality comparisons based on PISA test scores is the possibility of sample selection. The surveys are designed to be representative of the population of 15 year-olds enrolled in school, and attending grades 7 or above. While this stipulation covers most of the population of that age group in OECD countries, it purposively excludes substantial numbers in poorer countries. Selection into the sample is clearly correlated with determinants of test scores, leading to a classic problem of sample selection bias. Using information on characteristics of fifteen year-olds included in other, ancillary household surveys, we use sample re-weighting methods to assess the implications of the selection bias for our measures of educational inequality in achievement and opportunity. Results for Brazil, Indonesia, Mexico and Turkey suggest that the inequality measures are relatively robust to selection on the basis of three observed variables (gender, mother's

diverting some individuals from progress to tertiary education, it seems to reduce the impact of family background on adult literacy and promote further on-the-job training by offering more effective curricula to less well performing students.

education and father's occupation). Under a more stringent scenario of strong selection on unobservables with no common support, however, the current measures of educational inequality in these countries would appear to be substantially underestimated.

Finally, we also propose and compute a measure of inequality in educational opportunity. The measure is simply the share of the total variance in achievement that can be accounted for by pre-determined circumstance variables in a linear regression. The index is simple and intuitive, and provides a lower-bound estimate of the joint causal effect of all pre-determined circumstances on educational inequality. It is cardinally invariant to the standardization of test scores, and exactly additively decomposable into the partial shares accounted for by individual circumstance variables. It is also closely related to the origin-independence concept of inter-generational educational mobility.

Thus measured, inequality of opportunity in our sample of countries ranges from approximately 0.10 – 0.16 in Macao (China), Australia, and Hong Kong SAR, China, to 0.33 – 0.35 in Bulgaria, France and Germany. Although the measure is uncorrelated with average educational achievement and with GDP per capita, it appears to be higher in Latin America and parts of continental Europe (including France, Germany and Belgium). It is lower in Asia, the Nordic countries, and Australia. It is negatively correlated with the share of public educational spending allocated to primary schooling, and positively correlated with the extent of educational tracking, defined as the share of technical and/or vocational enrollment in secondary schools.

This paper has not reported on any causal analysis of specific policy determinants of educational inequality. Its aim was to place the measurement of these concepts on a sounder footing, given the specific characteristics of data on educational achievement. We hope that the measures proposed here, and the methods for assessing their sensitivity to sample selection, may be of use to other researchers interested in the determinants of educational achievement, and its distribution.

References

Ariga, K., G. Brunello, R. Iwahashi, and L. Rocco (2006): "On the Efficiency Costs of De-Tracking Secondary Schools". IZA Discussion Paper No. 2534.

Baker, F. (2001): *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD.

Bedard, K. and C. Ferrall (2003): "Wage and Test Score Dispersion: Some International Evidence" *Economics of Education Review,* **22**: 31-43.

Besley D., E. Kuh and R. Welsch (1980): *Regression Diagnostics: Identifying Influential Data and Sources of Colinearity*, New York, Wiley.

Birdsall, N. (1996): "Public Spending on Higher Education in Developing Countries: Too Much or Too Little?" *Economics of Education Review*, **15**(4): 407-19

Blau, Francine and Lawrence Kahn (2005): "Do Cognitive Test Scores Explain Higher US Wage Inequality?" *Review of Economics and Statistics*, **87**: 184-193.

Bourguignon, François, Francisco H.G. Ferreira and Marta Menéndez (2007): "Inequality of Opportunity in Brazil", *Review of Income Wealth,* **53** (4): 585-618.

Brown, G., J. Micklewright, S.V. Schnepf, and R. Waldmann, (2007), "International Surveys of Educational Achievement: How Robust are the Findings?" *Journal of the Royal Statistical Society*, **170** (3): 623-646

Brunello, G., K. Ariga and M. Giannini (2006): "The Optimal Timing of School Tracking", in P. Peterson and L.Wößmann, (eds), *Schools and the Equal Opportunity Problem*, MIT Press, Cambridge MA.

Brunello. G. and D. Checchi (2007): "Does School Tracking Affect Equality of Opportunity? New International Evidence", *Economic Policy*, **22**: 781-861.

Castelló, A. and R. Doménech (2002): "Human Capital Inequality and Economic Growth: Some New Evidence", *Economic Journal*, **112**: C187-200.

Castro-Leal, F., J. Dayton, L. Demery, and K. Mehra, (1999): "Public Social Spending in Africa: Do the Poor Benefit?", *World Bank Research Observer*, 14(1): 49-72.

Checchi, Daniele and Vito Peragine (2005): "Regional Disparities and Inequality of Opportunity: The Case of Italy", IZA Discussion Paper No. 1874.

Checchi, Daniele and Vito Peragine (2010): "Inequality of Opportunity in Italy", *Journal of Economic Inequality*, **8** (4): 429-450.

Cohen, Gerry A., (1989). "On the Currency of Egalitarian Justice", *Ethics*, **99**: 906-944.

DiNardo, John, Nicole Fortin and Thomas Lemieux (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semi-Parametric Approach", *Econometrica*, **64** (5): 1001-1044.

Dworkin, Ronald (1981), "What is Equality? Part 2: Equality of Resources". *Philosophy and Public Affairs*, **10**(4): 283-345.

Fernández, Raquel and Jordi Galí (1999): "To Each According to...? Markets, Tournaments, and the Matching Problem with Borrowing Constraints", *Review of Economic Studies*, **66**: 799-824.

Ferreira, Francisco and Jérémie Gignoux (forthcoming): "The Measurement of Inequality of Opportunity: Theory and an application to Latin America", *Review of Income Wealth*.

Ferreira, Francisco, Jérémie Gignoux and Meltem Aran (2011): "Measuring Inequality of Opportunity with Imperfect Data: the case of Turkey", *Journal of Economic Inequality* **9** (4): 651-680.

Fields, Gary S. & Ok, Efe A. (1996), "The Meaning and Measurement of Income Mobility," *Journal of Economic Theory,* **71** (2): 349-377.

Fleurbaey, Marc (2008): *Fairness, Responsibility, and Welfare*. Oxford: Oxford University Press.

Fleurbaey, Marc and Vito Peragine (forthcoming): "Ex ante versus ex post equality of opportunity", *Economica*.

Foster, James and Artyom Shneyerov (2000): "Path Independent Inequality Measures", *Journal of Economic Theory*, **91**: 199-222.

Gamboa, Luis Fernando and Fábio Waltenberg (2011): "Inequality of Opportunity in Educational Achievement in Latin America: Evidence from PISA 2006-2009". Unpublished manuscript. Universidad del Rosario, Bogotá, Colombia.

Hanushek, Eric and Ludger Woessmann (2006): "Does Educational Tracking Affect Performance and Inequality? Differences-In-Differences Evidence across Countries", *Economic Journal* 116: C63-C76.

Manning, A. and J.S. Piskhe (2006): "Comprehensive versus Selective Schooling in England in Wales: What Do We Know?" IZA Discussion Paper 2072.

Macdonald, Kevin, Felipe Barrera, Juliana Guaqueta, Harry Patrinos and Emilio Porta (2010): "The Determinants of Wealth and Gender Inequity in Cognitive Skills in Latin America", World Bank Policy Research Working Paper #5189.

Marks, G.N., (2005), "Cross-National Differences in Accounting for Social Class Inequalities in Education", *International Sociology*, **20** (4): 483-505.

Micklewright, John and Sylke Schnepf (2007): "Inequality of Learning in Industrialized Countries", Chapter 6 in S. Jenkins and J. Micklewright (eds.): *Inequality and Poverty Re-examined*. Oxford: Oxford University Press.

Mislevy, R. (1991): "Randomization Based Inference about Examinees in the Estimation of Item Parameters", *Psychometrika*, **56:** 177-196.

Mislevy, R., A. Beaton, B. Kaplan and K. Sheehan (1992): "Estimating Population Characteristics from Sparse Matrix Samples of Item Responses", *Journal of Educational Measurement*, **29** (2): 133-161.

Morrisson, Ch. and F., Murtin (2007): "Education inequalities and the Kuznets curves: a global perspective since 1870", PSE Working Paper 2007-12.

OECD (2006), *PISA 2006 technical report*.

Roemer, John E. (1998): *Equality of Opportunity*. Cambridge, MA: Harvard University Press.

Schultz, G., H.W. Ursprung, L. Wossmann (2008): "Education Policy and Equality of Opportunity", *Kyklos*, **61** (2): 279-308.

Sen, Amartya (1985): *Commodities and Capabilities*. Amsterdam: North-Holland.

Shorrocks, Anthony (1978): "The measurement of mobility", *Econometrica*, **46**: 1013-1024.

Shorrocks, Anthony (1999): "Decomposition Procedures for distributional analysis: a unified framework based on the Shapley Value", unpublished manuscript, University of Essex.

Thomas, V., Y. Wang and X. Fan (2001), "Measuring education inequality: Gini coefficients of education", Policy Research Working Paper 2525, Washington DC: The World Bank.

van de Gaer, Dirk (1993): *Equality of Opportunity and Investment in Human Capital*. PhD dissertation, Catholic University of Leuven, Belgium.

van de Walle, Dominique and Kimberly Nead (1995): *Public Spending and the Poor: Theory and Evidence*, Johns Hopkins and World Bank, Washington DC.

Zheng, B. (1994): "Can a Poverty Index be Both Relative and Absolute?" *Econometrica*, **62** (6): 1453-1458.

Table 1: Sample statistics, mean scores and the standard deviation in PISA test scores

	# Obs.	Coverage rate	Reading Mean	Reading SD	Reading (SE of SD)	Math Mean	Math SD	Math (SE of SD)	Science Mean	Science SD	Science (SE of SD)
Asia & North Africa											
Azerbaijan	5184	0.88	355.0	70.26	2.12	476.8	47.96	1.64	385.3	55.68	1.92
Hong Kong SAR, China	4645	0.97	538.9	81.79	1.92	551.4	93.39	2.31	546.1	91.71	1.92
Indonesia	10647	0.53	383.9	74.79	2.39	380.7	80.01	3.18	384.8	70.06	3.26
Israel	4584	0.76	441.3	119.34	2.79	443.3	107.33	3.20	455.6	111.45	1.92
Japan	5952	0.89	409.5	102.38	2.34	389.2	91.01	2.06	427.1	100.12	2.01
Jordan	6509	0.65	500.2	94.09	2.24	525.6	83.71	1.95	533.7	89.86	1.89
Korea	5176	0.87	290.5	88.29	2.68	315.9	92.59	3.12	326.3	90.06	2.35
Kyrgyzstan	5904	0.63	556.1	102.10	2.51	547.2	86.98	2.03	521.9	83.86	2.03
Macao-China	4760	0.73	490.6	76.36	2.26	524.4	83.90	1.51	509.5	77.83	1.58
Qatar	6265	0.90	312.5	108.12	1.15	317.7	90.24	1.39	349.1	83.29	1.37
Russian Federation	5799	0.81	442.4	93.23	1.87	478.7	89.53	1.58	481.5	89.57	1.33
Chinese Taipei	8812	0.88	506.7	84.38	1.73	562.7	103.11	2.16	543.7	94.45	1.63
Thailand	6192	0.72	425.2	81.85	1.73	425.5	81.43	1.57	429.7	77.17	1.45
Tunisia	4640	0.90	379.0	97.30	2.49	363.9	91.95	2.34	384.2	82.38	2.05
Turkey	4942	0.47	452.9	92.90	2.75	428.2	93.24	4.32	427.6	83.20	3.14
Latin America											
Argentina	4339	0.79	383.9	124.22	3.63	388.1	101.14	3.48	398.3	101.24	2.62
Brazil	9295	0.55	389.2	102.46	3.34	365.6	92.02	2.65	385.3	89.28	1.93
Chile	5233	0.78	447.9	103.24	2.44	417.1	87.44	2.17	443.1	91.68	1.72
Colombia	4478	0.60	390.3	107.83	2.38	373.8	88.04	2.42	391.9	84.81	1.81
Mexico	30971	0.54	427.4	95.68	2.27	420.7	85.27	2.16	422.6	80.70	1.47
Uruguay	4839	0.69	424.7	121.22	2.03	435.5	99.30	1.77	437.7	94.44	1.73
North America &											
Australia	22646	0.87	508.7	96.25	1.43	516.3	85.79	1.03	523.1	94.19	1.14
Canada	14170	0.87	512.3	93.79	1.00	517.4	88.03	1.09	522.5	100.23	1.02
New Zealand	4823	0.84	522.7	105.21	1.58	523.8	93.27	1.20	532.7	107.30	1.36
United States	5610	0.85				474.7	89.75	1.90	488.3	106.07	1.68
Eastern Europe											
Bulgaria	4498	0.83	406.8	117.51	4.00	417.4	101.10	3.65	439.1	106.72	3.20
Czech Republic	5932	1.01	509.6	111.21	2.90	536.0	103.14	2.08	537.6	98.41	2.00
Estonia	4865	0.94	502.4	85.19	1.87	516.8	80.68	1.54	533.7	83.75	1.09
Croatia	5213	0.85	477.6	88.83	2.12	467.3	83.31	1.50	493.7	85.72	1.44
Hungary	4490	0.85	488.1	94.39	2.37	496.2	91.04	1.94	508.7	88.20	1.53
Lithuania	4744	0.93	469.3	95.54	1.51	485.6	89.80	1.73	486.5	89.99	1.52

Latvia	4719	0.85	484.9	90.70	1.69	491.2	82.81	1.51	493.8	84.38	1.30
Montenegro	4455	0.84	388.2	89.41	1.64	395.8	84.45	1.80	408.8	79.69	1.19
Poland	5547	0.94	512.6	100.22	1.48	500.9	86.52	1.13	503.3	89.87	1.11
Romania	5118	0.66	392.0	91.86	2.93	415.0	83.97	2.85	416.6	81.16	2.37
Serbia	4798	0.83	402.9	91.84	1.69	436.6	91.76	1.77	436.9	85.15	1.56
Slovak Republic	4731	0.95	470.6	105.08	2.51	495.1	94.53	2.47	491.2	93.15	1.79
Slovenia	6595	0.88	468.6	87.97	2.47	482.2	89.25	1.36	494.2	98.11	1.35
Western Europe											
Austria	4927	0.92	494.0	108.16	3.16	509.5	98.06	2.29	513.9	97.83	2.41
Belgium	8857	0.99	507.1	110.02	2.81	526.9	106.13	3.31	516.3	99.70	2.00
Switzerland	12192	1.02	496.6	94.07	1.71	528.3	97.44	1.60	508.0	99.31	1.61
Germany	4891	0.95	496.5	111.95	2.67	504.3	99.08	2.53	516.2	99.98	1.99
Denmark	4532	0.85	493.8	89.30	1.63	512.2	84.85	1.53	494.7	93.13	1.42
Spain	19604	0.87	479.5	88.84	1.14	501.7	88.92	1.09	504.5	90.54	0.97
Finland	4714	0.93	547.1	81.23	1.08	549.0	80.87	1.01	563.4	85.62	1.00
France	4716	0.91	488.7	103.95	2.75	496.4	95.58	1.96	496.1	101.57	2.09
United Kingdom	13152	0.94	495.6	101.92	1.69	497.3	88.92	1.31	514.3	106.79	1.50
Greece	4873	0.90	461.9	102.61	2.92	462.0	92.30	2.37	476.6	92.12	2.03
Ireland	4585	0.94	518.6	92.39	1.86	502.3	81.99	1.50	509.5	94.35	1.50
Iceland	3789	0.96	485.0	97.09	1.23	505.6	88.08	0.89	491.0	96.87	0.95
Italy	21773	0.90	477.0	108.76	1.74	473.6	95.82	1.66	487.2	95.56	1.31
Liechtenstein	339	0.84	510.7	95.14	2.93	524.9	93.05	2.17	522.3	96.96	2.10
Luxembourg	4567	1.03	480.1	99.85	0.72	490.5	93.15	0.73	486.8	96.53	0.67
Netherlands	4871	0.96	513.9	96.62	2.47	537.4	88.60	2.18	530.8	95.63	1.64
Norway	4692	0.97	484.4	105.15	1.92	489.8	91.58	1.38	486.9	96.12	1.98
Portugal	5109	0.78	476.8	98.82	2.28	470.9	90.65	1.97	479.0	88.56	1.71
Sweden	4443	0.97	509.0	98.21	1.77	503.2	89.66	1.37	504.2	94.21	1.40

Note: The standard deviation (S.D.) of test scores is used as an ordinal measure of inequality in achievement, as discussed in the text. Standard errors reported in the columns next to the S.D. are bootstrapped.

Table 2: PISA Sample Coverage: Analysis for four developing countries

	Brazil	Indonesia	Mexico	Turkey
Expanded 15 year-old populations, using PISA data and weights				
Total population of 15-year-olds	3 390 471	4 238 600	2 200 916	1 423 514
Total enrolled population of 15-year-olds at grade 7 or above	2 374 044	3 119 393	1 383 364	800 968
Weighted number of students participating to the assessment	1 875 461	2 248 313	1 190 420	665 477
Coverage rate of the population of 15-year-olds, from PISA	55,3	53,0	54,1	46,7
Total missed children	44,7	47,0	45,9	53,3
Composition of those not covered by PISA samples				
Out-of-school children	10,2	25,5	24,1	21,6
Delays of more than two years	19,8	0,9	13,1	22,2
PISA sampling issues	14,7	20,6	8,8	9,5

Source: PISA 2006 surveys; PNAD 2006 for Brazil, Susenas 2005 for Indonesia; ENIGH 2006 for Mexico, and HBS 2006 for Turkey. The share of fifteen year-olds who are not enrolled in school comes from the ancillary household surveys. Those delayed by more than two years come from household surveys, and are checked with PISA administrative records. The last row is derived as a residual.

Table 3: Inequality of Achievement and Opportunity in Low-Coverage Countries: sensitivity to different assumptions on selection into the PISA sample

		PISA population without any correction			Correction assuming *selection on observables*			Correction assuming *strong selection on unobservables*		
		Reading	Math	Science	Reading	Math	Science	Reading	Math	Science
TURKEY										
	Inequality (SD)	92.90	93.24	83.20	98.38	91.43	82.58	155.67	134.04	121.61
		2.75	4.32	3.14						
	IOp	0.251	0.241	0.249	0.250	0.236	0.250	0.327	0.320	0.326
		0.026	0.033	0.032						
BRAZIL										
	Inequality (SD)	102.46	92.02	89.28	102.86	90.44	86.75	179.82	146.68	146.17
		3.34	2.65	1.93						
	IOp	0.268	0.318	0.286	0.265	0.309	0.262	0.404	0.404	0.385
		0.020	0.005	0.021						
MEXICO										
	Inequality (SD)	95.68	85.27	80.70	95.63	85.02	79.18	196.85	162.79	136.99
		2.27	2.16	1.47						
	IOp	0.278	0.261	0.271	0.267	0.242	0.255	0.256	0.250	0.228
		0.024	0.002	0.024						
INDONESIA										
	Inequality (SD)	74.79	80.01	70.06	71.03	76.27	65.74	130.56	135.89	112.79
		2.39	3.18	3.26						
	IOp	0.250	0.237	0.220	0.218	0.200	0.181	0.274	0.261	0.261
		0.038	0.042	0.045						

Note: IOp denotes the measure of inequality of educational opportunity, defined in equation (13). It is the share of the total variance in test scores which is accounted for by the student's pre-determined circumstance variables.

Table 4: Inequality of Educational Opportunity for three PISA subjects

	IOp Reading	Standard Error (Reading IOp)	IOp Mathematics	Standard Error (Math IOp)	IOp Science	Standard Error (Science IOp)
Asia & North Africa						
Azerbaijan	0.173	0.028	0.044	0.012	0.112	0.024
Hong Kong SAR, China	0.177	0.016	0.154	0.016	0.166	0.018
Indonesia	0.250	0.038	0.237	0.042	0.220	0.045
Israel	0.197	0.018	0.206	0.019	0.195	0.016
Japan	0.206	0.017	0.203	0.020	0.189	0.016
Jordan	0.346	0.024	0.272	0.024	0.271	0.019
Korea	0.214	0.022	0.209	0.021	0.173	0.019
Kyrgyzstan	0.314	0.023	0.306	0.027	0.269	0.023
Macao-China	0.127	0.012	0.102	0.009	0.111	0.008
Qatar	0.309	0.010	0.254	0.009	0.264	0.009
Russian Federation	0.238	0.021	0.165	0.020	0.183	0.020
Chinese Taipei	0.300	0.017	0.275	0.022	0.281	0.019
Thailand	0.325	0.023	0.230	0.021	0.265	0.022
Tunisia	0.215	0.024	0.273	0.031	0.191	0.026
Turkey	0.251	0.026	0.241	0.033	0.249	0.032
Latin America						
Argentina	0.289	0.024	0.315	0.007	0.312	0.026
Brazil	0.268	0.020	0.318	0.005	0.286	0.021
Chile	0.248	0.022	0.330	0.001	0.299	0.021
Colombia	0.181	0.018	0.216	0.007	0.193	0.018
Mexico	0.278	0.024	0.261	0.002	0.271	0.024
Uruguay	0.221	0.015	0.245	0.004	0.248	0.012
Australia	0.199	0.010	0.153	0.009	0.164	0.009
Canada	0.242	0.011	0.211	0.011	0.207	0.010
New Zealand	0.276	0.013	0.241	0.012	0.269	0.013
United States			0.279	0.020	0.282	0.019
Eastern Europe						
Bulgaria	0.377	0.028	0.331	0.030	0.364	0.030
Czech Republic	0.296	0.021	0.268	0.019	0.279	0.020
Estonia	0.271	0.013	0.206	0.013	0.208	0.012
Croatia	0.297	0.017	0.222	0.015	0.239	0.014
Hungary	0.345	0.023	0.326	0.022	0.326	0.019
Lithuania	0.318	0.017	0.279	0.017	0.262	0.016
Latvia	0.254	0.017	0.201	0.020	0.187	0.016
Montenegro	0.252	0.013	0.223	0.012	0.197	0.011
Poland	0.275	0.014	0.241	0.013	0.241	0.014
Romania	0.301	0.026	0.313	0.028	0.310	0.027
Serbia	0.311	0.018	0.276	0.017	0.255	0.016
Slovak Republic	0.292	0.026	0.317	0.030	0.297	0.024

Slovenia	0.336	0.018	0.263	0.016	0.268	0.014
Western Europe						
Austria	0.296	0.019	0.300	0.020	0.324	0.022
Belgium	0.335	0.015	0.329	0.018	0.338	0.015
Switzerland	0.313	0.013	0.282	0.013	0.322	0.012
Germany	0.368	0.021	0.351	0.018	0.352	0.019
Denmark	0.229	0.015	0.219	0.014	0.249	0.017
Spain	0.243	0.013	0.239	0.012	0.258	0.013
Finland	0.247	0.014	0.179	0.010	0.167	0.011
France	0.305	0.019	0.335	0.019	0.345	0.018
United Kingdom	0.274	0.014	0.258	0.012	0.275	0.012
Greece	0.261	0.023	0.228	0.022	0.245	0.019
Ireland	0.259	0.018	0.235	0.017	0.240	0.016
Iceland	0.234	0.009	0.167	0.009	0.184	0.009
Italy	0.207	0.015	0.178	0.014	0.206	0.014
Liechtenstein	0.388	0.031	0.323	0.034	0.379	0.030
Luxembourg	0.344	0.008	0.291	0.008	0.328	0.009
Netherlands	0.247	0.022	0.271	0.023	0.283	0.023
Norway	0.271	0.016	0.195	0.014	0.220	0.018
Portugal	0.303	0.021	0.274	0.019	0.267	0.020
Sweden	0.265	0.014	0.233	0.012	0.250	0.013

Note: IOp denotes the measure of inequality of educational opportunity, defined in equation (13). It is the share of the total variance in test scores which is accounted for by the student's pre-determined circumstance variables.

Table 5: A Decomposition of IOp (Mathematics) into Individual Circumstance Shares

	Total	Gender	Father's education	Mother's education	Father's occupa-tion	Area type	Language at home	Immi-gration status	Number of books	Durables	Cultural posses-sions
Asia & North Africa											
Azerbaijan	0.044	0.000	0.000	0.000	0.001	0.003	0.000	0.006	0.017	0.008	0.010
Hong Kong SAR, China	0.154	0.009	0.012	0.007	0.026	0.000	0.000	0.013	0.062	0.009	0.018
Indonesia	0.237	0.009	0.009	0.005	0.018	0.072	0.002	0.000	0.025	0.096	0.009
Israel	0.206	0.004	0.002	0.039	0.057	0.006	0.001	0.000	0.065	0.003	0.030
Japan	0.203	0.012	0.042	0.027	0.025	0.005	0.000	0.004	0.032	0.013	0.044
Jordan	0.272	0.001	0.030	0.029	0.043	0.022	0.007	0.000	0.021	0.103	0.016
Korea	0.209	0.004	0.017	0.011	0.000	0.019	0.000	0.001	0.086	0.014	0.061
Kyrgyzstan	0.306	0.000	0.002	0.012	0.014	0.107	0.008	0.007	0.066	0.053	0.037
Macao-China	0.102	0.006	0.008	0.001	0.007	0.003	0.005	0.003	0.010	0.021	0.039
Qatar	0.254	0.010	0.011	0.005	0.052	0.035	0.079	0.016	0.018	0.012	0.017
Russian Federation	0.165	0.001	0.001	0.009	0.030	0.009	0.004	0.003	0.046	0.037	0.024
Chinese Taipei	0.275	0.005	0.029	0.015	0.031	0.026	0.000	0.008	0.088	0.018	0.054
Thailand	0.230	0.001	0.023	0.026	0.048	0.028	0.001	0.000	0.024	0.079	0.000
Tunisia	0.273	0.009	0.001	0.000	0.072	0.032	0.005	0.000	0.046	0.077	0.034
Turkey	0.241	0.003	0.042	0.041	0.007	0.018	0.000	0.001	0.051	0.045	0.034
Latin America											
Argentina	0.315	0.004	0.014	0.026	0.024	0.022	0.000	0.003	0.079	0.114	0.029
Brazil	0.318	0.009	0.019	0.024	0.027	0.014	0.005	0.001	0.025	0.184	0.011
Chile	0.330	0.021	0.016	0.055	0.050	0.026	0.001	0.000	0.068	0.060	0.033
Colombia	0.216	0.017	0.009	0.015	0.014	0.014	0.003	0.000	0.049	0.085	0.010
Mexico	0.261	0.003	0.001	0.025	0.018	0.074	0.014	0.002	0.033	0.077	0.014
Uruguay	0.245	0.005	0.013	0.047	0.029	0.006	0.000	0.000	0.056	0.059	0.030
North America & Oceania											
Australia	0.153	0.008	0.007	0.009	0.044	0.002	0.000	0.000	0.055	0.011	0.016
Canada	0.211	0.008	0.029	0.011	0.035	0.017	0.003	0.000	0.078	0.013	0.018
New Zealand	0.241	0.005	0.036	0.016	0.036	0.003	0.000	0.000	0.074	0.034	0.037
United States	0.279	0.004	0.014	0.018	0.062	0.013	0.000	0.003	0.122	0.036	0.010
Eastern Europe											
Bulgaria	0.331	0.000	0.005	0.020	0.052	0.032	0.001	0.012	0.102	0.048	0.060
Czech Republic	0.268	0.004	0.010	0.035	0.045	0.007	0.001	0.001	0.089	0.052	0.024
Estonia	0.206	0.000	0.000	0.019	0.061	0.003	0.007	0.000	0.080	0.012	0.028
Croatia	0.222	0.011	0.006	0.000	0.041	0.007	0.000	0.004	0.060	0.046	0.048
Hungary	0.326	0.005	0.038	0.054	0.038	0.016	0.000	0.002	0.099	0.034	0.042
Lithuania	0.279	0.001	0.007	0.023	0.030	0.024	0.001	0.002	0.080	0.061	0.051
Latvia	0.201	0.002	0.000	0.025	0.028	0.007	0.000	0.000	0.069	0.048	0.024
Montenegro	0.223	0.006	0.000	0.014	0.025	0.002	0.001	0.007	0.071	0.021	0.081
Poland	0.241	0.004	0.014	0.035	0.019	0.008	0.000	0.000	0.078	0.030	0.051
Romania	0.313	0.004	0.000	0.006	0.057	0.022	0.000	0.001	0.084	0.062	0.078
Serbia	0.276	0.003	0.006	0.011	0.034	0.020	0.003	0.000	0.086	0.063	0.050

Slovak Republic	0.317	0.008	0.030	0.027	0.033	0.004	0.001	0.014	0.137	0.054	0.009
Slovenia	0.263	0.002	0.022	0.043	0.044	0.003	0.000	0.006	0.105	0.003	0.038
Western Europe											
Austria	0.300	0.017	0.003	0.017	0.026	0.006	0.018	0.008	0.144	0.017	0.044
Belgium	0.329	0.002	0.029	0.049	0.056	0.009	0.053	0.000	0.065	0.030	0.040
Switzerland	0.282	0.006	0.024	0.019	0.028	0.012	0.050	0.006	0.104	0.012	0.021
Germany	0.351	0.012	0.019	0.050	0.047	0.007	0.014	0.012	0.131	0.010	0.049
Denmark	0.219	0.005	0.018	0.020	0.028	0.002	0.015	0.013	0.064	0.008	0.047
Spain	0.239	0.004	0.014	0.026	0.028	0.002	0.010	0.001	0.103	0.032	0.020
Finland	0.179	0.008	0.011	0.018	0.019	0.000	0.009	0.004	0.073	0.006	0.033
France	0.335	0.002	0.034	0.025	0.059	0.000	0.007	0.008	0.104	0.028	0.069
United Kingdom	0.258	0.010	0.027	0.021	0.051	0.002	0.000	0.004	0.113	0.010	0.019
Greece	0.228	0.001	0.040	0.024	0.036	0.008	0.003	0.003	0.059	0.037	0.017
Ireland	0.235	0.006	0.011	0.024	0.025	0.001	0.001	0.006	0.103	0.017	0.040
Iceland	0.167	0.001	0.014	0.049	0.027	0.001	0.004	0.003	0.061	0.000	0.012
Italy	0.178	0.008	0.006	0.011	0.016	0.024	0.003	0.000	0.061	0.028	0.023
Liechtenstein	0.323	0.001	0.058	0.008	0.033	0.000	0.020	0.029	0.050	0.049	0.076
Luxembourg	0.291	0.010	0.007	0.011	0.072	0.009	0.018	0.007	0.102	0.013	0.041
Netherlands	0.271	0.006	0.009	0.020	0.065	0.010	0.018	0.004	0.111	0.004	0.024
Norway	0.195	0.002	0.010	0.013	0.050	0.000	0.006	0.003	0.063	0.006	0.041
Portugal	0.274	0.007	0.000	0.029	0.056	0.009	0.013	0.000	0.072	0.051	0.042
Sweden	0.233	0.001	0.002	0.020	0.052	0.004	0.011	0.004	0.095	0.009	0.034

Table 6: Coefficients on the primary share of public education expenditure in regressions of IOp on that variable; with and without controls.

	Reading		Math		Science	
No controls						
All countries	-0.00217***	(0.00092)	-0.00077	(0.00112)	-0.00152	(0.00105)
Excluding outliers	-0.00300***	(0.00078)	-0.00113	(0.00101)	-0.00172*	(0.00101)
Controlling for GDP and public expenditure in education per pupil						
All countries	-0.00197**	(0.00087)	-0.00013	(0.00120)	-0.00103	(0.00113)
Excluding outliers	-0.00184***	(0.00072)	-0.00181*	(0.00102)	-0.00185*	(0.00108)

*Notes: Regression coefficients of the share of public expenditure in education allocated to the primary level. Dependent variable: IOp in the subject at column header. Standard errors in parentheses. Where indicated, outliers are identified using the method proposed by Besley, Kuh and Welsch (1980). Data source: UNESCO Institute for Statistics database; ***/**/*: significant at 1/5/10%.*

Table 7: Coefficients on tracking in regressions of IOp on that variable; with and without controls.

	Reading		Math		Science	
No controls						
All countries	0.00106*	(0.00059)	0.00130*	(0.00070)	0.00179***	(0.00063)
Excluding outliers	0.00158**	(0.00060)	0.00109*	(0.00062)	0.00160***	(0.00059)
Controlling for GDP and public expenditure in education per pupil						
All countries	0.00148***	(0.00057)	0.00173***	(0.00074)	0.00214***	(0.00068)
Excluding outliers	0.00090*	(0.00047)	0.00175***	(0.00065)	0.00205***	(0.00067)

*Notes: Regression coefficients of tracking (measured as the share of technical and vocational enrollment at the secondary level). Dependent variable: IOp in the subject at column header. Standard errors in parentheses. Where indicated, outliers are identified using the method proposed by Besley, Kuh and Welsch (1980). Data source: UNESCO Institute for Statistics database; ***/**/*: significant at 1/5/10%.*

Figure 1: Inequality in Educational Achievement: countries ranked by standard deviation in Mathematics test scores.

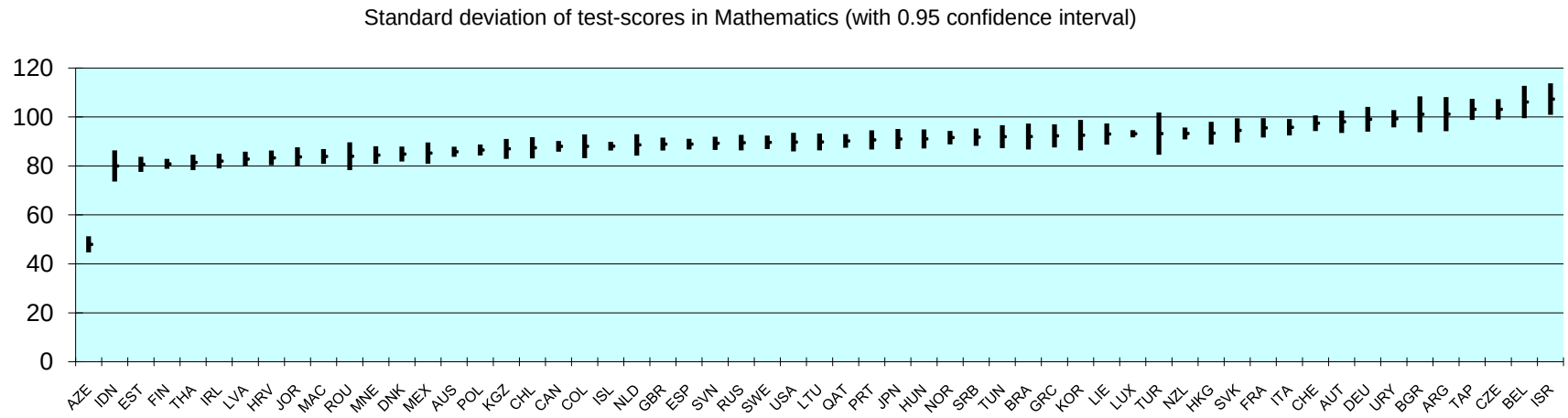


Figure 2: Inequality of Educational Opportunity (IOp): countries ranked by share of variance explained by circumstances.

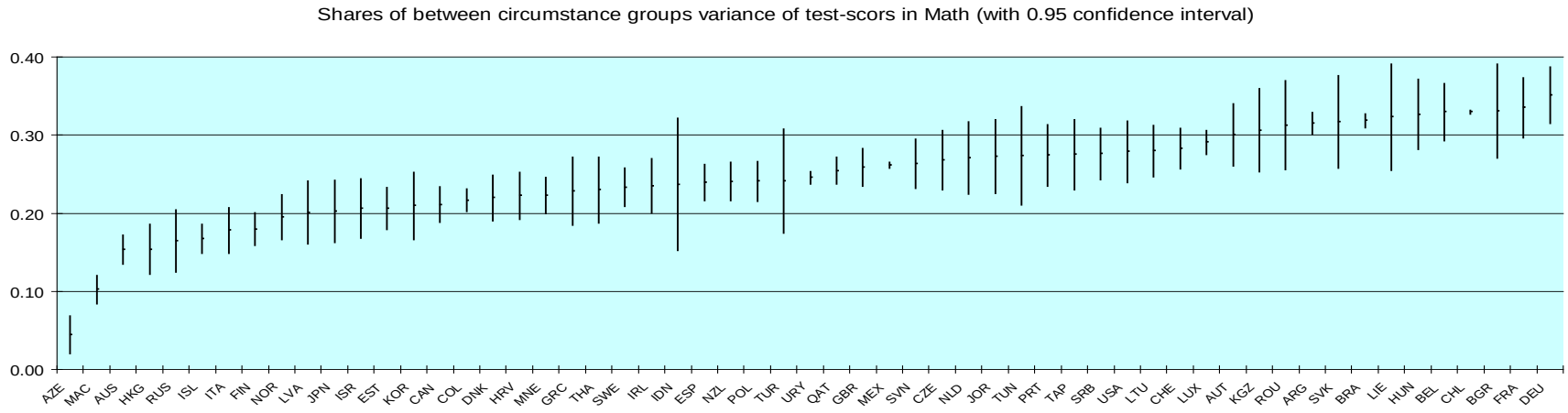
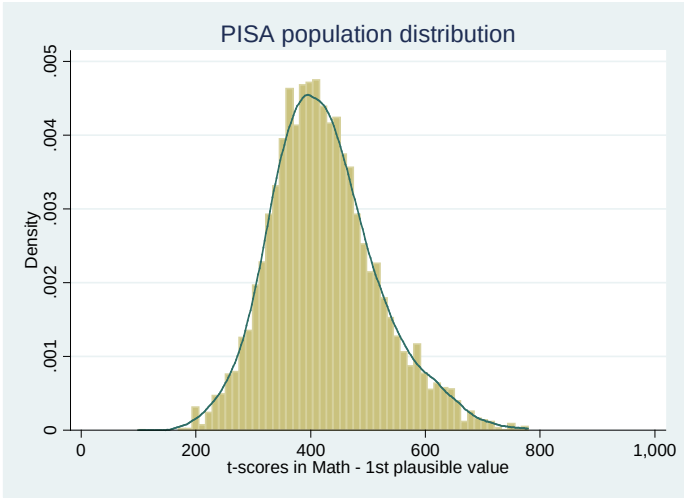
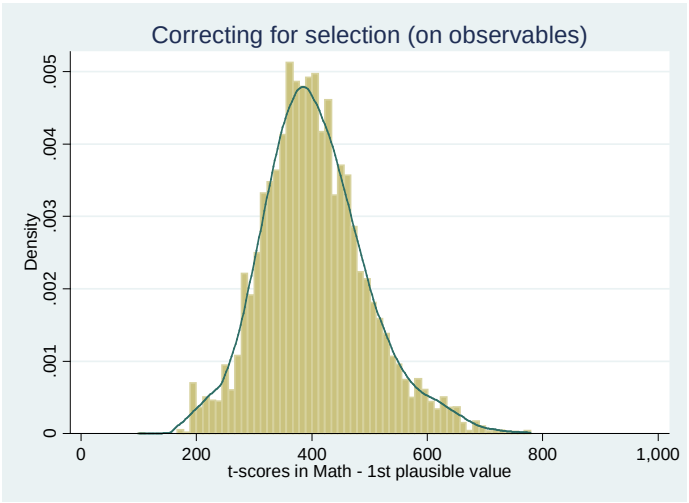


Figure 3: Distribution of standardized Turkish Mathematics test scores under three alternative assumptions about sample selection.





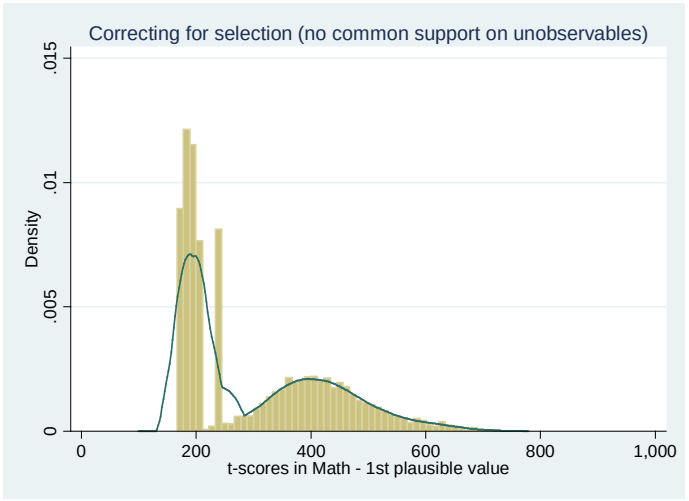


Figure 4: Inequality of educational opportunity and mean achievement

Figure 5: Inequality of educational opportunity and GDP per capita.

Figure 6: Inequality of educational opportunity and public expenditure at the primary level

Figure 7: Inequality of educational opportunity and tracking.

Note: Tracking is measured as the share of enrollment in technical or vocational curricula at the secondary level.