Session Number: Parallel Session 4B
Time: Tuesday, August 24, PM

Paper Prepared for the 31st General Conference of
The International Association for Research in Income and Wealth

St. Gallen, Switzerland, August 22-28, 2010

Educational Inequality and its Intergenerational Persistence: International Comparisons

Francisco H. G. Ferreira and Jérémie Gignoux

For additional information please contact:

Name: Francisco H. G. Ferreira
Affiliation: The World Bank

Email Address: fferreira@worldbank.org

Educational Inequality and its Intergenerational Persistence: International Comparisons

Francisco H. G. Ferreira and Jérémie Gignoux[1]

31 July 2010

Abstract: The standardization of test scores, which is a regular feature of most data on educational achievement, prevents a cardinal interpretation of inequality measures defined over those variables. Many common measures are not even ordinally equivalent to the original inequality in the underlying data. This paper presents comparable, ordinally equivalent measures of inequality in educational achievement for 57 countries for which PISA surveys were conducted in 2006. We also assess the robustness of the measures to sample selection biases arising from the PISA sampling frame. A simple measure of the intergenerational persistence of inequality (IPI) – or inequality of opportunity – is also computed for all countries. This measure is shown to be cardinally insensitive to standardization. As an application, we present its correlations with mean test scores, GDP per capita, the composition of educational spending, and tracking at the secondary school level. Continental European and Latin American countries display greater intergenerational transmission of inequality in educational achievement than nations in Asia, Scandinavia and North America. The share of spending in primary schooling (tracking in secondary schools) is negatively (positively) associated with IPI.

Keywords: Educational inequality; intergenerational transmission; inequality of opportunity; tracking.

JEL Codes: D39, D63, I29, O54

---

[1] Ferreira is the Deputy Chief Economist for Latin America and the Caribbean at the World Bank. Gignoux is at the Paris School of Economics. This paper was prepared for presentation at the 31st IARIW Conference, in St. Gallen, Switzerland (August 2010). We are grateful to participants at the 3rd ECINEQ conference in Buenos Aires in 2009 and, especially, to John Micklewright for helpful comments on an earlier draft. 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. Correspondence: fferreira@worldbank.org and gignoux@pse.ens.fr .

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. Recent research has also found that inequality in achievement and earnings inequality are correlated, both over time within the US 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. At a more aggregate level, it has recently been suggested that countries with higher levels of educational achievement tend to experience faster economic growth (Hanushek and Kimko, 2000; Hanushek and Woessmann, 2008).

For all of these reasons, people care about the distribution of education. Given the links between education and other dimensions of well-being, such as income and political participation, 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 education and its intergenerational transmission – 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". Is the value of one year in the middle of one's doctorate the same as the value of the final grade of high school? Does a student learn the same amount in 6^{th} grade in Zambia as in Finland? Is the value of one year of schooling the same even across schools in a single country or city?

The growing availability of data on student performance in standardized tests has confirmed what one already suspected: that the answer to the last two questions above is 'no'. The quality – and hence ultimately the value – of education varies considerably, both within and across countries. Over the last decade, a number of different projects have compiled school-based surveys that administer identical cognitive achievement tests to representative 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 (separate) 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.[2]

As anyone who has been to school may recall, performance in a test, while probably preferable to a simple measure 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 few recent 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 achievements obtained using different surveys, by comparing the measures and country rankings across surveys. They find broad agreement across surveys, but also some evidence that the specific statistical model used to estimate IRT adjustments does affect results, in particular for less developed countries. Turning to intergenerational persistence, there are also some cross-country comparisons of measures of the association between student achievement and a small number of family characteristics (e.g. Marks, 2005; and Schultz, Ursprung and Wossmann, 2008).

This paper seeks to contribute to that literature by addressing three issues which appear to have been neglected so far. The first is that the standardization process to which test scores are subjected creates some rather serious problems for the interpretation of inequality measures defined over them. Because standardization involves both a translation *and* a re-scaling of the distribution, no inequality index defined over the standardized distribution is *ever* cardinally identical to the same index defined over the pre-standardized (but post-IRT) distribution.[3] Furthermore, various commonly used inequality measures – such as the Gini coefficient or the Theil index – are not even ordinally equivalent in the two distributions: they may rank two countries in opposite ways in the pre- and post-standardization distributions. The variance, however, is shown to be ordinally invariant to standardization and, given the arbitrariness of the mean, its scale-dependence would seem to pose no real problem.

Second, we discuss the sample selection bias that may arise in the specific context of the PISA survey, from the fact that 15 year-olds no longer enrolled in school, or with a grade-age delay greater than two years, are excluded from the sample. We propose two alternative two-sample

[2] There is also an International Adult Literacy Survey (IALS), which is applied to adults long after they have left school.
[3] As we discuss further below, this is a direct implication of the Zheng (1994) impossibility theorem.

non-parametric procedures to assess the robustness of the inequality measure to these biases, and implement it in the four countries for which sample coverage (as a share of the total population of 15 year-olds) is smallest.

Finally, we propose a simple measure of intergenerational persistence of educational inequality which satisfies four desirable properties: (i) it utilizes information on student background more comprehensively than previous studies; (ii) it is cardinally invariant to standardization; (iii) it is directly related to existing measures of inequality of opportunity, and to (inverse) measures of socio-economic mobility; and (iv) it is additively decomposable.

We report our measures of educational inequality and intergenerational persistence of inequality 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.[4] We find considerable variation in the variance of achievement, from lows of around 6,000 (for Indonesia, Estonia and Finland) to highs of near 12,000 (in Belgium and Israel).[5] Similarly stark variation exists in our measure of intergenerational persistence, from 0.10 – 0.15 for Macau (China), Australia, and Hong-Kong (China), up to 0.33 – 0.35 in Bulgaria, France and Germany. Intergenerational persistence of inequality is uncorrelated with either mean scores or with GDP per capita. Broadly speaking, and within the PISA sample, 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. Intergenerational persistence 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 used in this paper. Section 3 considers the implications of test score standardization and of the PISA sampling frame for inequality measurement. Section 4 proposes a measure of intergenerational persistence of inequality that derives from both the mobility and the inequality of opportunity literatures, 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

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

[5] 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 assumption of "extreme" selection on unobservables, the variance of math scores for Indonesia triples. See below.

surveys in the two-sample non-parametric sample selection correction procedures described in Section 3. We briefly describe each of these in turn.

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). 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 Liechstenstein to 30,971 in Mexico. Table 1 lists all participating countries in the 2006 round, as well as their sample sizes.

In each country, the sample consisted of fifteen year-olds enrolled in any educational institution, and attending grade 7 or higher. All children surveyed took three tests: in reading, mathematics, and science.[6] 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 a statistical technique 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:

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

Equation (1) gives the probability of scoring *x* 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(x),* the IRT model can be used to back up a distribution of the latent variable θ.[7]

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 will now denote *x*) has been generated, this latter variable is standardized, according to a simple formula such as:

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

[6] The data for achievements in Reading for the United States were not issued after a problem occurred during the field operations in that country.
[7] See Mislevy (1991) and Mislevy et al. (1992) for a more detailed discussion.

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 (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 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. Since this paper is about educational inequality and its intergenerational persistence, we focus on a subset of these covariates that are informative of the family background and other inherited circumstances of the child. There are ten such 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 (used as an indicator or a rural or urban upbringing).[8]

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 measured 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;

[8] Information on a few additional variables was available but not used for the analysis, because of the likely simultaneous determination of students' achievements and those variables. In particular, mother's occupation and the availability of home educational resources (such as a desk or a computer devoted to home study) were not used. The inclusion of these variables does not qualitatively affect any of our results (results available from the authors on request).

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). School location information was not collected in France, Hong-Kong, 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 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 years olds arising from this process are not taken into account.

The PISA sample coverage rate varies considerably across countries, and is reported in column 3 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 problems. 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 samples. While these samples may have their own coverage 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 by each country's national statistical authority.

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) 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:

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

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:

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

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 its 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:

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

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.

Fortunately, this does not mean that the measurement of inequality in IRT-adjusted educational achievement must rely exclusively on rank-based measures such as the inter-quartile range or percentile differences which, because they do not use information on the entire distribution, do not satisfy symmetry or the transfer principle. It turns out that the variance of a post-standardized distribution (V_j^y) 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:

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

This gives rise to our third remark:

Remark 3: The variance is ordinally invariant to standardization.

The variance is seldom used as an inequality measure because it is scale-dependent: it increases with the mean. The coefficient of variation – the square root of the variance divided by the mean – is scale-invariant, and is often used instead. In addition, for skewed distributions like most income and wealth distributions, the variance (and the coefficient of variation) are often thought to place unduly large weight on observations at the top of the distribution. Neither of these features of the variance strikes us as problematic in the present context, however. IRT-adjusted distributions of test scores have an indeterminate metric, so that scale-dependence seems of little import. And the standardization procedure normalizes the distribution, so that weights on the right tail of the distribution are less of a problem than, say, in a lognormal income distribution. In addition, unlike percentile differences or ratios, the variance is a meaningful measure of inequality in the precise sense that it satisfies axioms (i)-(iii) above.

Columns 4-12 in Table 1 present the mean and variance 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 variance column reports its bootstrapped standard error. Among the countries with higher inequality in math scores are West 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 inequalities 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 variance (and its confidence interval) for the mathematics test scores in all countries in Table 1.

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.

We do not 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 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 to assess the sensitivity of our measure of inequality (and of our measure of its intergenerational persistence, discussed below) to alternative assumptions about the sample selection process.

[9] The inequality measures obtained for Azerbaijan seem particularly small and put the country as an outlier in all the analyses. It is unclear to us 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.

[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.

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:

$$f_j(y) = \iiint \Phi_{\mathrm{j}}(y, X)dX = \iiint g_j(y|X)\phi_j(X)dX \tag{7}$$

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:

$$f_j^{SO}(y) = \iiint \Phi_{\mathrm{j}}(y, X)dX = \iiint g_j(y|X)\psi_j(X)\phi_j(X)dX \tag{8}$$

Where

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

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. used the ratio of densities across different years, we use the ratio of densities across surveys. 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. Equation (9) used the ratios of densities in each cell in these partitions to construct the reweighting function, and both variances and IPI measures were computed over the counterfactual density of scores given by (8).

This procedure assumes that selection into the PISA sample (i.e. into the set of fifteen year olds that take the PISA tests in a particular country) 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

[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.

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 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 even 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 to 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 reading 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.[13] To economize on space, Table 3 reports the effects of these ‘selection correction’ procedures both on the variance of test scores and on our measure of intergenerational persistence, 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.

The results in Table 3 provide a mixed message. Somewhat surprisingly, both educational inequality (measured by the variance) and its inter-generational persistence 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, the variances in achievement increase by two to threefold in Brazil, Indonesia and Turkey, and by up to fourfold in Mexico. The intergenerational persistence of inequality 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

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

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 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 that we do not observe both in the PISA and the available ancillary household surveys. 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 Intergenerational Persistence of Educational Inequality

At least as important as the total level of inequality in education is the question of how persistent that inequality is across generations. While many find some inequality in achievement – that may reflect differences in effort, for instance, or perhaps even differences in innate ability – quite acceptable, it is common to come across arguments against the intergenerational transmission of that inequality. The normative argument against persistence is essentially that it is a form of inequality of opportunity: differences in achievement persist today 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.[14] 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]

Intergenerational persistence is usually seen as the converse of intergenerational mobility – and often measured as its complement. In the canonical Galton regression of a child's outcome (y_t) on the parent's outcome (y_{t-1}):

$$y_t = \beta y_{t-1} + \varepsilon_t \tag{10}$$

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

[14] The literature is vast, but see Cohen (1989), Dworkin (1981), Roemer (1998) and Fleurbaey (2008) for some of the classic references.

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

the R^2 of (10) which is, of course, also the square of the correlation coefficient between the two outcomes in the population: [16]

$$R^2 = \frac{Var(\beta y_{t-1})}{Var(y_t)} = \frac{Cov^2(y_{t-1}, y_t)}{Var(y_t)Var(y_{t-1})} = \rho^2_{t,t-1}$$

Although the PISA data sets do not contain a measure of scholastic achievement for the parents of the fifteen year-old test takers, they do contain information on various aspects of the child's background. In particular, as noted in Section 2, all surveys record information on 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.[17] We therefore use as our measure of intergenerational persistence of inequality the R^2 of a regression of the child's test score on the vector ***z*** of those ten circumstances:

$$y_t = z_t{}'\beta + \eta_t \tag{11}$$

Our measure of IPI is given by:[18]

$$\theta = \frac{Var(z_t{}'\hat{\beta})}{Var(y_t)} \tag{12}$$

In addition to its simplicity, and direct relation to the educational mobility literature, this measure of IPI has four attractive features. First, it allows for the use of more information on family background 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.[19]

Second, the IPI measure defined in (12) is a special case of a measure of inequality of opportunity proposed by Ferreira and Gignoux (2008), and closely related to those discussed by Bourguignon, Ferreira and Menéndez (2007) and Checchi and Peragine (2005). These authors note that a society can be partitioned into groups with identical circumstances (corresponding, in this case, to the elements of the ***z*** vector). 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 $z_i = z_j, \forall i, j | i \in T_k, j \in T_k, \forall k$. They calculate a lower-bound estimate of inequality of opportunity as $\theta_{IO} = \frac{I(\{\mu_i^k\})}{I(y)}$, where $\{\mu_i^k\}$ is the smoothed distribution corresponding to the distribution *y* and

[16] 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. Although the conceptual distinctions are important, we do not dwell on them here. By focusing on the correlation coefficient, we are implicitly adopting 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 a clear and appropriate meaning.

[17] The test-taker's gender is seldom used in mobility studies, but typically included as an important circumstance in studies of inequality of opportunity. Given the isomorphism between the two which is noted below, we include gender in this analysis.

[18] In (11), the hat denotes an OLS estimate.

[19] Schultz, Ursprung and Wossmann (2008), for example, focus on the number of books at home.

the partition *Π*.[20] Although θ_{IO} can be computed non-parametrically by means of a standard between-group inequality decomposition (provided the chosen inequality *I()* is properly decomposable), this procedure is data-intensive when the vector ***z*** 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_{IO}. Following Bourguignon et al. (2007), Ferreira and Gignoux (2008) then propose a parametric alternative for θ_{IO}, based on an OLS regression of y on z:

$$\theta_{IO} = \frac{I(z_t{}'\hat{\beta})}{I(y)} \tag{13}$$

Of course, our measure of intergenerational persistence of inequality in equation (12) is simply a special case of (13), when the variance is used as the inequality measure.

The third attractive feature of (12) as a measure of intergenerational persistence of inequality in the context of educational achievement is that, unlike any measure of the level of inequality (see Remark 1 above), this inequality ratio is actually cardinally invariant in the standardization. To see this, note that any sub-group mean is affected by standardization in a manner analogous to equation (3), so that

$$Var\{\mu_i^k(y)\} = \left(\frac{\hat{\sigma}}{\sigma}\right)^2 Var\{\mu_i^k(x)\} \tag{14}$$

Given (14) and equation (6), it follows that $\theta_{IO} = \frac{Var(\{\mu_i^k(y)\})}{Var(y)} = \frac{Var(\{\mu_i^k(x)\})}{Var(x)}$.

Finally, a fourth attractive feature of this IPI measure is that it is neatly decomposable into components for each individual variable in the vector z. Equation (12) can be rewritten as:

$$\theta = (\operatorname{var} y)^{-1}\left[\sum_j \beta_j^2 \operatorname{var} z_j + \frac{1}{2}\sum_k \sum_j \beta_k \beta_j \operatorname{cov}(z_k, z_j)\right] \tag{15}$$

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

$$\theta = \sum_j \theta^j = \sum_j (\operatorname{var} y)^{-1}\left[\beta_j^2 \operatorname{var} z_j + \frac{1}{2}\sum_k \beta_k \beta_j \operatorname{cov}(z_k, z_j)\right] \tag{16}$$

Having separately regressed test scores for each subject (in each country) on the vector ***z*** of family background variables (equation 11), and computed the R^2 of each regression, we report them on Table 4. These are our estimates of the intergenerational persistence of inequality (IPI) given by equation (12). 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

[20] 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).

accounted for by gender and family background in each country.[21] Bootstrapped standard errors are reported next to each IPI measure. The IPI 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.

Figure 2 provides the same results graphically for achievements in mathematics, after ranking the countries by the IPI measure. 95% confidence intervals are presented using the estimates of the 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 rates of intergenerational persistence of inequality, with shares above 30%, are Western European countries (such as Belgium, France, and Germany) but also Eastern European countries (such as Bulgaria, the Czech Republic, and Hungary), and Latin American countries (such as Argentina, Brazil and Chile). Among the countries with the lowest degrees of persistence, with shares below 20%, are Asian countries (such as Azerbaijan, Macao-China, Hong-Kong, and Japan), Nordic countries (such as Finland, Iceland, and Norway), Russia, Australia, but also other European countries (such as Italy, Estonia and Latvia). The United States, the UK, and Spain lie in an intermediate position with shares close to 25%.

One can use these results to make specific comparisons. For example, the degree of intergenerational transmission of educational inequality 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 IPI 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 the intergenerational persistence of inequality is mirrored in the absence of a correlation between IPI and either the level of educational achievement, as measured by mean test scores, or the level of economic development, as measured by GDP per capita.[22] Figure 4 plots the relationship between IPI and mean achievement in mathematics. The regression line and a 95% confidence

[21] The measure θ is a lower bound because there may be omitted family background (or circumstance) variables that are not included in z. Their inclusion might raise, but cannot lower, the R^2 of the regression. The existence of omitted variables of all sorts is also the reason why estimated coefficients in our test-score regressions are of very limited interest, and not presented. While individual coefficients may be biased either upwards or downwards, the R^2 can of course only rise as additional variables are included in the regression. See Ferreira and Gignoux (2008) for an analogous discussion on the measurement of inequality of opportunity.

[22] GDP per capita is measured in purchasing power parity in 2006 US dollars; the data are from the World Development Indicators (WDI) database.

interval are shown on the graphs. The regression coefficient is statistically insignificantly different from zero at the 10% level. Figure 5 plots IPI 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 overall persistence into partial shares by individual background variable, described in equation (16), is presented in Table 5 for mathematics scores. The shares of each of the nine background variables, and gender, add up to the total IPI given in the first column. As may be seen from inspection of equation (16), these partial shares are a function of individual regression coefficients from (11). Those 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 IPI.

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 has an important source of learning inequalities. Father's occupation and the three "asset" indicators 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.

5. A descriptive application: correlations between intergenerational persistence and education policies

As an illustration of potential applications, we now briefly investigate the cross-country correlation between the measure of intergenerational persistence of inequality developed in the previous section and two specific educational policies: 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 earlier than others, the allocation of resources to the primary level of schooling is generally thought more likely to have redistributive effects.

The impacts of tracking policies on the efficiency and equity in educational systems are another example of education policies having received attention in recent studies (Ariga et al. 2006, Brunello and Checchi 2007, Brunello et al. 2006, Hanushek and Woessman 2006, Manning and Pisckhe 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[23]. 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 exacerbate the intergenerational persistence of educational inequality.

We briefly examine the correlation between our measure of IPI and these two policies, using data on the policy indicators from the UNESCO Institute for Statistics (UIS).[24] 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, 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

[23] 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).

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

second place at 44.000 US dollars).

There is considerable variation in the share of expenditure allocated to the primary level of education in the remaining country sample. While the mean share is 25.8%, the lowest share is observed in Romania at 13.8% and the highest in Colombia at 41.8% (the first quartile is at 20.2% and the third quartile at 30.4%). Figure 6 provides an illustration of the relationship between the primary share of expenditures and IPI. 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 (in row 1) and controlling for per capita GDP and public education expenditure per pupil (row 2). Once outliers are excluded, significant negative correlations exist both for reading and science, with or without controls for per capita GDP and expenditure per pupil. 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 the share of educational inequality that is transmitted across generations.

There is also considerable heterogeneity in tracking (measured as the share of technical and vocational enrollment at secondary levels) in our country sample. The mean share is 19.8 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 above, Figure 7 provides a picture of the relationship between tracking and IPI in this sample, while Table 7 lists coefficients and standard errors, both without any controls (row 1) and controlling for per capita GDP and public education expenditure per pupil (row 2). There is a clear pattern of significant positive relationships across both subjects and regression specifications. Higher intergenerational persistence tends to be associated with higher shares of technical and vocational enrollment. The coefficients of correlations lie between 0.001 and 0.002 for IPI, 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 of the shares of between circumstance groups learning inequalities.

These simple correlations suggest that our measure of intergenerational persistence of inequality 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. The associations reported here 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.[25] These analyses remain descriptive in nature, and do nothing for controlling for the heterogeneity in the education systems or

[25] However, the long term effects of early tracking remain debated. For instance, Brunello and Checchi (2007) find that although it tends to increase the link between family background and educational attainments by diverting some individuals to progress to the tertiary level of education, it seems to reduce the impacts of family background on adult literacy and promote further on-the-job training by offering more effective curricula to less well performing students.

populations of pupils. However they illustrate the potential use of indicators of intergenerational transmission – or inequality of opportunity – in studying 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 – and others like them – 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, but also 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 of educational achievement, and which appear to have been overlooked so far. The first is standardization of test scores, to which all meaningful measures of inequality are cardinally sensitive. Compounding the problem, 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 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 variation in educational inequality thus measured. The variance in Math scores ranges from around 6,000 in Indonesia, Estonia and Finland, to nearly 12,000 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 grade 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 in the sample is clearly correlated with determinants of test scores, leading to a classic sample selection bias problem. 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 and intergenerational persistence, under different assumptions. 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 education and father’s occupation). Under a much more stringent scenario of strong selection on unobservables with no common support, however, the measures of inequality in these countries rise many-fold.

Finally, we also propose and apply a measure of the extent to which inequality in educational achievement is transmitted across generations. The measure is simply the share of the total variance in educational achievement that can be accounted for by family background variables (and gender) in a linear regression. It is closely related to the origin-independence concept of mobility, and to an existing measure of inequality of opportunity. It is cardinally invariant to the standardization of test scores, and exactly additively decomposable into the partial shares accounted for by individual circumstance variables.

Thus measured, the intergenerational persistence of inequality in our sample of countries ranges from approximately 0.10 – 0.15 in Macau (China), Australia, and Hong-Kong (China), up 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 and consistently correlated with the extent of educational tracking, defined as the share of technical and/or vocational enrollment in secondary schools.

This paper has not conducted any causal analysis of the determinants of inequality, or of its intergenerational transmission. 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", forthcoming 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 Vitoroco Peragine (2005): "Regional Disparities and Inequality of Opportunity: the Case of Italy", IZA Discussion Paper No. 1874/2005.

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 (2008): "The Measurement of Inequality of Opportunity: Theory and an application to Latin America", World Bank Policy Research Working Paper # 4659.

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.

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

Hanushek, Eric and Dennis Kimko (2000): "Schooling, Labor Force Quality, and the Growth of Nations" *American Economic Review*, **90** (5): 1184-1208.

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

Hanushek, Eric and Ludger Woessmann (2008): "The Role of Cognitive Skills in Economic Development" *Journal of Economic Literature*, **46** (3): 607-668.

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

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.

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 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 sizes, coverage rates, mean scores and overall inequality in PISA test scores

	# Obs.	Coverage rate	Mean in Reading	Variance in Reading (and SE)		Mean in Math	Variance in Math (and SE)		Mean in Science	Variance in Science (and SE)	
Asia & North Africa											
Azerbaijan	5184	0.88	355.0	4935.8	296.3	476.8	2300.5	157.6	385.3	3100.3	213.8
Hong Kong-China	4645	0.97	538.9	6689.3	314.6	551.4	8721.8	432.9	546.1	8410.8	351.6
Indonesia	10647	0.53	383.9	5593.2	357.5	380.7	6401.9	508.5	384.8	4908.9	457.4
Israel	4584	0.76	441.3	14242.3	665.2	443.3	11519.5	687.1	455.6	12421.4	428.9
Japan	5952	0.89	409.5	10483.2	480.4	389.2	8283.9	374.5	427.1	10024.8	403.0
Jordan	6509	0.65	500.2	8853.7	422.2	525.6	7006.7	326.8	533.7	8075.3	340.4
Korea	5176	0.87	290.5	7795.6	472.0	315.9	8572.4	576.8	326.3	8110.5	423.3
Kyrgyzstan	5904	0.63	556.1	10425.4	511.8	547.2	7566.1	351.7	521.9	7033.4	338.5
Macao-China	4760	0.73	490.6	5830.6	348.4	524.4	7038.8	255.5	509.5	6058.1	247.4
Qatar	6265	0.90	312.5	11690.6	248.9	317.7	8143.2	249.4	349.1	6937.7	227.9
Russian Federation	5799	0.81	442.4	8691.4	348.1	478.7	8016.0	283.3	481.5	8023.8	238.6
Chinese Taipei	8812	0.88	506.7	7119.8	292.3	562.7	10632.6	444.3	543.7	8920.7	307.1
Thailand	6192	0.72	425.2	6699.2	283.0	425.5	6631.0	255.1	429.7	5955.5	223.9
Tunisia	4640	0.90	379.0	9467.7	483.5	363.9	8454.5	430.6	384.2	6787.3	336.4
Turkey	4942	0.47	452.9	8631.1	510.5	428.2	8693.9	803.1	427.6	6923.2	521.8
Latin America											
Argentina	4339	0.79	383.9	15431.1	899.7	388.1	10230.3	192.6	398.3	10250.1	531.2
Brazil	9295	0.55	389.2	10497.6	685.4	365.6	8467.3	60.7	385.3	7970.9	344.2
Chile	5233	0.78	447.9	10658.3	503.6	417.1	7645.4	115.6	443.1	8405.1	316.5
Colombia	4478	0.60	390.3	11628.4	511.4	373.8	7750.9	118.5	391.9	7192.5	306.5
Mexico	30971	0.54	427.4	9154.5	433.9	420.7	7270.6	58.6	422.6	6513.4	237.3
Uruguay	4839	0.69	424.7	14694.7	491.4	435.5	9861.0	184.2	437.7	8919.0	326.9
North America & Oceania											
Australia	22646	0.87	508.7	9263.7	276.9	516.3	7359.7	176.2	523.1	8871.4	214.2
Canada	14170	0.87	512.3	8795.8	187.2	517.4	7749.0	191.2	522.5	10045.9	204.0
New Zealand	4823	0.84	522.7	11069.3	331.7	523.8	8700.0	224.1	532.7	11513.0	290.8
United States	missing	0.85				474.7	8055.6	341.7	488.3	11251.7	355.5
Eastern Europe											
Bulgaria	4498	0.83	406.8	13807.7	939.5	417.4	10221.9	736.9	439.1	11389.6	680.9
Czech Republic	5932	1.01	509.6	12368.4	643.7	536.0	10637.7	429.8	537.6	9685.0	394.0
Estonia	4865	0.94	502.4	7257.9	318.5	516.8	6509.3	247.9	533.7	7013.6	182.8
Croatia	5213	0.85	477.6	7890.4	376.7	467.3	6940.8	250.9	493.7	7347.1	247.5
Hungary	4490	0.85	488.1	8910.4	448.0	496.2	8288.2	351.8	508.7	7779.3	269.3
Lithuania	4744	0.93	469.3	9128.7	287.2	485.6	8064.5	309.5	486.5	8097.7	273.4
Latvia	4719	0.85	484.9	8226.9	306.0	491.2	6857.7	251.0	493.8	7119.4	218.7
Montenegro	4455	0.84	388.2	7994.3	295.4	395.8	7131.9	305.2	408.8	6350.7	191.2
Poland	5547	0.94	512.6	10043.6	296.2	500.9	7485.7	195.8	503.3	8077.1	199.1
Romania	5118	0.66	392.0	8438.8	539.4	415.0	7051.7	479.5	416.6	6586.6	385.3
Serbia	4798	0.83	402.9	8434.5	309.6	436.6	8419.1	323.8	436.9	7249.9	265.4
Slovak	4731	0.95	470.6	11042.1	526.1	495.1	8936.3	466.3	491.2	8676.5	332.2

Republic											
Slovenia	6595	0.88	468.6	7738.2	438.2	482.2	7965.2	243.2	494.2	9625.5	265.7
Western Europe											
Austria	4927	0.92	494.0	11698.0	686.2	509.5	9616.3	449.5	513.9	9570.0	471.2
Belgium	8857	0.99	507.1	12104.5	616.6	526.9	11263.4	703.3	516.3	9941.0	399.0
Switzerland	12192	1.02	496.6	8849.4	321.9	528.3	9495.2	311.6	508.0	9863.4	320.1
Germany	4891	0.95	496.5	12532.1	596.9	504.3	9817.2	501.6	516.2	9996.9	397.3
Denmark	4532	0.85	493.8	7975.3	290.1	512.2	7199.9	258.5	494.7	8673.2	264.0
Spain	19604	0.87	479.5	7892.9	203.1	501.7	7906.8	193.2	504.5	8197.2	176.4
Finland	4714	0.93	547.1	6597.8	175.8	549.0	6539.5	164.0	563.4	7331.6	171.1
France	4716	0.91	488.7	10805.1	571.5	496.4	9134.9	375.6	496.1	10317.6	426.2
United Kingdom	13152	0.94	495.6	10387.2	343.9	497.3	7906.7	233.8	514.3	11403.9	321.4
Greece	4873	0.90	461.9	10528.9	597.5	462.0	8518.9	437.1	476.6	8487.2	373.2
Ireland	4585	0.94	518.6	8536.7	343.2	502.3	6721.8	245.5	509.5	8902.1	284.1
Iceland	3789	0.96	485.0	9426.6	238.3	505.6	7758.6	156.7	491.0	9382.9	183.7
Italy	21773	0.90	477.0	11829.3	378.3	473.6	9181.5	318.7	487.2	9131.8	250.1
Liechtenstein	339	0.84	510.7	9054.6	556.2	524.9	8659.3	404.2	522.3	9401.1	406.4
Luxembourg	4567	1.03	480.1	9969.3	144.5	490.5	8677.1	135.6	486.8	9317.4	128.8
Netherlands	4871	0.96	513.9	9335.4	476.8	537.4	7850.6	385.7	530.8	9146.0	313.5
Norway	4692	0.97	484.4	11056.4	403.3	489.8	8386.6	253.3	486.9	9238.9	380.7
Portugal	5109	0.78	476.8	9765.1	449.7	470.9	8218.2	358.3	479.0	7843.2	303.5
Sweden	4443	0.97	509.0	9646.1	347.9	503.2	8038.8	246.4	504.2	8875.4	264.0

Note: The simple variance 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 variance 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 and its Intergenerational Persistence 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										
	Total variance	8631.1	8693.9	6923.2	9678.0	8360.1	6819.9	24231.6	17965.5	14790.0
		510.5	803.1	521.8						
	IPI	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										
	Total variance	10497.6	8467.3	7970.9	10579.2	8179.8	7525.6	32336.8	21514.8	21366.9
		685.4	60.7	344.2						
	IPI	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										
	Total variance	9154.5	7270.6	6513.4	9144.9	7228.4	6269.4	38749.9	26500.0	18766.3
		433.9	58.6	237.3						
	IPI	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										
	Total variance	5593.2	6401.9	4908.9	5045.1	5816.8	4322.3	17045.9	18466.6	12722.4
		357.5	508.5	457.4						
	IPI	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: IPI denotes the measure of intergenerational persistence of inequality, defined in equation (12). It is the share of the total variance in test scores which is accounted for by the student's gender and nine family background variables.

Table 4: Intergenerational Persistence of Inequality Measures: PISA test scores

	IPI Reading	Standard Error (Reading IPI)	IPI Mathematics	Standard Error (Math IPI)	IPI Science	Standard Error (Science IPI)
Asia & North Africa						
Azerbaijan	0.173	0.028	0.044	0.012	0.112	0.024
Hong Kong-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: IPI denotes the measure of intergenerational persistence of inequality, defined in equation (12). It is the share of the total variance in test scores which is accounted for by the student's gender and nine family background variables.

Table 5: Partial shares of the total variance in mathematics scores: decomposing IPI into individual components

	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-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 IPI 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: IPI 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 IPI 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: IPI 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 total variance in Mathematics test scores.

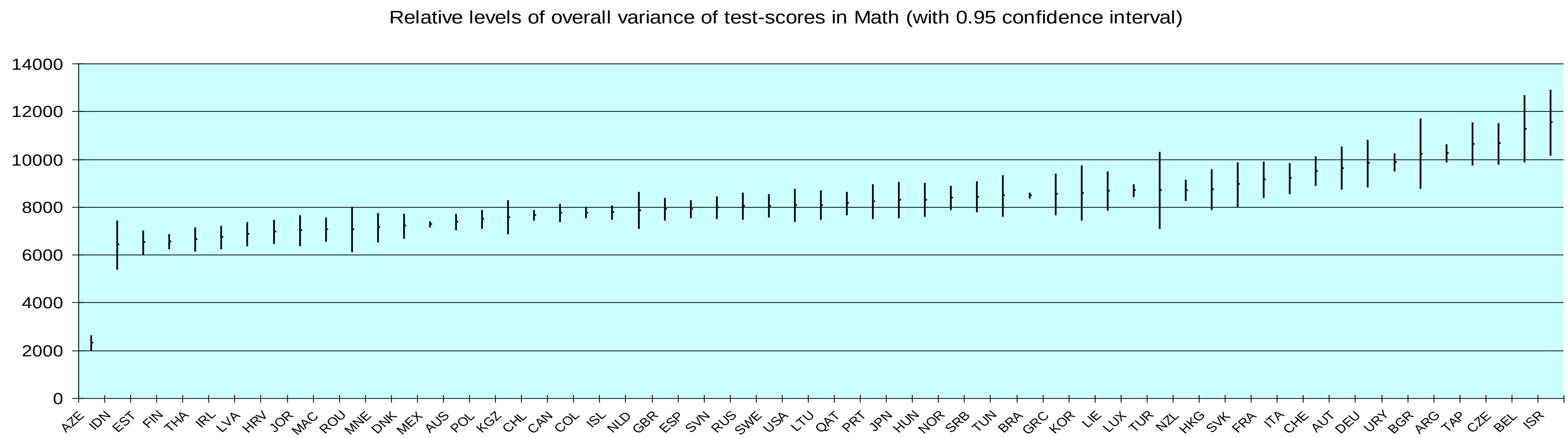


Figure 2: Intergenerational Persistence of Inequality: countries ranked by share of variance explained by background factors.

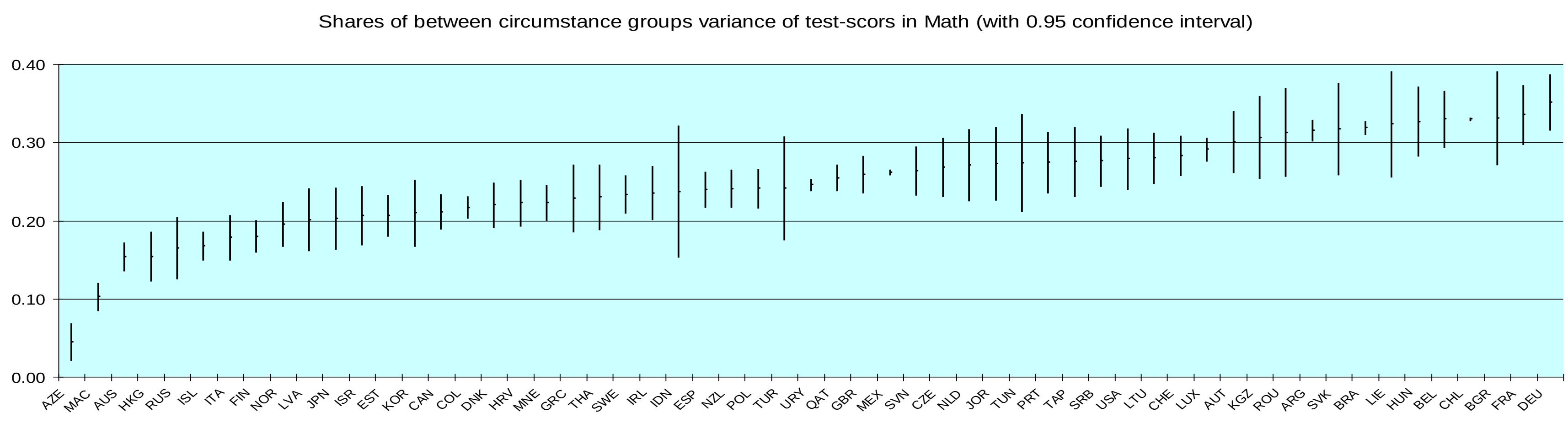


Figure 3: Distribution of standardized Turkish reading test scores under three alternative assumptions about selection into PISA participation

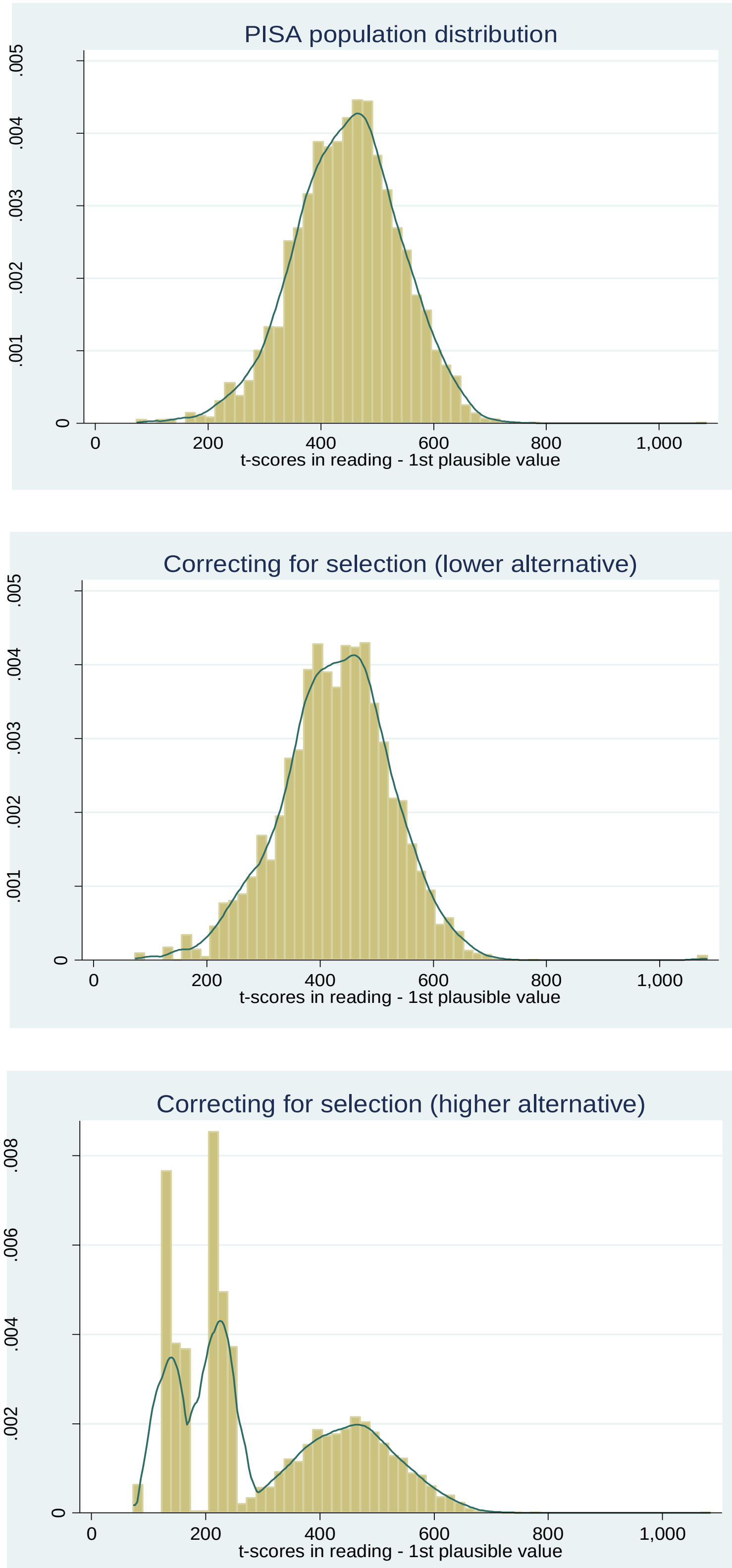


Figure 4: Intergenerational transmission of educational inequality and mean test scores

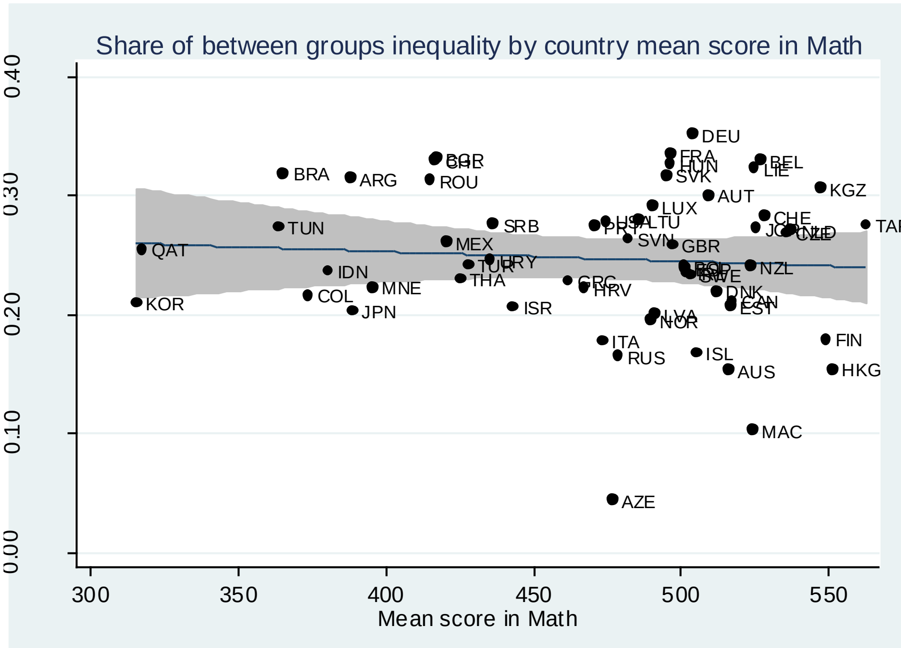


Figure 5: Intergenerational transmission of educational inequality and GDP per capita.

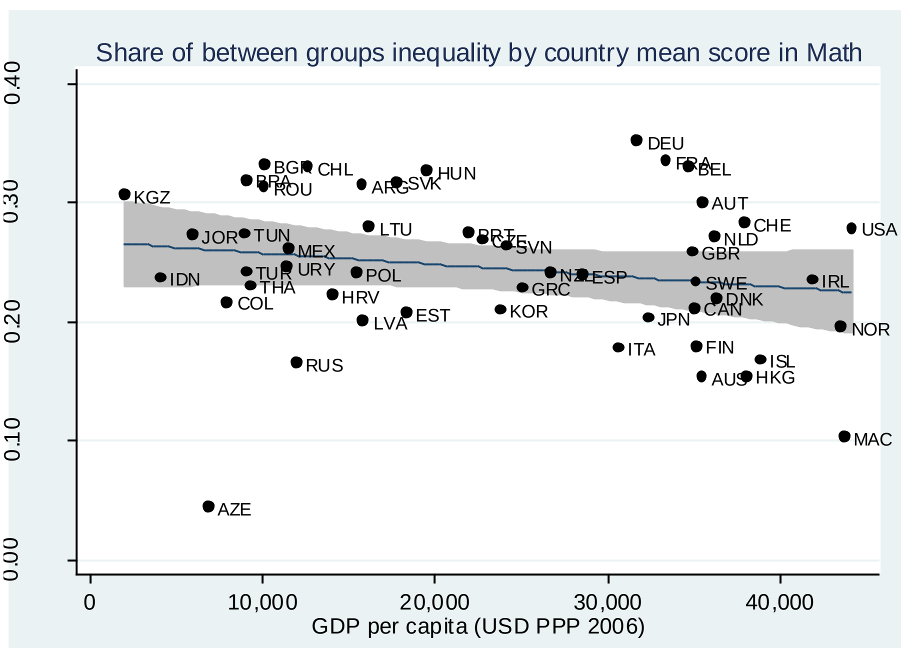


Figure 6: Intergenerational transmission of educational inequality and public expenditure at the primary level (as a share of total educational spending).

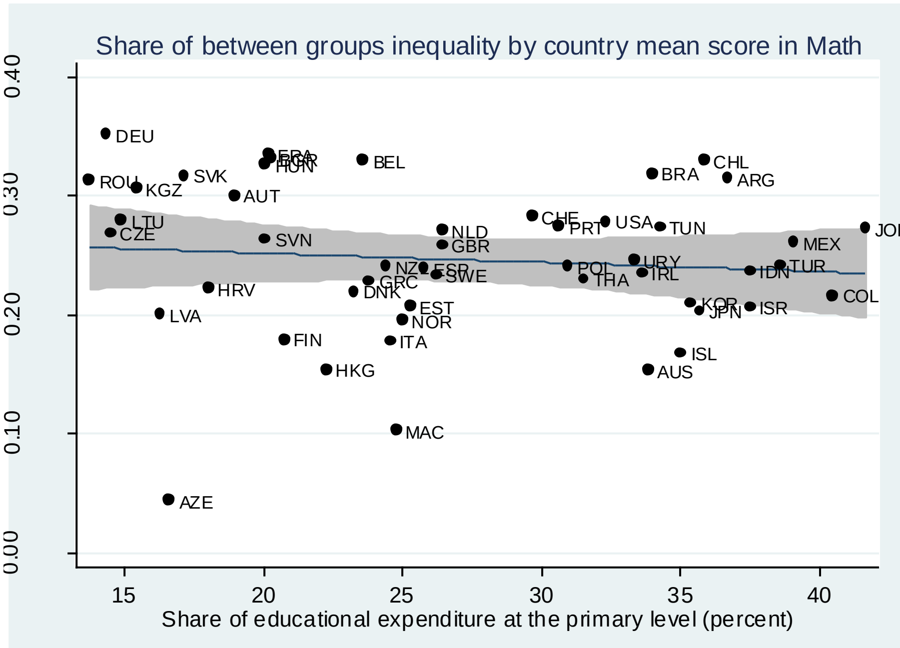
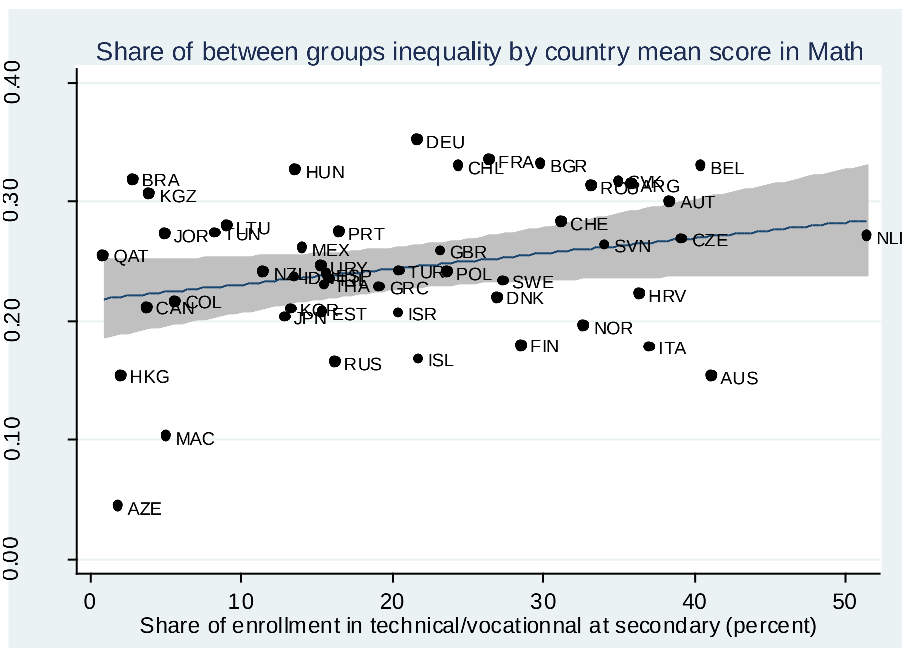


Figure 7: Intergenerational transmission of educational inequality and tracking.



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