Characteristics of teaching institutions and students' performance:

new empirical evidence from OECD data

Stephane R. Robin (corresponding author)
Researcher
CRESGE
1 rue Norbert Ségard
BP 109
59016 LILLE Cedex
FRANCE
Tel: +33.3.20.13.40.60.
Fax: +33.3.20.13.40.70.
E-mail: srobin@cresge.fr

Maresa Sprietsma
Research Assistant
IRES-UCL
Place Montesquieu 3
Boîte 1
1348 Louvain-la-Neuve
BELGIUM
Tel: +32.10.47.39.66.
Fax: +32.10.47.39.45
E-mail: sprietsma@ires.ucl.ac.be

Abstract:

A whole branch of the economic literature suggests that institutional differences between and inside educational systems may have a larger influence on students performance than the amount of resources devoted to schooling. In this paper, we use the PISA 2000 international OECD data to evaluate the impacts of organizational and institutional factors on students performance. We estimate an education production function with country fixed-effect and school random-effect.We find that, alongside individual characteristics, school autonomy in decisions regarding the recruitment of new personnel as well as pedagogical training strongly affect students performance. On the contrary, measures of school resources and standardised evaluation of students have no consistent effect.

Keywords: human capital formation, individual performance, school resources, school autonomy, institutional arrangements.

JEL codes: I2, I22, I28, H52, J24.

Over the last decades, multiple attempts have been made to estimate an *education production function*, linking various socio-economic factors to an indicator of students' performance. This empirical literature, however, has led to ambiguous results. In particular, no mechanical relationship between the amount of resources fed into the educational system and pupils' performances has been identified (e.g. Coleman, 1966; Hanushek, 2002; Woessmann, 2003). In this context, some authors suggested to pay more attention to the organizational aspects of the schooling process, and in particular to the specific incentive structures which prevail within schools (Hanushek 1997, Bishop *et al.* 2000).

The aim of the present paper is to provide an accurate evaluation of the impact of school organizational characteristics on pupils' results. Using recent international OECD data (which provides detailed information on both pupils and schools), we estimate an education production function, controlling for the differences in institutional and organizational arrangements that exist between countries and between schools. In a first section, we briefly summarize the literature on the education production function. In a second section, we present our data and explain our choice of explanatory variables. Section 3 is devoted to econometric modelling, with an emphasis put on the estimation strategy. Our main results are presented in Section 4. Our conclusions are given in a final section.

1. School characteristics as inputs in the education production function.

After Becker (1962) introduced the concept of human capital, economists began to regard education as a production process. In particular, schooling (initial training) came to be seen as the process by which individuals acquire their initial amount of human capital, prior to their entry on the labour market. Theory predicts that individuals with a higher human capital will obtain higher wages (or, loosely speaking, "better jobs"). At the macroeconomic level, the stock (and quality) of human capital in a country may influence its long-term economic growth (e.g. Lucas, 1988; Becker *et al.*, 1990; Mankiw *et al.*, 1992).

In that perspective, understanding the reasons why some pupils perform better than others (and are thus able to undergo longer schooling) has become of critical importance to economists over the last decades. Drawing a parallel between the process of human capital acquisition and a firm's production process, the concept of an education production function has become the primary tool to investigate this topic. A whole branch of the literature has been dedicated to the estimation of the education production function, defined as an efficient technology (Cooper & Cohn, 1997) turning a vector of inputs X into a vector of outputs Y:

(1) $Y = f(X)$

In Equation (1) above, X may include such inputs as a child's abilities, his/her family background, and/or educational resources, and where Y may be a set of test scores or exam results. In the literature, pupils' outcomes are generally function of two different types of inputs: variables describing the socio-economic profile of the pupils on the one hand, and variables describing the "quality" of schools on the other (Vignoles *et al.*, 2000).

In empirical analyses, the education production function thus generally takes the form:

(2) $y_{is} = f(S_{is}, P_{is})$

where, for an individual i trained in school s, y is the outcome, S the vector of school characteristics, and P the vector of variables describing the pupil's socio-economic profile. Proxies for "school quality" are very often based on expenditures (such as school budget, or expenditures per pupil). Most analyses relying on such proxies, however, have failed to prove the existence of a systematic relationship between schools resources and pupils performance (Hanushek, 1986, 1989, 1996, 1997; Gundlach *et al.*, 1999; Vignoles *et al.*, 2000).

For this reason, it has been suggested to concentrate the analysis on the organizational characteristics of schools (e.g. Chubb & Moe, 1990; Woessmann, 2000; Hanushek 2003). In the present research, we attempt to do so, by estimating an education production function where institutional and organizational settings are taken into account, alongside more "traditional" inputs. The details of our choice of variables and the practical constraints imposed by the data are developed extensively in the following sections.

2. Data and choice of variables.

2.1. THE PISA 2000 DATABASE

This paper uses cross-section data from the OECD survey conducted in 2000 as part of the Program for International Student Assessment (PISA). The PISA 2000 database contains math, science and reading test scores of a sample of 15 years-old pupils coming from 28 OECD and non-OECD countries[1] (cf. Table 1). These pupils are nested within schools, potentially attending different grades in countries where grade repetition is common practice.

The sampling process was rather specific (c.f. OECD, 2002, for technical details): all the students in the sample (i.e., approximately 175000) took a standard reading test, which provided the reading test score. Part of them also took a math test and/or a sciences test, which yielded two other smaller samples of approximately 97000 pupils each. The sampling procedure ensured that the three samples had the same by-country composition, and that each variable had the same mean and standard deviation in each sample (as can be seen in Tables 1 and 2 respectively). In our study, the same econometric models will be applied to each of these three samples separately.

For each pupil *i* trained in school *j*, the PISA data provides us with a large number of variables characterizing the student and/or the school, from which we selected the variables relevant to our analysis. Following a common practice in the literature (Jencks and Brown,

[1] Canada, Japan, Germany and Korea have been excluded due to missing variables (some variables used in the estimations are not available for these countries).

1999; Vignoles *et al.*, 2000), we use the test scores (normalized to mean 500 and variance 100 during the construction of the PISA 2000 database) as a measure of education output.

– TABLE 1 ABOUT HERE –

2.2. Choice of explanatory variables.

Table 2 gives summary statistics for our selection of explanatory variables, by discipline. They were chosen in order to avoid correlations, while staying as close as possible to the conceptual framework sketched in the theoretical literature[2] (e.g., Creemer, 1994; Scheerens, 1997; Creemers *et al.*, 2000). A small number of factors are herein put forward to explain pupils' success: (1) their initial aptitudes, (2) their ability to understand instructions, (3) their perseverance or effort, (4) the opportunities offered to them, and (5) the quality of instruction.

First of all, relying on an extensive literature (e.g. Brown, 1991; Ehrenberg, Goldhaber *et al* 1995; McNabb, Sarmistha *et al*, 2002), we control for gender and age. Indeed, although all pupils in PISA 2000 are aged 15 at the time of the survey, not all pupils are born the same month which allows for some variation.

– TABLE 2 ABOUT HERE –

We then chose relevant measures and/or proxies for the five groupes of factors listed above, starting with the opportunities offered to pupils.These are represented by the country of origin of the pupils, and a set of variables describing their family background: type of family (nuclear or not), OECD index of the father's education level (FISCED), and highest OECD socio-economic index in the household (HISEI). We also included the number of books in the house, several studies (e.g. Murnane *et al*, 1981; Todd and Wolpin, 2003) suggesting that it is a fairly reliable proxy to describe a family's inclination towards learning. Finally, we added a measure of the educational support the child receives from his/her parents (FAMEDSUP); this variable is a quantitative indicator specifically developed by the PISA team (cf. OECD, 2002 for details).

The quality of instruction was captured by several school proxies. We first introduced the students/teacher ratio, the condition of the buildings, the availability of teaching material, and the proportion of qualified teachers[3] in the school. The students/teacher ratio can be seen as an indicator of class size (Chubb and Moe, 1990; Card and Krueger, 1992). Whether or not class size has an effect on students' achievement is a hotly debated topic in the economic literature: some studies (e.g. Angrist and Lavy, 1999; Boozer and Rouse, 2001; Krueger, 1999, 2003) identify an inverse relationship (larger classes yielding a lower achievement).

Other studies, however, (Hanushek, 1986, 1996, 2003; Hoxby, 1996, 2000) underline that this is far from systematic, and suggest that institutional factors and incentives structures may have a stronger effect. Woessmann (2000), using data from a previous PISA survey, found that students' performance could stem primarily from centralized examinations and school organizational characteristics, such as autonomy in personnel decisions. To explore this possibility, we added three institutionnal indicators that complete our description of the quality of instruction: the degree of centralization in the hiring of teachers, the type of pupils' assessment (the use of standardised tests), and the percentage of public funding.

To some extent, the ability to understand instruction, as well as perseverance or effort, can be captured by the grade which the pupils are attending at the time of the survey. Since they are all aged 15 (with minor monthly variations as explained above), one can reasonably expect that lower grades correspond to less able and/or less perseverant pupils. In order to try

[2] In empirical applications, however, it is often difficult to find precise measures of, or even relevant proxies for, these five factors. Todd & Wolpin (2003) thus underline how data limitation may lead to various biases, especially in cases where the econometric analysis is only loosely related to a theoretical framework.

[3] So-called "qualified" teachers obtained an ISCED5 qualification in pedagogy.

and isolate more precisely effort *per se*, we added a synthetic index of the time each pupil spends doing homework each week (HMWKTIME). Again, more details on the construction of this index will be found in OECD (2000).

Finally, the initial aptitudes of the pupils were the most difficult factor to control for with the PISA2000 data. We used the daily time dedicated to "reading for enjoyment", hoping this measure can somehow be correlated to the pupils' taste for learning. This indicator may be far from perfect, however, as children's tastes can be strongly influenced by their parents (although none of the "family background" variables was correlated to the time spent reading for leisure). Further differences in initial aptitudes can only be captured by the residual term (random error). This naturally depends on the specification of our econometric model, which we will now examine in detail.

3. Econometric modelling and estimation strategy.

In theoretical works, the education production function *f* may appear as a "black box", with its the functional form left unspecified. Applied analyses, however, generally rely on a classical regression approach, which assumes a linear (or log-linear) shape for *f*. In this study, we expend the linear regression method by adding fixed and random effects. This allows us to take into account the hierarchical (or clustered) nature of the PISA 2000 data. Indeed students are grouped in schools that function in a particular country. The position in this hierarchy affects results and should be taken into account when estimating. The inclusion of both fixed and clustered random effects also gives more flexibility to the econometric model.

Our estimation strategy consist in three steps. We first estimate, as a "benchmark", a conventional linear regression model by the OLS technique:

(3) $$Y_i = \beta . X_i + \varepsilon_i$$

where Y_i is the test score of pupil *i* in a given discipline, X_i the vector of explanatory variables with β its associated vector of parameters (to be estimated), and ε_i a random error term. One can reasonably expect this model to be biased, since country-specific characteristics (which may be partially unobserved) are not controlled for. Thus, we may find a strong effect of, say, school characteristics, which is in fact an effect of specific national institutional arrangements or of other social and historical factors.

In order to control for such effects, we introduce a country fixed effect (i.e. a series of dummy indicators) in our linear model of the education production function, which leads to the following *fixed-effect model*:

(4) $Y_{ik} = \beta_k + \beta . X_{ik} + \varepsilon_{ik}$

where, for student *i* in country *k*, Y_{ik} is the test score in a given discipline, X_i the vector of explanatory variables, β its associated vector of parameters, β_k the country-specific fixed effect, and ε_{ik} a random error term.

Although this model is more sophisticated than the basic linear regression model, it may still be biased if pupils' performance is affected by unobserved school characteristics such as quality of management, security or teacher motivation. We thus need to control for unobserved school heterogeneity; Murnane *et al* (1981) achieved this by introducing a school fixed-effects in their regression model of the education production function. In our model, we want to take into account the stratified nature of the data, and thus introduce a nested[4] school random effect in our regression model. This modelling allows to control for sources of unobserved heterogeneity supposedly uncorrelated to the explanatory variables. The random effect is also nested in countries, since the distribution of the unknown school characteristics may be proper to the national environment.

[4] "Nested" means here that the distribution of the school random effect may vary across countries.

The resulting *mixed model* is written:

(5) $Y_{ijk} = \beta_k + X_{ijk}.\beta + \gamma_{jk} + \varepsilon_{iik}$

where, for student *i* in school *j* in country *k*, Y_{ijk} is the test score in a given discipline, X_{ijk} is the vector of explanatory variables, β its associated vector of parameters, β_k is the country fixed effect, γ_{jk} is the school random effect (nested within country *j*) and ε_{iik} the residual term. Similar models, also known as *multilevel models* (Yang, Goldstein *et al*, 2002), have been used to analyze examination results in a single country (the United Kingdom) and a single discipline (Mathematics). The model we applied here is both simple and convenient, and fitted to the analysis of several countries and disciplines.

Both the fixed effect and the mixed models are estimated using the Maximum Likelihood technique. The results of the estimations for the three models (linear, fixed effect and mixed) and the three disciplines (maths, reading and science) are given and commented in the next section.

4. Empirical results

Tables 3, 4 and 5 give the parameter and standard deviation estimates of the three models (OLS regression, simple fixed-effect and mixed) for mathematics, reading and science test scores respectively. Table 6 presents the details of the country fixed-effect, by discipline, for the last two models (simple fixed-effect and mixed). Recall from Section 2 that test scores are standardised at mean 500 and standard deviation 100 for all countries. This means we can interpret the coefficients as percentages of standard deviation. Goodness-of-fit statistics are featured at the bottom of each table. The results of the estimations appear to be quite consistent across disciplines (which, given the sampling scheme, does make sense).

In particular, individual characteristics (gender excepted), ability/effort, and family background variables have a similar impact on each test score. The case of gender is interesting, since it shows that female students can expect lower scores in mathematics and science, but higher scores in reading. This result is similar to what is generally observed in the literature (e.g., Ehrenberg, Goldhaber *et al.*, 1995); a (partial) explanation frequently proposed for this outcome is the absence of female role models in science and mathematics.

The inclusion of fixed- and random- effects in the model slightly affects the size of the impact of individual characteristics, effort, and family background, but not its significance (the exception being the country of origin, which doesn't affect the math test score in the OLS specification, although it has a significant effect in the two other specifications). For instance, an increment in the number of books at home explains 12 to 13% of the test score standard deviation in the OLS specification, 10 to 11% in the fixed-effect specification, and 8.2% in the mixed specification, no matter which discipline one considers. Similarly, each additional hour spent doing homework each week explains 9 to 10% of a standard deviation in the OLS model, 10 to 12% in the fixed-effect model, and 7% in the mixed model.

A surprising result, regarding family background, is the negative effect (across specifications and disciplines) of the "family educational support" (FAMEDSUP) variable on the test score. It may be that too much support from his/her family (parents, but also brothers and sisters) reduces the pupil's autonomy and ability to face a test on his/her own. This result could also mean that family support lacks (or even conflicts with) the pedagogy a professional teacher develops in the classroom, and has a counterproductive effect on pupils' performance.

Our measure of initial aptitudes clearly shows its limitations: although it has a positive influence on all test scores, its effect on maths is very small (it explains approximately 2% of the standard deviation in the fixed-effect and mixed models, and is not significant in the OLS specification). The effects on the science test score are somewhat larger (3% in the OLS, 5%

in the fixed-effect model, and 7% in the mixed model). Quite naturally, we observe the largest effects on the reading test score (6% in the OLS, 7% in the fixed-effect model, and 8% in the mixed model).These results simply suggest that the taste for learning (and the cognitive aptitudes) cannot be captured by the taste for reading only.

Let us now focus on the school characteristics, in order to examine the respective effects of institutional / organizational arrangements and monetary resources. The latter do not seem to have any consistent effect on test scores: the (commonly used) students/teacher ratio has a slightly negative impact (between -0.1% and -1.2%) on all test scores in the OLS and fixed-effect specifications. In the mixed model (our most reliable specification), the ratio has no significant effect on the math and reading test scores, and a very weak effect (at the 10% level of significance) on the science test score.

Similarly, the condition of the school buildings has no significant impact on the science test score (no matter what specification is used); if we rely on the mixed model, it has no effect on the other test scores either. Finally, the lack of teaching material has a significant negative impact on the math and reading test scores (the estimates given by the mixed model are 10% and -14% respectively). According to the mixed specification, it has no effect, however, on the science test score.

A more significant resource-related variable is the attainment of a degree in pedagogy by teachers. It is resource-related because of the cost related to hiring teachers with higher degrees. The proportion of teachers that have a minimum level of qualification[5] in pedagogy has a positive influence on students' test scores in all disciplines. When fixed and/or random effects are added in the regression, the size of the effect is halved; the most conservative estimates, given by the mixed model, are 12% of a standard deviation for the math test score, 10% for the reading test score, and 15% for the science test score. This result is in line with

[5] ISCED5

findings from the litterature. Many authors find that teacher qualification as measured by years of experience or the attainment of a university diploma is not a significant factor of student success at tests. It seems that specific pedagogical training is the more determinant teacher skill (Monk (1994), Angrist and Lavy (2001)).

Institutional and organizational variables also have a consistent effect. More precisely, school autonomy in the recruitment of teachers has a strong positive influence on students performance. This result is remarkably consistent over disciplines and specification: the most trustable estimates, given by the mixed model, suggest that a decentralized hiring process can increase students test scores by 12% in mathematics, 11% in reading, and 7% in science.

The results regarding the modes of assessment are less clear-cut, however: while the OLS and fixed-effect models indicate a small positive impact of *school-designed* (rather than standardised) tests on pupils' performance in all three disciplines, the mixed model suggests that this impact is not significant. A possible explanation for this outcome is that the type of examination procedure is often (although not systematically) adopted on a national basis; as such, its effect may be (partially) captured by the country fixed-effect. If that is effectively the case, our results would echo Woessmann (2000)'s findings (using the previous PISA survey data): assessment procedures matter. The main divergence between our findings and those of Woessmann (2000) is that the effect of standardized assesment procedures on students' performance is questionable when the PISA 2000 data is used.

Finally, if the amount of monetary resources spent at the school level does not seem to be the primary determinant of students' performance, the origin of these resources may nevertheless matter. Indeed, according to the mixed model estimates, more than 50 percent of *public* funding may lower the test scores by 11% in the case of mathematics, 14% in the case of reading, and 15% in science. Since a larger share public funding is often associated with more constraint (submission to standard rules regarding pedagogy and internal organization, for instance), this result again points out towards school autonomy as a determinant of students' performance.

– TABLE 3 ABOUT HERE –

– TABLE 4 ABOUT HERE –

– TABLE 5 ABOUT HERE –

– TABLE 6 ABOUT HERE –

5. Conclusions and policy implications.

In this paper, we used the PISA 2000 cross section data to estimate the impacts of organizational and institutional factors on students performance (controlling for individual characteristics). In order to take into account the hierarchical nature of the data, we introduced a country fixed-effect and a school random-effect in our regression model of the "education production function". Our results echoe the findings of Hanushek (1986, 1996, 1997, 2003), Hoxby (1996, 2000) and Woessmann (2000), in the sense that organizational and institutional factors may matter more than the amount of school resources.

More precisely, we found that school autonomy in the decision of hiring teachers significantly increases students performance, while a proportion of public funding higher than 50% tends to decrease performance. These results, however, should not be misinterpreted: they do not imply that public expenditures on education should be cut, but rather that more autonomy should be given to schools (including state-funded schools), especially in local and internal matters. Our results suggest that the mode of assessment also matters, but it is not clear whether standardized or school-designed tests are more helpful to students.

However, caution is needed, since the PISA 2000 data present a number of limitations that may condition our results. The main limitation comes from the cross-section nature of the data, which implies to relate test scores to contemporaneous measures of inputs in the education production function. Such a model rely on the implicit assumption that inputs are unchanging over time – which is obviously not true for all of them. Moreover, the PISA 2000 survey does not provide complementarity measures of students achievement, which could be related to the tests score in a "value added" model (Ehrenberg, Goldhaber, *et al.*, 1995; Todd and Wolpin, 2003). In the absence of longitudinal data, such a measure would be helpful in order to control more adequately for the pupils' initial aptitudes.

In spite of these shortcomings, the PISA 2000 data has several advantages: it is readily available, provides very detailed information, allows for comparisons across countries, and provides new directions for research. However, in order to implement relevant education policies, data allowing for longitudinal analysis is needed, in order to assess more firmly the type of results that have been highlighted in this paper.

References

Angrist, Joshua D. and Victor Lavy (1999), Using Maimonides' Rule to estimate the effect of class size on scholarly achievement, *Quarterly Journal of Economics*, 114 (May), pp. 533-575.

Angrist Joshua D. and Lavy Victor (2001), Does Teacher Training Affect Pupil Learning? Evidence from Matched Comparisons in Jerusalem Public Schools, *Journal of Labor Economics*, vol.19 No.2.

Becker Gary S. (1975), *Human capital: a theoretical and empirical analysis, with special reference to education* (Second Edition, The University of Chicago Press, Chicago).

Becker, Gary S., Murphy, Kevin M. and Tamura, Robert. (1990), Human capital, fertility, and economic growth, *Journal of Political Economy*, 98 (October, part 2), pp. 12-37.

Bishop John H., Ferran Mane and Bishop Michael (2000), "Secondary Education in the United States: What Can Others Learn from Our Mistakes?", *unpublished* (available on iadb.org).

Boozer, Michael and Cecilia Rouse (2001), Intraschool variation in class size: patterns and implications, *Journal of Urban Economics*, 50, pp. 163-189

Brown, Byron W. (1991), How gender and socioeconomic statuts affect reading and mathemaics achievement, *Economics of Education Review*, 10(4), pp. 343-57.

Card, David and Alan B. Krueger (1992), School quality and black-white relative earnings: a direct assessment, *The Quarterly Journal of Economics*, 107, pp. 151-200.

Chubb, John E. and Moe, Terry M. (1990), *Politics, Markets, and America's Schools (*The Brookings Institution, Washington D.C., 318 p).

Cooper, S.T. and Cohn, E. (1997), Estimation of a frontier production function for the South Carolina Educational Process, *Economics of Education Review*, 16(3), pp. 313-327.

Creemer, Bert P.M. (1994), *The effective classroom* (Cassell, London).

Creemer, Bert P.M., Scheerens, Jaap, and David Reynolds (2000), Theory development in school effectiveness research, in: Charles Teddlie and David Reynolds (Eds) *The International Handbook of School Effectiveness Research (*Fulmer, 411 p).

Ehrenberg, Ronald J., Daniel D. Goldhaber, and Dominic J. Brewer (1995), Do teachers' race, gender and ethnicity matter ? Evidence from the national educational longitudinal study of 1988, *Industrial and Labor Relations Review,* 48(3), pp. 547-561.

Gundlach, Erich, Woessmann, Ludger, and Gmelin, Jens (2001), The decline of schooling productivity in OECD country, *The Economic Journal*, 111 (May), pp. 135-147.

Hanushek, Eric A. (1986), The economics of schooling: production and efficiency in public schools, *Journal of Economic Literature*, 24(3), pp. 1141-1177

Hanushek, Eric A. (1989), Expenditures, Efficiency, and equity in education: the federal government's role, *American Economic Review,* 79(2), pp.46-51

Hanushek, Eric A. (1996), School resources and student performance, in: Gary Burtless (Ed.) *Does money matter? The effect of school resources on student achievement and adult success* (The Brookings Institution, Washington D.C., pp. 43-73).

Hanushek, Eric A. (1997), Assessing the effect of school resources on student performance: an update, *Educational evaluation and policy analysis,* 19(2), pp. 141-64.

Hanushek, Eric A.(2003), The failure of input-based schooling policies, *The Economic Journal*, 113(485), February, pp. F64-F98.

Hoxby, Caroline M. (1996), How Teachers' Unions Affect Education Production, *Quarterly Journal of Economics,* 111, August, pp. 671–718.

Hoxby, Caroline M. (2000), The effects of class size on student achievement: new evidence from population variation, *Quarterly Journal of Economics*, 116, November, pp. 1239-1286

Krueger, Alan B. (1999), Experimental estimates of educational production functions, *Quarterly Journal of Economics*, 114, May, pp. 497-532

Krueger, Alan B. (2003), Economic considerations and class size, *Economic Journal*, 113(485), February, pp. F34-F63

Lucas, Robert E., Jr. (1988), On the mechanics of economic development, *Journal of Monetary Economics*, 22 (July), pp. 3-42.

Mankiw, Gregory N., Romer, David, and Weil, David N. (1992), A contribution to the empirics of economic growth, *Quarterly Journal of Economics*, 107 (May): pp. 407-437.

McNabb, Robert, Pal Sarmistha and Peter Sloane (2002), Gender differences in educational attainment: the case of university students in England and Wales, *Economica*, 69(275), pp. 481-503.

Monk D.H. (1994), Subject area preparation of secondary mathematics and science teachers and student achievement, *Economics of Education Review* 13(2) pp.125-145.

Murnane R.J., R.A. Maynard and J.C. Ohls (1981), Home resources and children's achievement, *The Review of Economics and Statistics*, 63(3), pp. 369-377.

OECD (2002), *Manual for the PISA 2000 database*, 165 p.

Scheerens, Jaap (1997), Conceptual models and theory-embedded principles on effective schools, *School effectiveness and school improvement*, 8(3), pp. 269-310.

Vignoles Anna, Levacic Rosalind, Walker James, Machin Stephen, and Reynolds David (2000), *The relationship between resource allocation and pupil attainment: a review*, published by the Centre for the Economics of Education, London, 86 p.

Woessmann, Ludger (2000), Schooling resources, educational institutions, and student performance: the international evidence, *Kiel Working Paper*, n°983 (December), Kiel Institute of World Economics, Kiel.

Yang M., H. Goldstein , W. Browne and G. Woodhouse (2002), Multivariate Multilevel Analyses of Examination Results, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 165(1), pp. 137-53.

Table 1: composition of the sample by country and discipline

Country	Math		Reading		Sciences	
	Number	*% of sample*	*Number*	*% of sample*	*Number*	*% of sample*
Australia	2859	2.9	5176	3.0	2860	2.9
Austria	2640	2.7	4745	2.7	2669	2.7
Belgium	3784	3.9	6670	3.8	3722	3.8
Brazil	2717	2.8	4893	2.8	2710	2.8
Czech Republic	3066	3.1	5365	3.1	3062	3.1
Denmark	2382	2.4	4235	2.4	2346	2.4
Finland	2703	2.8	4864	2.8	2710	2.8
France	2597	2.7	4673	2.7	2592	2.7
Greece	2605	2.7	4672	2.7	2593	2.7
Hungary	2799	2.9	4887	2.8	2800	2.9
Iceland	1882	1.9	3372	1.9	1859	1.9
Ireland	2128	2.2	3854	2.2	2134	2.2
Italy	2765	2.8	4984	2.8	2766	2.8
Latvia	2149	2.2	3893	2.2	2157	2.2
Liechtenstein	175	0.2	314	0.2	176	0.2
Luxemburg	1959	2.0	3528	2.0	1950	2.0
Mexico	2567	2.6	4600	2.6	2548	2.6
Netherlands	1382	1.4	2503	1.4	1396	1.4
New Zealand	2048	2.1	3667	2.1	2029	2.1
Norway	2307	2.4	4147	2.4	2308	2.4
Poland	1976	2.0	3654	2.1	2043	2.1
Portugal	2545	2.6	4585	2.6	2552	2.6
Russian Federation	3719	3.8	6701	3.8	3719	3.8
Spain	3428	3.5	6214	3.6	3457	3.6
Sweden	2464	2.5	4416	2.5	2444	2.5
Switzerland	3396	3.5	6100	3.5	3397	3.5
UK	5195	5.3	9340	5.3	5179	5.3
USA	2135	2.2	3846	2.2	2129	2.2
Total	97384	100.0	174896	100.0	97321	100.0

Table 2: summary statistics by discipline (sub-sample)

Variable	Definition	Math	Reading	Science
		Mean (Std Dev)	Mean (Std Dev)	Mean (Std Dev)
Age	Age in months at time of survey	188.62 (3.44)	188.63 (3.44)	188.62 (3.44)
Gender	1 if female, 0 if male	0.50 (0.50)	0.50 (0.50)	0.50 (0.50)
Origin	1 if country of test, 0 if other	0.93 (0.25)	0.93 (0.25)	0.93 (0.25)
Family	1 if non-nuclear[6], 0 if nuclear	0.25 (0.43)	0.25 (0.43)	0.25 (0.43)
FISCED	OECD index of father's education	4.45 (1.43)	4.45 (1.43)	4.45 (1.43)
# of books	Number of books at home	4.47 (1.54)	4.47 (1.54)	4.47 (1.54)
HISEI	Highest ISEI in the family	49.01 (16.44)	49.00 (16.43)	48.93 (16.45)
FAMEDSUP	FAMily EDucational SUPport	0.00 (1.00)	-0.01 (1.00)	-0.01 (1.00)
S/T ratio	Student/teacher ratio	13.95 (6.61)	13.96 (6.60)	13.94 (6.61)
Buildings condition	2: good, 1: poor, 0: unknown	1.89 (0.41)	1.89 (0.42)	1.89 (0.42)
Teaching material	2: available, 1: lacking, 0: no information	1.88 (0.43)	1.88 (0.43)	1.88 (0.43)
% qualified	% of qualified professors in school	0.61 (0.40)	0.61 (0.40)	0.61 (0.40)
Hiring teachers	2: central, 1: decentralized, 0 : no info	1.23 (0.58)	1.23 (0.58)	1.22 (0.58)
Assessment	2: students assessed by std test at least once a year, 1: students assessed by school-designed (non-std) test, 0: no info	1.38 (0.69)	1.38 (0.69)	1.38 (0.69)
Funding	1 if more than 50% of school resources come from public authorities, 0 otherwise	0.91 (0.29)	0.91 (0.29)	0.91 (0.29)
Grade	Grade attented at time of survey	9.62 (0.75)	9.62 (0.75)	9.62 (0.75)
HMWKTIME	Index of weekly homework time	0.02 (1.01)	0.02 (1.01)	0.02 (1.01)
Reading as leisure	Number of hours dedicated daily to reading as leisure	2.27 (1.15)	2.27 (1.15)	2.28 (1.15)
Observations		97384	174896	97321

[6] E.g. single parent, recomposed family, etc.

Table 3: estimates for the math test score

Variables		OLS		Fixed-Effect Model		Mixed Model	
		Coeff.	S. E.	Coeff.	S. E.	Coeff.	S. E.
Intercept		270.12	(21.01)***	316.43	(20.60)***	302.71	(20.31)***
Age		-1.18	(0.11)***	-1.25	(0.11)***	-1.10	(0.11)***
Gender	*Female*	-19.20	(0.78)***	-20.03	(0.75)***	-19.55	(0.72)***
	Male						
Origin	*Country of test*	2.85	(2.04)	7.30	(1.99)***	4.44	(1.90)**
	Other country						
Family	*Non-Nuclear*	-14.68	(0.88)***	-13.75	(0.86)***	-9.51	(0.81)***
	Nuclear						
FISCED		5.38	(0.33)***	5.01	(0.32)***	3.13	(0.31)***
# of books		13.01	(0.28)***	10.84	(0.28)***	8.24	(0.27)***
HISEI		0.68	(0.03)***	0.73	(0.03)***	0.49	(0.03)***
FAMEDSUP		-11.87	(0.36)***	-11.60	(0.36)***	-9.77	(0.34)***
S/T ratio		-1.17	(0.04)***	-0.14	(0.05)***	-0.06	(0.13)
Buildings condition	*Poor*	-5.11	(1.84)***	-3.98	(1.80)**	-3.07	(3.59)
	Missing info	0.24	(4.62)	0.85	(4.48)	-5.16	(9.29)
	Good						
Teaching material	*Lacking*	-17.23	(1.52)***	-8.37	(1.51)***	-9.77	(3.26)***
	Missing info	-23.71	(3.98)***	-14.27	(3.88)***	-6.44	(8.90)
	Available						
%qualified		40.20	(1.15)***	16.66	(1.79)***	11.80	(3.40)***
Hiring	*Decentralized*	9.21	(0.99)***	11.06	(1.18)***	11.96	(2.45)***
	Missing info	6.75	(1.83)***	-8.25	(4.31)*	-7.42	(7.61)
	Centralized						
Pupils assessment	*Non-std test*	4.69	(0.98)***	4.17	(1.01)***	0.13	(2.01)
	Missing info	-21.96	(1.07)***	-6.36	(2.26)***	-1.87	(4.30)
	Std test						
Funding	*≥50% public*	-1.91	(1.21)	-8.38	(1.26)***	-10.70	(2.88)***
	< 50% public						
Grade		34.19	(0.54)***	31.42	(0.65)***	32.70	(0.74)***
HMWKTIME		9.49	(0.40)***	10.98	(0.39)***	7.84	(0.38)***
Reading as leisure		-0.06	(0.33)	2.20	(0.33)***	2.57	(0.31)***

Significance level: ***: 1%, **: 5%, *: 10%

Goodness-of-fit

OLS: a Fisher test led to the rejection of the null hypothesis H_0: "$\beta=0$" at the 1% level of significance ($R^2 = 0.40$, adjusted $R^2 = 0.40$)

Fixed-effect and mixed models: a LR test led to the rejection of the null hypothesis H_0: "$\beta=0$" at the 1% level of significance. A Fisher test shown the country fixed effect to be overall significant at the 1% level (c.f. Table 6 for details).

Table 4: estimates for the reading test score

Variables		OLS		Fixed-Effect Model		Mixed Model	
		Coeff.	S. E.	Coeff.	S. E.	Coeff.	S. E.
Intercept		209.36	(14.37)***	253.59	(14.29)***	267.33	(14.49)***
Age		-0.91	(0.08)***	-1.23	(0.08)***	-1.08	(0.07)***
Gender	*Female*	14.66	(0.53)***	13.28	(0.52)***	11.44	(0.50)***
	Male						
Origin	*Country of test*	9.35	(1.39)***	10.02	(1.37)***	7.83	(1.30)***
	Other country						
Family	*Non-Nuclear*	-12.72	(0.60)***	-13.17	(0.60)***	-9.95	(0.56)***
	Nuclear						
FISCED		4.43	(0.22)***	4.06	(0.22)***	1.36	(0.22)***
# of books		12.11	(0.19)***	11.37	(0.20)***	8.21	(0.19)***
HISEI		0.86	(0.02)***	0.82	(0.02)***	0.54	(0.02)***
FAMEDSUP		-13.25	(0.25)***	-12.86	(0.24)***	-10.70	(0.23)***
S/T ratio		-0.75	(0.03)***	-0.39	(0.04)***	-0.12	(0.12)
Buildings condition	*Poor*	-9.11	(1.25)***	-6.16	(1.24)***	-4.97	(3.22)
	Missing info	-0.05	(3.17)	4.34	(3.12)	-4.78	(8.31)
	Good						
Teaching material	*Lacking*	-22.05	(1.04)***	-12.17	(1.05)***	-13.99	(2.94)***
	Missing info	-9.33	(2.72)***	-4.40	(2.69)	3.80	(8.14)
	Available						
%qualified		23.19	(0.79)***	12.82	(1.24)***	10.07	(2.95)***
Hiring	*Decentralized*	4.15	(0.68)***	4.96	(0.82)***	10.85	(2.18)***
	Missing info	2.93	(1.24)**	-18.00	(2.97)***	-15.70	(6.79)**
	Centralized						
Pupils assessment	*Non-std test*	-0.01	(0.67)	2.58	(0.70)***	-1.21	(1.77)
	Missing info	-45.56	(0.73)***	-5.59	(1.56)***	-3.04	(3.82)
	Std test						
Funding	*≥50% public*	-2.60	(0.83)***	-8.62	(0.88)***	-14.15	(2.61)***
	< 50% public						
Grade		33.57	(0.37)***	37.17	(0.45)***	35.70	(0.52)***
HMWKTIME		10.05	(0.27)***	11.86	(0.27)***	7.72	(0.26)***
Reading as leisure		5.67	(0.23)***	6.94	(0.23)***	8.19	(0.21)***

Significance level: ***: 1%, **: 5%, *: 10%

<u>Goodness-of-fit</u>

OLS: a Fisher test led to the rejection of the null hypothesis H_0: "β=0" at the 1% level of significance (R^2 = 0.40, adjusted R^2 = 0.40)

Fixed-effect and mixed models: a LR test led to the rejection of the null hypothesis H_0: "β=0" at the 1% level of significance. A Fisher test shown the country fixed effect to be overall significant at the 1% level (c.f. Table 6 for details).

Table 5: estimates for the science test score

Variables		OLS		Fixed-Effect Model		Mixed Model	
		Coeff.	S. E.	Coeff.	S. E.	Coeff.	S. E.
Intercept		189.97	(21.80)***	209.40	(21.79)***	209.42	(21.75)***
Age		-0.48	(0.12)***	-0.35	(0.12)***	-0.37	(0.12)***
Gender	*Female*	-6.87	(0.81)***	-7.55	(0.80)***	-8.84	(0.77)***
	Male						
Origin	*Country of test*	6.38	(2.08)***	9.33	(2.07)***	10.43	(2.01)***
	Other country						
Family	*Non-Nuclear*	-8.61	(0.92)***	-9.25	(0.92)***	-6.31	(0.88)***
	Nuclear						
FISCED		4.85	(0.33)***	4.78	(0.34)***	3.43	(0.33)***
# of books		12.27	(0.30)***	10.72	(0.30)***	8.26	(0.30)***
HISEI		0.72	(0.03)***	0.74	(0.03)***	0.49	(0.03)***
FAMEDSUP		-12.83	(0.37)***	-12.62	(0.37)***	-11.55	(0.35)***
S/T ratio		-0.95	(0.04)***	-0.34	(0.05)***	-0.25	(0.13)*
Buildings condition	*Poor*	0.03	(1.90)	0.52	(1.89)	-0.67	(3.66)
	Missing info	5.20	(4.83)	7.30	(4.78)	-3.74	(9.53)
	Good						
Teaching material	*Lacking*	-13.87	(1.56)***	-6.49	(1.58)***	-4.51	(3.30)
	Missing info	-9.97	(4.13)**	-5.72	(4.10)	-0.40	(9.09)
	Available						
%qualified teachers		32.21	(1.19)***	18.50	(1.88)***	15.41	(3.50)***
Hiring	*Decentralized*	6.50	(1.03)***	4.61	(1.25)***	7.12	(2.50)***
	Missing info	7.00	(1.86)***	-8.52	(4.59)*	-10.29	(7.78)
	Centralized						
Pupils assessment	*Non-std test*	2.22	(1.01)**	3.68	(1.06)***	0.68	(2.05)
	Missing info	-42.71	(1.11)***	-10.73	(2.42)***	-5.22	(4.40)
	Std test						
Funding	*≥ 50% public*	-7.59	(1.26)***	-13.79	(1.33)***	-15.32	(2.92)***
	< 50% public						
Grade		28.84	(0.56)***	25.73	(0.69)***	27.71	(0.79)***
HMWKTIME		8.52	(0.41)***	9.99	(0.41)***	6.97	(0.41)***
Reading as leisure		3.44	(0.35)***	5.17	(0.35)***	6.58	(0.33)***

Significance level: ***: 1%, **: 5%, *: 10%

Goodness-of-fit

OLS: a Fisher test led to the rejection of the null hypothesis H_0: "$\beta=0$" at the 1% level of significance (R^2 = 0.33, adjusted R^2 = 0.33)

Fixed-effect and mixed models: a LR test led to the rejection of the null hypothesis H_0: "$\beta=0$" at the 1% level of significance. A Fisher test shown the country fixed effect to be overall significant at the 1% level (c.f. Table 6 for details).

Table 6: country effects by discipline

	MATH		READ		SCIENCE	
	Fixed-effect	Mixed model	Fixed-effect	Mixed model	Fixed-effect	Mixed model
	Estimate (Std Err)	Estimate (Std Err)	Estimate (Std Err)	Estimate (Std Err)	Estimate (Std Err)	Estimate (Std Err)
Australia	13.43 (2.90)***	23.49 (5.50)***	-8.20 (2.01)***	5.28 (5.02)	-0.18 (3.05)	11.28 (5.57)**
Austria	31.29 (4.91)***	40.32 (6.78)***	8.65 (3.44)**	18.37 (5.88)***	19.23 (5.21)***	29.78 (7.01)***
Belgium	35.33 (4.01)***	44.58 (6.21) ***	6.78 (2.78)**	17.50 (5.49)***	4.15 (4.24)	13.46 (6.36)**
Brazil	-51.15 (2.26)***	-51.55 (5.88)***	-22.74 (1.56)***	-30.39 (5.46)***	-41.82 (2.36)***	-40.51 (5.94)***
Czech Rep.	5.14 (3.28)	15.21 (5.24)***	-13.62 (2.28)***	-4.68 (4.72)	6.22 (3.46)*	16.57 (5.34)***
Denmark	40.50 (6.02)***	51.60 (7.04)***	12.66 (4.22)***	22.74 (5.85)***	-5.09 (6.41)	6.84 (7.38)
Finland	66.56 (4.78)***	76.37 (6.41)***	60.05 (3.32)***	70.39 (5.54)***	47.35 (5.03)***	58.73 (6.59)***
France	36.40 (2.08)***	42.14 (5.36)***	1.84 (1.44)	9.17 (5.07)*	-1.25 (2.20)	5.47 (5.40)
Greece	-51.29 (3.66)***	-45.13 (6.38)***	-49.27 (2.53)***	-40.02 (5.72)***	-46.24 (3.86)***	-40.34 (6.49)***
Hungary	-7.70 (3.61)**	3.13 (5.65)	-29.34 (2.50)***	-19.39 (5.05)***	-10.48 (3.79)***	0.52 (5.76)
Iceland	3.92 (21.59)	13.39 (20.28)	-24.84 (15.03)*	-13.03 (14.78)	-22.40 (22.97)	-12.44 (21.79)
Ireland	4.47 (4.74)	14.51 (6.27)**	11.37 (3.28)***	20.89 (5.41)***	3.60 (5.01)	13.87 (6.47)**
Italie	-26.57 (2.82)***	-21.01 (6.27)***	-25.29 (1.95)***	-15.07 (5.78)***	-21.92 (2.96)***	-15.69 (6.34)**
Latvia	-32.87 (8.07)***	-21.29 (8.85)**	-54.90 (5.53)***	-42.52 (7.05)***	-47.31 (8.39)***	-34.67 (9.20)***
Liechtenstein	58.24 (67.54)	69.65 (61.97)***	11.83 (45.47)	24.69 (42.95)	-1.11 (68.09)	12.11 (63.35)
Luxembourg	-4.64 (21.72)	9.17 (21.57)	-24.56 (15.14)	-6.50 (16.35)	-25.09 (23.16)	-9.84 (23.03)
Mexico	-45.18 (2.57)***	-48.97 (6.26)***	-41.91 (1.77)***	-48.48 (5.82)***	-34.67 (2.68)***	-35.21 (6.30)***
Netherlands	81.15 (3.94)***	86.94 (7.11)***	39.52 (2.73)***	46.06 (6.45)***	47.14 (4.19)***	55.11 (7.24)***
New Zealand	-1.39 (5.99)	8.15 (7.16)	-29.26 (4.13)***	-16.01 (5.93)***	-12.31 (6.27)*	-3.23 (7.40)
Norway	15.14 (7.48)**	20.46 (11.10)*	5.39 (5.21)	19.94 (9.67)**	7.59 (7.98)	17.03 (11.44)
Poland	20.04 (4.79)***	16.61 (9.26)*	16.06 (3.30)***	12.77 (8.42)	6.90 (5.07)	10.77 (9.42)
Portugal	-2.60 (4.06)	7.02 (6.54)	-11.16 (2.82)***	1.93 (5.82)	-14.17 (4.28)***	-3.54 (6.69)
Russia	-23.22 (2.60)***	-16.12 (6.16)***	-58.83 (1.80)***	-50.28 (5.71)***	-46.18 (2.77)***	-41.79 (6.23)***
Spain	-1.59 (2.84)	5.45 (6.05)	-11.58 (1.96)***	-0.90 (5.55)	-2.65 (2.97)	3.82 (6.12)
Sweden	32.51 (4.13)***	43.88 (6.00)***	26.60 (2.88)***	36.44 (5.31)***	16.80 (4.39)***	29.10 (6.17)***
Switzerland	59.41 (4.78)***	68.55 (6.37)***	15.50 (3.32)***	22.71 (5.46)***	13.20 (5.04)***	23.83 (6.57)***
UK	-3.97 (1.97)**	5.15 (4.71)	-27.17 (1.37)***	-16.37 (4.39)***	-3.20 (2.09)	2.70 (4.75)
USA (Ref)	0	0	0	0	0	0