KYKLOS, Vol. 61 – 2008 – No. 2, 279–308

Schütz, G., Ursprung, H. W., & Wößmann, L. (2008). Education policy and equality of opportunity. *KYKLOS: International Review for Social Sciences*, *61*(2), 279--308. https://doi.org/10.1111/j.1467-6435.2008.00402.x

Education Policy and Equality of Opportunity

Gabriela Schütz, Heinrich W. Ursprung and Ludger Wößmann*

I. INTRODUCTION

Equality in educational outcomes is a crucial determinant of the extent of equality of opportunity and intergenerational mobility achieved by societies. For example, Nickell (2004) shows that a large part of the existing cross-country variation in earnings inequality can be attributed to variation in skill dispersion. It is therefore of prime policy interest to understand the effects of education policies such as ability tracking, pre-school education, length of the school day and spending on the educational success of children from various family backgrounds. However, the direction of these effects is by no means straightforward from a theory point of view, and empirical evidence is limited. The lack of evidence derives from the lack of variation in the organization of education systems within most countries. Also, where such variation exists, it is unlikely exogenous to students' performance and backgrounds.

This paper makes two contributions to reduce this lack of evidence. First, it provides a comparable measure for 54 countries of how strongly children's educational performance is related to their family background (Section II). We interpret this measure as a proxy for the extent of inequality of opportunity. Second, we use the cross-country variation in education policies and its interaction with family background at the individual student level to identify the impact of education policies on equality of opportunity (Section III).

The database used combines two related extensive international student achievement tests, the Third International Mathematics and Science Study (TIMSS) and its replication for a partly different set of countries (TIMSS-

* Schütz: Ifo Institute for Economic Research, Poschingerstr. 5, 81679 Munich, Germany, schuetz@ifo.de; Ursprung: Department of Economics, University of Konstanz and CESifo, Box D-138, 78457 Konstanz, Germany, Heinrich.Ursprung@uni-konstanz.de; Wößmann: University of Munich, Ifo Institute for Economic Research, CESifo and IZA, Poschingerstr. 5, 81679 Munich, Germany, woessmann@ifo.de. For helpful comments and discussion, we would like to thank Steve Machin, Winfried Pohlmeier, Joachim Winter and the editors, as well as participants at the joint SOLE/EALE World Conference on Labor Economics in San Francisco, the conferences of the International Economic Association in Marrakech and the German Economic Association in Bonn, and at seminars at the universities of Louvain-La-Neuve, Dortmund, Zurich and St. Gallen.

Repeat) (Section II.1). These datasets provide information on students' educational performance, their family background and relevant control variables for individual students in each participating country. Our main indicator of family background is the number of books in the students' home. As suggested in the sociological literature, books at home provide a powerful proxy for the educational, social and economic background of the students' families. The validity of books at home as a family-background proxy in cross-country comparisons is corroborated by our separate evidence from a different dataset that there is no cross-country difference in the association between books at home and household income. Moreover, previous research suggests that in most countries, books at home are the single most important predictor of student performance, even surpassing parental education (Wößmann 2003, 2008). Furthermore, data coverage and cross-country comparability on the books indicator are superior to parental education.

This database allows us to estimate an index of equality of educational opportunity in 54 countries (Section II.2). More precisely, the index measures the inequality of educational outcomes for children from different family backgrounds. Given the strong relationship between education and economic outcomes, our measure can serve as an index of the inequality of opportunity for children from different family backgrounds later in life. To our knowledge, no previous evidence on the inequality of opportunity across countries has been available on a comparable scale. In estimating our index at the micro level, we make sure that it is not affected by cross-country differences in the immigrant population, but only reflects performance differences associated with socio-economic background.

Our results show that equality of opportunity as measured by our index varies considerably across countries (Section II.3). Among OECD countries, the impact of family background on student performance is largest in England, Scotland, Hungary and Germany and lowest in France, Canada, Portugal and Flemish Belgium. The family-background effect in the former countries is on average 2.7 times as large as in the latter countries. The United States falls in the top quarter of the most unequal OECD countries.

To understand what lies behind the substantial cross-country variation in the extent of equality of educational opportunity, we provide several hypotheses of how different organizational features of the education system may affect equality of educational opportunity (Section III.1). A formal model underpinning the main hypotheses is provided in Schütz et al. (2005, Section III).

Combining the observed variation in the index of educational inequality with country-level data on features of the education systems, we then test the hypotheses empirically. Our preferred empirical identification strategy is to estimate how the different country-level features of the school systems interact with the family-background measure at the student level in determining student performance, while at the same time controlling for unobserved country heterogeneity by country fixed effects (Section III.2).

Our results reveal that the family-background effect is larger (i.e. equality of opportunity is lower), the earlier a country tracks its students into different school types by ability (Section III.3). The family-background effect is also larger in countries with shorter pre-school education. With respect to pre-school enrollment, we find an inverted U-shaped relationship, with educational inequality increasing up to an enrollment of roughly 60 percent and decreasing thereafter. These results prove robust to more extensive model specifications, in which we do not find a statistically significant difference in the equality of opportunity by school starting age or between half-day and whole-day school systems. Neither does the observed equality of opportunity differ with average educational spending, nor with the country's level of economic development. At least in the OECD sample, there is also no significant association between equality of opportunity and a country's mean test score. Finally, the family-background effect is larger in countries with a larger share of private funding, but at the same time, it is smaller in countries with more private school provision.

The results on pre-school education and tracking relate to a substantial literature. Belfield et al. (2006), Garces et al. (2002) and Magnuson et al. (2007) present recent within-country evidence on the effects of early-childhood education, and Barnett (1992), Blau and Currie (2006), Currie (2001) and Cunha et al. (2006) provide extensive surveys of the existing evidence. Analyses of ability tracking within different countries include Bauer and Riphahn (2006), Betts and Shkolnik (2000), Dustmann (2004), Epple et al. (2002), Figlio and Page (2002), Galindo-Rueda and Vignoles (2007), Meghir and Palme (2005) and Pekkarinen et al. (2006). However, as the discussion of Manning and Pischke (2006) reveals, it is very hard to eliminate selection bias from different students attending different types of school even in the best within-country analyses. As a consequence, the first version of this paper has spurred a new literature on cross-country analyses of the effects of tracking. Hanushek and Wößmann (2006) develop a differences-in-differences approach using variation across countries and grades, but use only country-level data and thus cannot identify effects on equality of opportunity. By restricting the analysis to this differences-in-differences framework, the student-level analyses of Ammermüller (2005) and Waldinger (2006) are left with very limited country-level observations, which severely limits statistical inference relative to our 54-country sample. Brunello and Checchi (2007) explicitly adopt the approach developed in this paper, extending the analysis to outcomes later in life. Our paper goes beyond the subsequent contributions that it predates by estimating not only the effects of tracking, but also of pre-school education and several additional education policies in a unified framework, thereby reducing worries of omitted-variable bias from omitting other policies.

In effect, our empirical results show how school systems can facilitate intergenerational mobility. They suggest that education policies such as comprehensive school systems and extensive early-childhood education can increase the equality of educational opportunity for children from different family backgrounds. On the other hand, the empirical evidence suggests that extending the school day into the afternoon, lowering the starting age of compulsory education or increasing educational spending do not appear to be significantly related to equality of educational opportunity. Also, an efficiency-equity tradeoff does not seem to exist in education in the sense that more equal systems would systematically show different mean performance of their students, at least not among OECD countries. Relying on private spending to finance education is associated with increased inequality of educational opportunity, while relying on private production of schooling is associated with increased equality of educational opportunity.

II. ESTIMATING EQUALITY OF OPPORTUNITY ACROSS COUNTRIES

1. Data

The two TIMSS international student achievement studies: To derive estimates of the equality of educational opportunity, we employ student-level micro data from two extensive international student achievement tests. The first test is the Third International Mathematics and Science Study (TIMSS), conducted in 1995 (data released in 1997) by the International Association for the Evaluation of Educational Achievement (IEA), an independent cooperation of national, partly governmental, research agencies. The second test is the TIMSS-Repeat study, conducted in 1999 (data released in 2001) also by the IEA as a replication of the first study. All participating countries received the same test items, so that the ensuing measures of educational performance in math and science are directly comparable across countries.[1] Furthermore, both tests draw random samples of schools to provide representative samples of students in each participating country[2].

1. The development of the test contents was a cooperative process involving national research coordinators from all participating countries, and all participating countries endorsed the curriculum framework. Both studies also performed a test-curriculum matching analysis that restricted the analysis to items definitely covered in each country's curriculum, which made little difference for the overall achievement patterns.
2. Beaton et al. (1996), Gonzalez and Smith (1997) and Martin and Kelly (1996, 1997) provide detailed information on the TIMSS-95 database. For more information on the TIMSS-Repeat database, see Mullis et al. (2000), Martin et al. (2000) and Gonzalez and Miles (2001).

The target population of TIMSS-Repeat was the upper of the two adjacent grades with the largest share of 13-year-olds (usually eighth grade). While the TIMSS-95 study also targeted additional grade levels, we restrict our TIMSS-95 data to the eighth-grade students to ensure comparability. For our analyses, TIMSS-95 yielded internationally comparable data for representative samples of students in 40 countries, and TIMSS-Repeat for 38 countries[3]. Since the sample of participating countries differed considerably between the two tests, the pooled TIMSS-95/TIMSS-Repeat database contains data on more than 300,000 students from 54 different countries, which is the biggest sample of participating countries in comparable international tests to date (see *Table 1* for a list of the countries).

Both studies had basically the same design of a curriculum-valid test. Given that two-thirds of the test items of TIMSS-95 had been released to the public after the study, these items had to be replaced in TIMSS-Repeat by substitute items with similar content, format and level of difficulty. Because of the similarity of the test designs, it is possible to splice the eighth-grade data of the two TIMSS tests together. We do this by singling out the test scores of those 24 countries that participated in both studies and standardizing all scores according to the mean and the standard deviation of this sub-sample. We then standardize the test scores to have a standard deviation of 100 across all countries in the pooled dataset.

This paper uses students' mean test score of math and science as measure of educational performance, pooling the two TIMSS tests for those countries that participated in both studies. *Table 1* reports country mean performance and standard deviation on this variable.

In separate background questionnaires, students provide information on various features of their family background. These features include the number of books in their home (see next sub-section for details), whether they themselves, their mother and their father were born in the country, their family status (living together with both parents or not), their gender and age. *Table 2* contains descriptive statistics on these variables for the international dataset[4].

The family-background proxy: The proxy for the family background of students that we use in our study is the number of books in the students' home. 'Books at home' is a measure of family background proposed and frequently used in sociological research. A large number of books can be interpreted as an indicator for a family environment that highly esteems education and academic success and that will promote children's academic effort (cf. Beaton et al. 1996;

3. England and Scotland, as well as the Flemish and the French school system in Belgium, count as individual countries here as they have separate school systems that participated separately in the tests.
4. The small number of observations with missing data on these variables was dropped from the estimations in this paper.

Table 1

Educational Performance and Family Background by Country

	TIMSS performance		Books at home[a]					Mean category
	Mean	SD	1 (0–10)	2 (11–25)	3 (26–100)	4 (101–200)	5 (200+)	
Australia	541.4	87.1	3.3	7.0	24.0	25.8	40.0	3.9
Austria	546.9	88.7	10.5	17.5	31.4	16.7	23.8	3.3
Belgium (Flemish)	558.2	73.3	14.6	19.4	32.1	16.1	17.7	3.0
Belgium (French)	494.9	78.7	6.9	10.2	28.0	20.7	34.2	3.7
Bulgaria	529.2	83.1	8.7	11.6	23.8	19.2	36.6	3.6
Canada	537.0	75.6	4.5	10.8	28.1	24.6	31.9	3.7
Chile	423.4	78.2	20.0	31.6	28.2	11.0	9.2	2.6
Colombia	390.5	62.4	25.6	31.5	26.6	9.0	7.3	2.4
Cyprus	474.3	79.3	5.4	18.4	35.0	22.9	18.3	3.3
Czech Republic	555.7	79.5	0.8	5.8	31.9	30.7	30.9	3.9
Denmark	486.4	78.3	3.3	8.5	30.0	21.1	37.1	3.8
England	530.0	87.3	6.0	13.0	29.4	22.6	28.9	3.6
Finland	543.5	65.4	3.5	14.0	38.7	21.9	21.9	3.4
France	514.9	67.5	5.4	17.1	36.4	21.1	19.9	3.3
Germany	517.5	89.3	8.1	13.8	26.2	18.8	33.1	3.5
Greece	486.8	80.0	5.0	22.3	42.7	18.2	11.8	3.1
Hong Kong	562.1	77.7	24.8	28.3	28.2	9.6	9.1	2.5
Hungary	550.6	82.0	3.5	10.1	25.1	41.0	39.9	3.8
Iceland	486.3	70.5	0.7	5.2	29.0	28.4	36.7	4.0
Indonesia	433.2	86.5	25.7	38.6	25.8	5.3	4.5	2.2
Iran	447.6	68.0	33.2	32.3	19.5	7.1	8.0	2.2
Ireland	530.4	87.2	7.1	16.1	33.8	21.1	21.8	3.3
Israel	501.9	94.1	4.8	15.6	32.3	23.6	23.7	3.4
Italy	502.0	80.8	12.1	25.2	27.7	14.8	20.2	3.1
Japan	583.5	80.4	13.9	19.4	31.2	18.0	17.5	3.1
Jordan	455.1	94.8	21.3	30.5	28.4	29.7	10.0	2.6
Korea	584.7	85.5	9.7	11.2	34.8	23.4	20.8	3.3
Kuwait	403.9	57.0	22.3	26.8	38.3	10.2	12.5	2.6
Latvia	501.6	74.0	1.4	5.2	18.7	23.1	51.6	4.2
Lithuania	485.4	75.2	4.9	18.2	35.5	20.8	20.6	3.3
Macedonia	467.9	87.1	15.4	38.1	30.0	9.1	7.3	2.5
Malaysia	520.4	74.0	13.0	34.1	32.3	12.0	8.8	2.7
Moldova	478.5	82.4	20.2	32.7	27.5	11.0	8.5	2.5
Morocco	344.1	81.2	37.4	35.2	39.7	4.9	3.0	2.0
Netherlands	552.2	73.8	8.0	15.7	32.1	21.0	23.1	3.4
New Zealand	514.9	85.4	4.4	8.7	25.7	24.6	36.6	3.8
Norway	512.3	78.3	2.3	5.7	25.2	22.9	45.0	4.0
Philippines	376.4	93.7	37.4	30.5	19.3	6.2	6.5	2.1
Portugal	462.3	61.3	10.4	26.2	31.6	14.5	17.3	3.0
Romania	483.4	88.8	19.1	22.8	25.6	13.2	19.4	2.9
Russian Federation	539.0	84.7	3.1	12.0	33.2	26.7	24.9	3.6
Scotland	504.7	86.5	11.2	17.3	28.2	18.7	24.7	3.3
Singapore	613.4	83.0	11.8	22.0	40.6	13.8	11.8	2.9
Slovak Republic	546.6	77.9	2.1	12.5	43.7	23.8	17.8	3.4
Slovenia	547.8	78.5	3.1	15.8	42.1	21.1	17.9	3.3
South Africa	480.6	97.3	38.9	29.3	15.5	6.6	9.6	2.2
Spain	498.8	68.1	3.7	18.4	32.5	19.7	25.8	3.5
Sweden	524.5	80.5	3.1	8.0	24.3	23.9	40.7	3.9
Switzerland	531.4	81.9	7.9	16.2	30.3	20.1	25.5	3.4
Taiwan (Chinese Taipei)	592.8	90.2	17.5	23.2	31.2	11.9	16.3	2.9

Table 1. (Contd)

	TIMSS performance		Books at home[a]					Mean category
	Mean	SD	1 (0–10)	2 (11–25)	3 (26–100)	4 (101–200)	5 (200+)	
Thailand	505.9	74.2	20.1	33.7	30.1	8.6	7.5	2.5
Tunisia	452.5	56.9	21.4	35.9	24.5	9.3	8.9	2.5
Turkey	446.8	73.3	21.6	36.7	27.6	8.2	5.9	2.4
United States	518.9	89.2	8.1	13.2	28.3	21.0	29.4	3.5

Notes: Mean performance: TIMSS-95/TIMSS-Repeat international test score (mean of math and science), re-scaled, weighted by sampling probabilities. – SD: Standard deviation of the TIMSS-95/TIMSS-Repeat international test score. – Books at home: share of students in each category, weighted by sampling probabilities. – [a]1 = none or very few (0–10 books); 2 = enough to fill one shelf (11–25 books); 3 = enough to fill one bookcase (26–100 books); 4 = enough to fill two bookcases (101–200 books); 5 = enough to fill three or more bookcases (more than 200 books). – Mean category: mean of the books-at-home categories (1–5). – OECD member states marked in bold.

Table 2

Descriptive Statistics of the International Data

	Mean	Standard deviation	Minimum	Maximum	Number of countries
Books at home (1–5)	3.2	1.3	1	5	54
Student age[a]	14.3	0.8	13	17.3	54
Female student	0.499		0	1	54
Living with both parents	0.838		0	1	51
Student born in country	0.938		0	1	52
Mother born in country	0.892		0	1	53
Father born in country	0.886		0	1	53
Age of first tracking	15.2	2.5	10	19	54
Pre-school enrollment	60.1	29.4	5.1	115.9	53
Pre-school duration	2.7	1.0	1	4	53
Educational expenditure per student	4380.7	3874.9	189	14270.4	50
GNI per capita	14584.2	8020.8	1300	33160	53
Private enrollment share	15.1	20.0	0	77.5	43
Private expenditure share	13.7	11.5	1.2	44.9	32
Full-day schooling	0.559		0	1	34
Start of primary education	6.2	0.5	5	7	53

Notes: Mean: International mean, weighted by sampling probabilities. – Standard deviation: International standard deviation (only for discrete variables). – [a]Minimum and maximum are 1st and 99th percentile, respectively.

Mullis et al. 2000). Furthermore, the number of books at home proxies for the social background of the parents. It also proxies for their economic background, since books are goods that have to be paid for. Thus, the variable 'books at home' provides a proxy for the educational, social and economic background of the students' families.

In both TIMSS tests, the number of books at home was reported by the students themselves in the student background questionnaire according to the following question: 'About how many books are there in your home? (Do not count magazines, newspapers, or your school books)'. The following five answer categories were given: 1 – 'none or very few (0–10 books)'; 2 – 'enough to fill one shelf (11–25 books)'; 3 – 'enough to fill one bookcase (26–100 books)'; 4 – 'enough to fill two bookcases (101–200 books)'; 5 – 'enough to fill three or more bookcases (more than 200 books)'. *Table 1* reports the frequency with which each category was answered in each country.

An obvious alternative to books at home as a proxy for family background is parental education[5]. However, the books-at-home proxy is preferable for several reasons. First, previous results based on the TIMSS and the PISA 2000 dataset show that, on average, books at home are the single most important predictor of student performance, considerably stronger than parental education (see Wößmann 2003, 2008 for TIMSS; Fuchs and Wößmann 2007 for PISA)[6]. Second, even when the parents' level of education is reported using international standards such as the International Standard Classification of Education (ISCED), specific educational tracks differ considerably across countries. Thus, a given level of education does not imply the same knowledge in all countries, so that the cross-country comparability may be limited. In terms of mere units of measurement, the cross-country comparability of counting books at home is much more straightforward. Third, there is a TIMSS-specific data problem, since parental education is measured slightly differently in TIMSS-95 and TIMSS-Repeat (due to an interim adjustment of the ISCED classification), while the scaling of the books at home variable remained exactly the same. Finally, there is also a second, more general data problem, insofar as substantially more observations are missing for the parental education variable than for the books-at-home variable. E.g., about one third of the observations on parental education in the Western European countries are missing in the TIMSS dataset, while the average fraction of missing observations of the books-at-home variable in the same countries is less than three percent (cf. Wößmann 2008).

Despite these relative advantages of the books-at-home proxy, it is still only a proxy for family background and has, therefore, its limitations. In particular,

5. Other family-background proxies used in sociological research include parental occupation and indices of socio-economic status, usually also based on occupational status. For example, one widely used index is some form of the International Socio-Economic Index (ISEI), which derives from a classification of occupational status by Ganzeboom et al. (1992). Their mapping from occupational to socio-economic status is based on only 16 countries, however. The TIMSS studies do not provide data on parental occupation, and the cross-country comparability of occupational status may be more limited than that of books at home.
6. Results on the PISA data, which contain information on the work status and occupation of parents, also reveal that books at home are on average a stronger predictor of student performance than parental work status and occupation (Fuchs and Wößmann 2007). The PISA data also show that there are more missing observations on the occupation variable than on the books-at-home variable.

although it is reasonable to assume that the socio-economic position of a family is positively correlated with the number of books it owns, it is possible that this correlation varies across countries with different cultures' differing appreciation of books. The extent to which this is the case, and the extent to which this might bias the results reported in this study, is a priori unclear. We do not know of any study that has tried to validate the cross-cultural comparability of the number of books at home as a family-background proxy[7].

Taking these caveats into account, we suggest the following procedure to test for the cross-country validity of the books-at-home variable. We are aware of one international student achievement test which provides data on both books at home and household income (which unfortunately has much fewer participating countries than the pooled TIMSS tests). This is the Progress in International Reading Literacy Study (PIRLS), also conducted by the IEA, that tested fourth-grade students in 2001 (data released in 2003). PIRLS contained a home background questionnaire, completed by the parents of tested students, which asked both about the before-tax annual household income (in six bracketed categories) and about the number of books in their home (in five categories, in exactly the same way as in TIMSS). In a sense, household income may be viewed as the 'ideal' measure of family background, at least from an economic perspective. Given this dataset containing both income and books information, we can test whether the correlation between household income and books at home varies across countries. Unfortunately, there are only six OECD countries for which PIRLS provides income data in a comparable way (England, Germany, the Netherlands, Norway, the Slovak Republic and Sweden).

When we regress the categorically measured income variable on the books-at-home categorical variable, country dummies and interactions between books at home and the country dummies, none of the interactions is statistically or quantitatively significant. This is despite the fact that the coefficient on books at home is statistically and quantitatively highly significant: Given that the six-country sample contains 20,343 students, the effect is very precisely estimated, at 0.528 with a standard deviation of only 0.009 (t-statistic: 61.0). While this means that the 95-percent confidence interval ranges only from 0.511 to 0.545, still none of the country interactions is statistically significant, with their size varying only from -0.048 to 0.045. That is, the association between household income and books at home does not vary significantly between the countries. We view this as strong evidence in favor of

7. When experimenting with the parental-education measure as an alternative measure for socio-economic background, we found that the general pattern of results does look quite similar, although there are sizable differences for a few countries.

the validity of cross-country comparisons where the books-at-home variable proxies for family background.

2. *Empirical Specification*

Following Roemer's (1998) concept of equality of opportunity where equality means that outcomes differ only due to persons' different efforts but not due to circumstances beyond their control, our empirical measure of inequality of education opportunity is one of the association between students' educational outcomes and their family background. Thus, to estimate the extent of inequality of educational opportunity in country j, we regress the test performance of individual students on our proxy for family background, i.e. the number of books at home, and a set of control variables, separately for each country j:

$$\begin{aligned} T_{isj} = & \alpha_j + \beta_j B_{isj} + \gamma_{1j} A_{isj} + \gamma_{2j} G_{isj} + \gamma_{3j} F_{isj} + \gamma_{4j} S_{isj} \\ & + \gamma_{5j} I^i_{isj} + \gamma_{6j} I^m_{isj} + \gamma_{7j} I^f_{isj} + \gamma_{8j} \left(I^i_{isj} B_{isj} \right) + \gamma_{9j} \left(I^m_{isj} B_{isj} \right) \\ & + \gamma_{10j} \left(I^f_{isj} B_{isj} \right) + \varepsilon_{isj} \end{aligned} \tag{1}$$

where T_{isj} is test score of student i in school s in country j and B_{isj} is our measure of books in the individual student's home. The set of control variables includes: a constant α_j, student age A_{isj}, student gender G_{isj}, a dummy for family composition F_{isj} indicating whether the student lives together with both parents and, for the countries participating in both TIMSS studies, a study dummy S_{isj}. Furthermore, the regressions control for three immigration status dummies, indicating whether the student (I^i_{isj}), the mother (I^m_{isj}) and the father (I^f_{isj}) were born in the country, respectively. Finally, the regressions control for interaction terms between these three immigration dummies and books at home. ε_{isj} is the error term.

In the following, we discuss the functional form of the relation between test scores and books at home, the inclusion of control variables in general and the immigration controls in particular, and the specific structure of the error term.

We enter the books-at-home variable B_{isj} as a categorical variable ranging from 1 to 5 according to the five answer categories reported above. This approach is valid under the assumption that the performance differences of students between each of these categories are roughly the same. Specification tests suggest that this is a valid assumption, as the suggested functional form represents the data particularly well. That is, we initially estimated a form of equation (1) for the pooled sample of all countries (controlling for country fixed effects) in which we entered four dummies to represent the five available books-at-home categories individually (leaving the lowest one out as the residual category). This estimation does not place any restrictions on the functional form, as it uses all information available and allows the effect of books to vary in any possible way. The estimated coefficients on the four dummies showed a highly linear pattern. That is, when forcing the four steps to have equal length by applying the mean of the implied steps between the five categories (which was equal to 17.7) to each step, which implies the four linear steps of 17.7, 35.5, 53.2 and 71.0, these implied linear steps are very close to the actual estimates on the four dummies of 14.4, 38.4, 58.6 and 71.0. They all either fall within the 99% confidence interval of the four directly estimated individual dummies or are very close to the bounds of these intervals[8]. Thus, the linear functional form taking on values from 1 to 5 along the lines of the five answer categories reproduces the data considerably well.

The official TIMSS publications report simple (bivariate) comparisons of the average performance of students falling into different categories on a specific family-background variable (cf. Beaton et al. 1996; Mullis et al. 2000). However, such bivariate comparisons can easily be confounded by other basic characteristics of the students, such as age, gender, family composition and immigration status. Thus, our regressions include controls for these potentially confounding factors, thereby holding these factors constant when comparing the performance of students along the books-at-home dimension.

One particular criticism often raised against international comparisons of measures of achievement dispersion and bivariate estimates of family-background effects (FBEs) is that countries have different immigrant populations. Immigrant populations may cause a bias in these bivariate estimates for two reasons. First, if immigration status and family background (as proxied by books at home in our study) are correlated, international differences in estimated FBEs are biased when ignoring the immigration status. Second, the FBEs may be heterogeneous between native and immigrated families, which may introduce an additional bias to the cross-country pattern of estimated FBEs. Since we do not want our estimator of the strength of the influence of family background on student performance to be affected by the proportion of immigrant students in the respective countries, we calculate the FBEs net of immigration status. That is, we control for these potentially biasing effects already in the micro construction of our FBE measure. The three dummies for the immigration status of students, their mothers and their fathers ensure that the first possible cause for a bias does not affect our FBEs. The interaction terms

8. The two deviations where the implied steps do not fall within the confidence intervals are that the first implied step of 17.7 is slightly above the upper bound of 17.1 of the 99% confidence interval of the direct estimate on the first dummy, and that the third implied step of 53.2 is slightly below the lower bound of 55.3 of the 99% confidence interval of the direct estimate on the third dummy.

between the three immigration dummies and our family-background measure ensure that the second possible case does bias our FBEs. Thus, our estimated FBEs represent only the family-background influence of the native students in each country, which seems to be the best way to provide cross-country comparability despite cross-country differences in the immigrant population[9].

Under the assumption that any factors not controlled by the included explanatory variables, which therefore enter the error term, are not systematically related to the number of books in the students' home, least-squares estimation of equation (1) yields an estimate of the influence of family background (as proxied by books at home) on student performance.

In estimating the error term ε_{isj} of equation (1), it has to be recognized that the performance of students within the same school may not be independent from one another (cf. Moulton 1986 for this problem of hierarchical data structure). Furthermore, the TIMSS sampling procedure had a two-stage clustered sampling design within each country (cf. Martin and Kelly 1996; Martin et al. 2000). At the first stage, schools were sampled, and at the second stage classrooms were sampled within these schools. Thus, the primary sampling unit (PSU) in TIMSS was the school. This suggests that the independence assumption usually made with respect to individual observations in standard econometric methods should be relaxed in favor of the assumption that only the variation between schools (PSUs) provides independent variation. This is implemented by the clustering-robust linear regression (CRLR) method, which allows any given amount of correlation of the error terms within PSUs and requires only that observations be independent across PSUs (cf. White 1984; Deaton 1997).

Since TIMSS used a stratified sampling design within each country, sampling probabilities vary for different students (Martin and Kelly 1996; Martin et al. 2000). We obtain nationally representative coefficient estimates by employing weighted least squares (WLS) regressions, using the sampling probabilities as weights. WLS estimation guarantees that the proportional contribution to the parameter estimates of each stratum in the sample is the same as if a complete census had been obtained (cf. DuMouchel and Duncan 1983; Wooldridge 2001).

3. *Results on Family-Background Effects (FBEs)*

Our estimates of the FBEs in the 54 countries, estimated as β_j in equation (1), are reported in *Table 3*. The results suggest that in all countries student performance is influenced in a statistically significant way by our

9. One way in which the size of the immigrant population could still affect our FBEs is if the size of the FBE among native students is affected, for example, by how many immigrants are in their specific class.

Table 3

Family-Background Effects (FBEs) as an Index of Inequality of Educational Opportunity

		FBE	Std. Err.	Year(s)	Observations
1	**England**	28.81	*(1.70)*	1995 + 1999	4,388
2	Taiwan (Chinese Taipei)	27.91	*(1.31)*	1999	5,711
3	**Scotland**	26.95	*(1.70)*	1995	2,608
4	**Hungary**	25.84	*(1.29)*	1995 + 1999	5,727
5	**Germany**	25.57	*(1.94)*	1995	2,698
6	**Korea**	24.75	*(0.81)*	1995 + 1999	8,897
7	Macedonia	24.05	*(2.35)*	1999	3,683
8	**Slovak Rep.**	24.01	*(1.26)*	1995 + 1999	6,823
9	Bulgaria	23.32	*(2.94)*	1999	3,107
10	**United States**	23.13	*(0.81)*	1995 + 1999	14,877
11	Lithuania	23.09	*(1.32)*	1995 + 1999	4,749
12	**Ireland**	23.04	*(1.98)*	1995	2,989
13	**New Zealand**	22.80	*(1.40)*	1995 + 1999	6,974
14	**Czech Rep.**	22.45	*(1.67)*	1995 + 1999	6,604
15	Slovenia	22.25	*(1.51)*	1995 + 1999	5,676
16	Malaysia	22.18	*(2.06)*	1999	5,524
17	South Africa	22.05	*(2.67)*	1995 + 1999	10,307
18	Chile	21.95	*(1.78)*	1999	5,507
19	**Australia**	21.45	*(1.43)*	1995 + 1999	10,454
20	**Sweden**	21.24	*(1.59)*	1995	3,925
21	**Austria**	20.80	*(1.68)*	1995	2,632
22	Russian Fed.	20.27	*(1.78)*	1995 + 1999	7,941
23	**Norway**	20.00	*(1.68)*	1995	3,196
24	Romania	19.68	*(2.09)*	1995 + 1999	6,748
25	**Greece**	19.22	*(1.63)*	1995	3,898
26	Israel	18.84	*(2.17)*	1995 + 1999	4,965
27	Singapore	18.54	*(1.82)*	1995 + 1999	9,518
28	Jordan	17.92	*(2.40)*	1999	4,402
29	**Italy**	17.51	*(1.42)*	1999	3,300
30	**Netherlands**	17.03	*(2.08)*	1995 + 1999	4,729
31	**Belgium** (French)	16.77	*(1.77)*	1995	2,477
32	**Switzerland**	16.77	*(1.62)*	1995	4,720
33	Latvia	16.65	*(1.26)*	1995 + 1999	4,939
34	Philippines	16.53	*(2.09)*	1995 + 1999	11,693
35	Moldova	15.80	*(2.01)*	1999	3,342
36	**Spain**	15.26	*(1.06)*	1995	3,752
37	**Denmark**	15.08	*(1.62)*	1995	2,155
38	Cyprus	14.21	*(0.89)*	1995 + 1999	5,680
39	**Finland**	13.98	*(1.51)*	1999	2,828
40	**Japan**	13.50	*(1.13)*	1999[a]	4,608
41	Thailand	12.37	*(1.60)*	1995 + 1999	11,211
42	**Turkey**	11.77	*(1.35)*	1999	7,393
43	**Iceland**	11.42	*(2.55)*	1995	1,728
44	Iran	11.25	*(0.93)*	1995 + 1999	7,231
45	**Belgium** (Flemish)	10.95	*(1.11)*	1995 + 1999	7,657
46	Hong Kong	10.82	*(1.28)*	1995 + 1999	7,984
47	**Portugal**	10.40	*(1.05)*	1995	3,335
48	**Canada**	9.76	*(0.95)*	1995 + 1999	16,101
49	**France**	8.32	*(1.44)*	1995	2,770
50	Colombia	7.55	*(3.84)*	1995	2,455
51	Morocco	6.84	*(2.02)*	1999	2,383
52	Tunisia	6.32	*(0.72)*	1999	4,758

Table 3. (Contd)

		FBE	Std. Err.	Year(s)	Observations
53	Indonesia	4.83	*(1.81)*	1999	5,538
54	Kuwait	2.49	*(1.59)*	1995	1,442

Notes: Coefficient estimate on books at home. – Dependent variable: TIMSS-95/TIMSS-Repeat international test score (mean of math and science). – Regressions control for: student age, student gender, family status, student born in country, mother born in country, father born in country, interactions between the three immigration variables and books, a TIMSS-Repeat dummy and a constant. – Regressions weighted by students' sampling probabilities. – Clustering-robust standard errors (taking account of correlated error terms within schools) in parentheses. – [a]Japan also participated in 1995, but the question on books at home was not administered at an internationally comparable scale. – OECD member states marked in bold. – All estimates are statistically significantly different from zero at the 1 percent level, with the exceptions of Colombia (5 percent level) and Kuwait (14 percent level).

family-background variable[10]. The size of the estimated FBEs indicates how much students' test scores differ on average between the five categories of the variable 'books at home' reported above. Since the performance of students was measured by standardized test scores with an international standard deviation of 100, the coefficients can be interpreted as percentages of an international standard deviation by which test achievement increases when raising the number of books at home by one category[11].

In interpreting the estimated FBEs, first recall that initial evidence showed that each move from one books-at-home category to the next can be regarded as roughly equivalent in terms of its effect on test scores. The FBEs reported in *Table 3* show by how much one of these steps changes the test scores in each country. Consider as an example the estimated FBE for the United States of 23.1. A 1-point difference in the family-background proxy – e.g., the difference in social background of US students that is equivalent to the difference between having one bookcase and two bookcases of books at home – goes hand in hand with a difference of 23.1 percent of an international standard deviation in test scores.

With an FBE of 23.1, the United States falls in the top quarter of the most unequal countries – both in the overall sample and in the OECD sub-sample. As the results of *Table 3* show, England (28.8), Taiwan (27.9), Scotland (27.0), Hungary (25.8) and Germany (25.6) are the five countries with the largest estimated FBE in our sample of 54 countries. That is, they provide their students with the least equality of educational opportunity. On average, the impact of family background on student performance in these five countries is more than four times as large as in the countries with the smallest performance

10. The level at which the estimates are statistically significantly different from zero reaches 14 percent in Kuwait, 5 percent in Colombia, and 1 percent in all other countries.
11. As France and Japan did not collect information on the immigration status of the parents, and France also not for the student, their estimated FBEs do not control for these variables.

difference for students from different family backgrounds, Kuwait (2.5), Indonesia (4.8), Tunisia (6.3), Morocco (6.8) and Colombia (7.6). Their particularly low FBEs may be partly due to the fact that the average performance level of these countries is relatively low, so that the performance of all students is condensed at a rather low level. When looking at the more homogeneous sample of OECD countries, the countries that provide the largest extent of equality of educational opportunity to students from different family background are France (8.3), Canada (9.8), Portugal (10.4) and the Flemish school system in Belgium (11.0).

As an additional useful benchmark to assess the magnitude of the estimated effects, we can use the average performance difference between students in seventh grade and students in eighth grade, which were both tested in the TIMSS-95 study. This 'grade-level equivalent', which equals 35.4 test-score points (in terms of our scale) across the countries participating in the study, shows how much students learn on average during one school year. Thus, the difference in educational performance between children of families with more than two bookcases full of books and children of families with only very few books at home (the two extreme categories) in England (4 x 28.8 = 115.2) is more than three times of what students on average learn during a whole school year. Even in France, the OECD country with the lowest estimate, this socio-economic difference equals roughly one grade-level equivalent.

In sum, the estimated FBEs presented in *Table 3* constitute a cross-country index of the inequality of educational opportunity that the different school systems achieve for students from different family backgrounds, where a *high* FBE value is associated with *low* equality of educational opportunity. The results show that there is substantial variation across countries in the extent to which equality of educational opportunity is achieved.

III. SCHOOL SYSTEMS AND EQUALITY OF OPPORTUNITY

Combining the estimated FBEs with country-level data on different education policies, we can now analyze how the cross-country variation in equality of educational opportunity revealed in the previous section relates to several features of the different school systems. The results provide answers to the question of why it is that some countries achieve much higher equality of educational opportunity than others.

1. Hypotheses and Country-Level Data on Features of the School Systems

Several features of the different school systems may be hypothesized to be related to the estimated FBEs. We gather country-level data on these features to estimate whether such a relationship exists empirically. One systemic feature that shows considerable variation across countries is the age at which students are first tracked into different school types that serve students of differing ability. While school tracking in many countries does not occur at all before the age of the students tested in TIMSS (roughly 14 years), it occurs as early as age 10 in some other countries. The younger children are, the more likely is their educational performance to be strongly determined by their family background. Thus, as depicted by our formal model in Schütz et al. (2005, Section III), early tracking may harm the educational opportunities of low-background students, thereby reducing equality of educational opportunity.

Along similar lines of argumentation, one might expect that the earlier children – particularly from low family backgrounds – are exposed to formal education, the more equal educational opportunities are. Before children enter formal education, their educational performance is mostly determined by their families. Once they enter formal education, the educational institutions can also exert an effect, which – as an additional effect on top of the home-production effect – may be expected to exert equalizing influences. Thus, as again depicted by our formal model in Schütz et al. (2005, Section III), one might expect that the earlier the equalizing school influence sets in, the larger is the equality of educational opportunity for children of different family backgrounds. To test this hypothesis, we gathered information on enrollment in pre-primary education, on the duration of the pre-primary school cycle and on the age at which compulsory education starts in the different countries. As our formal model in Schütz et al. (2005, Section III) reveals, the relationship between FBEs and the enrollment in pre-primary education may be highly non-linear, because initially, it will be the children of relatively well-off families which start attending pre-primary education. Only once also children of less favorable family background also start enrolling in pre-primary education, so that a substantial part of the student population is enrolled, one might expect pre-primary enrollment to exert its equalizing effect.

According to the same argument of length of exposure to formal schooling, equality of educational opportunity may differ between system with full-day and half-day schooling. Other features which might be related to equality of educational opportunity are the level of educational expenditure per student in the country, the mean test-score performance of the country, the level of economic development and the share of the private sector in educational expenditure and in educational enrollment.

Country-level data on the different features of the school systems mainly come from statistical yearbooks and data collections by international organizations such as UNESCO and OECD, as well as detailed country-specific inquiries. *Table A1* in the appendix provides details on the definitions and sources of the different variables. We exercised great care to gather the information for the years relevant for the students tested in the two TIMSS tests in 1995 and 1999. Descriptive statistics of the data are reported in *Table 2*.

2. *Empirical Specifications*

There are two ways to identify how systemic features are related to the FBE. First, we can use the estimated country-level FBEs reported in *Table 3* directly as left-hand-side variables in regressions on the different systemic features. Second, we can identify the relationship by interacting the country-level systemic features with the individual-level family-background measure in cross-country microeconometric regressions that use the test scores as their dependent variable. We will pursue both avenues of investigation.

Country-level specification: In the first specification, we use the FBEs estimated in the previous section as the dependent variable in a country-level regression on several country-level systemic features; i.e., the FBE estimators β_j of equation (1) are simply regressed on the set of potentially determining features Z_j of the school systems:

$$\beta_j = \lambda + \theta Z_j + \mu_j. \tag{2}$$

Since the dependent variable in this regression is the outcome of an estimation rather than a precise observation, we have to account for the different standard errors with which the observations of the dependent variable are estimated. The error term from an ordinary-least-squares (OLS) estimation of equation (2) would be heteroscedastic with mean zero and a variance equal to the sum of the variance of the actual error term and the variance of the estimated FBE. We use the weighted estimation procedure proposed by Anderson (1993) which down-weights observations whose FBEs were relatively imprecisely measured: First, we compute the squared residuals of an OLS regression of equation (2). We then run a second regression of these squared residuals on the estimated variances of the FBEs, the variances squared and the variances cubed. The fitted values of this regression specify to what extend the squared residuals of the first regression can be explained by the variance of the FBE estimates. Finally, we use the inverse of these fitted values as weights in a WLS regression of equation (2), thereby giving lower weight to imprecisely estimated FBE observations.

Student-level interaction specifications: The country-level specification provides a rather ad-hoc partition of the estimation in two steps: in a first step, equation (1) is estimated using micro data within each country, and a second step then uses the resulting coefficient estimates in the country-level estimation of equation (2). This two-step estimation procedure places rather strong restrictions on the joint distribution of the variables used in the two different equations. We can relax these restrictions by pooling the micro data across countries and combining them with the additional system-level data. In this second set of specifications, the relationship between the country-level systemic features and the effect of family background on student performance can be identified by the interaction between the country-level features and the family-background measure in a student-level cross-country regression that has the individual test scores as its dependent variable.

The first of these student-level interaction specifications assumes that after having controlled for the observed systemic and individual effects, there is no unobserved heterogeneity left across countries which might bias the estimates. Under this assumption, the estimated coefficients η on the interaction terms between our family-background proxy B_{isj} (books at home, measured at the student level) and the vector of systemic features Z_j (measured at the country level) identify how the systemic features affect the FBE:

$$T_{isj} = \alpha + \beta B_{isj} + Z_j \rho + (B_{isj} Z_j)\eta + X_{isj}\gamma + \varpi_{isj}, \tag{3}$$

where X_{isj} is a vector combining all the (student-level) control variables of equation (1).

This specification does not only identify how the FBEs relate to the systemic features, but also the main effects of the systemic features on student test scores. However, the estimates of these main systemic effects in equation (3) will only be unbiased if there is no unobserved heterogeneity in the performance levels across the countries. Thus, this first student-level interaction specification still requires the assumption that there is no unobserved cross-country heterogeneity in student performance.

We can relax this assumption in a second student-level interaction specification that introduces a whole set of country fixed effects C_j, as well as interaction effects between all student-level controls X_{isj} and the country fixed effects:

$$T_{isj} = \alpha + \beta B_{isj} + (B_{isj} Z_j)\eta + C_j \delta_1 + (C_j X_{isj})\delta_2 + X_{isj}\gamma + \varpi_{isj}. \tag{4}$$

Due to the country fixed effects, this specification can no longer identify the main systemic effects that work at the country level. But despite the country fixed effect, the specification still identifies our main measure of interest, namely how the systemic features affect the FBE. This influence is captured by the coefficients η on the interaction terms at the student level.

For the identification of equation (4), the assumption of no unobserved cross-country heterogeneity can be replaced by the less restrictive assumption that any unobserved cross-country heterogeneity that may exist is unrelated to the size of the FBEs. Under this assumption, equation (4) can still identify how education policies relate to equality of educational opportunity. Since it requires the least restrictive assumptions on the cross-country distribution of test scores, the student-level interaction specification with country fixed effects of equation (4) is our preferred specification.

When estimating the student-level specifications of equations (3) and (4), one has to be aware that one part of the measures which identify our effect of interest, namely the systemic features of the school systems, are measured at the country level rather than the school level, since education policies are observed as system-level variables that vary across countries. Thus, the Moulton (1986) problem of a hierarchical data structure now applies at the country rather than the school level, requiring a higher-level (country) error component to avoid spurious results. We therefore use countries as PSUs in the CRLR estimation of equations (3) and (4), allowing any given interdependence of the error variance-covariance matrix within countries and requiring only that the observations are independent across countries. This basically means that the standard errors are measured as if there were only as many observations as there are countries in the regression. Furthermore, in addition to the within-country weighting of students according to their sampling weights, the observations are now weighted across countries such that each country in the sample has equal weight.

3. *Results on the Effect of Education Policy on Educational Inequality*

Tables 4–6 report the results for the three specifications. The country-level specification is reported in *Table 4*, the student-level interaction specification with main systemic effects in *Table 5* and the student-level interaction specification with country fixed effects in *Table 6*. The baseline model of each specification, reported in columns (1), (3) and (5) of *Tables 4–6*, includes the age of first tracking, enrollment in pre-school education and its square and the duration of pre-school education as potential determinants of the size of the FBE[12].

Tracking: In line with our hypotheses, the age at which students are first tracked into different school types is found to be statistically significantly negatively related to the FBEs in all three empirical specifications[13]. That is, the earlier an education system tracks its students into different types of schools according to their ability, the more unequal are educational opportunities. While the systemic effect is directly estimated in the country-level specification

12. Taiwan could not be included in any of the estimations due to lack of internationally comparable information on several school-system variables.
13. Our results are robust to capping the variable 'age of first tracking' at age 14.

Table 4

Education Policy and Inequality of Opportunity: Country-Level Specification

	(1)	(2)
Age of first tracking	− 1.225***	− 0.926***
	(0.337)	*(0.339)*
Pre-school enrollment	0.369***	0.213*
	(0.090)	*(0.115)*
Pre-school enrollment squared	− 0.003***	− 0.002*
	(0.001)	*(0.001)*
Pre-school duration	− 1.377*	− 1.317
	(0.781)	*(0.867)*
Educational expenditure per student /1000		− 0.140
		(0.411)
GNI per capita /1000		− 0.158
		(0.219)
Country mean test score /100		4.916***
		(1.714)
Observations (countries)	53	49
R^2 (adjusted)	0.400	0.268

Notes: Dependent variable: estimated family-background effect (FBE) of *Table 3*. – Regressions apply the Anderson (1993) weighted estimation procedure to account for estimated dependent variable. Standard errors in parentheses.
Significance level: ***1 percent. *10 percent.

of *Table 4*, where the FBE is the dependent variable, the systemic effect in the two student-level specifications is identified by the interaction effect. Under the assumption that there is no bias due to unobserved country heterogeneity, the student-level interaction specification with main systemic effects of *Table 5* yields also estimates of the main effects of the systemic features on student performance. The results suggest that tracking does not exert a statistically significant direct effect on student performance.

Table 6 reports the results of our preferred specification, which not only controls for a complete set of country fixed effects, but also for interaction effects between each of the student-level control variables and the complete set of country dummies, allowing the effects of the student-level controls to be country-specific. The size of the estimated interaction effect in column (5) suggests that for each additional year of earlier tracking, the estimated FBE (which has an international mean of 17.5 across all countries) increases by slightly more than one unit. Thus, a difference in the age of first tracking of four years is related to a difference in the FBE of an order of magnitude of roughly one quarter of the international mean of the whole FBE. Comparing the estimated effect size to the 'grade-level equivalent' discussed above reveals that postponing tracking by four years would be associated with a reduction in the performance difference between children from the first and the fifth category of

Table 5

Education Policy and Inequality of Opportunity: Student-Level Interaction Specification with Main Systemic Effects

	(3)	(4)
Books	75.530***	82.139***
	(13.791)	*(10.660)*
Age of first tracking * books	− 2.896***	− 2.624***
	(0.882)	*(0.701)*
Pre-school enrollment * books	0.579**	0.546**
	(0.244)	*(0.232)*
Pre-school enrollment squared * books	− 0.006***	− 0.005**
	(0.002)	*(0.002)*
Pre-school duration * books	− 6.875***	− 8.141***
	(1.977)	*(1.988)*
Educational expenditure per student /1000 * books		0.277
		(0.702)
GNI per capita /1000 * books		− 0.778*
		(0.450)
Age of first tracking	5.843	3.712
	(5.875)	*(4.937)*
Pre-school enrollment	− 0.105	− 0.490
	(1.025)	*(0.867)*
Pre-school enrollment squared	0.009	0.007
	(0.010)	*(0.008)*
Pre-school duration	25.747**	34.222***
	(11.887)	*(11.518)*
Educational expenditure per student /1000		− 2.372
		(4.045)
GNI per capita /1000		5.331*
		(2.702)
Observations (students)	295,026	276,577
Primary sampling units (countries)	53	50
R^2	0.217	0.241

Notes: Dependent variable: TIMSS-95/TIMSS-Repeat international test score (mean of math and science). – All regressions control for student age, student gender, family status, student born in country, mother born in country, father born in country, interactions between the three immigration variables and books, a TIMSS-Repeat dummy and a constant. – Regressions weighted by students' sampling probabilities. – Clustering-robust standard errors (taking account of correlated error terms within countries) in parentheses.
Significance level (based on clustering-robust standard errors): ***1 percent. **5 percent. *10 percent.

the family-background measure in a country by roughly half of what students on average learn during a whole school year.

The tracking effect is also directly observable in our estimated FBEs (*Table 3*). In our sample of countries, the education systems that track their students as early as age 10 – Austria (with an estimated FBE of 20.8), Germany (25.6), Hungary (25.8) and the Slovak Republic (24.0) – all show relatively large FBEs. By contrast, none of the OECD countries with relatively low FBEs – France (8.3), Canada (9.8) and Portugal (10.4) – track before an age of 15. The

Table 6

Education Policy and Inequality of Opportunity: Student-Level Interaction Specification with Country Fixed Effects, Full Sample

	(5)	(6)	(7)
Books	29.368***	10.612	30.612**
	(5.316)	*(8.534)*	*(11.973)*
Age of first tracking * books	− 1.074***	− 0.893***	− 0.968***
	(0.284)	*(0.277)*	*(0.341)*
Pre-school enrollment * books	0.336***	0.262***	0.041
	(0.085)	*(0.081)*	*(0.100)*
Pre-school enrollment squared * books	− 0.003***	− 0.002***	− 0.000
	(0.001)	*(0.001)*	*(0.001)*
Pre-school duration * books	− 1.303*	− 1.545**	− 2.103**
	(0.717)	*(0.689)*	*(1.018)*
Educational expenditure per student /1000 * books		− 0.136	− 0.178
		(0.396)	*(0.395)*
GNI per capita /1000 * books		− 0.144	0.161
		(0.225)	*(0.224)*
Country mean test score /100 * books		4.337**	1.016
		(1.753)	*(1.654)*
Private expenditure share * books			0.158**
			(0.060)
Private enrollment share * books			− 0.099***
			(0.031)
Observations (students)	295,026	276,577	156,412
Primary sampling units (countries)	53	50	27
R^2	0.414	0.417	0.294

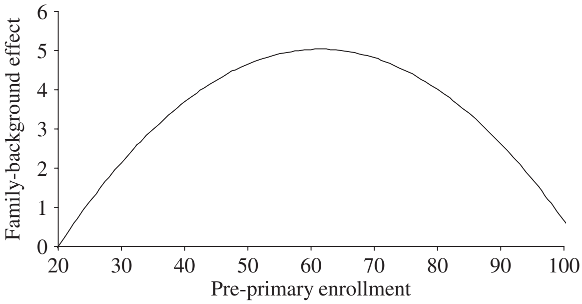
Notes: Dependent variable: TIMSS-95/TIMSS-Repeat international test score (mean of math and science). – All regressions control for: country fixed effects, student age, student gender, family status, student born in country, mother born in country, father born in country, interactions between the three immigration variables and books, interactions between all these previous variables and country dummies and a constant. – Regressions weighted by students' sampling probabilities. – Clustering-robust standard errors (taking account of correlated error terms within countries) in parentheses.
Significance level (based on clustering-robust standard errors): ***1 percent. **5 percent. *10 percent.

estimated tracking effect of column (5) in *Table 6* suggests that roughly one third of the 14.5-point difference in the FBE between these two groups of countries can be attributed to their divergent tracking policies.

Pre-school education: In all three empirical specifications (*Tables 4–6*), the relationship between the FBE and the enrollment share in pre-school education follows a statistically significant inverted U-shaped pattern, again strongly in line with our hypotheses. That is, as long as only a relatively small part of the student population is enrolled in pre-school, enrollment is positively related to the FBE, which may be due to unequalizing effects of non-random sorting of better-off students into pre-school. Only when a substantial share of students is enrolled in pre-school, so that less well-off students are also enrolled, do we find an equalizing effect of pre-school enrollment. *Figure 1* depicts this non-linear

Figure 1

Pre-School Enrollment and the Family-Background Effect.



Note: Estimated interaction effect of books at home with pre-school enrollment and its square in the student-level interaction specification with country fixed effects reported in column (5) of Table 6.

pattern graphically. Educational opportunities get more unequal with rising pre-school enrollment up to a maximum of 61 percent of pre-school enrollment. Only beyond this threshold is higher pre-school enrollment associated with more equal educational opportunities. As is evident from the figure, moving from a low pre-school enrollment of 20 percent to a medium enrollment of 60 percent increases the estimated FBE by roughly 5 units. Moving from 60 percent enrollment (as, e.g., in Switzerland or the United States) to full enrollment (as, e.g., in Denmark or the Netherlands) decreases the estimated FBE by about the same amount. Quantitatively, such a move can account for roughly one quarter of the whole FBE (on average across countries) and would reduce the performance difference between children from the highest and the lowest of the five family-background categories by roughly half of a grade-level equivalent.

Also in line with our hypotheses, we find that the official duration of pre-school education is statistically significantly negatively related to the estimated FBEs in all three specifications. A longer pre-school cycle is associated with lower inequality of educational opportunity. Each additional year of pre-school education is associated with a reduction of the performance difference between the highest and lowest category of family background equivalent to 15 percent of what students on average learn during one year.

While pre-school enrollment – like tracking – does not show a statistically significant direct effect on student performance in the specification with main systemic effects of *Table 5*, meaning that these systemic features exert their impact on student performance only through an indirect effect via family background, the duration of the pre-school cycle shows an additional statistically significant positive direct effect on the level of student performance.

The main results so far are borne out in all three empirical specifications: There is a negative relationship of the FBE with age of first tracking and pre-school duration, and an inverted U-shaped relationship between FBE and pre-school enrollment. In the country-level model of *Table 4*, these features of the education system can together account for 40 percent of the cross-country variation in our estimated FBEs. Combining the effects of a postponement of tracking by four years, of increasing pre-school enrollment from 60 percent to 100 percent, and of increasing the duration of pre-school education by one year is associated with a decrease in the FBE that is equivalent to more than half of the total FBE of the average country, or with decreasing the difference in educational performance between children from the extreme categories of our five-point-scale measure of family background by more than the equivalent of a whole year's average learning.

Robustness to alternative controls and samples: To test for the robustness of these findings and check for other potentially important effects, we extend the baseline model to include additional explanatory variables in columns (2), (4) and (6) of *Tables 4–6*. In the first and third specification, neither educational expenditure per student nor the gross national income (GNI) per capita are statistically significantly related to the FBEs[14]. The country-specific mean test-score performance, however, has a statistically significant and positive relationship with the FBEs[15]. In the specification with main systemic effects of *Table 5*, the GNI per capita displays a marginally significant negative interaction with family background, as well as a marginally significant positive direct relationship with student performance. Most importantly, the results for the features of the baseline model remain qualitatively the same in the extended model.

Column (7) of *Table 6* includes two further features of the education systems, namely the proportion of educational funds that stem from private sources and the share of enrollment in private schools. While the number of available country observations drops to 27 in this specification, the results still show statistically significant associations between the FBE and the private share in educational expenditure and enrollment. The FBE increases with private expenditure and decreases with private enrollment. That is, education systems that rely on private funding show larger inequality of educational opportunity, but more private schooling provision is associated with decreased inequality of opportunity.

14. Unfortunately, we do not have data on the within-country variance in educational expenditure.
15. To preserve the main-effects character of the specification of the first student-level specification, column (4) does not include the country mean test-score performance, as this would add aggregate values of the left-hand-side variable to the right-hand side.

Table 7

Education Policy and Inequality of Opportunity: Student-Level Interaction Specification with Country Fixed Effects, OECD Sample

	(8)	(9)	(10)
Books	29.003***	9.083	39.198**
	(8.731)	*(18.848)*	*(17.061)*
Age of first tracking * books	− 0.844**	− 0.488	− 0.932**
	(0.390)	*(0.421)*	*(0.397)*
Pre-school enrollment * books	0.329***	0.333**	0.126
	(0.110)	*(0.132)*	*(0.137)*
Pre-school enrollment squared * books	− 0.003***	− 0.003**	− 0.001
	(0.001)	*(0.001)*	*(0.001)*
Pre-school duration * books	− 2.280**	− 1.542	− 2.083*
	(0.967)	*(0.966)*	*(1.081)*
Educational expenditure per student /1000 * books		− 0.477	− 0.106
		(0.437)	*(0.444)*
GNI per capita /1000 * books		0.032	0.045
		(0.307)	*(0.282)*
Country mean test score /100 * books		− 2.742	− 1.081
		(2.893)	*(2.945)*
Private expenditure share * books			0.201**
			(0.078)
Private enrollment share * books			− 0.088***
			(0.030)
Observations (students)	154,243	154,243	135,292
Primary sampling units (countries)	29	29	24
R^2	0.256	0.257	0.257

Notes: Dependent variable: TIMSS-95/TIMSS-Repeat international test score (mean of math and science). – All regressions control for: country fixed effects, student age, student gender, family status, student born in country, mother born in country, father born in country, interactions between the three immigration variables and books, interactions between all these previous variables and country dummies and a constant. – Regressions weighted by students' sampling probabilities. Clustering-robust standard errors (taking account of correlated error terms within countries) in parentheses. Significance level (based on clustering-robust standard errors): ***1 percent. **5 percent.

The results so far are based on samples that include all countries for which the data are available. To make sure that they are not driven by variations between very heterogeneous countries, *Table 7* re-estimates the student-level interaction specification with country fixed effects for the more homogeneous sample of OECD countries. All substantive results prove robust in the OECD sample. The only exception is the effect of the country mean test score on the FBEs, which is now statistically insignificant and negative. That is, the tradeoff between equality of opportunity and a country's mean performance apparent in the previous specifications does not survive the scrutiny of the OECD sample. Countries that achieve a higher degree of equality of educational opportunity do not seem to have to compromise on the average performance level of their student population.

As a further robustness check, we drop France and Japan from the sample, because they did not provide data on parental immigration status (and France also not on student immigration status), which might affect the relative size of their estimated FBEs[16]. Again, our results are not sensitive to this reduction in sample size (details available from the authors).

We also experimented with two further systemic features of the education systems, namely half-day versus full-day schooling and the age at which compulsory education starts. A dummy for full-day schooling (available for 34 countries) is not statistically significantly related to the FBEs in any of our specifications. However, any effect of full-day schooling may be difficult to detect, since our data do not distinguish between countries that have afternoon classes and countries that only provide optional day care. The relationship between the FBE and the starting age of compulsory education proves very sensitive, mainly because the starting age does not show much variation across countries (lying between 5 and 7 years) and because effects of pre-school enrollment and duration are already controlled for.

IV. CONCLUSION

In this paper, we developed an index of the inequality of educational opportunity in 54 countries, based on estimates of the effect of family background on students' educational performance. The family-background effects (FBEs) reveal substantial variation in the extent to which different countries achieve equality of educational opportunity for children from different family backgrounds.

Our results suggest that this variation is related to certain systemic features of the countries' education systems, but not to others. Different specifications consistently indicate the importance of extensive early-childhood education and late tracking.

The earlier an education system tracks its students into different school types, the larger are the performance differences along the family-background dimension. The FBEs increase with enrollment in pre-primary education until slightly more than half of the students attend pre-primary education, and decrease afterwards up to universal pre-primary enrollment. The longer the pre-primary education cycle in a country, the smaller are the FBEs. Thus, comprehensive school systems and extensive early-childhood education seem to increase the equality of educational opportunity for children from different

16. In *Tables 5–7*, the immigration dummies in the two countries were set to 1; in our preferred specification, any systematic effect that this constant imputation might have relative to other countries would be captured by the country fixed effects.

family backgrounds. The estimated FBEs also increase with the share of private expenditure in total educational expenditure, but decrease with the share of privately managed schools.

By contrast, several other country-specific features do not seem to be significantly related to the degree of equality of educational opportunity achieved. This is true for the level of educational expenditure, GNI per capita, the length of the average school day and the official school starting age. In the sample of OECD countries, there is also no evidence for an efficiency-equity tradeoff in education, in the sense that more equal education systems would systematically show lower mean performance of their students.

APPENDIX: SOURCES OF COUNTRY-LEVEL DATA

Table A1

Definition and Source of Data

Variable	Definition and Year of Observation	Source
Age of first tracking	Age of students at the time of first streaming, different years since 1999	European Commission (1999, 2000), detailed country-specific inquiries
Pre-school enrollment	Gross enrollment rate in pre-primary education, 1987 and 1991, respectively	UNESCO Institute for Statistics (2003)
Pre-school duration	Duration of pre-primary education, 1988 and 1992, respectively	UNESCO Institute for Statistics (2003)
Educational expenditure per student	Current expenditure per student in pre-primary, primary and secondary schooling, 1995	UNESCO (1998, 2000)
GNI per capita	GNI per capita in PPP (current international $), 1995 and 1999, respectively	World Development Indicators database
Private expenditure share	Proportion of private sources of funds for educational institutions (after transfers from public sources), all levels of education, 1999	OECD (2002: 212), Table B4.1
Private enrollment share	Private enrollment as percentage of total enrollment, general secondary education, 1995 and 1996, resp.	UNESCO (1998, 2000)
Full-day schooling	Dummy: 1 if in the respective country, full-time schooling is the rule, different years since 1999	Renz (1994), detailed country-specific inquiries
Start of primary education	Age of students at start of primary education, 1988 and 1992, respectively	UNESCO Institute for Statistics (2003)

REFERENCES

Ammermüller, Andreas (2005). Educational Opportunities and the Role of Institutions. ZEW Discussion Paper 05-44. Mannheim: Centre for European Economic Research.

Anderson, Patricia M. (1993). Linear Adjustment Costs and Seasonal Labor Demand: Evidence from Retail Trade Firms, *Quarterly Journal of Economics*. 108(4): 1015–1042.

Barnett, Steven W. (1992). Benefits of Compulsory Preschool Education, *Journal of Human Resources*. 27(2): 279–312.

Bauer, Philipp and Regina T. Riphahn (2006). Timing of School Tracking as a Determinant of Intergenerational Transmission of Education, *Economics Letters*. 91(1): 90–97.

Beaton, Albert E. and Ina V.S. Mullis, et al. (1996). *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.

Belfield, Clive R., Milagros Nores, Steve Barnett and Lawrence Schweinhart (2006). The High/Scope Perry Preschool Program, *Journal of Human Resources*. 41(1): 162–190.

Betts, Julian R. and Jamie L. Shkolnik (2000). The Effects of Ability Grouping on Student Achievement and Resource Allocation in Secondary Schools, *Economics of Education Review*. 19(1): 1–15.

Blau, David and Janet Currie (2006). Pre-School, Day Care, and After-School Care: Who's Minding the Kids?, In: Eric A. Hanushek and Finis Welch (eds.), *Handbook of the Economics of Education*. Vol. 2, pp. 1163–1278. Amsterdam: North-Holland.

Brunello, Giorgio and Daniele Checchi (2007). Does School Tracking Affect Equality of Opportunity? New International Evidence, *Economic Policy*. 22(52): 781–861.

Chiswick, Barry and Noyna DebBurman (2006). Pre-School Enrolment: An Analysis by Immigrant Generation, *Social Science Research*. 35(1): 60–87.

Cunha, Flavio, James J. Heckman, Lance Lochner and Dimitriy V. Masterov (2006). Interpreting the Evidence on Life Cycle Skill Formation, In: Eric A. Hanushek and Finis Welch (eds.), *Handbook of the Economics of Education*. Vol. 1, pp. 697–812. Amsterdam: North-Holland.

Currie, Janet (2001). Early Childhood Education Programs, *Journal of Economic Perspectives*. 15(2): 213–238.

Deaton, Angus (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Baltimore: The Johns Hopkins University Press.

DuMouchel, William H. and Greg J. Duncan (1983). Using Sample Survey Weights in Multiple Regression Analyses of Stratified Samples, *Journal of the American Statistical Association*. 78(383): 535–543.

Dustmann, Christian (2004). Parental Background, Secondary School Track Choice, and Wages, *Oxford Economic Papers*. 56(2): 209–230.

Epple, Dennis, Elizabeth Newlon and Richard Romano (2002). Ability Tracking, School Competition, and the Distribution of Educational Benefits, *Journal of Public Economics*. 83(1): 1–48.

European Commission (1999, 2000). *Key Data on Education in Europe*. Luxembourg: Office for Official Publications of the European Communities.

Feinstein, Leon, Donald Robertson and James Symons (1999). Pre-School Education and Attainment in the National Child Development Study and British Cohort Study, *Education Economics*. 7(3): 209–234.

Figlio, David N. and Marianne E. Page (2002). School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Inequality?, *Journal of Urban Economics*. 51(3): 497–514.

Fuchs, Thomas and Ludger Wößmann (2007). What Accounts for International Differences in Student Performance? A Re-examination using PISA Data, *Empirical Economics*. 32(2–3): 433–464.

Galindo-Rueda, Fernando and Anna Vignoles (2007). The Heterogeneous Effect of Selection in UK Secondary Schools, In: Ludger Wößmann and Paul E. Peterson (eds.), *Schools and the Equal Opportunity Problem*. pp. 103–128. Cambridge, MA: MIT Press.

Ganzeboom, Harry B.G., Paul M. De Graaf and Donald J. Treiman (1992). A Standard International Socio-Economic Index of Occupational Status, *Social Science Research*. 21(1): 1–56.

Garces, Eliana, Duncan Thomas and Janet Currie (2002). Longer-Term Effects of Head Start, *American Economic Review*. 92(4): 999–1012.

Gonzalez, Eugenio J. and Julie A. Miles (eds.) (2001). *User Guide for the TIMSS 1999 International Database*. Chestnut Hill, MA: Boston College.

Gonzalez, Eugenio J. and Teresa A. Smith (eds.) (1997). *User Guide for the TIMSS International Database: Primary and Middle School Years*. Chestnut Hill, MA: Boston College.

Hanushek, Eric A. and Ludger Wößmann (2006). Does Early Tracking Affect Educational Inequality and Performance? Differences-in-Differences Evidence across Countries, *Economic Journal*. 116(510): C63–C76.

Magnuson, Katherine A., Christopher J. Ruhm and Jane Waldfogel (2007). Does Prekindergarten Improve School Preparation and Performance?, *Economics of Education Review*. 26(1): 33–51.

Manning, Alan and Jörn-Steffen Pischke (2006). Comprehensive versus Selective Schooling in England and Wales: What Do We Know? IZA Discussion Paper 2072. Bonn: Institute for the Study of Labor.

Martin, Michael O. and Dana L. Kelly (eds.) (1996). *Third International Mathematics and Science Study Technical Report, Volume I: Design and Development*. Chestnut Hill, MA: Boston College.

Martin, Michael O. and Dana L. Kelly (eds.) (1997). *Third International Mathematics and Science Study Technical Report, Volume II: Implementation and Analysis – Primary and Middle School Years*. Chestnut Hill, MA: Boston College.

Martin, Michael O., Kelvin D. Gregory and Steven E. Stemler (eds.) (2000). *TIMSS 1999 Technical Report*. Chestnut Hill, MA: Boston College.

Meghir, Costas and Mårten Palme (2005). Educational Reform, Ability, and Family Background, *American Economic Review*. 95(1): 414–424.

Moulton, Brent R. (1986). Random Group Effects and the Precision of Regression Estimates, *Journal of Econometrics*. 32(3): 385–397.

Mullis, Ina V.S. and Michael O. Martin, et al. (2000). *TIMSS 1999 International Mathematics Report: Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*. Chestnut Hill, MA: Boston College.

Nickell, Stephen (2004). Poverty and Worklessness in Britain, *Economic Journal*. 114(494): C1–C25.

Organisation for Economic Co-Operation and Development (OECD) (2002). *Education at a Glance: OECD Indicators 2002*. Paris: OECD.

Pekkarinen, Tuomas, Roope Uusitalo and Sari Pekkala (2006). Education Policy and Intergenerational Income Mobility: Evidence from the Finnish Comprehensive School Reform. IZA Discussion Paper 2204. Bonn: Institute for the Study of Labor.

Renz, Monika (1994). Ganztagsschule in anderen EG-Ländern. Wie sieht die Realität aus? *Die Ganztagsschule* (1/2).

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

Schütz, Gabriela, Heinrich W. Ursprung and Ludger Wößmann (2005). Education Policy and Equality of Opportunity. CESifo Working Paper 1518. Munich: CESifo. (available online at http://www.cesifo-group.de/~DocCIDL/cesifo1_wp1518.pdf)

United Nations Educational, Scientific and Cultural Organization (UNESCO) (1998, 2000). *World Education Report*. Paris: UNESCO Publishing.

UNESCO Institute for Statistics (2003). Data on National Education Systems. http://www.uis.unesco.org/pagesen/DBSysCri.asp (05.12.2003).

Waldinger, Fabian (2006). Does Tracking Affect the Importance of Family Background on Students' Test Scores? Mimeo, London School of Economics.

White, Halbert (1984). *Asymptotic Theory for Econometricians*. Orlando: Academic Press.

Wooldridge, Jeffrey M. (2001). Asymptotic Properties of Weighted *M*-Estimators for Standard Stratified Samples, *Econometric Theory*. 17(2): 451–470.

World Bank (2005). World Development Indicators. http://devdata.worldbank.org/dataonline

Wößmann, Ludger (2003). Schooling Resources, Educational Institutions and Student Performance: The International Evidence, *Oxford Bulletin of Economics and Statistics*. 65(2): 117–170.

Wößmann, Ludger (2008). How Equal Are Educational Opportunities? Family Background and Student Achievement in Europe and the United States. *Zeitschrift für Betriebswirtschaft* 78 (1): 45–70 .

SUMMARY

We provide a measure of equality of educational opportunity in 54 countries, estimated as the effect of family background on student performance in two international TIMSS tests. Using cross-country variation in education policies and its interaction with family background at the student level, we then estimate how equality is related to organizational features of the education system. We find that equality of opportunity is positively related to late tracking into different school types and to longer pre-school education. Pre-school enrollment has an inverted U-shaped relationship with equality. Equality is negatively related to private school financing, but positively to private provision.

