EUROPEAN ECONOMY

EUROPEAN COMMISSION
DIRECTORATE-GENERAL FOR ECONOMIC AND FINANCIAL AFFAIRS

ECONOMIC PAPERS

ISSN 1725-3187
http://europa.eu.int/comm/economy_finance

N° 190 September 2003

European "education production functions": what makes a difference for student achievement in Europe?

by
Ludger Wößmann
CESifo Munich

Economic Papers are written by the Staff of the Directorate-General for Economic and Financial Affairs, or by experts working in association with them. The "Papers" are intended to increase awareness of the technical work being done by the staff and to seek comments and suggestions for further analyses. Views expressed represent exclusively the positions of the author and do not necessarily correspond to those of the European Commission. Comments and enquiries should be addressed to the:

European Commission
Directorate-General for Economic and Financial Affairs
Publications
BU1 - -1/180
B - 1049 Brussels, Belgium

ISBN 92-894-6194-2

KC-AI-03-011-EN-C

ECFIN/349/03-EN

This paper only exists in English.

European "Education Production Functions":
What Makes a Difference for Student Achievement in Europe?

Ludger Wößmann

CESifo Munich
ifo Institute for Economic Research
Poschingerstr. 5
81679 Munich, Germany
Phone: (+49) 89 / 9224-1699
E-mail: woessmann@ifo.de
Internet: www.cesifo.de/link/woessmann_1.htm

Abstract

This paper estimates the effects of family background, resources, and institutions on student performance in 17 Western European school systems. Family background has a strong effect in Europe, remarkably similar in size to the United States. France and Flemish Belgium achieve the most equitable performance for students from different family backgrounds, and Britain and Germany the least. Equality of opportunities is unrelated to countries' mean performance. Quantile regressions show little variation in family-background effects across the ability distribution. There is little evidence of substantial class-size effects, but slight evidence of effects of material shortage and teacher experience in some countries. Stronger evidence exists for effects of within-country variations in schools' hiring autonomy, testing, and homework.

JEL Classification: I21, H52

Keywords: Educational production, student performance, Western Europe, family background, class size, school-system institutions, TIMSS, effect heterogeneity, equity-efficiency tradeoff, United States

*Paper prepared for and finished during my Visiting Fellowship at the Directorate General for Economic and Financial Affairs (DG ECFIN), European Commission, Brussels, whose hospitality I gratefully acknowledge. During the initial stages of this research, I also benefited from the hospitality of PEPG at Harvard University and the NBER in Cambridge MA. I would like to thank Sascha Becker, Charlotte Lauer, George Psacharopoulos and Marty West as well as seminar participants at the European Commission for helpful comments and discussion, and Andreas Ammermüller for research assistance in the construction of the country-specific databases.

1. Introduction

Empirical estimates of education production functions inform about the relationship between students' educational performance and their family background as well as their schools' resource endowment. Such evidence provides the foundation for many policy discussions and initiatives in the United States (see, e.g., the recent large-scale class-size reduction program in California (cf. Jepsen and Rivkin 2002)). Leading scholars in the economics of education have lamented that comparable evidence is lacking for European countries. For example, George Psacharopoulos (2000, p. 84) asks, "Where are the educational production function results that would explain within-country educational achievement [in Europe]?" Likewise, Eric Hanushek (2002, p. 2049) deplores that the restriction of knowledge on educational production to the United States "is an unfortunate limitation forced by the availability of comparable studies from other countries." While some evidence exists by now for several developing (e.g., Glewwe 2002), newly industrializing (e.g., Wößmann 2003a), and Eastern European transition countries (Ammermüller et al. 2003), estimates of education production functions for other developed, and notably Western European, countries are nearly not existent.

This paper starts to fill this gap by estimating European education production functions, using data on representative samples of middle-school students in 17 Western European school systems (in 15 countries). For many of the countries, this is the first time that such evidence is available. Using data from a large international comparative student achievement test – the Third International Mathematics and Science Study (TIMSS) – the estimates presented in this paper have the additional advantage of being directly comparable across countries. This allows a comparison of the determinants of student performance both across European countries and with the United States.

The paper deals with three research questions and three groups of potential determinants. The research questions addressed are: What are the main determinants of student performance in each system? How do the results compare across the different European systems? And, is educational production in Europe different from or similar to the United States? These questions are applied to three groups of potential determinants: family background measures, which reveal the extent of equality in opportunity achieved in a school system; resource measures, which inform about the possibility of education policy to further student performance through class-size reductions, improved resource endowments, and a higher-quality teaching force; and institutional measures, which constitute other potential options for policy to affect performance.

The European evidence is particularly revealing because abundant US evidence on the determinants of student learning has been around for a long time, but nobody knew whether the US findings are valid in the European setting. European researchers have always wondered whether the US evidence, as well as the evidence from several other countries, applies in a similar way to Europe. School systems differ in many respects, which might have an impact on the importance of the different potential determinants of student performance. For example, systems differ in the extent of centralization, the competitiveness and flexibility of the teacher labor market, the tracking of students, and the performance information available due to central exams, league tables, and the like. Because of such differences, it was never quite clear what one country can learn from findings based on other systems in general, and what Europe can learn from findings based on the US system in particular.

Probably the main reason for the lack of European evidence was that large-scale testing and thus extensive micro data on student performance simply did not exist in many European countries. Thus, in one of the few previous cross-country studies on resource endowments as a potential determinant of student performance, Gundlach et al. (2001) use aggregate time-series data to extract evidence on expenditure effects in several European countries. Now, more extensive international comparative student achievement tests have become available, which allow the lack of microeconometric evidence for Europe to be overcome. In TIMSS, representative samples of students in lower secondary education have been tested in math and science in 17 Western European school systems, and abundant data on family background, resource endowments, and institutional features are available from student, teacher, and school background questionnaires.

The standard method applied to these data in order to address the research questions are microeconometric student-level least-squares regressions, weighted by sampling probabilities and adjusted for clustering within schools. As least-squares estimates of the effect of class size on student performance may be particularly prone to endogeneity bias, given that students are regularly allocated to classes in compensatory ways, following Wößmann and West (2002) I also present estimates of class-size effects based on a quasi-experimental strategy that excludes effects of between- and within-school sorting through a combination of school fixed effects and instrumental variables, in effect exploiting between-grade variations in individual schools made possible by the particular design of the TIMSS database.

The paper is structured as follows. Section 2 places the paper into the existent literature on education production functions. Section 3 introduces the database and some descriptive statistics. As a general rule, the discussion in this paper will focus on student performance in math, as this is generally viewed as being most easily comparable across countries. Results on science performance are presented in appendix tables. Section 4 presents evidence on the effects of family background on student performance in the 17 Western European school systems, compares the European evidence to results for the United States, performs quantile regressions to analyze effect heterogeneity, and discusses whether the results imply an equity-efficiency tradeoff in educational production. Section 5 deals with the estimation of resource effects, particularly focusing on a quasi-experimental design to estimate class-size effects. Section 6 presents evidence on the relationship between variations in several institutional features and student performance within the European countries. Section 7 summarizes the results and concludes on their limitations.

2. Previous Research

After presenting a basic framework for the estimation of education production functions, this section gives a brief overview of the vast literature for the United States and of studies for less developed countries, as a background for the European discussion. It then briefly reviews studies that exist for European countries.

The basic theoretical foundation is a straightforward production-function framework. In educational production, the educational output – measured by test score T – is produced using several inputs:

$$T = f\left(F, R, I\right) \quad . \tag{1}$$

The inputs employed are not only schooling resources R, but also the inputs of the "home production" part of the educational process, captured by measures of family background F. Furthermore, institutional features I of the school system may influence the efficiency of educational production. While family-background effects are generally found to be of utmost importance for students' educational performance, resource inputs have received much more attention in the empirical literature. The econometric specification of the specific equation to be estimated, as well as the interpretation of ensuing results, comprises many econometric problems, which will be discussed in the relevant sections below.[1]

Since the mid-1960s, a vast literature has unfolded in the United States that tries to estimate some version of equation (1). Hanushek (2002) counts 376 published estimates of US education production functions from 89 separate articles published in refereed journals or books until 1994. In general, the US research has focused on resource effects, and on class-size effects in particular. Although extensive reviews of this literature exist (cf. Hanushek 1986; 2002; Card and Krueger 1996), much controversy remains what conventional estimates of education production functions can and do tell about the causal determination of student performance in the United States, particularly with respect to the effectiveness of class-size reductions (cf. the controversy between Hanushek (2003) and Krueger (2003)). More recent influential studies have tried to avoid endogeneity biases in the class-size estimates by using experimental evidence from the Tennessee Project STAR (cf. Krueger 1999) and quasi-experimental research design for data from Connecticut (Hoxby 2000), one of them finding substantial class-size effects, the other none at all. While most US studies include some control for family background in their estimation, there is much less direct focus on family-background and other effects in educational production (cf. Hanushek 2002, p. 2082).

Apart from the US evidence, there are also several studies on education production functions in developing countries, summarized e.g. in Hanushek (1995) and Glewwe (2002). While many of the older studies for developing countries have been criticized for their lack of methodological and data quality, several more recent studies have presented convincing evidence on education production functions for countries in Latin America, Africa, and Asia (see the references in Glewwe 2002). In particular, several studies use data from randomized trials to estimate the impact of distinct educational policies on student performance, most recently in Kenya (see the overview by Kremer 2003), and quasi-experimental evidence on class-size effects exists for South Africa (Case and Deaton 1999), as well as Israel (Angrist and Lavy 1999). While nearly all of these studies are confined to individual countries, recent internationally comparable evidence exists for several countries in francophone sub-Saharan

[1] For a general discussion of the specification and interpretation of education production functions, see e.g. the reviews by Hanushek (1986; 2002).

Africa (Michaelowa 2001) and for several newly industrializing countries in East Asia (Wößmann 2003a).

In Europe, there is nothing to match the US literature, and not even the literature for developing countries, with the possible limited exception of the United Kingdom. This is particularly disappointing as European evidence would seem much more directly comparable to the US setting than developing-country evidence, given the relative levels of economic development and educational attainment.

The evidence for the United Kingdom is capably reviewed in Vignoles et al. (2000) and Levaĉić and Vignoles (2002),[2] who conclude that "the UK literature has relatively few methodologically strong studies [… and] lacks both depth and breadth of coverage" (Levaĉić and Vignoles 2002, p.323). They argue that the limited amount of high quality research in the UK can only be overcome once higher-quality datasets become available. They count only four UK studies that estimate education production functions on student-level data, all of which use the National Child Development Study (NCDS) dataset: Feinstein and Symons (1999), Dolton and Vignoles (2003), Dearden et al. (2002), and Dustmann et al. (2003). These studies produce some evidence of small resource effects, but little evidence of strong resource effects. Another recent study, applying an instrumental-variable strategy to the same dataset, finds significant class-size effects (Iacovou 2002). Feinstein and Symons (1999) also explicitly look at family-background effects and show that these dwarf the resource effects.

Apart from the UK evidence, there are scattered studies for the Scandinavian countries. Bonesrønning (1996) estimates the effect on student performance of the number and size of departments in Norwegian schools, and Bonesrønning (2003) employs Angrist and Lavy's (1999) method of using a maximum class size rule to obtain quasi-experimental evidence on class-size effects in Norway, finding small negative class-size effects and analyzing how these vary across student subgroups. Häkkinen et al. (2003) estimate fixed-effect panel data models of education production functions for Finland, finding strong effects of parental education on student achievement, but no effect of teaching expenditure. Björklund et al. (2003) study the effect of family background on student performance in Sweden, finding a relatively stable effect over the 1990s despite major school reforms. Applying a quasi-experimental approach exploiting variations between school-time and summer learning to a sample of 16 schools in

2 The former study also includes another informative review of the international (mostly US) evidence.

Stockholm, Lindahl (2001a,b) finds significant class-size effects, as well as attenuating effects of schooling on the performance lag of students with non-native parents. Based on rich panel data from Danish administrative registers and using the probability of attaining a post-compulsory education as a measure educational output, Heinesen and Graversen (2001) find small positive effects of educational expenditures and less evidence for effects of teacher-student ratios, extensively controlling for family-background effects. Also for Denmark, and using yet different measures of educational output – namely students' own assessments from survey data on issues such as their satisfaction with their school, their self-esteem, and their social skills – Schindler Rangvid (2001) finds no evidence that school resources significantly affect these outcomes.

A few recent further European studies exist. Levin (2001) and Dobbelsteen et al. (2002) estimate class-size effects in the Netherlands, using an instrumental-variable strategy basically based on a maximum class-size rule similar to Angrist and Lavy's (1999), and finding no evidence of direct class-size effects, but an indirect *positive* effect of larger classes on student performance working through the increased likelihood of having peers of comparable ability in a class who further students' performance. Brunello and Checchi (2003) use regional resource data and quantitative educational attainment as their output measure to find a stronger effect of parental education than of lower student-teacher ratios. In a recent study applying quantile regressions to the German sample of the PISA data, Fertig (2003) finds strong effects of parental education and counterintuitive class-size effects, among other family-background and school effects. Wolter and Vellacott (2002) use the Swiss sample of the PISA data to estimate the effect of sibling rivalry on educational performance, and Wolter (2003) extents the sibling analysis to six countries, five of which are European. Wößmann and West's (2002) quasi-experimental class-size evidence based on within-school between-grade variations in TIMSS contains results for eight Western European school systems, finding class-size effects in a few countries but rejecting substantial class-size effects in most. Hanushek and Luque (2003) also estimate education production functions using TIMSS data, but detailed results for individual European countries are not reported. Ammermüller et al. (2003) use the TIMSS data to estimate education production functions for seven Eastern European countries.

In sum, there is not much evidence on education production functions in Europe in general, and virtually none for continental European countries. Some of the European studies that exist are not based on representative samples of students, and given the data limitations, some use educational output measures with limited variation such as whether students attain a post-compulsory education. Furthermore, several studies do not really focus on family-background, resource, or institutional effects in general, but on specific questions such as sibling rivalry, effects of department size, or class-size effects. In addition to these limitations, the existing European evidence focuses on specific countries, but no detailed education production function evidence exists that is readily comparable across countries. Thus, to cite Psacharopoulos (2000, p. 92) again, "European research on the economics of education lags considerably behind the US," and he even ventures to say that "more research has been done on the economics of education in developing countries than in Europe."

3. Database for Education Production Functions in Europe

This section discusses the TIMSS student performance database used in this paper and presents some comparative descriptive statistics on the Western European education systems both from TIMSS and from other sources.

3.1. The TIMSS Student Performance Data

The Third International Mathematics and Science Study (TIMSS) is a large-scale cross-country comparative test of student achievement, conducted in 1995 by the International Association for the Evaluation of Educational Achievement (IEA), an independent international cooperative of national research institutions and government agencies. From Western Europe, 17 separate school systems in 15 countries participated in TIMSS: Austria, Flemish and French Belgium, Denmark, England and Scotland, France, Germany, Greece, Iceland, Ireland, the Netherlands, Norway, Portugal, Spain, Sweden, and Switzerland.[3] The TIMSS database combines individual student-level performance data in math and science with extensive information from student, teacher, and school-principal background questionnaires for nationally representative samples of students in each of the countries.

Two other comparable large-scale international student achievement tests have been conducted since the original TIMSS test. The repeat study of TIMSS in 1999 included only

3 Thus, among the bigger Western European countries, only Finland and Italy are missing in TIMSS.

five Western European countries, as compared to the original 17.[4] The Programme for International Student Assessment (PISA), conducted in 2001 by the Organisation for Economic Co-operation and Development (OECD), included basically all Western European countries. There are three main differences between TIMSS and PISA. First, TIMSS follows a curriculum-based test design, so that its questions are meant to reflect the students' current curriculum and are thus appropriate for them (cf. Martin and Kelly 1996; 1997). Thus, TIMSS covers the core material common to the curricula in the majority of participating countries, and every effort was made to help ensure that the items did not exhibit any bias towards or against particular countries. A matching analysis later performed between the test and the actual curriculum in each country showed that disregarding any items not covered in the country's intended curriculum barely affected the results. In contrast to this, PISA was constructed to cover "the range of skills and competencies that were … considered to be crucial to an individual's capacity to fully participate in, and contribute meaningfully to, a successful modern society" (OECD 2001, p. 27). Thus, PISA is meant to assess how well students approaching the end of compulsory schooling are prepared to meet real-life challenges, rather than to master their curriculum. While without doubt this focus on preparedness for everyday life is of high interest, for the assessment of the effectiveness of schooling policies it might be better warranted to focus on the achievement of the curricular content, as it might be a bit unfair to schools to ask them for outcomes that they were never meant to achieve.

A second difference between TIMSS and PISA are their target populations. In TIMSS, the target population of middle school students to which each participating country administered the test was defined as those students enrolled in the two adjacent grades with the largest proportion of 13-year-olds at the time of testing.[5] By contrast, PISA's target population were the 15-year-olds of each country, regardless of the specific grade they may currently be attending. Again, the grade-based target population may be better suited than the age-based one to assess the effectiveness of particular schooling policies. Furthermore, self-selection into the sample of students might already constitute a problem with 15-year-olds in some

4 These were: Flemish Belgium, England, Finland, Italy, and the Netherlands.

5 In most countries, this corresponded to 7th and 8th grade of formal education (excluding pre-primary education). Exceptions are Denmark, Norway, Sweden, and the German-speaking cantons of Switzerland, where tested students had 6/7 years of formal education, as well as England and Scotland with 8/9 years (Martin and Kelly 1997).

countries, while schooling is compulsory for 13-year-olds in all countries. Also, as will be discussed in detail in Section 5.2, the two-grade design of the TIMSS population offers a unique possibility for the identification of causal class-size effects that is given neither in PISA nor in the TIMSS-Repeat study, which tested only the upper of the two grade levels tested in TIMSS.

Third, TIMSS includes an extensive teacher background questionnaire which is not available in PISA, and which provides detailed information on teachers' gender, education, and experience, as well as their influence in different areas of decision-making and their testing and homework-assignment practices, among other things. Finally, it should be borne in mind that TIMSS can build on a long experience of the IEA, which has been developing and conducting cross-national studies of student achievement since 1959. Thus, for the purposes of this paper, the TIMSS test has several particular advantages over PISA. Surely, any future analysis of PISA could yield interesting additional insights that would complement the findings presented here.

In TIMSS, each participating country randomly sampled the schools to be tested in a stratified sampling design, and within each of these schools, generally one class was randomly chosen from each of the two grades and all of its students were tested in both math and science, yielding a representative sample of students within each country.[6] Schools in geographically remote regions, extremely small schools, and schools for students with special needs were excluded from the target population. Within sampled schools, disabled students who were unable to follow even the test instructions were excluded; students who merely exhibited poor academic performance or discipline problems were required to participate (Martin and Kelley 1996). The overall exclusion rate was not to exceed 10% of the total student population.[7]

As a general rule, each country was meant to sample at least 150 schools (Gonzalez and Smith 1997). Response rates were not perfect (even after taking into account randomly

---

6 In Germany, one Bundesland (Baden-Württemberg) did not participate in TIMSS, and in Switzerland, four cantons did not participate. Denmark and Greece had unapproved sampling procedures at the classroom level (Beaton et al. 1996).

7 England was the only country not to meet this general rule with an overall exclusion rate of 11.3% because it excluded schools that were selected for specific national evaluation samples.

selected replacement schools), though, while some countries sampled even more schools.[8] As reported in Table 1, Switzerland features the largest sample size across the Western European countries with 11,722 students in 613 classrooms in 327 schools, while England has the smallest number of sampled students at 3,579 and the Netherlands has the smallest number of sampled classrooms (187) and schools (95).

TIMSS gave rigorous attention to quality control, using standardized procedures to ensure comparability in school and student sampling, to prevent bias, and to assure quality in test design and development, data collection, scoring procedures, and analysis. The TIMSS achievement tests were developed through an international consensus-building process involving inputs from international experts in math, science, and measurement, and were endorsed by all participating countries. Given the international standardization of the test results, the cooperative nature of the test development, its endorsement by all participating countries, and the substantial efforts to ensure high-quality sampling and testing in all countries, the TIMSS student performance and background data should be comparable across countries. This should also make the empirical estimates presented in this paper directly comparable across the different countries. This makes the database uniquely capable of using student, class, and school level data to analyze the determination of student performance across the Western European countries.

The performance data are merged with background data from three different TIMSS background questionnaires for each individual student.[9] I draw information on student- and family-background characteristics from the student background questionnaires, where students report the level of their parents' education, the number of books in their home, whether they live with both parents and were born in the country, and their sex and age. The math and science teacher background questionnaires contain data on the actual sizes of the respective math and science classes, as well as on teacher characteristics such as sex, experience, and education. They also report the teachers' influence on the subject matter taught, on the textbook used, and on what supplies to be purchased, as well as their time spent on preparing and grading exams, the amount of homework assignments per week, and whether teaching

8 As a general guideline, TIMSS required participation rates of 85% or higher. This was not met by Austria, French Belgium, the Netherlands, and Scotland. Flemish Belgium, England, and Germany met the participation rates only after replacement schools were included, as did the United States (Beaton et al. 1996).

9 The background questionnaires are available at http://timss.bc.edu/timss1995i/database/UG1_Sup2.pdf.

was thought to be limited by uninterested or interested parents. The school-principal background questionnaires provide information on the community location of the school, on shortage of materials, instruction time, average class sizes in the two relevant grades, and on whether the school had responsibility for hiring its teachers. Most of these background variables based on qualitative survey data were transformed into dummy variables for the estimations of this paper.

Complete performance data are available for all the students participating in TIMSS. As can be seen from Table A1 in the appendix, though, some countries did not administer certain questions in an internationally comparable way, so that these variables are missing in these countries. This case is evident most severely for data on teachers' education, which is not available in Austria, the two Belgian systems, and Denmark, and on school responsibility for hiring teachers, not available in England, Norway, and Scotland.

Furthermore, some students, teachers, and school principals failed to answer some items in the background questionnaires. While in general, response rates are reasonable, they are pretty bad for some variables in some countries. Most evidently, some of the resource and institutional data taken from the teacher and school questionnaires are missing for up to two thirds of the students in England and Germany. By contrast, data availability is very good in other countries, such as Portugal and Spain. Response rates are also very good for nearly all of the family background data, particularly in the case of student characteristics. The only exceptions here are parental education, where England has no internationally comparable data and several other countries have data missing for more than a third of the students; France has no internationally comparable data on whether the student was born in the country; and in Germany, 40% of the data on community location are missing. All in all, England has the worst data situation, closely followed by Germany. In these two countries, many estimates will not be based on samples large enough to yield statistically significant and representative results.

In cases of missing data, I chose to impute missing values within each country for the analyses in this paper. This seems superior to dropping all students with missing data on some explanatory variables altogether, because the latter would disregard the information available for these students on other explanatory variables, would severely reduce the sample size in many countries, and would probably introduce bias because it seems likely that observations are not missing at random. Appendix 1 presents the details of the method used to impute missing data, which is based on least-squares, probit, and ordered-probit models relating the observations from students with original data to a set of "fundamental" explanatory variables available for all students.

3.2. Comparative Descriptive Statistics on European Education

To give a background for the following econometric estimations, this section presents country means of TIMSS performance and background data, as well as some further descriptive statistics on education in Europe. (Readers only interested in the estimation results may want to have a quick glance at Tables 1, 2, and A2a-A2c and jump to Section 4.)

TIMSS measured student performance separately in math and science, using an international achievement scale with scores having an international mean of 500 and an international standard deviation of 100. As shown in Table 1, the mean math performance in TIMSS of the Western European countries ranges from 438.3 in Portugal to 561.7 in Flemish Belgium. As a comparison, Singapore was the international top performer at 622.3, South Africa was the lowest performer at 351, and the United States scored an average of 487.8 test-score points in math. The cross-country distribution of science performance in Western Europe is broadly comparable to the math performance, with the same top and low performers. Some countries still feature quite a difference in their performance across subjects, and the cross-country correlation of mean performance stands at 0.64. Portugal has the lowest variation in math performance among its students, both when measured by the absolute standard deviation in test scores (63.7) and relative to its mean performance (14.5). On the relative measure, Flemish Belgium and France also feature relatively low performance variations. At the other extreme, England has the highest standard deviation in absolute terms (92.7), while in relative terms it is Greece (19.4). The United States also shows a relatively large performance variation among its students.

In addition to the performance data, data from the TIMSS background questionnaires reveal a broader picture of the educational production process in the different European countries. Tables A2a-A2c in the appendix present country means of the data on family background, resources, and institutional features used in this paper. The average parental education varies widely among European countries (Table A2a). On the one extreme, 57% of Portuguese parents do not have any secondary education. On the other extreme, 40% of parents in French Belgium have a finished university education. Some of the cross-country variation in parental education levels may reflect differences in the education systems, though. The number of books in the students' home may be a proxy for the educational background of the family that is more readily comparable across countries. Portugal again fares relatively low on this measure, while Norway features relatively high values. The distribution across the five categories of books at home in countries such as England, Germany, and the United States is actually very close together.

Across Europe, the Netherlands had the lowest share of students not living with both parents at 9%, while Scotland had the highest share at 19%. The share was even higher in the United States at 21%. The share of immigrant students ranges from 3% in Flemish Belgium, Ireland, and Spain to 12% in Switzerland. In all countries, students are roughly equally divided into boys and girls. The average age of the tested students ranges from 13.1 years in Greece and Iceland to 14.3 years in Germany.[10] In all countries, the student population is roughly equally divided between the two tested grades. Sweden and Switzerland tested students in a third grade above the other two, with roughly one third of their samples coming from each grade. In terms of the community in which the schools are located, Norway features a relatively high share of geographically isolated locations (12%) and a relatively low share of schools located close to the center of a town (also 12%). By contrast, several countries do not feature tested school in isolated areas, and 62% of Portuguese students visit a school close to a town center.

The lowest mean class sizes at around 20 students per class can be found in Denmark, the two Belgian systems, and Switzerland (Table A2b). Specifically in math, Austria has by far the smallest classes at 10.6 students on average, which is due to specific streaming of math classes.[11] The largest average class sizes at around 28 can be found in Greece and Spain. Schools in Denmark feel particularly strongly that their capacity to provide instruction is affected by shortage or inadequacy of instructional materials, with 35% of Danish schools reporting a lot of shortage and only 9% no shortage of materials. On the other hand, about three quarters of schools in Austria and Flemish Belgium did not feel affected by inadequacy

---

10 Germany actually chose to test its 7^{th} and 8^{th} grade level to provide a decent curricular match, although that meant not testing the two grades with the most 13-year-olds. This breach of the official TIMSS age/grade specification led to German students being substantially older than those in most other countries.

11 The specific data on science teaching from the science teacher questionnaires are taken from the teacher who taught the most lessons of science per week to the class, which may not perfectly reflect all science teaching in systems where science is taught in several separate subject lessons. This is another reason for focusing on math in this paper and presenting science results only in the appendix.

of materials, and less than 2% of the schools in Austria, Flemish Belgium, Ireland, and the Netherlands reported that they had a lot of shortage. This shows that there is basically no variation in some countries in some of the measures analyzed in this paper. Average instruction time ranges from 606 hours per year in Iceland to 971 hours in Switzerland. More than two third of math teachers in Portugal (and the United States) are female, while less than one quarter are female in Switzerland and the Netherlands. Portuguese math teachers have also relatively few years of teaching experience (less than 8), while math teachers in Flemish Belgium and Spain have on average more than 21 years of experience. In many countries, the education level of teachers does not feature much variation, and the cross-country comparability of these data on education levels may again be limited.

The extent of school autonomy for hiring teachers does not vary much within most countries (Table A2c), limiting the possibility of estimating the effect of this institutional feature of the school system on the basis of single-country estimations relative to cross-country estimations as in Wößmann (2003b). There are systems where nearly all schools have this responsibility (such as the Belgian systems, Iceland, the Netherlands, Sweden, and the United States), while only very few schools have this responsibility in other systems (Germany and Greece). Nearly two thirds of Spanish math teachers report that they have a lot of influence on the subject matter to be taught, while this is the case for less than 10% in French Belgium and France. In terms of the specific textbooks to be used, few Greek and Norwegian teachers report having a lot of influence, while about half of the math teachers in Flemish Belgium, Ireland, and Spain do so. And while few math teachers in Austria, French Belgium, and Greece can strongly influence what supplies are purchased, 37% of German math teachers can. The time that teachers spend outside the formal school day on preparing or grading student tests or exams ranges from 1.5 hours per week in Scotland to 4.3 hours per week in France. Reported math homework assignments range from less than three quarters of an hour in Scotland and Sweden to more than 3 hours in Greece. In most countries, very few teachers report that their teaching is limited either by particularly uninterested or by particularly interested parents. However, in Spain, 22% of math teachers report uninterested parents as being limiting, and in Iceland, 26% report interested parents as being limiting.

Some further aggregate descriptive statistics on education in Europe are presented in Table 2. Obviously, the absolute size of the relevant student population varies widely across the European countries. While there are less than 10,000 students in 7th and 8th grade in Iceland

and about 100,000 in Denmark, this stands at nearly 1.5 million in Germany and more than 1.6 million in France. In the United States, the student population in the two grades is 6.3 million. All countries have virtually universal school enrollment of children aged 5-14. Table 2 also reports the age range at which 90% of an age group are enrolled in school, average school expectancy, and entry rates to university. Educational expenditure per student, measured in US dollars converted using purchasing power parities, ranges from 1,950 in Greece to 7,601 in Switzerland. Relative to GDP per capita, Austria features the highest expenditure per student at 35%. Teacher salary levels are highest in Switzerland and lowest in Greece, Norway, and Sweden. As a measure of the educational background of the adult population, average years of schooling range from 4.5 in Portugal to 11.8 in Norway (and 12.2 in the United States). Recent comparable estimates show that education pays most on the labor markets in Portugal and the United Kingdom, with rates of return to education of 9.7% and 9.4% (and 10.0% in the United States), and least in Sweden and Norway at 4.1% and 4.6%. Also, unemployment rates differ by education levels achieved, with Ireland having the largest difference between university graduates and people with less than upper secondary education at 13%. Greece is the only country where university graduates do not face substantially lower unemployment rates than people with less than upper secondary education.

4. Family Background and Student Performance

In this section, after discussing the empirical model, I present results of regressions of student performance on family-background and related measures for the 17 Western European education systems. This is followed by a comparison of the European results to results for the United States. Next, quantile regression estimates investigate heterogeneity of the family-background effects by the underlying ability of students. Finally, I analyze whether the European cross-country pattern of results suggests that there is a tradeoff between equity and efficiency in educational production or not. Here as well as in the remaining sections, the discussion will focus on results for math performance. Results for science performance are relegated to Tables A3-A7 in the appendix without further discussion.

4.1. *Empirical Model*

To assess the influence of students' family background on their educational performance in the different Western European countries, I estimate education production functions for each country of the following form:

$$T_{ics} = F_{ics}\alpha_1 + D_{ics}^F\delta_1 + \left(D_{ics}^F F_{ics}\right)\delta_2 + \varepsilon_{ics} \quad , \qquad (2)$$

where T is the test score of student i in class c in school s and F is the vector of family-background variables. The coefficient vectors α_1, δ_1, and δ_2 are to be estimated. Since family-background characteristics such as parental education can clearly be viewed as being exogenous to student performance, and since there is no self-selection into the sample as 7^{th}- and 8^{th}-grade schooling is compulsory, least-squares estimates of α_1 should represent causal family-background effects. The inclusion of the imputation controls D^F and the structure of the error term ε are discussed below. The estimation does not control for school characteristics, such as schools' resource endowments or institutional features as in equation (1), because in this section I am interested in the total impact of family background on student performance, including any effect that might work through families' differential access to schools or their influence on school policies.

It helps to clarify in advance what the estimates of the coefficients α_1 on the family-background variables (and of the coefficients on the other explanatory variables in later sections), and especially differences in the estimates across countries, mean and do not mean. Because the TIMSS data were generated by the same data-generating process in the different countries and are therefore directly comparable across countries, and given the technical constraints on the pedagogical process, the prior should be that the size of the *effect* of any family-background characteristic on students' educational performance should be the same everywhere. If this is not the case, this implies that there must be differences in how the school systems work. This does *not* reflect different *distributions* of family-background characteristics in the different populations, as they have become apparent in Section 3.2, which are no a priori reason for the performance gap between students with two different characteristics to differ. For example, the performance gap between students with more than two bookcases at home and students with less than one shelf of books at home may be expected to be independent of the relative share of homes with different numbers of books in the population. If this gap is 25 TIMSS test-score points in one country but 50 points in another country, this would rather be a sign that the school systems work differently in the two countries.

As discussed in the previous section, some of the data are imputed rather than original. Generally, data imputation introduces measurement error in the explanatory variables, which should make it more difficult to observe statistically significant effects. Still, to make sure that the results are not driven by imputed data, a vector of dummy variables D^F is included as controls in the estimation. The vector D^F contains one dummy for each variable in the family-background vector F that takes the value of 1 for observations with missing and thus imputed data and 0 for observations with original data. The inclusion of D^F as controls in the estimation allows the observations with missing data on each variable to have their own intercepts. The inclusion of the interaction term between imputation dummies and background data, $D^F F$, allows them to also have their own slopes for the respective variable. These imputation controls for every variable with missing values ensure that the results are robust against possible bias arising from data imputation.

Further problems in the econometric estimation of equation (2) are that the explanatory variables in this study are measured at different levels, with some of them not varying within classes or schools; that the performance of students within the same school may not be independent from one another; and that the primary sampling unit (PSU) of the two-stage clustered sampling design in TIMSS was the school, not the individual student (see Section 3.1). As shown by Moulton (1986), a hierarchical structure of the data requires the addition of higher-level error components to avoid spurious results. Therefore, the error term ε of equation (2) has a school-level and a class-level element in addition to the individual-student element:

$$\varepsilon_{ics} = \eta_s + \nu_c + \upsilon_i \quad , \tag{3}$$

where η is a school-specific error component, ν is a class-specific error component, and υ is a student-specific error component. Clustering-robust linear regression (CRLR) is used to estimate standard errors that recognize this clustering of the survey design. The CRLR method relaxes the independence assumption and requires only that the observations be independent across the PSUs, i.e. across schools. By allowing any given amount of correlation within the PSUs, CRLR estimates appropriate standard errors when many observations share the same value on some but not all independent variables (cf. Deaton 1997).

Finally, TIMSS used a stratified sampling design within each country, producing varying sampling probabilities for different students (Martin and Kelly 1997). To obtain nationally representative coefficient estimates from the stratified survey data, weighted least squares (WLS) estimation using the sampling probabilities as weights is employed. The WLS estimation ensures that the proportional contribution to the parameter estimates of each stratum in the sample is the same as would have been obtained in a complete census enumeration (DuMouchel and Duncan 1983; Wooldridge 2001).

4.2. Basic Results

Table 3 reports the results of the family-background regression in math for the different Western European countries and the United States. It starts with several measures of the educational background of the family, followed by student characteristics and the community location.

The education level attained by the parents is strongly related to student performance in all European countries, as well as the United States. In the estimations, I do not make the assumption of a linear relationship, e.g. by using parents' years of education as a single explanatory measure. Rather, the estimations use all the information available in the background questionnaires and allow a more flexible functional form by including a dummy for each category, namely parents whose highest education level is some secondary, finished secondary, some after secondary, and finished university, with parents with no secondary education at all as the residual category.

Across the European countries, the relationship between parents' education and their children's math performance is particularly low in French Belgium, the only country where the difference in math performance between students whose parents finished university and students whose parents had no secondary education is not statistically significant (although it is in science, cf. Table A3). At the other extreme, this difference is largest in Greece at 51.3 test-score points. The effect size is smaller in all Western European countries than the one observed in the United States (52.7), where basically all previous studies found important family-background effects (cf. Hanushek 2002). Effect sizes of 40-50 test-score points, as observed in many European countries, are very large indeed: They are of the same order of magnitude as the performance difference between 7^{th} and 8^{th} graders in most countries, and equal about half a standard deviation in TIMSS test scores (cf. Table 1). When considering the performance difference between children of parents who finished university and children of parents who finished secondary education, there is basically no difference in Spain, and the largest difference is observed in Scotland at 39.7 test-score points.

A limitation to the comparability of the effects of parental education across countries is that the categories in the different countries may not be easily comparable. Therefore, the second indicator of family background, the number of books in the students' home, may be particularly interesting for cross-country comparisons. Students were asked to report the total number of books that are in their home, not counting newspapers, magazines, or their school books. This measure can be viewed as a proxy of the educational, social, and economic background of the students' families that is readily comparable across countries. Again, this measure was collected in five categories. In the regressions reported in Table 3, England appears to feature the largest effect based on this measure. However, England is also the only country without data on parental education, which might affect the estimates on books at home. Therefore, I did the regressions without controlling for the parental education data also for all other countries to see whether the relative size of the coefficient on books at home in England can be attributed to the missing parental-education data. The coefficients on books at home actually do increase in all other countries once parental education is excluded.

However, England still has the largest effect among all countries, with students from homes with more than two bookcases full of books performing 104.4 math test-score points better than students from homes with less than one shelf of books. Scotland (93.6) and Germany (84.9) also feature very large family-background effects on this account, followed by Austria, Ireland, and Sweden. In all these countries, the effect of family background as measured by the number of books in a student's home on student performance is comparable to or larger than the effect found in the United States. At the other extreme, France features the lowest effect (20.9), and Flemish Belgium has the lowest effects when looked at relative to one shelf of books as the lowest category. A third country with relatively small effects is Portugal. This general pattern of results is basically unaffected by which specific difference one looks at; e.g. it also holds for two bookcases relative to less than one shelf, more than two bookcases relative to one shelf, more than two bookcases relative to one bookcase, or two bookcases relative to one shelf.

The relatively low intergenerational *educational* mobility in Britain mirrors previous findings of relatively low intergenerational *earnings* mobility in this country (cf. Dearden et al. 1997; Solon 2002). However, the Swedish results contrast with studies finding relatively high intergenerational earnings mobility in Sweden (cf. Björklund and Jäntti 1997; Solon 2002), suggesting that the latter is not predominantly driven by particularly strong equality in educational opportunities.

Two neighboring countries of which one belongs to the group of countries with small family-background impacts and the other to the group with large family-background impacts are France and Germany. In a French-German comparison of the impact of family background on educational attainment, Lauer (2003) hypothesizes that family background should have a larger impact on children's educational prospects in Germany than in France, mainly based on the observation that streaming takes place much earlier in the German school system than in the French one (cf. also Schnepf 2002). However, her evidence suggests that the two countries prove surprisingly similar with regard to the impact of family background on educational attainment, i.e. on a measure of educational *quantity*. By contrast, my results show that when using a measure of educational *quality* instead, her hypothesis actually holds true. That is, the larger family-background effect in Germany materializes in terms of better educational performance of the children rather than higher educational attainment. One way to reconcile the two findings would be that certification regimes differ, so that the attainment certificates represent different ranges of actual skills in the two countries. The extent to which the differing finding with regard to performance is due to the different streaming practices in the two systems is an open question, though, because the two systems also differ in several other important dimensions. Most notably, France features a much more widespread prevalence of pre-primary education than Germany. As seen in Table 2, starting at the age of 3, more than 90% of children are enrolled in the French education system, while in Germany, this is not the case until the age of 7. Of the children aged 2-4, 80% are enrolled in France but only 45% in Germany (OECD 1998). This may be another impact factor lying behind the large French-German difference in family-background effects, and it could also account for the low family-background effect in Flemish Belgium where the 90% enrollment range also starts at the age of 3 and where 79% of children aged 2-4 are enrolled.

In interpreting the results, one has to bear in mind that family-background measures such as parental education or the number of books at home will be correlated with other, unobserved aspects of family background, such as parents' motivation or their willingness and capability to help with homework. Therefore, the estimated coefficients in this model – as well as in the following models – have to be interpreted as the effect of the respective variable and anything else that goes with it. In that sense, both parental education and the number of books at home serve as general proxies for the educational, social and economic background in the students' homes.[12]

The next measure of family background is whether students live with both parents. In no country do students living with both parents perform statistically significantly worse than students not living with both parents, and they perform statistically significantly better in 8 Western European school systems (as well as in the United States). Only in Norway (18.7) is the effect larger than in the United States (15.5), and the next-largest effects are in Ireland, Greece, and Germany.

The other considered student characteristics are also significantly related to student performance in many Western European countries. Native students perform statistically significantly better than immigrated students in 10 European countries, but not in the United States (at least in math). The estimated effect is largest in Sweden (30.3), followed by Austria (25.9), Switzerland, Norway, and Denmark. By contrast, immigrated students actually perform statistically significantly (at least at the 5% level) better than native students in Ireland (13.2) and Portugal (6.1).[13] With regard to students' gender, Flemish Belgium, Iceland, and Sweden are the only countries where girls do not perform statistically significantly worse than boys. The boys' lead is largest by far in Ireland (19.7), followed by Switzerland (13.3), French Belgium (13.1), and Denmark (12.6).

Controlling for grade differences, older students perform statistically significantly worse in most countries, which presumably reflects a grade repetition effect. In all countries, 8th-graders perform statistically significantly higher than 7th-graders. The performance difference ranges from 23.0 in England to 67.4 in France. In countries with small performance differences between grades (e.g., England, Iceland, Scotland, and to a lesser extent Norway

---

12 The estimates may also to some extent reflect heritable ability. More able parents, who obtained more education, may have more able children, who then perform better on the performance tests. Behrman and Rosenzweig (2002) show for Minnesota twins that heritable ability is a likely source of the correlation between the quantitative educational attainment of mothers and their children. This is not found for fathers, though, and Behrman et al. (1999) find a causal impact of mothers' schooling on their children's schooling, working through home teaching, in rural India during the green revolution. The extent to which heritability can account for the cross-country variation in the presented findings should be limited, though.

13 Because the strong correlation between the immigration status of parents and children regularly causes problems of multicollinearity, only the latter was included in the regressions and the former left out, with the evident ensuing consequences for the interpretation of results.

and Sweden), the age effect tends to be low (in absolute terms), while it tends to be high in countries with strongly negative age effects (e.g., France, the Netherlands, and the two Belgian systems). This may reflect differences in the policies of how to deal with low-performing students. In the latter countries, the lowest-performing students regularly seem to be relegated into lower grades, so that the grade difference is relatively large, and at the same time the age effect is strongly negative since the relegated students are older but relatively low performing. By contrast, in the former countries, low-performing students generally seem not to be relegated to lower grades, so that the performance difference between grades is relatively low as the low-performers remain in upper grade, and age does not make much of a difference within grades and might even have a slightly positive effect on performance.

Finally, there is no clear pattern in the relative performance of schools by community location. In some countries, schools in geographically isolated locations perform much worse, while in others, they perform much better. Here, it should also be noted that most of the statistically significant estimates are based on very few observations, because only 2 schools in Austria and Portugal, and only 1 in Spain, are reported to be geographically isolated (cf. Table A2a). There is also no clear pattern of results for schools located in communities close to the center of a town.

The explanatory power of the student and family background characteristics, when measured as the proportion of the test-score variation accounted for by the model (the R^2), is largest in Switzerland at 26.4% (disregarding the part "explained" by imputation controls), followed by French Belgium, Sweden, and France. At the lower end, only 9.2% of the performance variation can be attributed to family background in Iceland, followed by Denmark, Norway, and Flemish Belgium.

4.3. Family-Background Effects in Europe vs. the United States

The previous discussion already allowed a comparison of results for European countries with the United States at an individual-country level. To estimate more formally how family-background effects in educational production compare between the United States and Europe as an aggregate, Table 4 reports the same model specification used above in Table 3 for the pooled sample of all students from the European countries and the United States. For each explanatory variable, the pooled regression includes an interaction term with a Europe dummy to test whether the family-background effect differs between Europe and the United States.

The regression also controls for a whole set of country dummies to account for country fixed effects.

Comparing Europe as an aggregate to the United States, there is remarkably little difference in the direction, size, and significance of the family-background effects. Particularly when considering parental education and books at home as the main measures of family background, their effects on student performance in math actually barely differ between the "average" European country and the United States. At low levels of significance, there are slight differences in the size of the performance difference between the lowest two categories, and the effect of parents having finished university is somewhat larger in the United States. As a general pattern, however, family background has a strong and comparable impact on student performance both in Europe and the United States.

The effect of students living with both parents is statistically significantly smaller in Europe than in the United States, but it is still statistically significantly positive. Europe differs from the United States in that immigrant students perform statistically significantly worse than native students in math, which is not the case in the United States. There is also a difference in that students from schools located in geographically isolated areas perform statistically significantly worse in the United States, but not in Europe. However, the effects of students' age, gender, grade, and town-center location of schools do not differ significantly between Europe and the United States. All in all, when taking Western Europe as a whole, there is remarkably little difference in family-background effects to the United States, not even in their size.

4.4. Quantile Regressions: Heterogeneity of Family-Background Effects by Ability

Given that 75-90% of the test-score variation in the different countries remains unexplained, a large part of the performance variation across students has to be attributed to omitted variables (in addition to measurement errors). This is not unexpected, because the basic learning ability of students – e.g., their innate ability or their learning motivation – remains unobserved and thus enters the error term ε of equation (2) (or, more precisely, the student-specific error component υ in equation (3)). However, it begs the question whether the family-background effects vary across the ability distribution of students. E.g., it may be of different importance to low-ability students compared to high-ability students whether they come from a disadvantaged family background. Such heterogeneity in the family-background effects can be detected by quantile regressions (cf. Koenker and Bassett 1978).

Student ability itself remains unmeasured, virtually by definition. However, once family-background effects are controlled for, the conditional performance distribution should be strongly correlated with ability (or, more precisely, with that part of ability that is not correlated with family background). In the following, I just term this conditional performance distribution as "ability." Quantile regressions estimate the effect of family background on student performance for students at different points on this ability distribution. While Eide and Showalter (1998) and Levin (2001) have applied the quantile regression method to the estimation of resource effects in education, I am unaware of any country study except for Fertig (2003) that uses quantile regressions to estimate family-background effects, where the distinction between family background and underlying ability seems to offer a particularly appealing analytical background for the estimation and interpretation of quantile estimates. The family-background proxy used in this section is the number of books at home, which has the advantage over parental education to be readily comparable across countries and whose effect is considerably stronger than that of any other variable.

The horizontal lines of Figure 1 replicate the least-squares coefficients on the four books-at-home dummies from Table 3 for each country. The relatively large effects in England and the relatively small effects in France are immediately apparent. The curved lines with dots report quantile-regression coefficients of the same model for 19 quantiles ranging from 0.05 to 0.95.[14] As a general pattern, these quantile-regression estimates are relatively constant across quantiles in almost all countries. This remarkable uniformity suggests that the family-background effects do *not* vary strongly with the underlying ability of students.

To the extent that they do, they tend to *increase* with quantiles in the vast majority of countries. This pattern of family-background effects slightly increasing with ability is particularly notable in Austria, England, Ireland, Norway, Spain, Sweden, and the United States. To a lesser extent, it may also be observed in French Belgium, France, Germany, Greece, Iceland, Portugal, Scotland, and Switzerland. However, England is the only country

---

[14] Confidence intervals are not shown for expositional reasons. For the least-squares estimates, standard errors are reported in Table 3. The precision of the quantile-regression coefficients is roughly comparable in size.

where a really strong pattern of effect heterogeneity exists.[15] In all these countries – albeit to varying degrees – a good family background is most important for high-ability children. Thus, for instance, while the performance lead of students from homes with more than two bookcases over students from homes with less than one shelf of books is 69.4 for students at the 10th percentile of the conditional performance distribution in England, it is 146.5 for students at the 90th percentile. This effect difference of 77.1 between the 10th and the 90th percentile in England compares, e.g., to a difference of only 5.3 in Switzerland (59.1 at the 90th percentile versus 53.7 at the 10th percentile). In the countries where no increasing pattern can be observed, the family-background effect seems to slightly shrink with ability in the Netherlands, make roughly a U-shape in Flemish Belgium, and an inverted-U-shape in Denmark.

The median regression can also be viewed as a test of the least-squares result for robustness against outliers. The conditional median function estimated by median regression minimizes the sum of the absolute residuals and is thus less prone to outliers than the conditional mean function estimated by least-squares regressions. In this sense, median regressions may better describe the central tendency of the data. As is evident from Figure 1, in most countries the least-squares and the median estimates do roughly coincide. In these countries, the initial least-squares results do not seem to be substantially biased by outlying observations. A difference between the two estimates in some countries suggests that the least-squares estimates seem to be biased to an extent by outliers. In Flemish Belgium, the median estimates are somewhat smaller than the least-squares estimates, suggesting that outliers seem to bias the least-squares coefficients upward. By contrast, in Austria, Denmark, and Norway, the median estimates are slightly larger than the least-squares estimates, suggesting that outliers seem to bias the least-squares coefficients downward.

4.5. Is There a Tradeoff Between Equity and Efficiency?

The size of the family-background effects can be viewed as a measure of the equality of educational opportunities for children from different backgrounds. In this regard, the French

---

[15] One might advance the thought that the conditional performance may reflect other, unmeasured family-background characteristics as much as underlying ability. At a first glance, this would suit to the fact that England is the only country with a strong positive link between family-background effects and underlying ability and that it is also the only country with no data available to control for parental education as a second family-background measure. However, performing the same quantile regressions under negligence of the parental education variables for the other countries reveals that this is not driving the exceptional English result.

school system seems to achieve the highest extent of equal opportunity, and the British systems the lowest. From a policy viewpoint, this begs the question whether equalizing educational opportunities must be bought by having to accept a lower overall level of performance, i.e. whether a tradeoff exists between educational equity and efficiency.

To give a cross-country view on this question, Figure 2 plots the mean of the math achievement in each country against a measure of the family-background effect in the country, namely the performance difference between students with more than two bookcases at home and students with less than one shelf of books at home. From this scatter plot, it is immediately evident that there is no direct relationship between average performance and our measure of equality of opportunity across countries. The cross-country correlation coefficient between the two measures is 0.002.

When using other proxies for family-background effects, the correlation with mean performance is similarly low, and often negative. For example, the correlation coefficient is –0.03 for the performance difference between students with more than two bookcases relative to students with one shelf as the proxy for family-background effects, –0.15 for more than two bookcases relative to one bookcase, and 0.11 for two bookcases relative to less than one shelf as well as one shelf. Likewise, using parental education as the family-background proxy, the correlation coefficient with mean math performance is –0.11 when considering the performance difference between students whose parents finished university relative to students whose parents had no secondary education, and –0.10 when considering finished university relative to finished secondary. Thus, the cross-country pattern of mean performance and family-background effects in Europe suggests that there is no apparent tradeoff between achieving efficiency in educational production and equality of educational opportunity.

5. Resource Endowments and Student Performance

This section starts by presenting results of least-squares regressions of student performance on several measures of schools' resource endowments and of teacher characteristics, everything controlling for family-background influences. Given that such least-squares estimates might suffer from endogeneity bias, I then present results on class-size effects based on a quasi-experimental design that exploits between-grade variations within schools by combining school fixed effects with an instrumental-variable estimation.

5.1. *Least-Squares Coefficients on Resources and Teacher Characteristics*

The conventional procedure to estimate the relationship between schools' resource endowments and their students' performance is to simply introduce resource measures into the previously estimated equation:

$$T_{ics} = F_{ics}\alpha_2 + R_{cs}\beta_1 + D_{ics}^F\delta_3 + \left(D_{ics}^F F_{ics}\right)\delta_4 + D_{cs}^R\delta_5 + \left(D_{cs}^R R_{cs}\right)\delta_6 + \varepsilon_{ics} \quad , \tag{4}$$

where *R* is a vector of resource measures such as class size, the availability of instructional materials, and teacher characteristics. The imputation controls D^R again ensure that the results are robust against possible bias arising from missing and thus imputed data in the resource variables. Under the assumption that the resource endowment is exogenous to student performance, the coefficient vector β_1 estimated in a least-squares regression would reflect the impact of resources on student performance.

However, the main problem with equation (4) is that the resource endowments that students face may be particularly prone to endogeneity, in that resources and student performance are jointly determined by other factors or that the latter is even directly used to determine the former (cf. Hoxby 2000). For example, school systems may sort weaker students into smaller classes, giving rise to a negative correlation between performance and the teacher-student ratio in classes (cf. West and Wößmann 2003). Also, parents of high-performing students may choose to move into areas where schools are well equipped with resources, giving rise to a positive performance-resource correlation. None of these correlations says something about the causal effect of resource endowments on student performance. In short, parents, teachers, schools, and administrators all make choices that may give rise to a non-causal relationship between resources and student performance, so that least-squares estimates of resource effects may be biased.

Hoxby (2000) shows that simple least-squares estimates of resource effects are biased upwards in the case of class-size effects in the US state of Connecticut – i.e., the least-squares estimates are biased in the direction of finding positive effects of smaller classes on student performance. While the simple least-squares coefficients reported in column I of her Table II are negative and statistically highly significant, her identification strategies that use exogenous class-size variation to estimate the causal class-size effect yield "rather precisely estimated zeros" (Hoxby 2000, p. 1280; cf. her Tables IV and VII). A closer look at Hoxby's findings reveals, however, that once the least-squares regressions control for the effects of several family-background measures (column II of her Table II), they basically already yield as precisely estimated zeros as the more sophisticated identification strategies reported in her Tables IV and VII. Thus, in this particular setting, any bias in the resource effects seems to be related to family background, so that least-squares estimation of education production functions such as equation (4) that control for family-background effects yield virtually unbiased estimates of resource effects. In light of this, I start by presenting least-squares results on resource effects in this section.

The least-squares results on resource effects are presented in Table 5. The regressions control for all the family-background variables of Table 3. Following Hoxby (2002), class size as the first resource measure is measured in natural logs because a change in class size by 1 student is proportionately larger for lower class sizes. The least-squares estimate on log class size for the United States is very similar to Hoxby's result: namely, a rather precisely estimated effect that is statistically indistinguishable from zero. However, the results are very different in most European countries. In 12 European school systems, class size is statistically significantly *positively* related to math performance. That is, students in bigger classes perform significantly better, even after family-background effects are controlled for. A statistically significant negative relationship, which might have been expected on the basis of conventional wisdom, does not exist in any country. The positive relationships raise doubt that the least-squares estimates yield unbiased estimates of class-size effects in most European countries. While in the United States, the resource allocation is strongly linked to family background in a regressive way due to the unique US features of decentralized education finance and considerable residential mobility, many other forces seem to be at play in Europe, largely unrelated to family background, that give rise to a compensatory resource allocation (cf. the findings by West and Wößmann 2003). In the following section, I will present results of a particular estimation strategy that tries to eliminate the remaining biases in the class-size effects.

The second measure of resource endowment is based on a school questionnaire item where heads of school report whether their school's capacity to provide instruction is affected by the shortage or inadequacy of instructional materials, e.g. textbooks.[16] In four European school

[16] Given that this measure is based on subjective assessments, it may to some extent suffer from assessments of resource shortages that are endogenous to observed student performance. Also, the cross-country comparability of this measure may be limited, as assessments of what might constitute a resource shortage may

systems (England, Ireland, the Netherlands, and Portugal), students in schools without any shortage of materials perform statistically significantly better than students in schools with a little or some shortage. In three European school systems (Austria, Norway, and Switzerland) and in the United States, students in schools with considerable shortage of materials perform statistically significantly worse. However, in two countries (the Netherlands and Spain), students in schools whose heads report considerable shortage of materials actually exhibit a statistically significant better performance.[17] On some occasions, the effect size is quite substantial. For example, students without shortage of materials perform 37.4 test-score points better in England, and students with much shortage perform 30.5 test-score points worse in Switzerland. Annual instruction time is statistically significantly related to math performance in only one country, Switzerland.

In terms of teacher characteristics, the teachers' gender does not make a difference for student performance in most countries. By contrast, teachers' experience, measured as the logarithm of years of experience, is statistically significantly related to better student performance in 9 European school systems.[18] The largest statistically significantly estimated coefficient is 10 in the Netherlands (the lowest one is 3 in Portugal). This means that students of teachers with 1 year's experience perform 6.9 (2.1) test-score points lower than students of teachers with 2 years' experience, 23.0 (6.9) lower than students of teachers with 10 years' experience, and 30.0 (9.0) lower than students of a teacher with 20 years' experience. Starting with the 11th (4th) year of experience, the effect of one additional year of experience is smaller than 1 test-score point.

There is not much evidence that teacher education substantially affects student performance in Europe. The statistically significant coefficient estimates in France, Iceland, Ireland, and the Netherlands are all just driven by very small cells where less than 2% of students fall into the relevant category (cf. Table A2b). Among the European countries, this leaves only Switzerland with convincing evidence of performance differences between students of teacher with different educational degrees. Here, the effect is quite substantial,

---

differ between countries. In these respects, absolute measures of resource endowments would be preferable, but are not satisfactorily available in the TIMSS database.

17 In Austria and the Netherlands, the statistically significant estimates are based on only 1 school.

18 The logarithmic shape of the function allows for some concavity in the effect of teacher experience, as is generally found in the literature, but future research might want to analyze this effect in far greater detail, allowing for additional non-linearities.

though, with students of teachers with a BA performing 42.6 points better than students of teachers with only secondary education. In no European country exists a significant performance effect of teachers with an MA relative to teachers with a BA. By contrast, such an effect is found in the United States at 10.0 points (statistically significant at the 10% level).

To the extent that the least-squares estimates can be viewed as reasonable estimates of the effect of resource endowments on student performance, material endowment and teacher experience seem to be relevant for student performance in several European school systems (but by no means in all), while instruction time and teachers' gender and education do not. The results of this standard procedure may be substantially biased, though. The change in explanatory power (ΔR^2) of the model including the resource measures over the model including only the family-background measures is very small in most countries, and where it is not, this is virtually exclusively due to the positive correlation of student performance with class size.

5.2. School-Fixed-Effects Instrumental-Variables Estimates of Class-Size Effects

As discussed above, the quantitative estimates of the resource effects may be biased by the potential endogeneity of schooling resources *R*. In the case of the estimated coefficients on class size, I can exploit specific characteristics of the TIMSS data in a quasi-experimental estimation design in order to obtain unbiased estimates of the effects of class size on student performance. Akerhielm (1995) suggested to instrument the actual class size C_{cs} (one vector in the resource matrix R_{cs} of equation (4)) by the average class size in the school A_s in a two-stage least-squares estimation to control for the problem of endogenous resource allocation within schools.[19] The grade-average class size promises to be a valid instrument for actual class size: It is generally strongly linked to the size of the class actually tested in TIMSS; within each school, it is exogenous to the performance of the students (although this might not be the case between schools, a fact that I will return to shortly); and there is no reason to expect that it affects students' performance in any other way than through the size of the class

19 Akerhielm (1995) also used the overall grade-level enrollment of a school as an additional instrument. However, this may be an invalid instrument because a direct relationship might exist between overall enrollment and student performance that is unrelated to differences in class size (Angrist and Lavy 1999). Moreover, none of the coefficients on enrollment in Akerhielm's first-stage regressions are statistically significant, suggesting that it is anyway not a good instrument.

in which they are actually taught.[20] The first-stage estimation regresses the log of the actual class size, C_{cs}, on the log of the grade-average class size, A_s, and all other exogenous variables X_{ics}:

$$C_{cs} = \chi_1 A_s + X_{ics}\chi_2 + \mu_{ics} \quad , \tag{5}$$

where X_{ics} includes the family-background measures and the imputation controls. The second stage then employs $\hat{C}_{cs} = C_{cs} - \mu_{ics}$ instead of C_{cs} in lieu of R_{cs} in the estimation of equation (4). This specification eliminates any bias in the estimated class-size effects that would result from within-school sorting of differently performing students, at a given grade level, to differently sized classes.

However, these IV estimates may still be biased by between-school sorting effects. These may be particularly relevant where parental residential decisions are related to their children's performance, or where the school systems tracks students into different schools according to their ability, as is the case, e.g., in Germany. In order to exclude any effects of either within- or between-school sorting from the estimates of class-size effects, Wößmann and West (2002) suggest an identification strategy specifically designed to exploit the multi-grade nature of the TIMSS database. They combine the aforementioned IV strategy with a school-fixed-effects estimation that disregards any between-school variation, as this may reflect between-school sorting effects. The combined school-fixed-effects instrumental-variables (SFE-IV) estimation then is:

$$T_{ics} = F_{ics}\alpha_3 + \hat{C}_{cs}\beta_2 + S_s\varphi + D^F_{ics}\delta_7 + \left(D^F_{ics}F_{ics}\right)\delta_8 + D^C_{cs}\delta_9 + \left(D^C_{cs}C_{cs}\right)\delta_{10} + \varepsilon_{ics} \quad , \tag{6}$$

where S_s is a complete set of school dummies and $\hat{C}_{cs}$ is again the result of a first-stage regression that instruments actual class size by grade-average class size and all other exogenous variables as in equation (5).[21] Because equation (6) includes school fixed effects, and because every class size at a given grade level is instrumented by the same average class

[20] See Wößmann and West (2002) for a more detailed discussion of the validity of the instrument.

[21] The imputation dummies D^C for the class-size variable used in this section equal 1 if either the observation on actual class size or the observation on grade-average class size (the instrument) is imputed. In the IV and SFE-IV regressions, in addition to instrumenting class size, the interaction term $D^C C$ between the imputation dummy and actual class size is also instrumented, using the interaction term $D^C A$ between the imputation dummy and grade-average class size as an additional instrument.

size, this SFE-IV strategy requires comparable information on student performance from more than one grade level in each school. This is exactly the structure of the TIMSS data.

The grade-level dummy included in the background measures F controls for the average difference in performance between students from the two adjacent grades. Therefore, the remaining performance difference between students from the different grades is idiosyncratic to each school. Equation (6) relates this idiosyncratic between-grade variation in student performance to that part of the actual class-size difference between the two grades that is due to differences in grade-average class sizes. Thereby, the SFE-IV identification strategy effectively excludes both between-school and within-school sources of student sorting. Between-school sorting is eliminated by controlling for school fixed effects. Within-school sorting is filtered out by instrumenting actual class size by grade-average class size. Arguably, the remaining variation in class size between classes at different grades of a school is caused by random fluctuations in cohort sizes between the two adjacent grades, presumably reflecting natural fluctuations in student enrollment. The coefficient estimate β_2 can thus be interpreted as an unbiased estimate of the causal impact of class size on student performance.

As there is no comparable quasi-experimental identification strategy for the other resource measures, these are not included in equations (5) and (6). Therefore, the resulting coefficient estimates on class size should be interpreted as the effect on student performance of class size and any other resource with which class size may be associated.

Wößmann and West (2002) have presented estimates of class-size effects based on this identification strategy for several Western European countries. However, they simply drop any observations of schools with missing data on (actual or average) class size, substantially reducing the sample size. Instead, I use the imputation strategy detailed above to be able to retain observations with missing data in the sample, while at the same time using the imputation controls to ensure that the estimated results are not driven by the imputation. This strategy should be superior both in terms of robustness of the estimation and in terms of avoiding bias due to dropping observations. This new data strategy also allows me to perform the estimations for an additional 9 school systems in Western Europe, 3 of which (Denmark, Germany, and the Netherlands) yield acceptably precise SFE-IV estimates to warrant an assessment of causal class-size effects.

The first row of Table 6 replicates the least-squares (LS) estimation of Table 5 under omission of the other resource variables and with a slightly different sample. (Schools which did not have both a 7th-grade class and an 8th-grade class tested were excluded from the sample because these schools do not have any of the variation necessary for instrumenting.) In this specification, the coefficient estimates on class size are statistically significantly *positive* in 11 Western European school systems.

The second row of Table 6 presents results of an instrumental-variable (IV) estimation that does not control for school fixed effects. In England, the instrumenting cannot be implemented as data on the instrument, the grade-average class size, are not available in the school background questionnaire. The difference between the LS and the IV results reveal that substantial within-school sorting is going on in many school systems that biases the LS estimates of class-size effects. For example, the fact that in Norway the LS estimate is 27.9 and the IV estimate is 98.9 implies that within Norwegian schools, relatively low-performing students are sorted into larger classes.

Next, the third row of Table 6 reports results of a specification that does not use instrumental variables, but includes a whole set of school dummies to control for school fixed effects (SFE). This specification cannot be implemented in Sweden because 7th and 8th grade are taught in different schools in Sweden (7th grade in elementary and 8th grade in secondary schools), so that no class-size variation is left after controlling for school fixed effects. Again, the inclusion of school fixed effects makes substantial difference for the class-size estimate relative to the LS estimate in many Western European countries, suggesting that between-school sorting introduces a significant bias to the LS class-size coefficient. For example, the large and statistically significant difference in Germany between the LS estimate of 93.4 and the SFE estimate of –1.0 implies that relatively low-performing students are regularly sorted into schools with smaller classes in the tracked German secondary-education system. By contrast, the between-school sorting pattern in England is the other way round, with relatively high-performing students ending up being taught in schools that feature smaller classes.

Finally, results of the combined school-fixed-effects instrumental-variables (SFE-IV) strategy, which accounts for both within- and between-school sorting, are reported in the fourth row of Table 6. It should be obvious that the SFE-IV strategy is very demanding in terms of data requirements, because it disregards both all class-size variations between schools and all class-size variations within individual grades in each school. This is reflected in a low degree of precision of the SFE-IV estimate in several countries. As the *F*-statistic of instrument relevance in the first-stage regression (reported below the SFE-IV estimates in

Table 6) shows, the SFE-IV estimations in Austria, Ireland, and Portugal do not meet the weak-instruments test of an *F*-statistic bigger than 10, so that the estimates are biased and the standard errors unreliable in these countries.

For some of the remaining countries, the precision of the SFE-IV estimate is too low to allow assessments of the existence of causal class-size effects. As a benchmark for reasonable precision, I take Krueger's (1999) estimate from the Project STAR class-size experiment for primary schools in Tennessee. Krueger found that a 10% reduction in class size for one year led to an improvement in test scores of about 10% of a standard deviation. In our case, 10% of an international standard deviation in TIMSS test scores equals 10 test-score points, so that a coefficient on log class size comparable in size to Krueger's estimate would equal roughly –100. Thus, in order for such a class-size effect to be detectable at the 95% level of statistical confidence, the standard error would have to be smaller than 50. This suggests that standard errors bigger than 50 may be viewed as too imprecise to detect notable class-size effects.

The SFE-IV estimates find a statistically significant negative causal effect of class size on student performance in math – i.e., students perform better because they are taught in smaller classes – only in one Western European country, Iceland. The effect size is only half as large as the one Krueger (1999) found for primary school, though, and given the precision of the estimate in Iceland, a class-size effect of Krueger's magnitude can even be rejected using a standard Wald test (also reported in Table 6). The effect size of –53.1 means that for every 10% reduction in class size, students perform 5.3 test-score points (or 5.3% of an international standard deviation) higher. Particularly when compared to the previous family-background effects, this seems small, and given that Krueger presents a rough cost-benefit analysis suggesting that the effect size he found would roughly equal the economic costs and benefits of the class-size reduction, it seems unlikely that class-size reductions would be advisable in economic terms even in Iceland. In France, an effect of Krueger's size cannot be rejected, but the effect is also not statistically significantly different from zero at conventional significance levels, only at 13.5%.[22]

Altogether, the SFE-IV estimates rule out the existence of notable causal class-size effects of Krueger's magnitude in 8 Western European school systems – the two Belgian systems,

[22] With their different specification, Wößmann and West (2002) found statistically significantly negative coefficients for Iceland and France. The statistically significant *positive* coefficient they found for Portugal does not survive in our alternative specification.

Denmark, Germany, Greece, Iceland, the Netherlands, and Scotland – and in the United States (see the Wald tests reported in Table 6).[23] In the Netherlands, Scotland, and the United States, Wald tests even reject effects of –50, i.e. of half Krueger's size found in Tennessee. Thus, while the precision of the SFE-IV estimates does not allow to detect as "precisely estimated zeros" as Hoxby's (2000) estimates for Connecticut, whose precision would equal standard errors of about 10-20 in our case, it still allows for a meaningful assessment of the existence of class-size effects in these countries.

There is only one country, the Netherlands (not analyzed in Wößmann and West 2002), where the SFE-IV estimation yields a statistically significantly positive estimate, as was the case for the least-squares estimate in 11 countries. That is, in the Netherlands, *bigger* classes seem to cause better student performance. Intriguingly, this is the same country where Dobbelsteen et al. (2002) found the same result before, based on a completely different data set and identification method. Dobbelsteen et al. use an argument from the social cognitive learning literature to suggest that the positive class-size effect may be a result of the increased probability in larger classes of having classmates with a similar level of ability, which seem to be particularly important for student learning. All in all, these results cast strong doubt on whether class-size reductions are a promising strategy to boost student performance in the Western European school systems.

6. Institutional Features and Student Performance

Wößmann (2003b) showed that institutional features of school systems are strongly related to student performance at the between-country level.[24] While many institutional features – e.g., the existence of central examinations – are basically constant within countries (cf. also Table A2c), rendering no variation that would enable an estimation of institutional effects within countries, there is some institutional variation in some countries that seems worth exploiting. Thus, in this section, I present European evidence on the relationship between the within-country variation in student performance and the within-country variation in institutional features that exhibit noteworthy variation at least within a couple of countries. In particular,

23 The same is true in Portugal, but this estimate suffers from a weak instruments problem.

24 See Hanushek et al. (1994) for incentive-based arguments and Bishop and Wößmann (2003) for a theoretical model why institutional features affect student performance.

these institutions are school autonomy in hiring teachers, the extent of teacher influence in several areas, testing policies and homework assignment, and parental involvement.

Most of these institutional features may be assumed largely exogenous to student performance. In as far as this assumption holds for a particular institution, its effect on student performance can be estimated by adding the vector of institutional measures *I* as explanatory variables to the education production function of equation (4):

$$\begin{aligned} T_{ics} &= F_{ics}\alpha_4 + R_{cs}\beta_3 + I_{cs}\gamma \\ &\quad + D^F_{ics}\delta_{11} + \left(D^F_{ics}F_{ics}\right)\delta_{12} + D^R_{cs}\delta_{13} + \left(D^R_{cs}R_{cs}\right)\delta_{14} + D^I_{cs}\delta_{15} + \left(D^I_{cs}I_{cs}\right)\delta_{16} + \varepsilon_{ics} \end{aligned} \quad . \qquad (7)$$

D^I is again a set of imputation dummies to control for possible effects of the data imputation. It should be borne in mind that particularly in England and Germany, a large share of the data on institutions is missing in the TIMSS background questionnaires. Table 7 reports the results on the institutional variables. The estimations keep controlling for all family-background and resource variables of Tables 3 and 5, as well as for their respective imputation controls.[25]

6.1. School Autonomy in Hiring Teachers

When schools have autonomy in hiring their teachers, they may achieve a better match to local needs compared to centralized hiring decisions. Decentralized personnel decisions may also allow a higher degree of performance-conducive competition between schools. In line with this argument, students in schools that had primary responsibility for hiring teachers performed statistically significantly better than students in schools that did not have primary responsibility in 4 Western European school systems (out of 12 with available data and variation in this institutional feature), as well as in the United States. In no country is the effect statistically significantly negative.

The effect size of school autonomy in hiring teachers is quite substantial, at 19.5 test-score points in Flemish Belgium, 17.5 in Sweden, and 16.4 in Ireland. In the United States, it is even larger at 60.2. This effect size dwarfs the resource and class-size effects found in these countries, and it even holds up against many family-background effects. Thus, this institutional effect in educational production can be quite substantial in determining performance differences within countries.

---

[25] Excluding the resource variables and their imputation controls, because their estimation may be biased by sorting effects, does not make any qualitative difference to the estimated coefficients on the institutional variables.

Still, the within-country variation in this institutional feature is limited. The variation in the extent of school autonomy in hiring teachers between countries is much bigger. At the one extreme, all schools in Iceland and the Netherlands can hire their own teachers, and the estimate in the United States is based on only 2 schools that report not having responsibility in hiring teachers. At the other extreme, only 4 schools in Germany and 6 in Greece in our sample have autonomy in hiring their teachers (cf. Table A2c). In light of this, it seems obvious that such institutional variation should be much more relevant in the determination of performance differences between countries, and thus should be much more readily detectable using the cross-country variation (cf. Wößmann 2003b).

6.2. Teacher Influence

A second institutional feature of the school system is how much influence teachers have in different decision-making areas. Teacher influence will generally increase the specific local knowledge that goes into a decision, but it also opens the opportunity for opportunistic behavior where teachers' own interests contrast with the furthering of student performance (cf. Bishop and Wößmann 2003). As a consequence, the net effect of teacher influence may be expected to be negative in decision-making areas where teachers' work load is involved, such as decisions on the subject matter to be covered in class, but positive in procedural matters, such as choice of a specific textbook.

Roughly in line with this general argument, students whose math teacher reported having considerable influence on the subject matter to be taught scored statistically significantly lower than students of teachers without considerable influence in this area in 4 Western European countries – Germany, Portugal, Sweden, and Switzerland. This effect was largest in Germany at 16.8 test-score points. By contrast, the opposite effect of a statistically significant higher performance of students of teachers with considerable influence on the subject matter is detected in Austria and Spain.

Students whose teacher reported having considerable influence in the choice of the specific textbook to be used scored statistically significantly higher in 4 Western European countries – French Belgium, France, Greece, and Scotland. In Scotland, the size of the performance difference is 14.3 test-score points. In no country does teacher influence on textbook choice have a statistically significantly negative effect.

Teacher influence on what supplies are purchased does not seem to have a consistent effect on student performance. Only in two countries is the estimate statistically significant; in Flemish Belgium, the estimate is negative, while in Norway, it is positive. However, the within-country variation in the extent of teacher influence on purchasing supplies is particularly low in most countries.

6.3. Testing and Homework Assignment

Two institutional features of the school system that should affect how much students spend studying are the regularity of testing students' performance and the amount of homework assignments. Both features should positively impact the relative benefits of studying for students, and thus their incentives to learn. In the TIMSS background questionnaires, testing can be measured by the hours per week outside the formal school day spent by the class teacher on preparing or grading student tests or exams. Students' homework assignments are also measured in hours per week, based on reports from each class teacher.

The time teachers spend on testing their students' achievement is statistically significantly positively related to student performance in 4 Western European countries – Austria, Germany, Greece, and Norway. No country exhibits a statistically significantly negative estimate. At about 5 additional test-score points for each additional hour per week spent by the teacher on preparing or grading exams, the effect size is not too large, though.

Homework assignments are statistically significantly positively related to student performance in 7 European school systems – Flemish Belgium, England, Germany, Iceland, Norway, Scotland, and Sweden – as well as in the United States. Again, there is no case of a statistically significantly negative estimate. In most countries, the effect size is substantial. For example, each additional hour of homework assignment in math goes with additional 28.7 math test-score points in England, i.e. more than a quarter of a standard deviation in test scores. The estimates are also large in Sweden at 16.4, in the United States at 14.9, in Scotland at 12.5, and in Flemish Belgium at 9.6.

Some caution is warranted in the interpretation of these two findings, though, because both measures might suffer from endogeneity bias. If teachers tend to assign more homework to high-performing students and if they tend to treat high-performing students with more testing, the least-squares coefficients on homework and testing will be upwardly biased estimates of their causal effect on student performance. However, the opposite might just as well be true: If teachers tend to try to particularly foster low-performing students by giving them more homework and testing them more regularly, the reported estimates will be biased downward. Lacking convincing instruments that might yield exogenous variation in testing and homework assignment, the possibility of bias must remain an open issue.

6.4. Parental Involvement

As a final institutional feature of schooling, the extent of the involvement of parents in the formal teaching process might have an influence of the educational performance achieved. In the TIMSS teacher background questionnaires, class teachers report whether parents uninterested in their children's learning and progress considerably limit how they teach their math class. They also report the same question for interested parents.

Students whose teachers report limitations through uninterested parents perform statistically significantly worse in 8 of the European school systems (Austria, French Belgium, England, France, Germany, Greece, Ireland, and Scotland), as well as in the United States. This finding relates to the previous evidence of considerable effects of family background in general. Only in Iceland is the estimate statistically significantly positive.

The same general pattern of negative effects on performance was not detected in the case where teachers deemed the involvement of interested parents as limiting for their teaching. Quite to the contrary, in French Belgium, Germany, and the United States, students whose teachers report being limited by interested parents perform statistically significantly *better*. Even though teachers deem the involvement of interested parents as limiting to their teaching, these "limitations" affect student performance positively in these countries. However, there is not much variation in this variable in many countries. The statistically significantly positive coefficient in French Belgium is based on only 6 teachers reporting limitations, and in Germany, it is based on only 1 teacher. This casts some doubt on the generalizability of this finding. Likewise, the statistically significant negative coefficients found in Austria, England, and Sweden are based on only 2, 3, and 3 teachers, respectively. Still, the pattern that limitations by uninterested parents are harmful for student performance but "limitations" by interested parents are actually conducive to performance, found in French Belgium, Germany, and the United States, is remarkable. Also, the estimated effect sizes are substantial, at –53.4 for uninterested parents and 48.2 for interested parents in Germany, –15.8 and 30.1 in the

United States, and –12.4 and 27.6 in French Belgium. Again, potential endogeneity considerations apply.

6.5. Additional Explanatory Power of the Institutional Features

The change in the explained test-score variation (ΔR^2) between the model reported in Table 7 and the one reported in Table 5 informs how much of the within-country performance variation can be attributed to the analyzed institutional factors, after influences of family background and resource endowments have been accounted for. This additional proportion of the test-score variation accounted for by the 8 institutional variables is larger than 3% in Austria and Germany, and larger than 2% in Flemish Belgium and the United States. For the other Western European countries, it is smaller than 2%, and in Portugal, it is as low as 0.3%. Relative to the explained test-score variation of the model presented in Table 5 – which includes 25 grade, student, family, and resource measures – the 8 institutional features added in Table 7 increase the explained test-score variation by 20% in Austria, 16% in Iceland, 15% in Germany, and 13% in Flemish Belgium. At the other extreme, the addition is only 1.5% in Portugal, 2.4% in Switzerland, and 2.5% in Denmark.

However, the institutional features are to some extent correlated with the family-background and resource measures. The above calculations attribute any explanatory power that is accounted for by this joint variation to the family-background and resource measures, and none to the institutional measures. It is not clear, though, whether the driving force of this jointly explained part are the family-background and resource measures or the institutions. Going to the other extreme of attributing the whole jointly explained part to institutions – i.e., looking at the R^2 of a model that includes only the 8 institutional variables – leads to a proportion of the total test-score variation attributed to institutional differences of 11.3% in Germany and 8.0% in Austria. Even in Portugal, 0.9% are attributed to institutions in this case. As a share of the total test-score variation explained by the model in Table 7 that includes all the explanatory variables, this is as much as 47.3% in Germany, 35.9% in Austria, and 34.8% in the United States. It is more than 10% in 13 of the 17 Western European school system, and even in Portugal, it is 4.6% (3.5% in the Netherlands).

7. Summary and Conclusion

This paper has analyzed the effects of family background, school resources, and institutional features on the educational performance of students in 17 Western European school systems. The general pattern suggests that family background – as measured by parental education, number of books at home, and whether students live with both parents and are native – is clearly the strongest predictor of student performance in all European countries. Resources, and particularly class sizes, play a small role at best in the success of educational production in Europe, while within-country variations in several institutional features are related to student performance in many European countries. With regard to this general pattern of results, and even with regard to the particular size of most family-background effects, educational production in Western Europe (as an aggregate) seems remarkably similar to educational production in the United States.

Still, there are also noteworthy variations between European countries. Family background, proxied by the number of books in the students' home, seems to exert the largest effect on students' math performance in Britain and Germany, and the lowest in France and Flemish Belgium. In most countries, the family-background effect does not vary strongly for students of different underlying ability. If anything, the effects tend to be stronger for higher-ability children in some countries, most notably in England. Across countries, the size of the family-background effects is basically unrelated to mean performance, suggesting that there is no inevitable equity-efficiency tradeoff in educational production – more equal opportunities for students from different family backgrounds do not necessarily have to be bought by lower average performance. Thus, any policy initiative should be aware of the importance of family effects and of the possibilities for equalizing opportunities.

In terms of more direct options for schooling policy, there is not much evidence that smaller classes cause better performance in Europe. In 8 European school systems, effects of substantial magnitude can be ruled out. Only in one country, Iceland, statistically significant (but relatively small) positive effects of smaller classes are found, while in the Netherlands, *bigger* classes cause better performance. Least-squares estimates – which might be biased, though – suggest that there is also little evidence of positive effects of instruction time and teachers' education and gender, while at least in some countries, shortage of materials and inexperienced teachers seem to exert negative effects on student outcomes.

In contrast to the humble effects of resource policies, there is some evidence that several institutional measures have a noteworthy impact on student performance in many European countries. Thus, within-country variations in the autonomy of schools to hire their teachers, in the regularity of student testing, and in the extent of homework assignments are quite strongly related to performance differences in several European school systems. In a few countries, within-country performance variation is also related to the extent of parental involvement and teacher influence. Altogether, in many countries a substantial part of the total test-score variation can be accounted for by within-country variations in the institutional features analyzed in this paper.

For many of the European countries, the results presented here are the first solid evidence on educational production functions ever. As is evident from the vast US literature on education production functions, though, many interesting and important features remain unresolved in the European setting, and others have been left out of the analysis altogether. Not all potential endogeneity problems have been resolved in this paper. This limitation is particularly severe in the case of most measures of resource endowments, where the variation given in observational data seems unlikely to be exogenous to student performance. As long as we do not have convincingly exogenous variation in these measures, the question of the importance of resource effects in educational production in Europe must remain open. In this regard, some of the presented findings must be viewed as hints towards possible effects to be scrutinized by future research, rather than definite results as to the causal determination of student performance.

Many more detailed questions have not been covered in this paper. For example, the effect of teacher gender may depend on the gender match between students and teachers, there may be other interaction effects between the different variables, and the question of possible peer effects in educational production is neglected altogether. Still, the presented results lay a sound foundation for future research to build on. As a general outlook, there is substantial room for future research to check and refine the results presented here, both for individual European countries and as an international comparison. For example, it remains open how to identify class-size effects in the countries where this paper did not get far, and how to identify causal resource effects more generally. Other international datasets such as the TIMSS-Repeat and the PISA studies may offer additional evidence, but are unlikely to help in this regard. Datasets to estimate value-added specifications are not available at this broad international spectrum, but might advance the knowledge on educational production in individual European countries. Even more promising in terms of detecting causal effects would be well founded and thoroughly researched experimental studies, of which I am unaware in Europe. Also, while this paper has focused mainly on basic cognitive skills in math as an educational output, only briefly reporting results in science, there are many other outcomes of the educational production process that should be considered in the future, such as social capabilities, motivation, and labor-market outcomes.

The most auspicious and proximate road for future research, however, seems to be to look at what might lie behind the cross-country differences in the results presented in this paper, particularly in terms of the findings on family-background effects. The results can feed future work trying to explain differences in educational equity achieved in different countries, by relating the family-background effects to characteristic features of the school systems across countries.

In light of all this, the presented results can only be viewed as a first step towards analyzing educational production in Europe, and much more research will be needed before definite answers can be given on the most important determinants of student performance in the different European school systems. But the future educational human capital in Europe seems like a topic too important to be as neglected as it has been in the past.

References

Akerhielm, Karen (1995). Does Class Size Matter? *Economics of Education Review* 14 (3): 229-241.

Ammermüller, Andreas, Hans Heijke, Ludger Wößmann (2003). Schooling Quality in Eastern Europe: Educational Production During Transition. Research Memorandum ROA-RM-2003/2E. Maastricht: Research Centre for Education and the Labour Market.

Angrist, Joshua D., Victor Lavy (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *Quarterly Journal of Economics* 114 (2): 533-575.

Barro, Robert J., Jong-Wha Lee (2001). International Data on Educational Attainment: Updates and Implications. *Oxford Economic Papers* 53 (3): 541-563.

Beaton, Albert E., Ina V.S. Mullis, Michael O. Martin, Eugenio J. Gonzalez, Dana L. Kelly, Teresa A. Smith (1996). *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.

Behrman, Jere R., Andrew D. Foster, Mark R. Rosenzweig, Prem Vashishtha (1999). Women's Schooling, Home Teaching, and Economic Growth. *Journal of Political Economy* 107 (4): 682-714.

Behrman, Jere R., Mark R. Rosenzweig (2002). Does Increasing Women's Schooling Raise the Schooling of the Next Generation? *American Economic Review* 92 (1): 323-334.

Bishop, John H., Ludger Wößmann (2003). Institutional Effects in a Simple Model of Educational Production. *Education Economics*: forthcoming (Kiel Working Paper 1085. Kiel: Institute for World Economics, 2001).

Björklund, Anders, Markus Jäntti (1997). Intergenerational Income Mobility in Sweden Compared to the United States. *American Economic Review* 87 (5): 1009-1018.

Björklund, Anders, Mikael Lindahl, Kristen Sund (2003). Family Background and School Performance During a Turbulent Era of School Reforms. Paper prepared for a conference organized by the Swedish Economic Council, March 2003, Stockholm.

Bonesrønning, Hans (1996). School Characteristics and Student Achievement: Evidence from Combined Upper Secondary Schools in Norway. *Education Economics* 4 (2): 143-160.

Bonesrønning, Hans (2003). Class Size Effects on Student Achievement in Norway: Patterns and Explanations. *Southern Economic Journal* 69 (4): 952-965.

Brunello, Giorgio, Daniele Checchi (2003). School Quality and Family Background in Italy. IZA Discussion Paper 705. Bonn: Institute for the Study of Labor.

Case, Anne, Angus Deaton (1999). School Inputs and Educational Outcomes in South Africa. *Quarterly Journal of Economics* 114 (3): 1047-1084.

Card, David (1999). The Causal Effect of Education on Earnings. In: Orley Ashenfelter, David Card (eds.), *Handbook of Labor Economics, Volume 3A*: 1801–1863. Amsterdam: Elsevier.

Card, David, Alan B. Krueger (1996). School Resources and Student Outcomes: An Overview of the Literature and New Evidence from North and South Carolina. *Journal of Economic Perspectives* 10 (4): 31-50.

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

Dearden, Lorraine, Javier Ferri, Costas Meghir (2002). The Effect of School Quality on Educational Attainment and Wages. *Review of Economics and Statistics* 84 (1): 1-20.

Dearden, Lorraine, Stephen Machin, Howard Reed (1997). Intergenerational Mobility in Britain. *Economic Journal* 107 (1): 47-66.

Dobbelsteen, Simone, Jesse Levin, Hessel Oosterbeek (2002). The Causal Effect of Class Size on Scholastic Achievement: Distinguishing the Pure Class Size Effect from the Effect of Changes in Class Composition. *Oxford Bulletin of Economic Research* 64 (1): 17-38.

Dolton, Peter J., Anna Vignoles (2003). The Impact of School Quality on Labour Market Success in the United Kingdom. *Bulletin of Economic Research*: forthcoming (Discussion Paper 98-03. Newcastle: University of Newcastle, 1999).

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

Dustmann, Christian, Najma Rajah, Arthur van Soest (2003). Class Size, Education, and Wages. *Economic Journal* 133 (485): F99-F120.

Eide, Eric, Mark H. Showalter (1998). The Effect of School Quality on Student Performance: A Quantile Regression Approach. *Economics Letters* 58 (3): 345-350.

Feinstein, Leon, James Symons (1999). Attainment in Secondary School. *Oxford Economic Papers* 51 (2): 300-321.

Fertig, Michael (2003). Who's to Blame? The Determinants of German Students' Achievement in the PISA 2000 Study. IZA Discussion Paper 739. Bonn: Institute for the Study of Labor.

Glewwe, Paul (2002). Schools and Skills in Developing Countries: Education Policies and Socioeconomic Outcomes. *Journal of Economic Literature* 40 (2): 436-482.

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

Gundlach, Erich, Ludger Wößmann, Jens Gmelin (2001). The Decline of Schooling Productivity in OECD Countries. *Economic Journal* 111 (471): C135-C147.

Häkkinen, Iida, Tanja Kirjavainen, Roope Uusitalo (2003). School Resources and Student Achievement Revisited: New Evidence from Panel Data. *Economics of Education Review* 22 (3): 329-335.

Hanushek, Eric A. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature* 24 (3): 1141-1177.

Hanushek, Eric A. (1995). Interpreting Recent Research on Schooling in Developing Countries. *World Bank Research Observer* 10 (2): 227-246.

Hanushek, Eric A. (2002). Publicly Provided Education. In: Alan J. Auerbach, Martin Feldstein (eds.), *Handbook of Public Economics, Volume 4*: 2045–2141. Amsterdam: Elsevier.

Hanushek, Eric A. (2003). The Failure of Input-Based Schooling Policies. *Economic Journal* 113 (485): F64-F98.

Hanushek, Eric A., with Charles S. Benson, Richard B. Freeman, Dean T. Jamison, Henry M. Levin, Rebecca A. Maynard, Richard J. Murnane, Steven G. Rivkin, Richard H. Sabot, Lewis C. Solmon, Anita A. Summers, Finis Welch, Barbara L. Wolfe (1994). *Making Schools Work: Improving Performance and Controlling Costs*. Washington, D.C.: The Brookings Institution.

Hanushek, Eric A., Javier A. Luque (2003). Efficiency and Equity in Schools around the World. *Economics of Education Review*: forthcoming (NBER Working Paper 8949. Cambridge, MA: National Bureau of Economic Research, 2002).

Harmon, Colm, Ian Walker, Niels Westergaard-Nielsen (2001). Introduction. In: Colm Harmon, Ian Walker, Niels Westergaard-Nielsen (eds.), *Education and Earnings in Europe: A Cross Country Analysis of the Returns to Education*. Cheltenham: Edward Elgar Publishing.

Heinesen, Eskil, Brian Krogh Graversen (2001). The Effect of School Resources on Educational Attainment: Evidence from Denmark. Paper presented at the Annual Conference of the European Society for Population Economics, Athens.

Hoxby, Caroline M. (2000). The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *Quarterly Journal of Economics* 115 (4): 1239-1285.

Iacovou, Maria (2002). Class Size in the Early Years: Is Smaller Really Better? *Education Economics* 10 (3): 261-290.

Jepsen, Christopher, Steven Rivkin (2002). What Is the Tradeoff Between Smaller Classes and Teacher Quality? NBER Working Paper 9205. Cambridge, MA: National Bureau of Economic Research.

Koenker, Roger, Gilbert Bassett, Jr. (1978). Regression Quantiles. *Econometrica* 46 (1): 33-50.

Kremer, Michael (2003). Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons. *American Economic Review* 93 (2): 102-106.

Krueger, Alan B. (1999). Experimental Estimates of Education Production Functions. *Quarterly Journal of Economics* 114 (2): 497-532.

Krueger, Alan B. (2003). Economic Considerations and Class Size. *Economic Journal* 113 (485): F34-F63.

Lauer, Charlotte (2003). Family Background, Cohort and Education: A French-German Comparison Based on a Multivariate Ordered Probit Model of Educational Attainment. *Labour Economics* 10 (2): 231-251.

Levačić, Rosalind, Anna Vignoles (2002). Researching the Link between School Resources and Student Outcomes in the UK: A Review of Issues and Evidence. *Education Economics* 10 (3): 313-331.

Levin, Jesse (2001). For Whom the Reductions Count: A Quantile Regression Analysis of Class Size and Peer Effects on Scholastic Achievement. *Empirical Economics* 26 (1): 221-246.

Lindahl, Mikael (2001a). Home versus School Learning: A New Approach to Estimating the Effect of Class Size on Achievement. IZA Discussion Paper 261. Bonn: Institute for the Study of Labor.

Lindahl, Mikael (2001b). Summer Learning and the Effect of Schooling: Evidence from Sweden. IZA Discussion Paper 262. Bonn: Institute for the Study of Labor.

Martin, Michael O., 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., 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.

Michaelowa, Katharina (2001). Primary Education Quality in Francophone Sub-Saharan Africa: Determinants of Learning Achievement and Efficiency Considerations. *World Development* 29 (10): 1699-1716.

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

Organisation for Economic Co-operation and Development (OECD) (1997), (1998). *Education at a Glance: OECD Indicators*. Paris: OECD.

Organisation for Economic Co-operation and Development (OECD) (2001). *Knowledge and Skills for Life: First Results from the OECD Programme for International Student Assessment (PISA) 2000*. Paris: OECD.

Psacharopoulos, George (2000). Economics of Education à la Euro. *European Journal of Education* 35 (1): 81-95.

Schindler Rangvid, Beatrice (2001). Resource Use and Complementary Measures of School Quality. Paper presented at the Annual Conference of the European Society for Population Economics, Athens.

Schnepf, Sylke Viola (2002). A Sorting Hat That Fails? The Transition from Primary to Secondary School in Germany. Innocenti Working Papers 92. Florence: UNICEF Innocenti Research Centre.

Solon, Gary (2002). Cross-Country Differences in Intergenerational Earnings Mobility. *Journal of Economic Perspectives* 16 (3): 59-66.

Vignoles, Anna, Rosalind Levaĉić, James Walker, Stephen Machin, David Reynolds (2000). The Relationship Between Resource Allocation and Pupil Attainment: A Review. Centre for the Economics of Education, Discussion Paper DP 02. London: London School of Economics and Political Science.

West, Martin R., Ludger Wößmann (2003). Which School Systems Sort Weaker Students into Smaller Classes? International Evidence. Kiel Working Paper 1145. Kiel: Institute for World Economics.

Wolter, Stefan C. (2003). Sibling Rivalry: A Six Country Comparison. IZA Discussion Paper 734. Bonn: Institute for the Study of Labor.

Wolter, Stefan C., Maja Coradi Vellacot (2002). Sibling Rivalry: A Look at Switzerland with PISA Data. IZA Discussion Paper 594. Bonn: Institute for the Study of Labor.

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

Wößmann, Ludger (2003a). Educational Production in East Asia: The Impact of Family Background and Schooling Policies on Student Performance. Kiel Working Paper 1152. Kiel: Institute for World Economics.

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

Wößmann Ludger, Martin R. West (2002). Class-Size Effects in School Systems Around the World: Evidence from Between-Grade Variation in TIMSS. Harvard University, Program on Education Policy and Governance Research Paper PEPG/02-02.

Table 1: Student Performance and Sample Size

Student performance: International test scores. Standard deviation in parentheses.
Standard deviation in percent of country mean test score in brackets. – Sample size: Absolute numbers.

	AUS	BFL	BFR	DEN	ENG	FRA	GER	GRE	ICE
Student performance									
Math score	524.1	561.7	517.5	485.3	491.3	514.4	496.6	461.2	473.1
Standard deviation	*(89.9)*	*(85.3)*	*(83.1)*	*(83.1)*	*(92.7)*	*(78.3)*	*(88.2)*	*(89.6)*	*(73.3)*
Std. dev./score (in percent)	*[17.2]*	*[15.2]*	*[16.1]*	*[17.1]*	*[18.9]*	*[15.2]*	*[17.8]*	*[19.4]*	*[15.5]*
Science score	537.9	540.3	457.5	460.5	532.5	473.9	515.2	472.2	477.8
Standard deviation	*(98.0)*	*(78.3)*	*(83.7)*	*(89.4)*	*(105.5)*	*(78.9)*	*(99.8)*	*(89.6)*	*(78.7)*
Std. dev./score (in percent)	*[18.2]*	*[14.5]*	*[18.3]*	*[19.4]*	*[19.8]*	*[16.7]*	*[19.4]*	*[19.0]*	*[16.5]*
Sample size									
Students	5,786	5,662	4,883	4,370	3,579	6,014	5,763	7,921	3,730
Classes	249	296	261	282	243	253	272	312	274
Schools	129	141	120	153	127	134	137	156	155

Table 1 (continued)

	IRE	NET	NOR	POR	SCO	SPA	SWE	SWI	USA
Student performance									
Math score	513.5	529.0	481.8	438.3	481.1	467.6	517.1	544.6	487.8
Standard deviation	*(90.8)*	*(85.2)*	*(82.8)*	*(63.7)*	*(86.5)*	*(74.2)*	*(90.2)*	*(91.2)*	*(90.9)*
Std. dev./score (in percent)	*[17.7]*	*[16.1]*	*[17.2]*	*[14.5]*	*[18.0]*	*[15.9]*	*[17.5]*	*[16.8]*	*[18.6]*
Science score	516.4	539.6	505.0	452.9	493.2	497.1	532.0	521.0	521.4
Standard deviation	*(95.7)*	*(85.0)*	*(88.9)*	*(77.1)*	*(99.9)*	*(81.4)*	*(95.7)*	*(93.9)*	*(106.2)*
Std. dev./score (in percent)	*[18.5]*	*[15.7]*	*[17.6]*	*[17.0]*	*[20.3]*	*[16.4]*	*[18.0]*	*[18.0]*	*[20.4]*
Sample size									
Students	6,203	4,084	5,736	6,753	5,776	7,596	8,855	11,722	10,973
Classes	261	187	287	283	257	309	535	613	529
Schools	132	95	249	142	128	154	270	327	183

Table 2: Descriptive Statistics on European Education Systems

	AUS	BFL	BFR	DEN	ENG	FRA	GER	GRE	ICE
(1) Size of student population	176,332	139,246	109,168	99,152	950,738	1,676,167	1,468,434	252,132	8,447
(2) Enrollment rate	98.6	99.6[1]	99.6[1]	97.4	98.9[2]	100.1	95.9	96.7	98.4[3]
(3) 90% enrollment range	5-16	3-17[1]	3-17[1]	6-16	4-15[2]	3-17	7-17	6-14	6-15
(4) School expectancy	15.2	17.6[1]	17.6[1]	16.3	15.3[2]	16.3	16.2	14.0	17.5[3]
(5) Entry rates to university	26	–	–	31	43[2]	33	27	16	39
(6) Expenditure per student	7,118	5,770	–	6,247	4,246[2]	5,182	6,254	1,950	–
(7) Exp. p. stud. rel. to GDP per capita	35	27	–	29	24[2]	31	29[4]	16	–
(8) Educat. expenditure relative to GDP	2.7	–	–	2.6	–	3.2	3.8[5]	1.5	3.6[5]
(9) Private funds	2	0	–	2	–	7	24	–	12[4]
(10) Teacher salary	26,200	28,360[1]	28,360[1]	28,990	29,948[2,3]	28,210	37,060	16,420	–
(11) Tea. salary rel. to GDP per capita	1.3	1.4[1]	1.4[1]	1.3	1.6[2,3]	1.4	1.8	1.3	–
(12) Education years of adult population	8.4	8.5[1]	8.5[1]	9.9	9.0[2]	7.9	9.6	8.0	8.3
(13) Return to education	6.9	–	–	6.4	9.4[2]	7.5	7.9	6.3	–
(14) Rel. unemployment by education	3.6	9.8[1]	9.8[1]	10.3	8.7[2]	7.0	8.6	-0.8	–

Table 2 (continued)

	IRE	NET	NOR	POR	SCO	SPA	SWE	SWI	USA
(1) Size of student population	136,121	367,082	101,389	284,341	126,576	1,096,146	194,687	136,414	6,345,143
(2) Enrollment rate	100.0	99.3	95.1	103.1	98.9[2]	104.0	94.7	97.5	101.9
(3) 90% enrollment range	5-15	4-17	6-17	6-15	4-15[2]	4-15	6-18	6-15	5-15
(4) School expectancy	15.2	16.9	16.2	15.7	15.3[2]	16.1	15.8	15.4	15.8
(5) Entry rates to university	27	34	25	–	43[2]	–	–	15	52
(6) Expenditure per student	3,395	4,351	6,360[4]	–	4,246[2]	3,455	5,643	7,601	6,812
(7) Exp. P. stud. rel. to GDP per capita	20	22	28[4]	–	24[2]	24	30	30	26
(8) Educat. expenditure relative to GDP	2.1	2.0	–	2.3	–	2.7	2.5	–	2.0
(9) Private funds	4	6	–	–	–	13	0	–	10
(10) Teacher salary	33,840	35,340	21,120	24,560	29,948[2,3]	28,020	20,310	50,400	30,460
(11) Tea. salary rel. to GDP per capita	2.0	1.8	0.9	2.0	1.6[2,3]	2.0	1.1	2.0	1.2
(12) Education year of adult population	8.8	9.0	11.8	4.5	9.0[2]	6.6	11.2	10.2	12.2
(13) Return to education	9.0	6.3	4.6	9.7	9.4[2]	7.2	4.1	9.0	10.0
(14) Rel. unemployment by education	13.0	3.8	4.8	2.9	8.7[2]	6.8	5.9	3.2	7.5

Definitions and sources (1995 unless noted otherwise):

(1) Size of student population in the two sampled grades, TIMSS estimate (Gonzalez and Smith 1997).

(2) Enrollment rate of students aged 5-14 as percentage of population aged 5-14, public + private institut. (based on head counts) (OECD 1997).

(3) Age range at which over 90% of the population are enrolled, in public + private institut., full-time + part-time students, 1996 (OECD 1998).

(4) Years of school expectancy for a 5 year-old child, in public + private institutions (based on head counts) (OECD 1997).

(5) Net entry rates for university-level education (OECD 1997).

(6) Expenditure per student (US dollars converted using PPPs), public + private institut. (full-time equivalents), secondary education (OECD 1998).

(7) = (6) relative to GDP per capita, in percent (OECD 1998).

(8) Educational expenditure from public + private sources for educational institutions as a percentage of GDP, secondary education (OECD 1998).

(9) Private sources of final (after transfers from public sources) funds for educ. institut., in percent, primary + secondary education (OECD 1998).

(10) Annual statutory teachers' salary after 15 years of experience with minimum training in public institutions at the lower secondary level of education, in equivalent US dollars converted using PPPs (OECD 1997).

(11) Ratio of (10) to GDP per capita (OECD 1997).

(12) Average schooling years in the total population aged 25 and over (Barro and Lee 2001).

(13) Rate of return to education (wage effect), men, in percent, year closest to 1995 (Harmon et al. 2001; USA: Card 1999).

(14) Difference in unemployment rate between 25-64 year-olds with university-level educ. + with educ. below upper secondary level (OECD 1995).

[1] Belgium. [2] United Kingdom. [3] 1996 (from OECD 1998). [4] All levels of education combined. [5] Primary and secondary education. [6] Public education only. [7] All secondary education.

Table 3: Family Background and Math Performance

Least-squares regression within each country, weighted by students' sampling probabilities.
Dependent variable: TIMSS math test score. Clustering-robust standard errors in parentheses.

	AUS	BFL	BFR	DEN	ENG	FRA	GER	GRE	ICE
Parents' education									
Some secondary	–	6.98	-19.79^{+}	–	–	8.38	–	9.24*	25.83*
		(12.03)	*(9.40)*			*(6.64)*		*(3.48)*	*(7.82)*
Finished secondary	16.64^{+}	20.37°	-11.04	14.71°	–	19.63*	12.42*	23.00*	30.75*
	(6.45)	*(11.08)*	*(9.32)*	*(7.53)*		*(6.88)*	*(3.44)*	*(3.44)*	*(8.14)*
Some after secondary	–	37.87*	2.02	24.97*	–	21.24*	11.37°	23.01*	34.81*
		(13.27)	*(10.15)*	*(7.60)*		*(7.16)*	*(6.26)*	*(5.19)*	*(6.89)*
Finished university	31.62*	41.39*	9.13	32.54*	–	38.25*	29.51*	51.28*	43.36*
	(8.27)	*(12.18)*	*(9.71)*	*(7.39)*		*(6.94)*	*(5.25)*	*(4.44)*	*(7.19)*
Books at home									
One shelf (11-25)	18.22*	15.86*	9.66	14.74°	24.41*	-4.62	20.06*	11.15*	19.23
	(4.26)	*(5.51)*	*(6.96)*	*(8.02)*	*(5.96)*	*(4.98)*	*(4.28)*	*(3.85)*	*(11.90)*
One bookcase (26-100)	37.70*	26.56*	29.53*	33.08*	60.29*	8.75°	42.78*	34.16*	30.46*
	(4.67)	*(6.20)*	*(5.00)*	*(7.73)*	*(6.45)*	*(4.67)*	*(4.79)*	*(4.58)*	*(10.43)*
Two bookcases (101-200)	62.83*	33.74*	36.69*	45.48*	81.22*	16.63*	65.12*	44.70*	44.44*
	(5.91)	*(6.19)*	*(5.25)*	*(8.12)*	*(6.62)*	*(4.93)*	*(5.86)*	*(5.76)*	*(10.47)*
More than two bookcases	71.83*	29.97*	44.68*	53.45*	104.43*	11.17^{+}	75.24*	44.66*	54.54*
(>200)	*(5.25)*	*(6.45)*	*(5.45)*	*(7.63)*	*(7.77)*	*(5.23)*	*(5.34)*	*(5.64)*	*(11.39)*
Living with both parents	-2.84	-2.28	-6.97	2.46	4.51	7.82*	9.37*	13.85*	5.37
	(2.99)	*(4.56)*	*(4.36)*	*(3.82)*	*(4.27)*	*(2.46)*	*(3.10)*	*(2.46)*	*(3.98)*
Born in country	25.93*	8.54	18.96*	20.40*	10.59	–	16.07*	9.47^{+}	10.32^{+}
	(6.79)	*(8.28)*	*(5.61)*	*(6.29)*	*(7.42)*		*(4.89)*	*(4.47)*	*(5.22)*
Female	-10.22*	0.34	-13.05*	-12.56*	-9.13°	-10.69*	-8.51*	-9.73*	-3.21
	(2.95)	*(6.42)*	*(3.42)*	*(2.60)*	*(4.87)*	*(2.04)*	*(3.28)*	*(1.92)*	*(2.83)*
Age	-22.79*	-36.08*	-31.36*	-19.39*	7.03	-24.74*	-18.67*	-13.42*	-0.70
	(2.68)	*(3.05)*	*(1.97)*	*(3.37)*	*(4.99)*	*(2.36)*	*(2.50)*	*(3.34)*	*(4.48)*
Upper grade	52.27*	52.46*	55.82*	57.79*	23.00*	67.39*	42.31*	55.59*	26.80*
	(3.73)	*(5.14)*	*(3.92)*	*(4.68)*	*(5.30)*	*(4.36)*	*(3.78)*	*(4.62)*	*(7.57)*
Above upper grade	–	–	–	–	–	–	–	–	–
Community location									
Geographically isolated	50.49*	–	–	-26.10^{+}	-11.25	–	27.11	-7.39	-9.71
	(10.42)			*(10.19)*	*(11.42)*		*(31.11)*	*(5.61)*	*(5.99)*
Close to town center	15.70^{+}	-4.48	12.82^{+}	-12.55^{+}	-12.30	2.25	4.65	15.28*	2.24
	(7.13)	*(8.53)*	*(6.38)*	*(6.11)*	*(8.51)*	*(5.23)*	*(9.81)*	*(5.17)*	*(4.77)*
Imputation controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Students [Unit of observation]	5786	5662	4883	4370	3579	6014	5763	7921	3730
Schools [Unit of clustering]	129	141	120	153	127	134	137	156	155
R^2	0.187	0.149	0.243	0.142	0.157	0.230	0.218	0.201	0.118
R^2 (without imputation controls)	0.173	0.137	0.221	0.123	0.148	0.211	0.187	0.179	0.092

Significance levels (based on clustering-robust standard errors): * 1%. – $^{+}$ 5%. – ° 10%.

Table 3 (continued)

	IRE	NET	NOR	POR	SCO	SPA	SWE	SWI	USA
Parents' education									
Some secondary	7.65	25.63^{+}	–	5.36^{+}	–	7.82*	–	–	11.06
	(5.27)	*(10.05)*		*(2.19)*		*(2.57)*			*(8.63)*
Finished secondary	15.60*	24.46^{+}	4.62	13.92*	-1.58	22.19*	13.63*	16.01*	17.20°
	(5.49)	*(10.43)*	*(9.10)*	*(3.32)*	*(4.46)*	*(2.99)*	*(4.18)*	*(3.66)*	*(8.83)*
Some after secondary	25.20*	44.27*	20.28^{+}	24.64*	14.36*	8.90*	22.42*	24.41*	31.48*
	(5.60)	*(9.44)*	*(8.20)*	*(4.15)*	*(5.33)*	*(3.19)*	*(5.00)*	*(4.71)*	*(8.29)*
Finished university	36.57*	40.71*	27.46*	37.52*	38.07*	19.78*	29.54*	34.98*	52.66*
	(6.40)	*(10.50)*	*(7.88)*	*(3.87)*	*(5.09)*	*(3.67)*	*(4.70)*	*(4.95)*	*(9.16)*
Books at home									
One shelf (11-25)	13.99*	17.34^{+}	10.20	8.81*	26.02*	15.33*	3.66	18.97*	9.75^{+}
	(4.89)	*(7.49)*	*(9.58)*	*(2.22)*	*(3.76)*	*(4.48)*	*(5.55)*	*(4.48)*	*(3.78)*
One bookcase (26-100)	47.36*	37.53*	23.33*	17.93*	48.05*	31.12*	36.84*	39.69*	34.57*
	(5.06)	*(6.77)*	*(8.81)*	*(2.55)*	*(3.87)*	*(4.28)*	*(4.99)*	*(4.66)*	*(3.56)*
Two bookcases (101-200)	63.70*	51.37*	37.23*	27.44*	69.99*	42.60*	55.93*	57.89*	53.48*
	(5.30)	*(7.04)*	*(9.34)*	*(2.99)*	*(4.79)*	*(4.75)*	*(5.01)*	*(5.30)*	*(4.23)*
More than two bookcases (>200)	65.35*	60.37*	54.96*	24.77*	78.56*	49.81*	67.75*	65.88*	62.61*
	(5.48)	*(7.61)*	*(8.85)*	*(3.31)*	*(4.62)*	*(4.51)*	*(5.06)*	*(5.26)*	*(4.75)*
Living with both parents	14.04*	-1.86	18.66*	1.79	7.18^{+}	0.50	5.19°	8.34*	15.48*
	(3.59)	*(4.90)*	*(5.33)*	*(2.23)*	*(2.89)*	*(2.57)*	*(3.14)*	*(2.45)*	*(2.89)*
Born in country	-13.16^{+}	8.71	22.34*	-6.07^{+}	2.05	14.37*	30.30*	24.27*	1.56
	(5.76)	*(6.18)*	*(4.85)*	*(3.06)*	*(4.31)*	*(5.06)*	*(4.01)*	*(4.06)*	*(4.57)*
Female	-19.67*	-10.68*	-5.59^{+}	-9.79*	-10.56*	-9.70*	-2.26	-13.26*	-9.01*
	(4.93)	*(3.01)*	*(2.27)*	*(1.48)*	*(2.23)*	*(2.05)*	*(2.02)*	*(2.01)*	*(2.33)*
Age	-26.67*	-29.41*	2.42	-12.85*	8.79^{+}	-20.97*	-1.63	-17.11*	-22.09*
	(2.45)	*(2.99)*	*(3.81)*	*(0.90)*	*(4.14)*	*(1.44)*	*(3.13)*	*(3.47)*	*(2.61)*
Upper grade	54.68*	60.93*	40.33*	43.20*	27.18*	59.25*	39.56*	60.23*	45.99*
	(5.32)	*(6.80)*	*(4.54)*	*(2.22)*	*(5.84)*	*(2.45)*	*(4.55)*	*(5.51)*	*(4.44)*
Above upper grade	–	–	–	–	–	–	74.06*	118.93*	–
							(7.54)	*(9.23)*	
Community location									
Geographically isolated	6.71	–	6.98	-20.87*	12.15°	-46.10*	-19.35	4.85	-28.90*
	(12.83)		*(7.42)*	*(4.99)*	*(6.71)*	*(2.22)*	*(13.76)*	*(14.67)*	*(7.95)*
Close to town center	5.62	15.01	7.11	0.38	-3.31	2.37	10.04*	-5.74	-4.11
	(6.47)	*(10.67)*	*(5.74)*	*(3.00)*	*(5.08)*	*(4.52)*	*(3.88)*	*(8.85)*	*(6.64)*
Imputation controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Students [Unit of observation]	6203	4084	5736	6753	5776	7596	8855	11722	10973
Schools [Unit of clustering]	132	95	249	142	128	154	270	327	183
R^2	0.180	0.207	0.147	0.200	0.214	0.205	0.232	0.272	0.185
R^2 (without imputation controls)	0.162	0.187	0.135	0.196	0.200	0.199	0.219	0.264	0.175

Significance levels (based on clustering-robust standard errors): * 1%. – $^{+}$ 5%. – $^{\circ}$ 10%.

Table 4: Family-Background Effects in Europe vs. the United States in Math

Least-squares regression within each country, weighted by students' sampling probabilities. Dependent variable: TIMSS math test score. Clustering-robust standard errors in parentheses.

	All countries	Interaction with Europe dummy
Parents' education		
Some secondary	16.56+	-12.26°
	(7.22)	*(7.34)*
Finished secondary	17.27+	-4.39
	(7.33)	*(7.41)*
Some after secondary	30.83*	-8.73
	(7.01)	*(7.17)*
Finished university	52.01*	-20.08+
	(8.02)	*(8.15)*
Books at home		
One shelf (11-25)	9.25+	7.60°
	(3.70)	*(3.94)*
One bookcase (26-100)	34.82*	2.97
	(3.58)	*(3.84)*
Two bookcases (101-200)	53.42*	-0.39
	(4.20)	*(4.46)*
More than two bookcases (>200)	62.41*	-0.87
	(4.67)	*(4.90)*
Living with both parents	15.40*	-9.69*
	(2.82)	*(2.96)*
Born in country	2.12	12.63*
	(4.69)	*(4.92)*
Female	-8.92*	0.00
	(2.32)	*(2.48)*
Age	-22.21*	3.47
	(2.64)	*(2.73)*
Upper grade	46.15*	4.96
	(4.42)	*(4.56)*
Above upper grade	111.60*	–
	(3.29)	
Community location		
Geographically isolated	-28.59*	25.77*
	(7.68)	*(8.56)*
Close to town center	-6.65	9.43
	(6.30)	*(6.50)*
Imputation controls	yes	
Country dummies	yes	
Students [Unit of observation]	115406	
Schools [Unit of clustering]	2932	
R^2	0.262	
R^2 (without imputation controls)	0.253	

Significance levels (based on clustering-robust standard errors): * 1%. – + 5%. – ° 10%.

Table 5: Resources, Teacher Characteristics, and Math Performance

Least-squares regression within each country, weighted by students' sampling probabilities.
Dependent variable: TIMSS math test score. Clustering-robust standard errors in parentheses.

	AUS	BFL	BFR	DEN	ENG	FRA	GER	GRE	ICE
Class size (log)	3.24	66.76*	60.19*	9.22	87.94*	63.96*	64.97*	-8.94	11.54+
	(6.13)	*(14.59)*	*(11.89)*	*(7.16)*	*(18.74)*	*(18.84)*	*(14.73)*	*(12.78)*	*(4.88)*
Shortage of materials									
None	-10.00	8.27	6.83	-1.21	37.40*	7.89	-1.90	-7.83	-5.66
	(6.89)	*(9.16)*	*(6.77)*	*(6.88)*	*(11.02)*	*(5.08)*	*(10.29)*	*(5.11)*	*(3.73)*
A lot	-41.19*	–	3.41	4.44	-5.07	4.84	-4.43	-1.26	4.95
	(8.39)		*(7.26)*	*(4.58)*	*(7.07)*	*(5.85)*	*(13.47)*	*(4.82)*	*(6.97)*
Instruction time	–	–	–	–	-10.33	1.03	-3.27	–	0.09
					(8.81)	*(1.79)*	*(6.32)*		*(1.79)*
Teacher characteristics									
Female teacher	-15.16+	-6.07	10.69+	-1.40	-7.44	5.56	-1.06	3.27	-1.48
	(6.24)	*(5.89)*	*(4.30)*	*(3.84)*	*(6.90)*	*(3.94)*	*(6.88)*	*(4.30)*	*(5.07)*
Teacher's experience (log)	5.69	4.73	7.77*	9.08*	-1.91	2.37	-4.73	4.52	4.70°
	(3.79)	*(4.48)*	*(2.96)*	*(2.58)*	*(3.65)*	*(2.29)*	*(6.11)*	*(3.14)*	*(2.50)*
Teacher's education									
Secondary only	–	–	–	–	–	59.80*	–	–	-3.52
						(13.94)			*(6.42)*
BA or equivalent	–	–	–	–	6.48	52.56*	–	–	1.48
					(7.41)	*(14.67)*			*(5.23)*
MA/PhD	–	–	–	–	7.98	53.27*	–	14.11	-26.69+
					(11.03)	*(15.37)*		*(18.49)*	*(12.83)*
Family-background controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Imputation controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Students [Unit of observation]	5786	5662	4883	4370	3579	6014	5763	7921	3730
Schools [Unit of clustering]	129	141	120	153	127	134	137	156	155
R^2	0.243	0.194	0.277	0.149	0.216	0.259	0.253	0.206	0.132
R^2 (without imputation controls)	0.186	0.161	0.235	0.127	0.203	0.229	0.208	0.183	0.100

Significance levels (based on clustering-robust standard errors): * 1%. – + 5%. – ° 10%.

Table 5 (cont.)

	IRE	NET	NOR	POR	SCO	SPA	SWE	SWI	USA
Class size (log)	121.86*	125.34*	11.14	21.74*	64.50*	9.47°	42.83*	3.60	-3.72
	(13.43)	*(18.55)*	*(8.23)*	*(5.87)*	*(12.10)*	*(4.97)*	*(6.87)*	*(7.32)*	*(6.44)*
Shortage of materials									
None	9.39°	24.78*	-0.78	7.38^{+}	–	5.01	-2.45	6.43	-1.67
	(5.66)	*(7.24)*	*(3.60)*	*(2.99)*		*(3.85)*	*(4.10)*	*(5.52)*	*(6.06)*
A lot	-2.41	33.24*	-18.39^{+}	-1.25	–	20.35°	-5.12	-30.50*	-28.58^{+}
	(7.29)	*(6.83)*	*(7.27)*	*(3.20)*		*(11.92)*	*(6.47)*	*(10.17)*	*(11.64)*
Instruction time	-0.76	1.36	–	-0.93	-0.47	0.62	0.38	5.93^{+}	-1.94
	(1.93)	*(7.51)*		*(0.89)*	*(2.02)*	*(1.13)*	*(1.24)*	*(2.60)*	*(1.61)*
Teacher characteristics									
Female teacher	-7.67	8.22	4.18	2.93	4.85	-1.58	3.61	-2.83	8.82°
	(5.72)	*(9.60)*	*(3.32)*	*(2.58)*	*(4.40)*	*(3.48)*	*(3.56)*	*(6.52)*	*(5.28)*
Teacher's experience (log)	-1.98	9.93^{+}	4.41^{+}	2.87°	3.87°	1.67	4.14°	5.12°	2.87
	(3.23)	*(4.25)*	*(2.13)*	*(1.46)*	*(2.29)*	*(2.98)*	*(2.18)*	*(2.74)*	*(2.98)*
Teacher's education									
Secondary only	–	53.61*	–	–	–	9.71	–	–	–
		(9.34)				*(6.33)*			
BA or equivalent	30.83*	36.61°	-0.88	–	2.58	7.86	8.98	42.60*	–
	(9.92)	*(20.12)*	*(3.61)*		*(9.53)*	*(5.75)*	*(6.55)*	*(4.98)*	
MA/PhD	33.18^{+}	75.66*	-5.43	–	8.06	-9.92	15.77	53.61*	9.95°
	(15.18)	*(10.46)*	*(7.59)*		*(10.34)*	*(8.68)*	*(14.07)*	*(8.56)*	*(5.88)*
Family-background controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Imputation controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Students [Unit of observation]	6203	4084	5736	6753	5776	7596	8855	11722	10973
Schools [Unit of clustering]	132	95	249	142	128	154	270	327	183
R^2	0.2438	0.3276	0.158	0.2107	0.238	0.2163	0.2513	0.3349	0.2034
R^2 (without imputation controls)	0.216	0.297	0.141	0.204	0.219	0.208	0.238	0.320	0.187

Significance levels (based on clustering-robust standard errors): * 1%. – $^{+}$ 5%. – ° 10%.

Table 6: The Coefficient on Log Class Size in Math

Regressions within each country, weighted by students' sampling probabilities. Dependent variable: TIMSS math test score. Controlling for family-background variables and imputation controls. Clustering-robust standard errors in parentheses.

	AUS	BFL	BFR	DEN	ENG	FRA	GER	GRE	ICE
LS	9.92	61.53*	55.23*	5.79	79.67*	60.82*	93.39*	-1.55	0.97
	(7.36)	*(15.27)*	*(10.60)*	*(8.27)*	*(18.81)*	*(21.42)*	*(19.65)*	*(11.04)*	*(5.35)*
IV	236.45	129.67*	120.23*	36.48°	–	-13.59	137.72*	-15.64	-5.26
	(1326.38)	*(29.79)*	*(44.40)*	*(18.67)*		*(32.48)*	*(49.38)*	*(20.62)*	*(4.45)*
SFE	12.04	33.23°	26.44	-20.83*	120.41*	43.02°	-0.99	-19.75	-5.26
	(7.83)	*(19.66)*	*(18.24)*	*(6.90)*	*(18.58)*	*(21.84)*	*(17.71)*	*(18.62)*	*(9.26)*
SFE-IV	120.80	63.27	21.59	-26.60	–	-81.21	-3.95	-5.24	-53.09*
	(135.77)	*(78.17)*	*(46.20)*	*(27.21)*		*(54.00)*	*(33.47)*	*(47.67)*	*(17.71)*
F (instr. 1st stage)	2.63	12.99	17.08	37.17	–	30.12	10.76	42.03	11.76
F ($\beta_2 = -100$)	2.64	4.36	6.93	7.28	–	0.12	8.24	3.95	7.02
Prob. > *F*	*(0.107)*	*(0.039)*	*(0.010)*	*(0.008)*		*(0.729)*	*(0.005)*	*(0.049)*	*(0.010)*
Students	5599	5649	4866	3996	3270	5669	5620	7921	3119
Schools	120	140	119	128	116	119	130	156	118

Methods of estimation: LS = Least squares. – IV = Instrumental variables. – SFE = School fixed effects. – SFE-IV = Combination of school fixed effects and instrumental variables. See text for details on the four methods of estimation.

Significance levels (based on clustering-robust standard errors): * 1%. – $^{+}$ 5%. – ° 10%.

Table 6 (cont.)

	IRE	NET	NOR	POR	SCO	SPA	SWE	SWI	USA
LS	120.08*	108.34*	27.88*	20.62*	52.23*	9.46	39.52*	12.81	-3.29
	(15.22)	*(14.49)*	*(9.05)*	*(5.94)*	*(13.98)*	*(5.97)*	*(7.84)*	*(12.44)*	*(6.66)*
IV	109.85*	299.24*	98.91*	7.16	36.76^{+}	18.77	21.06	85.35	-25.98
	(37.16)	*(104.46)*	*(32.20)*	*(19.05)*	*(17.66)*	*(11.67)*	*(26.75)*	*(64.87)*	*(25.67)*
SFE	142.47*	69.24*	3.75	24.43*	58.81^{+}	-2.18	–	18.64*	-0.81
	(26.83)	*(24.53)*	*(13.03)*	*(8.41)*	*(23.01)*	*(5.77)*		*(7.03)*	*(7.90)*
SFE-IV	-53.20	88.04*	167.86	31.34	6.91	-12.25	–	323.14	52.38
	(191.11)	*(30.12)*	*(237.07)*	*(28.08)*	*(30.27)*	*(88.90)*		*(2228.14)*	*(42.66)*
F (instr. 1st stage)	7.16	20.21	18.13	1.28	36.49	16.32	–	13.53	10.86
F ($\beta_2 = -100$)	0.06	38.98	1.28	21.87	12.48	0.97	–	0.04	12.76
Prob. > *F*	*(0.807)*	*(0.000)*	*(0.266)*	*(0.000)*	*(0.001)*	*(0.325)*		*(0.850)*	*(0.001)*
Students	6152	4025	1351	6721	5667	7556	8855	6159	10831
Schools	129	92	38	141	124	152	270	131	179

Methods of estimation: LS = Least squares. – IV = Instrumental variables. – SFE = School fixed effects. – SFE-IV = Combination of school fixed effects and instrumental variables. See text for details on the four methods of estimation.

Significance levels (based on clustering-robust standard errors): * 1%. – $^{+}$ 5%. – ° 10%.

Table 7: Institutions and Math Performance

Least-squares regression within each country, weighted by students' sampling probabilities.
Dependent variable: TIMSS math test score. Clustering-robust standard errors in parentheses.

	AUS	BFL	BFR	DEN	ENG	FRA	GER	GRE	ICE
School responsibility	7.76	19.52°	8.39	6.70	–	2.91	3.16	-19.44	–
for hiring teachers	*(8.76)*	*(10.67)*	*(7.65)*	*(6.65)*		*(6.04)*	*(21.60)*	*(13.18)*	
Teacher influence									
Subject matter	11.19°	-3.03	-11.62	-2.66	–	6.36	-16.82°	-0.60	-4.19
to be taught	*(6.65)*	*(11.43)*	*(10.07)*	*(3.68)*		*(4.92)*	*(8.98)*	*(4.61)*	*(6.01)*
Specific textbooks	-8.60	-4.53	10.66^{+}	6.98	–	9.94^{+}	-7.26	9.16°	1.12
to be used	*(6.02)*	*(6.39)*	*(5.34)*	*(4.78)*		*(4.63)*	*(8.11)*	*(5.05)*	*(6.95)*
What supplies	-6.08	-14.91^{+}	-2.46	5.35	–	1.58	-2.50	-0.75	2.85
are purchased	*(11.44)*	*(6.62)*	*(9.62)*	*(3.69)*		*(4.99)*	*(7.28)*	*(7.22)*	*(5.82)*
Testing and homework									
Tests/exams	4.89*	-0.41	1.56	–	-0.94	2.00	5.11*	4.70*	0.44
	(1.73)	*(1.90)*	*(1.33)*		*(1.57)*	*(1.55)*	*(1.94)*	*(1.28)*	*(1.98)*
Homework	5.28	9.62^{+}	1.95	-1.16	28.73*	3.21	8.07°	-1.49	4.75^{+}
	(3.22)	*(4.34)*	*(2.87)*	*(2.04)*	*(8.36)*	*(2.85)*	*(4.37)*	*(1.00)*	*(2.27)*
Teaching limited by									
Uninterested parents	-47.22*	4.85	-12.38°	12.46	-28.47^{+}	-17.99^{+}	-53.36*	-11.47^{+}	27.65^{+}
	(12.57)	*(9.83)*	*(6.96)*	*(9.93)*	*(11.90)*	*(7.18)*	*(8.57)*	*(5.52)*	*(10.70)*
Interested parents	-54.99*	-1.35	27.56^{+}	2.15	-48.14*	–	48.22^{+}	11.75	-8.69
	(15.97)	*(11.93)*	*(11.37)*	*(6.94)*	*(15.78)*		*(19.04)*	*(10.95)*	*(8.02)*
Family-background controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Resource controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Imputation controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Students [Unit of observation]	5786	5662	4883	4370	3579	6014	5763	7921	3730
Schools [Unit of clustering]	129	141	120	153	127	134	137	156	155
R^2	0.274	0.212	0.286	0.156	0.227	0.275	0.282	0.218	0.153
R^2 (without imputation controls)	0.223	0.181	0.244	0.130	0.210	0.240	0.240	0.193	0.116

Significance levels (based on clustering-robust standard errors): * 1%. – $^{+}$ 5%. – $^{\circ}$ 10%.

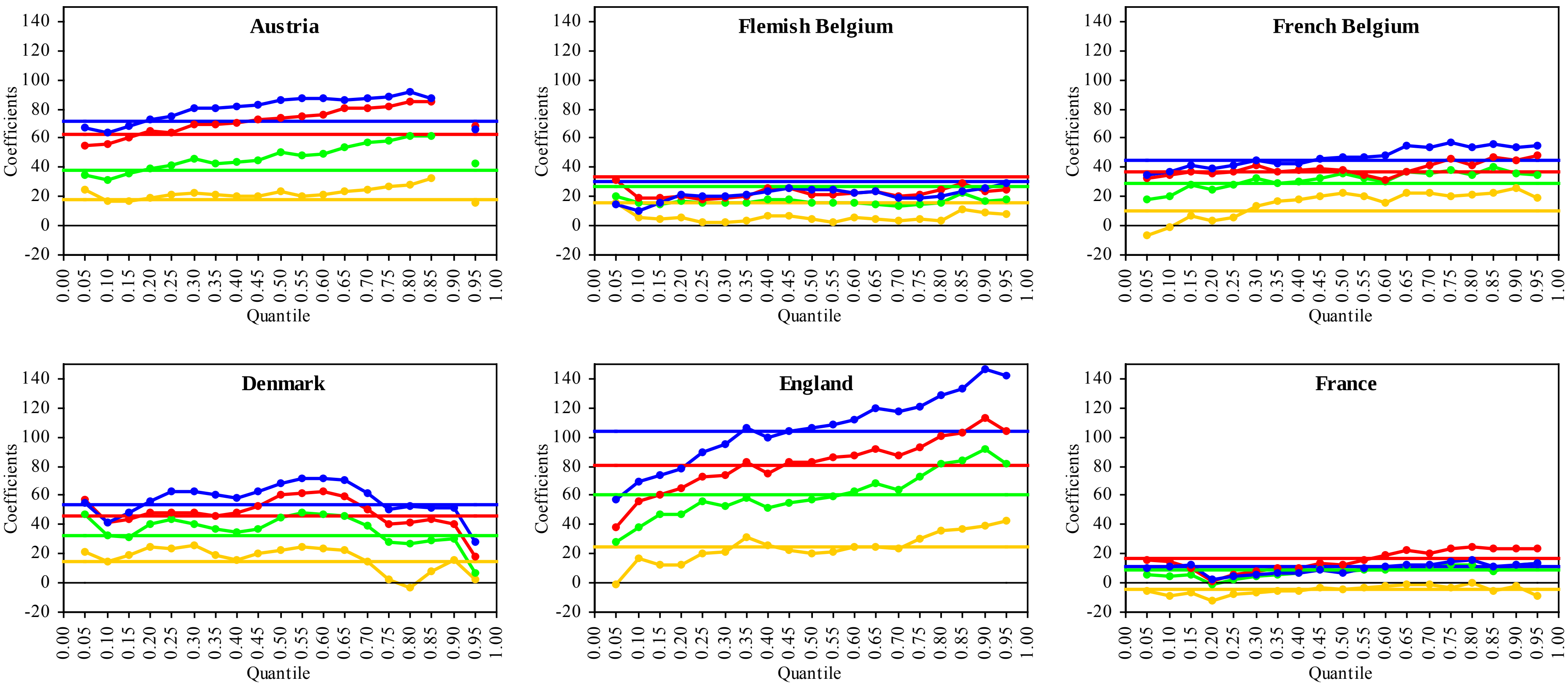
Table 7 (cont.)

	IRE	NET	NOR	POR	SCO	SPA	SWE	SWI	USA
School responsibility	16.36°	–	–	0.57	–	8.50°	17.45+	-11.86	60.16*
for hiring teachers	*(8.89)*			*(2.83)*		*(4.59)*	*(7.80)*	*(8.07)*	*(12.22)*
Teacher influence									
Subject matter	-4.68	9.78	1.24	-6.82°	-10.80	11.24*	-10.81*	-11.60+	-2.75
to be taught	*(5.86)*	*(8.80)*	*(3.61)*	*(3.52)*	*(7.04)*	*(3.95)*	*(3.61)*	*(5.08)*	*(5.98)*
Specific textbooks	7.42	-5.74	-3.07	0.19	14.25+	-2.02	-3.07	-4.56	-1.88
to be used	*(5.26)*	*(9.46)*	*(4.99)*	*(3.01)*	*(6.79)*	*(3.60)*	*(4.45)*	*(7.03)*	*(6.57)*
What supplies	-2.15	3.44	8.35°	-3.31	-11.74	-3.25	4.62	-0.94	3.66
are purchased	*(6.45)*	*(12.08)*	*(4.39)*	*(4.81)*	*(8.50)*	*(3.82)*	*(4.54)*	*(5.84)*	*(6.07)*
Testing and homework									
Tests/exams	2.60	-0.56	2.23+	-0.36	-0.46	1.40	-0.09	2.06	-0.02
	(1.65)	*(2.30)*	*(1.01)*	*(0.69)*	*(1.47)*	*(1.50)*	*(1.29)*	*(1.41)*	*(1.60)*
Homework	3.63	8.27	3.68°	0.06	12.47*	-0.47	16.35*	1.26	14.94*
	(3.74)	*(6.85)*	*(1.97)*	*(1.40)*	*(3.84)*	*(1.46)*	*(4.03)*	*(3.31)*	*(2.35)*
Teaching limited by									
Uninterested parents	-34.21*	–	16.91	-7.26	-17.44+	0.04	6.09	-6.58	-15.75+
	(8.24)		*(11.06)*	*(4.97)*	*(7.83)*	*(6.37)*	*(10.70)*	*(28.55)*	*(7.33)*
Interested parents	-7.90	6.90	2.64	3.40	–	7.32	-21.92+	-19.24	30.08°
	(12.32)	*(19.36)*	*(6.26)*	*(4.36)*		*(6.90)*	*(11.05)*	*(12.18)*	*(16.64)*
Family-background controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Resource controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Imputation controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Students [Unit of observation]	6203	4084	5736	6753	5776	7596	8855	11722	10973
Schools [Unit of clustering]	132	95	249	142	128	154	270	327	183
R^2	0.269	0.338	0.162	0.217	0.248	0.228	0.265	0.343	0.235
R^2 (without imputation controls)	0.235	0.307	0.145	0.207	0.228	0.216	0.248	0.327	0.209

Significance levels (based on clustering-robust standard errors): * 1%. – + 5%. – ° 10%.

Figure 1: Quantile Regression Estimates of Family-Background Effects

Coefficients on the four books-at-home categories in the model of Table 3. – Horizontal lines: WLS estimates. – Curved lines with dots: Quantile regression estimates for 0.05 intervals. – Categories (residual = less than one shelf): –●– One shelf. –●– One bookcase. –●– Two bookcases. –●– More than two bookcases.



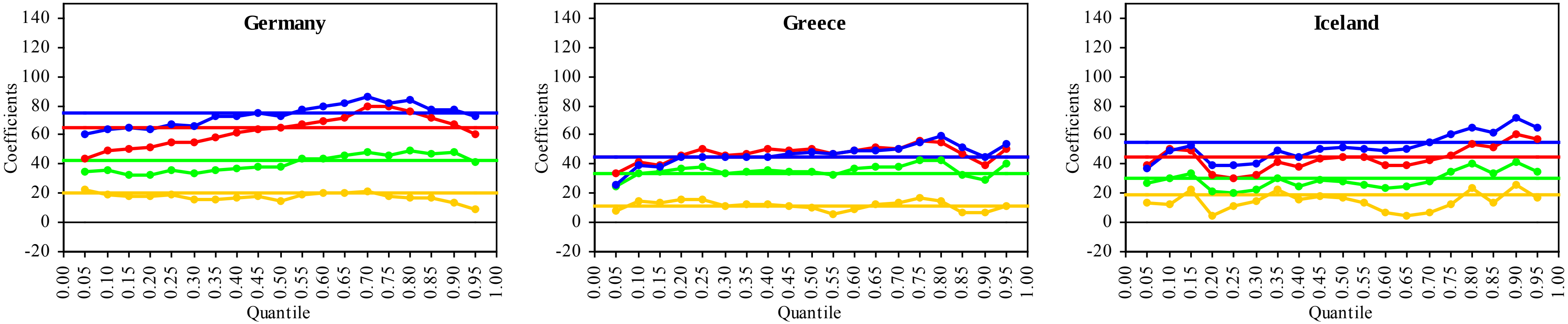
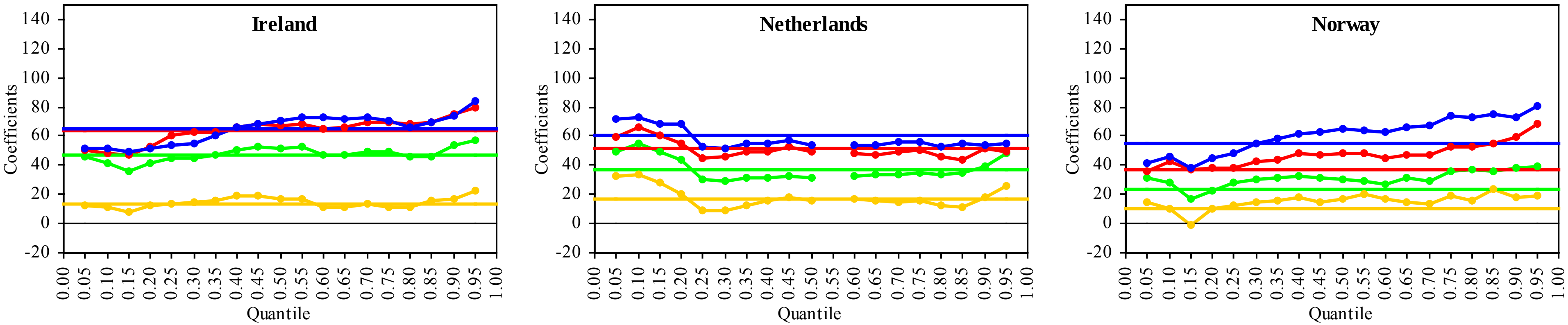


Figure 1 (cont.)



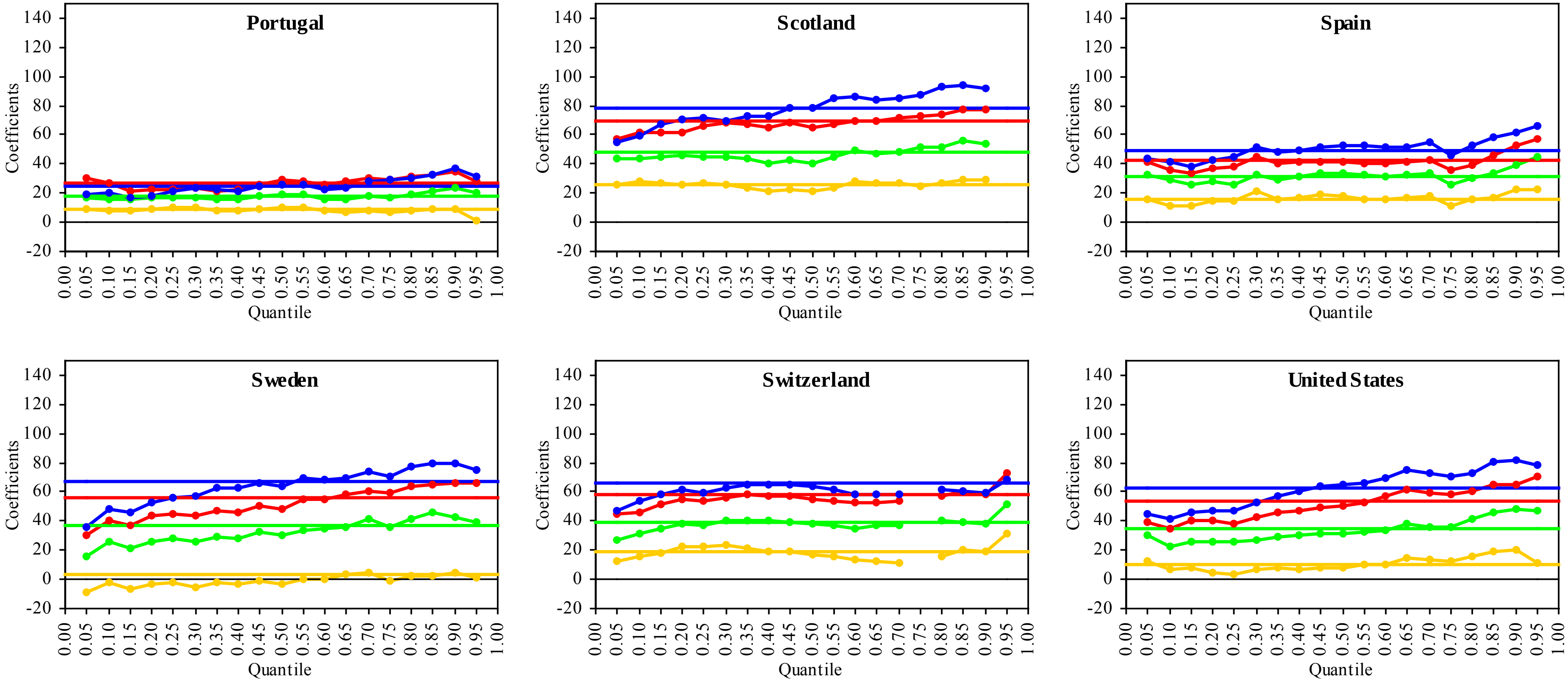
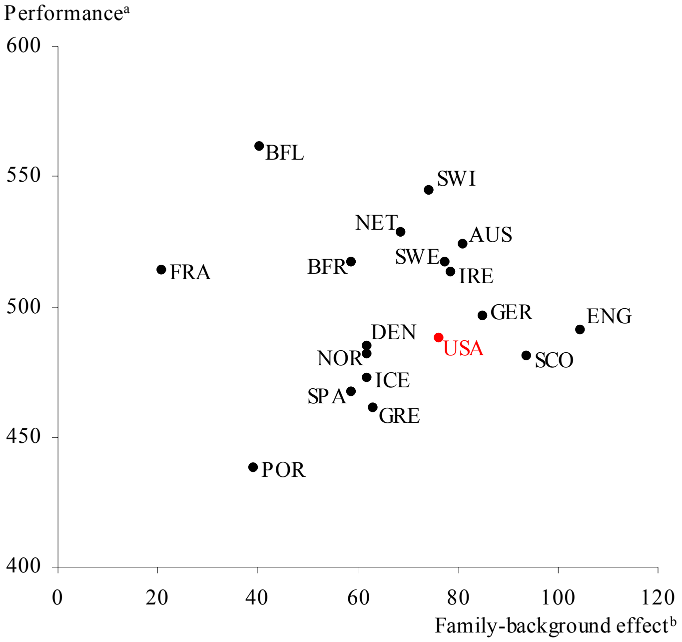

Portugal
Scotland
Spain
Sweden
Switzerland
United States
Coefficients
Quantile
140
120
100
80
60
40
20
0
-20
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00

Figure 2: Family-Background Effects and Mean Performance



Notes: [a] Mean performance on the TIMSS math test (cf. Table 1). – [b] Performance difference between students with more than two bookcases at home and students with less than one shelf of books at home, controlling for the variables reported in Table 3 except for parental education.

Appendix 1: Imputation of Missing Data within Each Country

In the background questionnaires, some students, teachers, and school principals failed to answer some questionnaire items. Table A1 reports the share of missing values for each variable in each country. For the analyses in this paper, data for missing responses in the TIMSS questionnaires were imputed within each country based on a method that relates the observations from students with original data to a set of "fundamental" explanatory variables available for all students.

Specifically, I first chose the following set of "fundamental" variables F with data available for virtually all students: grade level; student sex; student age; four dummies for the parents' education level; four dummies for the number of books in the student's home; and three dummies for the community location of the schools. The small amount of missing data within F was imputed by the median category observed at the lowest level available, i.e., either the class median, the school median, or the country median.

The variables in F were then used to impute missing data on each variable M for each student i within each country. Let S denote the set of students j with available data for M. Using the students in S, the variable M was regressed on F:

$$M_{j \in S} = F_{j \in S} \phi + \varepsilon_{j \in S} \tag{A1}$$

For M being a continuous variable, the regression model was a least-squares estimation, weighting each student by her sampling probability. For M being a dichotomous (binary) variable, an equivalently weighted probit model was used. For M being a polychotomous qualitative variable with multiple categories, a weighted ordered-probit model was estimated.

Finally, the coefficients ϕ from these regressions and the data on F_i were used to impute the value of M_i for the students with missing data:

$$\widetilde{M}_{i \notin S} = F_{i \notin S} \phi \tag{A2}$$

For the probit models, the estimated coefficients were used to forecast the probability of occurrence associated with each category for the students with missing data, and the category with the highest probability was imputed. This data imputation technique was applied within each country individually, resulting in a complete data set for all the students sampled in TIMSS.

Appendix 2: Appendix Tables

Table A1: Missing Values

Fraction of students with missing data, not weighted.

	AUS	BFL	BFR	DEN	ENG	FRA	GER	GRE	ICE
Family-background data									
Parents' education	0.22	0.33	0.35	0.42	1.00	0.45	0.29	0.17	0.23
Books at home	0.04	0.01	0.02	0.06	0.02	0.04	0.02	0.02	0.02
Living with both parents	0.07	0.02	0.02	0.06	0.02	0.05	0.03	0.04	0.05
Born in country	0.03	0.00	0.02	0.05	0.02	1.00	0.01	0.01	0.01
Sex	0.03	0.00	0.02	0.02	0.00	0.04	0.01	0.00	0.00
Age	0.03	0.00	0.03	0.02	0.00	0.07	0.03	0.00	0.00
Community location	0.13	0.10	0.17	0.19	0.15	0.11	0.40	0.03	0.10
Resource data									
Math class size	0.68	0.13	0.26	0.32	0.60	0.13	0.38	0.14	0.22
Science class size	0.27	0.23	0.29	0.48	0.63	0.17	0.44	0.11	0.29
Grade-average class size	0.35	0.10	0.21	0.23	1.00	0.12	0.43	0.06	0.18
Shortage of materials	0.15	0.08	0.19	0.23	0.18	0.10	0.41	0.05	0.09
Instruction time	0.12	0.07	0.00	0.19	0.22	0.40	0.55	1.00	0.25
Math teacher characteristics									
Teacher's sex	0.24	0.02	0.21	0.29	0.58	0.07	0.33	0.07	0.17
Teacher's experience	0.24	0.03	0.23	0.30	0.58	0.10	0.34	0.08	0.17
Teacher's education	1.00	1.00	1.00	1.00	0.58	0.09	0.05	0.08	0.18
Science teacher characteristics									
Teacher's sex	0.16	0.04	0.23	0.46	0.60	0.07	0.40	0.05	0.19
Teacher's experience	0.19	0.05	0.23	0.45	0.60	0.08	0.40	0.06	0.18
Teacher's education	1.00	1.00	1.00	1.00	0.60	0.08	0.08	0.05	0.18
Institutional data									
School responsibility for hiring teachers	0.16	0.07	0.23	0.30	1.00	0.11	0.41	0.03	0.11
Math teacher influence									
Subject matter	0.24	0.03	0.23	0.30	1.00	0.08	0.33	0.10	0.18
Specific textbooks	0.24	0.03	0.23	0.30	1.00	0.08	0.33	0.12	0.19
Supplies purchased	0.24	0.03	0.24	0.30	1.00	0.10	0.33	0.12	0.21
Math									
Tests/exams	0.25	0.03	0.22	1.00	0.59	0.09	0.33	0.08	0.23
Homework	0.25	0.13	0.24	0.30	0.58	0.12	0.36	0.09	0.26
Uninterested parents limit	0.24	0.13	0.25	0.29	0.59	0.10	0.36	0.10	0.22
Interested parents limit	0.25	0.13	0.26	0.30	0.59	1.00	0.36	0.11	0.21
Science teacher influence									
Subject matter	0.19	0.05	0.26	0.43	1.00	0.07	0.40	0.07	0.18
Specific textbooks	0.18	0.04	0.28	0.43	1.00	0.07	0.40	0.09	0.18
Supplies purchased	0.18	0.04	0.26	0.44	1.00	0.07	0.41	0.09	0.20
Science									
Tests/exams	0.20	0.05	0.24	1.00	0.61	0.07	0.40	0.04	0.22
Homework	1.00	0.25	0.29	0.46	0.60	0.17	0.44	0.08	0.25
Uninterested parents limit	0.18	0.25	0.30	0.44	0.61	0.17	0.43	0.05	0.25
Interested parents limit	0.17	0.26	0.32	0.44	0.60	1.00	0.43	0.08	0.26

Table A1 (cont.)

	IRE	NET	NOR	POR	SCO	SPA	SWE	SWI	USA
Family-background data									
Parents' education	0.15	0.29	0.35	0.10	0.49	0.16	0.38	0.22	0.10
Books at home	0.02	0.04	0.02	0.01	0.07	0.01	0.02	0.02	0.02
Living with both parents	0.03	0.04	0.01	0.02	0.08	0.03	0.03	0.02	0.03
Born in country	0.01	0.04	0.01	0.01	0.06	0.01	0.01	0.01	0.02
Sex	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Age	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Community location	0.04	0.19	0.14	0.06	0.18	0.06	0.08	0.13	0.15
Resource data									
Math class size	0.29	0.19	0.30	0.11	0.21	0.23	0.31	0.45	0.32
Science class size	0.52	0.23	0.41	0.03	0.47	0.22	0.44	0.51	0.57
Grade-average class size	0.06	0.24	0.21	0.07	0.19	0.12	0.11	0.16	0.20
Shortage of materials	0.07	0.23	0.15	0.07	1.00	0.07	0.09	0.13	0.15
Instruction time	0.10	0.43	0.12	0.22	0.21	0.35	0.23	0.28	0.34
Math teacher characteristics									
Teacher's sex	0.07	0.05	0.04	0.04	0.11	0.05	0.07	0.18	0.14
Teacher's experience	0.08	0.05	0.05	0.06	0.11	0.07	0.07	0.18	0.14
Teacher's education	0.07	0.08	0.16	0.05	0.11	0.06	0.08	0.18	0.14
Science teacher characteristics									
Teacher's sex	0.20	0.08	0.07	0.02	0.40	0.05	0.06	0.20	0.23
Teacher's experience	0.20	0.08	0.08	0.03	0.40	0.06	0.06	0.20	0.25
Teacher's education	0.20	0.09	0.21	0.05	0.41	0.06	0.08	0.20	0.23
Institutional data									
School responsibility for hiring teachers	0.07	0.24	1.00	0.07	1.00	0.12	0.08	0.16	0.21
Math teacher influence									
Subject matter	0.08	0.06	0.04	0.04	0.12	0.06	0.06	0.19	0.13
Specific textbooks	0.07	0.07	0.04	0.04	0.12	0.07	0.07	0.19	0.13
Supplies purchased	0.08	0.07	0.04	0.04	0.12	0.06	0.07	0.19	0.13
Math									
Tests/exams	0.08	0.06	0.07	0.06	0.12	0.06	0.09	0.18	0.13
Homework	0.21	0.20	0.31	0.04	0.25	0.19	0.30	0.45	0.28
Uninterested parents limit	0.22	0.18	0.28	0.07	0.19	0.19	0.29	0.45	0.29
Interested parents limit	0.24	0.17	0.28	0.07	0.18	0.19	0.29	0.45	0.29
Science teacher influence									
Subject matter	0.21	0.07	0.07	0.03	0.40	0.06	0.06	0.20	0.23
Specific textbooks	0.20	0.08	0.07	0.03	0.40	0.06	0.06	0.21	0.23
Supplies purchased	0.20	0.09	0.07	0.02	0.41	0.06	0.06	0.20	0.23
Science									
Tests/exams	0.21	0.08	0.10	0.03	0.42	0.06	0.08	0.21	0.23
Homework	0.40	0.21	0.40	0.04	0.44	0.21	0.41	0.51	0.52
Uninterested parents limit	0.40	0.20	0.39	0.02	0.46	0.21	0.42	0.50	0.53
Interested parents limit	0.41	0.20	0.38	0.02	0.45	0.22	0.41	0.49	0.53

Table A2a: Descriptive Statistics – Family Background

Country means. Standard deviations in parentheses for continuous variable (the std. dev. for dummies simply equals roughly 2*mean*(1–mean), neglecting the weighting). Only non-imputed data for each variable. Weighted by sampling probabilities.

	AUS	BFL	BFR	DEN	ENG	FRA	GER	GRE	ICE
Parents' education									
Primary	–	0.08	0.06	–	–	0.09	–	0.23	0.08
Some secondary	0.10	0.18	0.09	0.11	–	0.25	0.47	0.18	0.09
Finished secondary	0.78	0.40	0.20	0.35	–	0.33	0.34	0.17	0.12
Some after secondary	–	0.06	0.25	0.33	–	0.15	0.06	0.22	0.39
Finished university	0.13	0.29	0.40	0.21	–	0.18	0.14	0.20	0.32
Books at home									
Less than one shelf (<=10)	0.11	0.09	0.06	0.03	0.06	0.05	0.08	0.05	0.01
One shelf (11-25)	0.16	0.16	0.10	0.09	0.13	0.19	0.14	0.23	0.05
One bookcase (26-100)	0.32	0.34	0.29	0.29	0.27	0.36	0.28	0.41	0.30
Two bookcases (101-200)	0.18	0.19	0.21	0.23	0.23	0.20	0.19	0.19	0.28
More than two bookcases (>200)	0.24	0.22	0.35	0.36	0.32	0.20	0.32	0.12	0.36
Living with both parents	0.85	0.90	0.85	0.86	0.83	0.86	0.83	0.84	0.84
Born in country	0.94	0.97	0.91	0.94	0.95	–	0.90	0.94	0.93
Sex (female)	0.52	0.50	0.52	0.51	0.47	0.50	0.51	0.48	0.49
Age	13.81	13.64	13.79	13.43	13.57	13.81	14.27	13.12	13.14
	(0.71)	*(0.76)*	*(0.91)*	*(0.64)*	*(0.58)*	*(0.91)*	*(0.82)*	*(0.74)*	*(0.58)*
Upper grade	0.49	0.54	0.54	0.55	0.51	0.49	0.49	0.48	0.50
Above upper grade	–	–	–	–	–	–	–	–	–
Community location									
Geographically isolated	0.00	0.00	0.00	0.07	0.04	0.00	0.02	0.03	0.07
Close to town center	0.43	0.39	0.56	0.23	0.33	0.39	0.48	0.52	0.42

Table A2a (continued)

	IRE	NET	NOR	POR	SCO	SPA	SWE	SWI	USA
Parents' education									
Primary	0.07	0.04	–	0.57	–	0.43	–	–	0.02
Some secondary	0.21	0.09	0.12	0.21	0.23	0.16	0.14	0.15	0.06
Finished secondary	0.30	0.31	0.12	0.09	0.36	0.10	0.44	0.61	0.19
Some after secondary	0.23	0.39	0.40	0.05	0.18	0.12	0.10	0.11	0.38
Finished university	0.19	0.18	0.36	0.09	0.23	0.19	0.32	0.13	0.36
Books at home									
Less than one shelf (<=10)	0.07	0.07	0.02	0.12	0.11	0.05	0.03	0.07	0.08
One shelf (11-25)	0.16	0.15	0.06	0.27	0.17	0.18	0.08	0.15	0.12
One bookcase (26-100)	0.34	0.34	0.25	0.31	0.29	0.33	0.24	0.31	0.28
Two bookcases (101-200)	0.21	0.20	0.23	0.13	0.19	0.19	0.24	0.21	0.21
More than two bookcases (>200)	0.22	0.24	0.44	0.17	0.25	0.25	0.41	0.27	0.31
Living with both parents	0.90	0.91	0.85	0.88	0.81	0.89	0.86	0.84	0.79
Born in country	0.97	0.95	0.95	0.92	0.92	0.97	0.92	0.88	0.93
Sex (female)	0.52	0.50	0.50	0.50	0.49	0.50	0.49	0.50	0.50
Age	13.93	13.82	13.37	13.97	13.22	13.76	13.94	14.07	13.74
	(0.70)	*(0.80)*	*(0.60)*	*(1.11)*	*(0.60)*	*(0.86)*	*(0.88)*	*(1.01)*	*(0.72)*
Upper grade	0.50	0.52	0.50	0.48	0.50	0.50	0.33	0.36	0.50
Above upper grade	–	–	–	–	–	–	0.34	0.29	–
Community location									
Geographically isolated	0.06	0.00	0.12	0.01	0.05	0.01	0.04	0.04	0.03
Close to town center	0.35	0.22	0.12	0.62	0.44	0.23	0.46	0.23	0.44

Table A2b: Descriptive Statistics – Resources

Country means. Standard deviations in parentheses for continuous variables (the std. dev. for dummies simply equals roughly 2*mean*(1–mean), neglecting the weighting). Only non-imputed data for each variable. Weighted by sampling probabilities.

	AUS	BFL	BFR	DEN	ENG	FRA	GER	GRE	ICE
Math class size	10.60	20.32	20.61	19.94	25.59	25.38	24.64	27.44	21.68
	(6.01)	*(4.22)*	*(3.26)*	*(4.03)*	*(5.77)*	*(3.28)*	*(4.56)*	*(4.62)*	*(5.96)*
Science class size	24.73	20.64	20.93	19.70	24.69	24.99	25.84	28.31	21.83
	(3.77)	*(4.61)*	*(3.79)*	*(3.60)*	*(5.36)*	*(3.89)*	*(4.27)*	*(9.62)*	*(6.49)*
Grade-average class size	27.96	19.94	20.45	20.08	–	25.36	25.74	28.29	21.51
	(15.06)	*(3.82)*	*(3.06)*	*(3.46)*		*(2.57)*	*(4.10)*	*(7.64)*	*(6.33)*
No shortage of materials	0.72	0.76	0.23	0.09	0.23	0.39	0.41	0.43	0.15
A lot shortage of materials	0.01	0.00	0.17	0.35	0.11	0.18	0.07	0.24	0.09
Instruction time (in 100 hours	7.13	8.09	8.09	7.14	7.81	7.04	7.65	–	6.06
of 60 minutes per year)	*(0.13)*	*(0.00)*	*(0.00)*	*(0.13)*	*(0.52)*	*(1.51)*	*(0.77)*		*(0.98)*
Math teacher's sex (female)	0.52	0.62	0.56	0.40	0.46	0.48	0.35	0.31	0.54
Math teacher's experience	16.82	21.60	20.21	20.38	15.59	19.78	18.71	12.72	15.32
(in years)	*(7.92)*	*(9.08)*	*(8.72)*	*(7.98)*	*(8.72)*	*(10.30)*	*(8.27)*	*(6.36)*	*(10.15)*
Math teacher's education									
Less than secondary	–	–	–	–	0.00	0.01	0.00	–	0.34
Secondary only	–	–	–	–	0.36	0.34	0.00	–	0.09
BA or equivalent	–	–	–	–	0.54	0.40	0.00	0.98	0.56
MA/PhD	–	–	–	–	0.11	0.26	1.00	0.02	0.01
Science teacher's sex (female)	0.56	0.58	0.60	0.31	0.36	0.53	0.43	0.42	0.43
Science teacher's	15.90	19.25	17.70	18.37	15.18	19.74	19.23	12.89	13.22
experience (in years)	*(7.31)*	*(10.04)*	*(9.14)*	*(9.86)*	*(9.36)*	*(10.45)*	*(9.11)*	*(7.38)*	*(9.66)*
Science teacher's education									
Less than secondary	–	–	–	–	0.00	0.00	0.00	–	0.32
Secondary only	–	–	–	–	0.20	0.24	0.00	–	0.10
BA or equivalent	–	–	–	–	0.63	0.30	0.00	0.95	0.55
MA/PhD	–	–	–	–	0.17	0.46	1.00	0.05	0.03

Table A2b (continued)

	IRE	NET	NOR	POR	SCO	SPA	SWE	SWI	USA
Math class size	26.52	24.97	21.48	25.20	25.80	28.40	21.33	19.58	27.40
	(4.70)	*(4.64)*	*(6.29)*	*(4.74)*	*(4.15)*	*(8.85)*	*(5.69)*	*(4.74)*	*(15.64)*
Science class size	25.29	25.00	21.57	25.22	18.29	29.57	23.75	20.30	38.25
	(4.41)	*(4.75)*	*(5.75)*	*(4.17)*	*(5.91)*	*(9.65)*	*(6.62)*	*(3.98)*	*(34.86)*
Grade-average class size	26.52	25.18	21.55	25.80	26.16	28.27	24.69	19.95	25.62
	(3.56)	*(3.12)*	*(7.40)*	*(3.87)*	*(3.30)*	*(6.91)*	*(3.99)*	*(3.71)*	*(4.54)*
No shortage of materials	0.52	0.63	0.33	0.35	–	0.54	0.39	0.45	0.46
A lot shortage of materials	0.01	0.02	0.05	0.08	–	0.04	0.07	0.05	0.06
Instruction time (in 100 hours	7.76	8.61	6.70	8.10	8.62	7.13	7.17	9.71	7.68
of 60 minutes per year)	*(1.28)*	*(0.71)*	*(0.23)*	*(1.79)*	*(0.74)*	*(1.39)*	*(1.58)*	*(1.03)*	*(2.23)*
Math teacher's sex (female)	0.56	0.23	0.44	0.67	0.44	0.41	0.46	0.18	0.69
Math teacher's experience	15.81	15.13	18.15	7.94	15.53	21.39	18.77	18.33	15.08
(in years)	*(9.16)*	*(8.78)*	*(8.94)*	*(7.24)*	*(8.91)*	*(8.74)*	*(10.10)*	*(10.01)*	*(9.75)*
Math teacher's education									
Less than secondary	–	0.02	–	–	–	0.10	–	–	–
Secondary only	0.02	0.74	0.63	0.00	0.03	0.20	0.34	0.47	–
BA or equivalent	0.93	0.04	0.33	1.00	0.83	0.67	0.65	0.44	0.57
MA/PhD	0.06	0.21	0.04	–	0.14	0.04	0.02	0.09	0.43
Science teacher's sex (female)	0.52	0.18	0.44	0.77	0.35	0.44	0.46	0.17	0.57
Science teacher's	15.18	17.36	17.22	8.53	15.66	18.48	18.44	17.36	13.91
experience (in years)	*(9.06)*	*(9.06)*	*(9.01)*	*(8.37)*	*(8.22)*	*(8.95)*	*(10.34)*	*(10.19)*	*(9.84)*
Science teacher's education									
Less than secondary	–	0.02	–	–	–	0.08	–	–	–
Secondary only	0.03	0.52	0.60	0.00	0.04	0.22	0.32	0.47	–
BA or equivalent	0.89	0.06	0.35	1.00	0.80	0.66	0.66	0.42	0.58
MA/PhD	0.08	0.40	0.05	–	0.17	0.05	0.02	0.11	0.43

Table A2c: Descriptive Statistics – Institutional Features

Country means. Standard deviations in parentheses for continuous variables (the std. dev. for dummies simply equals roughly 2*mean*(1–mean), neglecting the weighting). Only non-imputed data for each variable. Weighted by sampling probabilities.

	AUS	BFL	BFR	DEN	ENG	FRA	GER	GRE	ICE
School responsibility for hiring teachers	0.12	0.97	0.96	0.89	–	0.23	0.05	0.07	1.00
Math teacher influence									
Subject matter to be taught	0.38	0.11	0.07	0.54	–	0.09	0.24	0.25	0.32
Specific textbooks to be used	0.30	0.50	0.27	0.34	–	0.37	0.29	0.09	0.28
What supplies are purchased	0.05	0.17	0.09	0.26	–	0.24	0.37	0.06	0.19
Math exam preparation/grading	2.40	3.92	3.77	–	2.14	4.29	3.16	2.44	1.79
(in hours per week)	*(1.57)*	*(1.65)*	*(1.66)*		*(1.71)*	*(1.43)*	*(1.71)*	*(1.54)*	*(1.25)*
Math homework assignment	1.35	0.80	1.24	0.83	0.82	1.54	1.35	3.35	1.86
(in hours per week)	*(0.87)*	*(0.78)*	*(0.94)*	*(0.71)*	*(0.37)*	*(0.69)*	*(0.74)*	*(1.59)*	*(0.95)*
Math teaching limited by									
Uninterested parents	0.03	0.04	0.18	0.05	0.03	0.13	0.03	0.09	0.18
Interested parents	0.00	0.02	0.01	0.01	0.01	–	0.00	0.06	0.26
Science teacher influence									
Subject matter to be taught	0.54	0.14	0.12	0.76	–	0.07	0.34	0.21	0.43
Specific textbooks to be used	0.40	0.61	0.27	0.49	–	0.49	0.38	0.10	0.38
What supplies are purchased	0.13	0.28	0.16	0.23	–	0.47	0.47	0.08	0.30
Science exam preparation/grading	1.55	3.53	3.26	–	2.16	4.03	2.95	2.88	1.61
(in hours per week)	*(1.22)*	*(1.63)*	*(1.54)*		*(1.62)*	*(1.52)*	*(1.77)*	*(1.63)*	*(1.24)*
Science homework assignment	–	0.18	0.35	0.19	0.69	0.47	0.47	1.41	0.50
(in hours per week)		*(0.25)*	*(0.51)*	*(0.27)*	*(0.37)*	*(0.40)*	*(0.57)*	*(1.25)*	*(0.45)*
Science teaching limited by									
Uninterested parents	0.03	0.04	0.15	0.06	0.04	0.09	0.02	0.05	0.11
Interested parents	0.00	0.02	0.02	0.00	0.01	–	0.00	0.01	0.15

Table A2c (continued)

	IRE	NET	NOR	POR	SCO	SPA	SWE	SWI	USA
School responsibility for hiring teachers	0.91	1.00	–	0.50	–	0.31	0.97	0.89	0.99
Math teacher influence									
Subject matter to be taught	0.25	0.25	0.35	0.20	0.13	0.62	0.51	0.47	0.41
Specific textbooks to be used	0.47	0.36	0.08	0.34	0.23	0.50	0.36	0.15	0.25
What supplies are purchased	0.19	0.11	0.13	0.13	0.15	0.26	0.24	0.18	0.25
Math exam preparation/grading	2.36	3.87	2.19	2.89	1.54	2.10	2.02	3.09	2.74
(in hours per week)	*(1.52)*	*(1.58)*	*(1.48)*	*(1.62)*	*(1.22)*	*(1.24)*	*(1.44)*	*(1.55)*	*(1.64)*
Math homework assignment	1.79	1.23	1.57	1.12	0.66	1.63	0.68	1.24	1.65
(in hours per week)	*(0.70)*	*(0.60)*	*(0.76)*	*(0.85)*	*(0.54)*	*(1.14)*	*(0.46)*	*(0.82)*	*(1.07)*
Math teaching limited by									
Uninterested parents	0.12	0.00	0.02	0.11	0.02	0.22	0.06	0.02	0.15
Interested parents	0.05	0.01	0.01	0.12	0.00	0.16	0.01	0.01	0.04
Science teacher influence									
Subject matter to be taught	0.26	0.46	0.40	0.27	0.15	0.62	0.56	0.54	0.48
Specific textbooks to be used	0.47	0.53	0.10	0.38	0.29	0.45	0.40	0.18	0.37
What supplies are purchased	0.35	0.34	0.15	0.11	0.31	0.23	0.30	0.20	0.41
Science exam preparation/grading	2.07	4.03	2.04	3.13	1.45	2.15	2.20	3.04	2.28
(in hours per week)	*(1.33)*	*(1.61)*	*(1.37)*	*(1.66)*	*(0.98)*	*(1.25)*	*(1.59)*	*(1.58)*	*(1.55)*
Science homework assignment	1.07	0.67	0.67	0.61	0.30	1.00	0.49	0.47	0.96
(in hours per week)	*(0.74)*	*(0.52)*	*(0.51)*	*(0.54)*	*(0.28)*	*(0.83)*	*(0.37)*	*(0.53)*	*(0.92)*
Science teaching limited by									
Uninterested parents	0.09	0.01	0.02	0.16	0.03	0.17	0.04	0.03	0.09
Interested parents	0.07	0.00	0.01	0.16	0.01	0.14	0.00	0.03	0.03

Table A3: Family Background and Science Performance

Least-squares regression within each country, weighted by students' sampling probabilities.
Dependent variable: TIMSS science test score. Clustering-robust standard errors in parentheses.

	AUS	BFL	BFR	DEN	ENG	FRA	GER	GRE	ICE
Parents' education									
Some secondary	–	-3.11	2.99	–	–	5.60	–	8.71^{+}	17.65^{+}
		(7.17)	*(11.01)*			*(5.91)*		*(3.55)*	*(8.27)*
Finished secondary	18.70*	7.10	15.35	10.23	–	12.39°	7.59°	19.85*	16.24°
	(5.95)	*(6.12)*	*(9.91)*	*(7.93)*		*(6.38)*	*(3.90)*	*(3.42)*	*(8.53)*
Some after secondary	–	28.68*	21.34^{+}	29.79*	–	21.30*	18.88*	19.48*	29.29*
		(8.68)	*(10.73)*	*(8.01)*		*(6.86)*	*(5.54)*	*(4.32)*	*(7.41)*
Finished university	30.50*	24.05*	29.42*	29.17*	–	26.84*	22.81*	37.39*	33.61*
	(7.45)	*(6.46)*	*(10.83)*	*(7.88)*		*(6.25)*	*(5.53)*	*(4.37)*	*(8.15)*
Books at home									
One shelf (11-25)	17.80*	14.85^{+}	19.37*	20.54^{+}	30.03*	8.95	32.28*	13.99*	-5.48
	(5.34)	*(5.92)*	*(5.86)*	*(8.89)*	*(6.30)*	*(5.47)*	*(5.71)*	*(4.40)*	*(15.91)*
One bookcase (26-100)	38.07*	24.08*	37.04*	34.82*	63.98*	24.64*	56.43*	34.94*	12.82
	(5.48)	*(5.62)*	*(5.81)*	*(7.44)*	*(6.45)*	*(4.88)*	*(5.30)*	*(4.69)*	*(14.14)*
Two bookcases (101-200)	59.14*	35.01*	50.94*	51.75*	91.62*	33.23*	78.56*	42.92*	22.93
	(5.58)	*(5.07)*	*(5.54)*	*(7.91)*	*(6.91)*	*(5.36)*	*(6.70)*	*(5.39)*	*(15.28)*
More than two bookcases	69.18*	34.66*	55.32*	61.86*	121.06*	29.40*	90.76*	44.35*	37.13^{+}
(>200)	*(5.95)*	*(5.57)*	*(5.12)*	*(7.44)*	*(7.60)*	*(5.27)*	*(5.61)*	*(5.23)*	*(14.39)*
Living with both parents	-3.61	-1.47	-7.89^{+}	-2.12	0.75	4.94°	1.94	6.91*	7.34°
	(3.66)	*(3.54)*	*(3.19)*	*(4.05)*	*(4.63)*	*(2.97)*	*(3.73)*	*(2.60)*	*(3.77)*
Born in country	44.14*	2.57	8.60	25.41*	22.17^{+}	–	33.02*	5.88	6.91
	(6.88)	*(8.02)*	*(5.78)*	*(7.95)*	*(8.73)*		*(5.22)*	*(4.67)*	*(5.49)*
Female	-17.05*	-16.49*	-19.74*	-27.73*	-19.89*	-19.69*	-20.34*	-15.02*	-15.10*
	(3.46)	*(4.45)*	*(3.23)*	*(2.53)*	*(5.04)*	*(2.37)*	*(3.49)*	*(2.35)*	*(2.84)*
Age	-18.97*	-18.63*	-17.13*	-15.04*	19.33*	-16.41*	-15.86*	-10.02*	7.43°
	(2.92)	*(2.93)*	*(1.90)*	*(3.24)*	*(5.25)*	*(1.77)*	*(2.73)*	*(3.64)*	*(4.48)*
Upper grade	57.24*	46.26*	48.94*	55.15*	21.66*	60.68*	46.93*	56.24*	23.19*
	(4.51)	*(3.91)*	*(3.52)*	*(4.40)*	*(5.45)*	*(2.89)*	*(4.29)*	*(4.61)*	*(7.29)*
Above upper grade	–	–	–	–	–	–	–	–	–
Community location									
Geographically isolated	47.91*	–	–	-12.44	-9.77	–	10.52	-13.38°	-15.48^{+}
	(5.40)			*(9.47)*	*(8.54)*		*(26.34)*	*(7.86)*	*(6.77)*
Close to town center	14.76^{+}	-6.42	7.45	-10.60^{+}	-12.15	2.81	6.84	11.98*	-2.69
	(6.58)	*(5.26)*	*(5.07)*	*(5.21)*	*(7.70)*	*(3.70)*	*(8.73)*	*(4.27)*	*(5.10)*
Imputation controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Students [Unit of observation]	5786	5662	4883	4370	3579	6014	5763	7921	3730
Schools [Unit of clustering]	129	141	120	153	127	134	137	156	155
R^2	0.175	0.113	0.183	0.155	0.178	0.183	0.210	0.180	0.114
R^2 (without imputation controls)	0.162	0.108	0.170	0.133	0.172	0.171	0.181	0.156	0.092

Significance levels (based on clustering-robust standard errors): * 1%. – $^{+}$ 5%. – ° 10%.

Table A3 (continued)

	IRE	NET	NOR	POR	SCO	SPA	SWE	SWI	USA
Parents' education									
Some secondary	7.84	26.82*	–	7.17*	–	10.60*	–	–	-9.81
	(5.70)	*(8.70)*		*(2.44)*		*(2.87)*			*(10.16)*
Finished secondary	11.82+	31.63*	5.80	12.94*	-4.73	19.57*	13.69*	21.65*	5.22
	(5.51)	*(9.51)*	*(6.45)*	*(4.27)*	*(5.36)*	*(3.74)*	*(4.99)*	*(3.12)*	*(10.18)*
Some after secondary	22.41*	49.55*	21.69*	23.73*	12.30+	9.63*	13.93+	22.85*	18.55°
	(5.22)	*(8.43)*	*(5.39)*	*(4.87)*	*(5.37)*	*(3.53)*	*(5.55)*	*(4.52)*	*(9.68)*
Finished university	30.81*	42.27*	26.60*	36.18*	31.25*	19.17*	29.44*	38.55*	34.47*
	(5.67)	*(9.64)*	*(6.03)*	*(4.46)*	*(5.76)*	*(3.93)*	*(5.10)*	*(4.58)*	*(10.08)*
Books at home									
One shelf (11-25)	24.35*	6.06	20.25+	6.53°	30.64*	12.23+	17.13+	19.75*	15.78*
	(5.12)	*(6.05)*	*(10.03)*	*(3.31)*	*(5.05)*	*(4.91)*	*(6.79)*	*(5.27)*	*(4.82)*
One bookcase (26-100)	51.87*	22.67*	34.54*	14.96*	50.76*	28.58*	43.75*	39.45*	48.24*
	(5.27)	*(5.61)*	*(9.35)*	*(3.30)*	*(4.34)*	*(4.68)*	*(6.39)*	*(4.54)*	*(4.81)*
Two bookcases (101-200)	71.79*	40.33*	51.32*	21.78*	82.09*	38.80*	66.53*	55.97*	67.22*
	(5.14)	*(6.57)*	*(9.03)*	*(3.49)*	*(5.27)*	*(4.98)*	*(6.72)*	*(4.84)*	*(5.16)*
More than two bookcases	74.97*	46.82*	66.05*	30.70*	93.53*	45.33*	78.55*	69.22*	79.10*
(>200)	*(5.94)*	*(6.36)*	*(9.10)*	*(3.99)*	*(5.10)*	*(5.07)*	*(6.30)*	*(5.05)*	*(6.01)*
Living with both parents	4.69	-3.23	6.98°	-1.95	6.68+	0.23	2.16	-3.92	14.37*
	(3.36)	*(4.83)*	*(3.60)*	*(2.60)*	*(3.30)*	*(3.21)*	*(3.36)*	*(2.48)*	*(3.51)*
Born in country	-16.06*	23.94*	36.11*	-2.90	-9.03°	12.50+	41.81*	34.64*	17.61*
	(5.39)	*(6.07)*	*(5.99)*	*(3.89)*	*(5.12)*	*(5.63)*	*(4.16)*	*(3.63)*	*(4.97)*
Female	-19.10*	-18.68*	-14.77*	-19.92*	-20.40*	-21.64*	-11.02*	-20.47*	-15.14*
	(4.02)	*(2.98)*	*(2.56)*	*(1.79)*	*(2.49)*	*(1.96)*	*(1.97)*	*(2.10)*	*(2.68)*
Age	-16.55*	-17.98*	4.72	-11.68*	12.60*	-15.60*	0.73	-10.03*	-17.24*
	(2.42)	*(3.25)*	*(3.73)*	*(1.13)*	*(4.13)*	*(1.78)*	*(2.93)*	*(3.06)*	*(2.92)*
Upper grade	58.96*	65.69*	39.59*	62.00*	36.45*	54.49*	43.23*	50.53*	43.01*
	(5.08)	*(5.53)*	*(4.70)*	*(2.27)*	*(5.56)*	*(2.64)*	*(4.39)*	*(4.60)*	*(4.92)*
Above upper grade	–	–	–	–	–	–	74.97*	100.93*	–
							(7.00)	*(8.00)*	
Community location									
Geographically isolated	-0.67	–	16.44*	-13.90*	15.60°	-37.70*	-5.03	-8.93	-40.68*
	(13.34)		*(5.04)*	*(3.92)*	*(9.16)*	*(1.92)*	*(8.23)*	*(10.68)*	*(9.11)*
Close to town center	2.86	13.58°	1.63	3.35	-8.06	2.86	8.83+	-5.22	-13.79+
	(5.30)	*(8.13)*	*(5.13)*	*(3.03)*	*(5.08)*	*(4.34)*	*(3.59)*	*(6.40)*	*(6.07)*
Imputation controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Students [Unit of observation]	6203	4084	5736	6753	5776	7596	8855	11722	10973
Schools [Unit of clustering]	132	95	249	142	128	154	270	327	183
R^2	0.178	0.203	0.140	0.216	0.231	0.165	0.236	0.255	0.174
R^2 (without imputation controls)	0.160	0.185	0.127	0.215	0.217	0.158	0.225	0.247	0.161

Significance levels (based on clustering-robust standard errors): * 1%. – + 5%. – ° 10%.

Table A4: Family-Background Effects in Europe vs. the United States in Science

Least-squares regression within each country, weighted by students' sampling probabilities.
Dependent variable: TIMSS science test score. Clustering-robust standard errors in parentheses.

	All countries	Interaction with Europe dummy
Parents' education		
Some secondary	-4.79	10.29
	(9.26)	*(9.35)*
Finished secondary	7.16	4.49
	(8.81)	*(8.88)*
Some after secondary	20.01^{+}	1.41
	(8.62)	*(8.73)*
Finished university	36.79*	-9.61
	(9.17)	*(9.28)*
Books at home		
One shelf (11-25)	15.03*	5.09
	(4.76)	*(4.98)*
One bookcase (26-100)	49.17*	-8.94°
	(4.86)	*(5.08)*
Two bookcases (101-200)	67.25*	-9.83°
	(5.04)	*(5.26)*
More than two bookcases (>200)	79.62*	-11.59°
	(5.79)	*(5.99)*
Living with both parents	14.78*	-13.16*
	(3.55)	*(3.66)*
Born in country	18.57*	0.96
	(5.17)	*(5.40)*
Female	-15.13*	-3.10
	(2.64)	*(2.76)*
Age	-17.49*	5.62°
	(2.94)	*(3.02)*
Upper grade	43.31*	8.72°
	(4.90)	*(5.02)*
Above upper grade	102.21*	–
	(2.96)	
Community location		
Geographically isolated	-40.19*	40.05*
	(8.96)	*(9.62)*
Close to town center	-18.15*	19.55*
	(5.98)	*(6.15)*
Imputation controls	yes	
Country dummies	yes	
Students [Unit of observation]	115406	
Schools [Unit of clustering]	2932	
R^2	0.243	
R^2 (without imputation controls)	0.235	

Significance levels (based on clustering-robust standard errors): * 1%. – $^{+}$ 5%. – ° 10%.

Table A5: Resources, Teacher Characteristics, and Science Performance

Least-squares regression within each country, weighted by students' sampling probabilities.
Dependent variable: TIMSS science test score. Clustering-robust standard errors in parentheses.

	AUS	BFL	BFR	DEN	ENG	FRA	GER	GRE	ICE
Class size (log)	10.31	20.10°	8.28	9.19	32.92°	9.45	56.85*	9.90+	-5.09
	(17.10)	*(12.03)*	*(12.40)*	*(9.90)*	*(17.72)*	*(9.82)*	*(21.31)*	*(4.49)*	*(4.90)*
Shortage of materials									
None	-7.63	2.96	9.26	1.30	21.34+	1.33	8.92	-7.56+	-6.90
	(8.56)	*(6.02)*	*(5.79)*	*(6.32)*	*(10.52)*	*(3.95)*	*(9.84)*	*(3.61)*	*(5.87)*
A lot	-42.69*	–	-3.49	-1.70	-10.95	-2.52	21.79+	1.44	0.27
	(10.01)		*(5.48)*	*(4.62)*	*(9.58)*	*(4.05)*	*(10.02)*	*(4.11)*	*(6.15)*
Instruction time	–	–	–	–	-2.72	-1.52	-4.62	–	-4.22+
					(6.31)	*(1.88)*	*(5.45)*		*(1.67)*
Teacher characteristics									
Female teacher	-6.13	0.82	0.89	-0.61	0.33	4.21	10.30	9.03*	9.46+
	(5.56)	*(4.34)*	*(3.59)*	*(4.62)*	*(6.99)*	*(3.48)*	*(7.48)*	*(3.10)*	*(4.70)*
Teacher's experience (log)	8.71+	0.25	-0.28	3.26	2.79	1.59	-3.01	7.15*	3.14
	(4.01)	*(2.24)*	*(2.87)*	*(2.24)*	*(3.75)*	*(2.25)*	*(5.72)*	*(1.80)*	*(2.68)*
Teacher's education									
Secondary only	–	–	–	–	–	–	–	–	1.62
									(6.03)
BA or equivalent	–	–	–	–	6.64	-1.42	–	–	-1.10
					(8.78)	*(3.82)*			*(4.85)*
MA/PhD	–	–	–	–	16.42	1.88	–	18.30*	1.15
					(10.07)	*(4.44)*		*(5.99)*	*(9.30)*
Family-background controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Imputation controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Students [Unit of observation]	5786	5662	4883	4370	3579	6014	5721	7921	3730
Schools [Unit of clustering]	129	141	120	153	127	134	137	156	155
R^2	0.186	0.131	0.197	0.160	0.198	0.189	0.227	0.194	0.133
R^2 (without imputation controls)	0.168	0.110	0.172	0.134	0.186	0.173	0.191	0.164	0.097

Significance levels (based on clustering-robust standard errors): * 1%. – + 5%. – ° 10%.

Table A5 (cont.)

	IRE	NET	NOR	POR	SCO	SPA	SWE	SWI	USA
Class size (log)	99.95*	84.53*	6.34	-1.38	6.13	5.60	1.07	5.04	3.99
	(17.76)	*(15.84)*	*(6.62)*	*(6.90)*	*(11.42)*	*(4.62)*	*(6.90)*	*(12.26)*	*(6.59)*
Shortage of materials									
None	11.68^{+}	11.52°	0.76	2.57	–	2.42	-0.14	2.24	-3.80
	(5.14)	*(6.51)*	*(3.70)*	*(3.02)*		*(3.36)*	*(3.57)*	*(4.26)*	*(5.65)*
A lot	1.26	2.70	-12.94°	-3.66	–	6.80	-3.08	-18.21^{+}	-27.76*
	(6.34)	*(6.51)*	*(7.22)*	*(5.60)*		*(10.69)*	*(8.36)*	*(8.86)*	*(10.03)*
Instruction time	2.15	5.17	–	-0.38	-1.90	0.28	2.04^{+}	3.43°	-0.51
	(1.95)	*(4.82)*		*(0.86)*	*(1.82)*	*(0.97)*	*(0.91)*	*(1.93)*	*(1.41)*
Teacher characteristics									
Female teacher	7.04	6.86	2.42	1.72	-2.62	4.38	2.40	-1.15	2.88
	(5.02)	*(6.88)*	*(3.33)*	*(3.02)*	*(4.05)*	*(3.14)*	*(2.99)*	*(4.13)*	*(5.19)*
Teacher's experience (log)	-3.89	2.17	1.79	3.82*	4.52°	-2.32	2.77°	2.86	1.72
	(3.00)	*(2.48)*	*(1.77)*	*(1.33)*	*(2.51)*	*(2.33)*	*(1.62)*	*(2.45)*	*(3.03)*
Teacher's education									
Secondary only	–	-20.18*	–	–	–	-9.73°	–	–	–
		(7.28)				*(5.42)*			
BA or equivalent	-9.80	-40.66*	-4.57	–	-4.57	-11.04^{+}	0.10	40.16*	–
	(11.13)	*(9.68)*	*(3.80)*		*(13.63)*	*(4.80)*	*(4.54)*	*(3.99)*	
MA/PhD	-2.12	5.44	-10.52°	–	-6.06	-1.59	7.95	44.05*	-6.63
	(12.81)	*(8.25)*	*(5.90)*		*(14.18)*	*(8.92)*	*(11.29)*	*(5.09)*	*(5.34)*
Family-background controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Imputation controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Students [Unit of observation]	6203	4084	5736	6753	5776	7596	8855	11722	10973
Schools [Unit of clustering]	132	95	249	142	128	154	270	327	183
R^2	0.2078	0.2795	0.1468	0.2219	0.2363	0.1701	0.2404	0.2982	0.1819
R^2 (without imputation controls)	0.182	0.251	0.129	0.217	0.218	0.161	0.227	0.285	0.166

Significance levels (based on clustering-robust standard errors): * 1%. – $^{+}$ 5%. – ° 10%.

Table A6: The Coefficient on Log Class Size in Science

Regressions within each country, weighted by students' sampling probabilities. Dependent variable: TIMSS science test score. Controlling for family-background variables and imputation controls. Clustering-robust standard errors in parentheses.

	AUS	BFL	BFR	DEN	ENG	FRA	GER	GRE	ICE
LS	7.20	19.13	11.53	5.37	44.96*	8.38	67.19*	9.39^{+}	-9.65°
	(21.34)	*(12.37)*	*(12.86)*	*(10.70)*	*(17.05)*	*(9.84)*	*(24.56)*	*(4.70)*	*(5.79)*
IV	-36.46	215.73°	46.88^{+}	37.80	–	4.33	358.33	-6.28	-13.74^{+}
	(98.83)	*(111.78)*	*(22.93)*	*(29.52)*		*(28.69)*	*(711.10)*	*(25.69)*	*(5.77)*
SFE	30.47	9.16	0.29	-7.65	89.51*	-5.06	-22.87	-5.25	-9.32
	(18.43)	*(16.13)*	*(12.81)*	*(18.24)*	*(15.95)*	*(15.24)*	*(23.60)*	*(7.99)*	*(12.36)*
SFE-IV	479.12	125.09	-14.74	27.76	–	-42.51	-68.94	-61.01	-8.56
	(1066.82)	*(144.20)*	*(29.39)*	*(29.21)*		*(40.46)*	*(109.65)*	*(40.10)*	*(26.80)*
F (instr. 1st stage)	0.22	2.08	14.75	46.85	–	14.74	6.56	5.23	17.95
F ($\beta_2 = -100$)	0.29	2.44	8.42	19.14	–	2.02	0.08	0.95	11.64
Prob. > *F*	*(0.588)*	*(0.121)*	*(0.004)*	*(0.000)*		*(0.158)*	*(0.777)*	*(0.332)*	*(0.001)*
Students	5599	5649	4866	3996	3270	5669	5620	7921	3119
Schools	120	140	119	128	116	119	130	156	118

Methods of estimation: LS = Least squares. – IV = Instrumental variables. – SFE = School fixed effects. – SFE-IV = Combination of school fixed effects and instrumental variables. See text for details on the four methods of estimation.

Significance levels (based on clustering-robust standard errors): * 1%. – $^{+}$ 5%. – ° 10%.

Table A6 (cont.)

	IRE	NET	NOR	POR	SCO	SPA	SWE	SWI	USA
LS	100.33*	82.12*	1.41	1.31	-4.91	3.22	-0.17	32.11	7.48
	(18.27)	*(14.32)*	*(10.51)*	*(7.02)*	*(17.86)*	*(5.13)*	*(7.16)*	*(19.60)*	*(5.76)*
IV	383.69	162.78*	38.03^{+}	-23.63^{+}	31.13	-21.43	28.26	80.23	10.17
	(346.63)	*(57.37)*	*(17.78)*	*(11.78)*	*(141.13)*	*(108.01)*	*(19.04)*	*(94.25)*	*(25.15)*
SFE	117.59*	65.28*	-23.23	2.09	-6.84	1.47	–	-8.62	9.73
	(24.32)	*(18.34)*	*(14.50)*	*(9.03)*	*(16.05)*	*(7.23)*		*(21.63)*	*(6.00)*
SFE-IV	-59.45	59.74^{+}	44.75	-13.22	-75.66	58.54	–	-19.84	340.33
	(93.11)	*(29.24)*	*(45.23)*	*(32.11)*	*(160.33)*	*(329.42)*		*(31.69)*	*(2410.34)*
F (instr. 1st stage)	10.59	17.60	42.52	1.19	2.01	23.20	–	3.72	2.07
F ($\beta_2 = -100$)	0.19	29.84	10.24	7.30	0.02	0.23	–	6.40	0.03
Prob. > *F*	*(0.664)*	*(0.000)*	*(0.003)*	*(0.008)*	*(0.880)*	*(0.631)*		*(0.013)*	*(0.855)*
Students	6152	4025	1351	6721	5667	7556	8855	6159	10831
Schools	129	92	38	141	124	152	270	131	179

Methods of estimation: LS = Least squares. – IV = Instrumental variables. – SFE = School fixed effects. – SFE-IV = Combination of school fixed effects and instrumental variables. See text for details on the four methods of estimation.

Significance levels (based on clustering-robust standard errors): * 1%. – $^{+}$ 5%. – ° 10%.

Table A7: Institutions and Science Performance

Least-squares regression within each country, weighted by students' sampling probabilities.
Dependent variable: TIMSS science test score. Clustering-robust standard errors in parentheses.

	AUS	BFL	BFR	DEN	ENG	FRA	GER	GRE	ICE
School responsibility	7.50	13.68	11.51	-1.60	–	10.20^{+}	1.06	-2.89	–
for hiring teachers	*(8.67)*	*(8.28)*	*(7.86)*	*(7.66)*		*(5.00)*	*(20.61)*	*(6.64)*	
Teacher influence									
Subject matter	9.20	-6.90	1.14	-4.47	–	-3.46	-15.90*	4.14	3.49
to be taught	*(5.67)*	*(8.86)*	*(5.35)*	*(4.45)*		*(5.37)*	*(5.51)*	*(4.26)*	*(5.59)*
Specific textbooks	-10.39^{+}	6.11	10.69^{+}	1.46	–	0.26	-15.79^{+}	2.09	1.13
to be used	*(5.25)*	*(4.85)*	*(4.84)*	*(3.70)*		*(3.79)*	*(6.93)*	*(7.69)*	*(5.37)*
What supplies	-5.76	4.40	2.03	1.69	–	3.32	1.84	-2.07	-0.53
are purchased	*(7.00)*	*(5.56)*	*(6.90)*	*(4.93)*		*(3.43)*	*(5.92)*	*(5.24)*	*(4.91)*
Testing and homework									
Tests/exams	-3.57^{+}	-0.34	-1.84	–	0.39	-0.60	2.40	-0.15	1.11
	(1.76)	*(1.23)*	*(1.16)*		*(1.87)*	*(1.15)*	*(1.63)*	*(0.95)*	*(1.86)*
Homework	–	0.19	-5.44^{+}	-2.40	28.55*	2.05	13.53°	-0.16	-8.32
		(6.97)	*(2.36)*	*(6.80)*	*(8.23)*	*(3.01)*	*(7.39)*	*(1.09)*	*(5.51)*
Teaching limited by									
Uninterested parents	1.44	-16.99	-13.67^{+}	-9.20	7.80	-4.97	-65.46*	-7.58	-7.14
	(20.94)	*(14.24)*	*(6.28)*	*(10.65)*	*(15.15)*	*(6.47)*	*(15.05)*	*(5.23)*	*(6.78)*
Interested parents	23.43°	-12.43	16.32	–	18.13	–	–	9.42	6.91
	(11.87)	*(27.33)*	*(13.45)*		*(23.85)*			*(9.88)*	*(5.05)*
Family-background controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Resource controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Imputation controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Students [Unit of observation]	5786	5662	4883	4370	3579	6014	5721	7921	3730
Schools [Unit of clustering]	129	141	120	153	127	134	137	156	155
R^2	0.208	0.142	0.207	0.162	0.205	0.196	0.250	0.202	0.145
R^2 (without imputation controls)	0.173	0.115	0.178	0.136	0.191	0.175	0.212	0.165	0.103

Significance levels (based on clustering-robust standard errors): * 1%. – $^{+}$ 5%. – $^{\circ}$ 10%.

Table A7 (cont.)

	IRE	NET	NOR	POR	SCO	SPA	SWE	SWI	USA
School responsibility	15.17°	–	–	-0.93	–	3.02	19.42*	-3.40	84.87*
for hiring teachers	*(8.13)*			*(2.46)*		*(4.47)*	*(5.27)*	*(5.59)*	*(29.09)*
Teacher influence									
Subject matter	-3.69	-3.75	1.11	-2.83	-4.75	6.86⁺	-8.46⁺	-6.35°	2.08
to be taught	*(4.69)*	*(4.94)*	*(3.23)*	*(2.67)*	*(6.92)*	*(3.39)*	*(3.40)*	*(3.69)*	*(5.56)*
Specific textbooks	5.04	-5.11	6.21	3.41	-1.24	-0.23	0.32	-2.53	20.12*
to be used	*(5.00)*	*(6.29)*	*(5.88)*	*(2.66)*	*(4.75)*	*(3.34)*	*(3.17)*	*(4.37)*	*(5.24)*
What supplies	-4.06	9.18	6.91	0.86	4.30	1.33	0.26	5.30	-0.40
are purchased	*(4.07)*	*(5.60)*	*(4.36)*	*(3.96)*	*(5.24)*	*(3.56)*	*(3.63)*	*(4.13)*	*(5.02)*
Testing and homework									
Tests/exams	-1.18	0.58	1.65	-0.39	-0.31	-0.25	0.53	1.99°	-0.17
	(1.39)	*(1.77)*	*(1.09)*	*(0.75)*	*(2.21)*	*(1.12)*	*(1.06)*	*(1.06)*	*(1.68)*
Homework	8.46⁺	-9.99°	1.15	-3.51	10.77	-1.45	4.67	2.16	4.02
	(3.39)	*(5.51)*	*(3.33)*	*(2.62)*	*(11.59)*	*(1.76)*	*(4.38)*	*(4.76)*	*(4.20)*
Teaching limited by									
Uninterested parents	-27.08⁺	4.23	-0.60	-3.11	-2.27	-3.25	-12.85	-20.71°	-6.33
	(12.34)	*(12.31)*	*(12.16)*	*(5.00)*	*(11.09)*	*(4.70)*	*(9.04)*	*(12.23)*	*(11.40)*
Interested parents	6.23	–	-28.41*	-3.82	-11.69	-7.54°	-15.45⁺	-1.71	35.39
	(12.21)		*(6.70)*	*(5.34)*	*(9.10)*	*(4.43)*	*(6.77)*	*(8.02)*	*(22.90)*
Family-background controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Resource controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Imputation controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Students [Unit of observation]	6203	4084	5736	6753	5776	7596	8855	11722	10973
Schools [Unit of clustering]	132	95	249	142	128	154	270	327	183
R^2	0.221	0.288	0.151	0.228	0.241	0.176	0.246	0.303	0.201
R^2 (without imputation controls)	0.193	0.258	0.132	0.219	0.219	0.165	0.230	0.288	0.178

Significance levels (based on clustering-robust standard errors): * 1%. – ⁺ 5%. – ° 10%.