ISSN 1995-2848
OECD Journal: Economic Studies
Volume 2009

The Role of Teacher Compensation and Selected Accountability Policies for Learning Outcomes: An Empirical Analysis for OECD Countries

by

Romina Boarini and Elke Lüdemann

Romina Boarini, OECD Economics Department: *romina.boarini@oecd.org*; Elke Lüdemann, Ifo Institute for Economic Research: *luedemann@ifo.de*. The authors wish to thank Paul O'Brien, Patrick Lenain, Jørgen Elmeskov and Andrew Dean for their helpful comments on a previous draft. They would also like to thank Paulo Santiago and Corinne Prost for their excellent remarks and valuable suggestions for revision. Excellent editorial assistance was provided by Lyn Urmston.

Introduction and motivation

Educational outcomes are shaped by a wide range of factors, including innate students' characteristics, family and school background and other environmental factors. But a key-question for policy-makers is what schools and school policies can do to raise overall student performance.

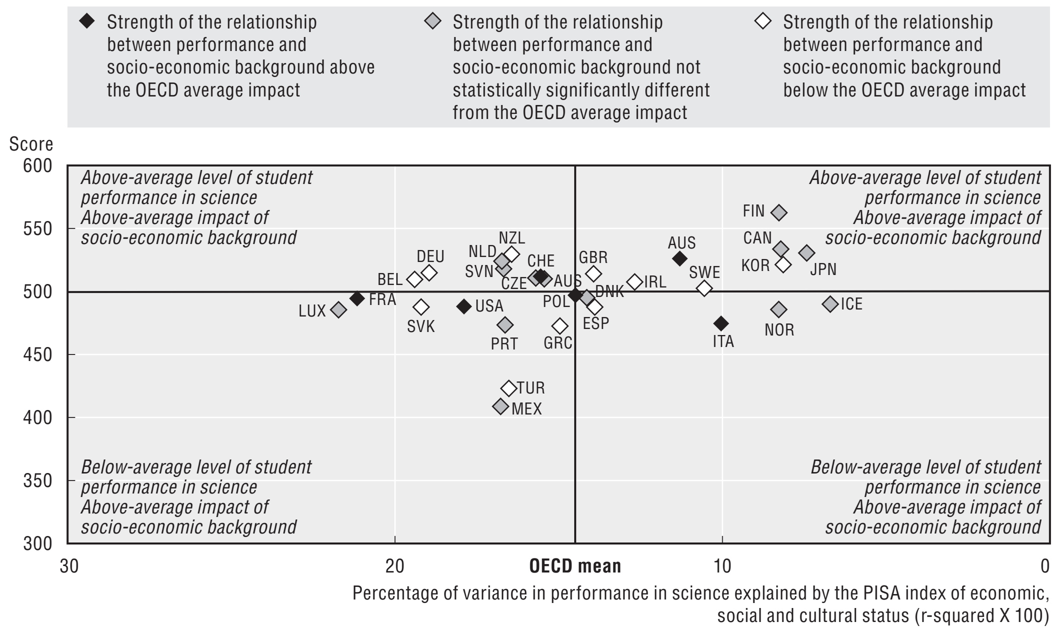
Several studies have indeed shown the positive effect of an increase in cognitive skills and competencies on both social and individual economic welfare. From a macroeconomic viewpoint, international differences in student achievement tests have been shown to increase long-run economic growth (*cf.* Hanushek and Kimko 2000; Woessmann 2002; Hanushek and Woessmann 2007, 2008). From a microeconomic viewpoint, investment in human capital increases labour productivity. There is abundant evidence that, at the individual level, higher cognitive skills measured in tests like PISA lead to higher earnings (*e.g.* Mulligan 1999; Murnane, Willett, Duhaldeborde and Tyler 2000; Lazear 2003), as well as to lower probability of being unemployed across developed countries (*e.g.* Bishop 1992; OECD 2000; McIntosh and Vignoles 2001). Moreover, several studies have found a causal effect of skill differences at younger ages on earnings inequality at later ages (*e.g.* Nickell 2004, Juhn, Murphy and Pierce 1993). Finally, research has shown that higher education leads to gains in health (Kenkel, Lillard and Mathios, 2006), subjective well-being, stronger participation in the civil society (Dee 2004; Milligan, Moretti and Oreopoulos 2004) and to lower criminality rates (Lochner and Moretti 2004).

A vast literature exists on determinants of learning outcomes (see *e.g.* Fuchs and Woessmann, 2007; *cf.* Woessmann *et al.* 2009 for a review). Family background generally has a strong, positive effect on student performance in all countries. In fact, in international student achievement tests such as PISA, TIMSS or PIRLS, students' family background is commonly found to be one of the most important factors influencing student outcomes (*e.g.* OECD 2007a). Features of the school system, institutional characteristics like accountability and autonomy or resource endowments are generally found to be less important. Given that students' family background is less easily policy amenable (at least in the short run), our study concentrates on the net association of institutional features and resources with student performance (*i.e.* we control, in all specifications, for several variables reflecting students' family background).

Our analysis focuses on the quality of school systems, *i.e.* on the association of educational policies with *average* learning outcomes. This policy objective is sometimes also referred to as "effectiveness" of educational policies. A different policy objective that national governments often pursue is that of equity or equality of opportunities. In cross-sectional international student achievement tests like PISA, TIMSS or PIRLS, equality of opportunities is commonly measured as the degree to which students' family backgrounds are related to learning outcomes in a school system. An interesting question in this context is whether quality and equity are indeed conflicting policy objectives. Figure 1 shows that some OECD countries, namely Finland, Canada, Japan and Korea, manage to achieve both

Figure 1. **Average performance in science and the impact of socio-economic background**

Average performance of countries on the PISA science scale and the relationship between performance and the index of economic, social and cultural status



Note: OECD mean used in this figure is the arithmetic average of all OECD countries.

Source: OECD PISA 2006 database, Table 4.4a.

a high degree of equity and good average quality within the school system, suggesting that quality and equity need not be considered as competing policy objectives. OECD (2007a) also finds that there is no obvious trade-off between quality and equity of achievements. A similar conclusion is reached by Woessmann *et al.* (2009), who, based on an analysis of PISA 2003 data, conclude that "there is very little evidence that those aspects of accountability, autonomy and choice that are associated with higher levels of student achievement across countries have adverse consequences for the equity of student achievement". These findings are both obtained for OECD countries and on a larger sample including non-member countries.

Though quality and equity seem not be conflicting policy objectives, various combinations of quality-equity outcomes would likely call for a different mix of policies. For instance, a country with a poor average performance but a relative high level of equity might want to prioritise policy efforts towards the improvement of the former. In our analysis we are concerned with average learning outcomes, and thus our policy conclusions are especially relevant for countries which need to raise the average quality of their results across the distribution of schools and students.

Figure 1, which contrasts the average performance in science (as shown on the vertical axis) with the strength of the relationship between socio-economic background and science performance (a proxy of distribution of learning opportunities), shows that three OECD countries are in the situation of having poor average results but a relatively weak influence of the socio-economic background: Iceland, Italy[1] and Norway. Sweden is in a somewhat similar situation, having an average performance level around the OECD mean, and below average impact of students' socio-economic background. Denmark and Spain are not significantly different from the OECD mean in terms of neither equity nor quality. Our analysis is thus therefore particularly interesting for these six countries that do not face an equity problem but have below average or average science performance.

This work follows very closely the PISA 2006 Initial report (OECD, 2007a and b), which has examined various aspects of school systems and their importance on students' performance. The main results of this report are that ability grouping is associated with lower science performance, while high academic selectivity of schools is associated with higher science performance. Public posting of achievement data is also associated with higher science performance, and so is learning time in mathematics and science.

The main innovation of this paper is the test of the importance of the composition of spending for learning outcomes. Many papers have looked at the role played by various types of resources by concentrating on national data (Hanushek, 2006 for a review); some papers have extended this analysis to international dataset as TIMMS and previous PISA cycles (Woessmann, 2005 and Woessmann *et al.,* 2007). But there is no recent international evidence shedding some light on the role of teachers' wages when a number of other determinants of students learning are controlled for.

Our analysis follows closely some previous OECD work (OECD, 2007a and b). It however differs in several respects: first, it only considers OECD countries, (as opposed to the 55 country sample tested in the PISA 2006 Initial Report [OECD 2007a; b]), under the assumption that learning drivers are relatively homogenous for countries which are at a similar stage of economic development; this assumption is partly supported by the literature, which shows that resources may matter only at an early stage of development (Hanushek, 2007). Second, the analysis does not include three groups of educational policies tested in OECD 2007a and b, *i.e.* admission, grouping and selection; school management and funding; parental pressure and choice. This choice is motivated by the fact that many of these policies are in fact correlated at school and country level and it is thus not possible to test them all in the same specification. OECD 2007a and b shows that the policies omitted in our analysis can be excluded from a model that tests accountability and resources, without loss of generality. In addition, some of these policies, especially selection and grouping particularly matter for the distribution of outcomes but not for the average outcome, which is the main focus of this paper.

While this paper follows a relatively standard empirical approach for analysing learning outcomes, this approach has some limitations that have to be borne in mind when interpreting the results. In particular, given the cross-sectional nature of data, the estimation results cannot be read in terms of causality, but only in terms of correlation. In terms of policy conclusions, this notably means that educational policies which are found to be associated to higher learning outcomes may in fact not be sufficient to improve the quality of education system. The results nevertheless illustrate some international best-practices that are relevant for policy makers when (re)designing education systems with the view of increasing their effectiveness.

This paper is organised as follows. The next section shortly reviews research on the impact of educational policies on learning outcomes, followed by a section presenting the empirical method and the dataset used. The penultimate section discusses the main results and the final section concludes.

Previous research

Most of the research to date suggests that, in industrialised countries with an already high level of spending in education, resources do not substantially improve student achievement. Putting in place the right institutions in school systems can however substantially improve student performance. This section reviews the theoretical background for why institutions should matter and then describes the main findings of previous studies on the role of school accountability and school autonomy.

Principal agent theory

From the point of view of education economics, the provision of schooling can be understood as a network of principal-agent relationships in which a principal (*e.g.* the parent) commissions an agent (*e.g.* the school's head) to perform a service (the education of the child) on her behalf (see *e.g.* Bishop and Woessmann 2004; Woessmann *et al.* 2009). According to principal-agent theory, asymmetric information and divergent interests are the fundamental sources of failures in optimal principal-agent relationships. That is, if the agent's interests diverge from those of the principal, and if the information on the agent's real performance is asymmetric (available only to the agent), then the agent may pursue her own interests instead of those of the principal; the principal will remain unaware of this behaviour, being unable to sanction it. Such principal-agent problems are pervasive in school systems, since there is usually little institutional control of schools' functioning.

Accountability: public posting of achievement data

Our empirical analysis focuses on two specific measures of accountability, namely public posting of achievement data and external exit exams. Because of the problem of asymmetric information about schools' and teachers' effort, setting clear performance standards and providing performance information can give the incentives to raise student achievement (*cf., e.g.* Costrell 1994; Betts 1998). Public posting of achievement data tackles the problem of asymmetric information because the outcome of school production becomes transparent and is, ideally, comparable across schools, so that parents can monitor school's success in transmitting knowledge and competencies. Thus, public posting of achievement data increases the schools' incentives to raise students' cognitive skills. The specific design through which achievement data are collected and disseminated matters however. For instance, value-added measures are superior to "gross" results because they allow to separate school's contribution from students' innate abilities and other environmental factors.

Accountability: external exams

External exams are another accountability device. External exams help resolve the problem of incomplete monitoring of schools' effort by supplying information about the performance of individual students relative to the national (or regional) student population. This information is unavailable in the absence of external exams, when grades assigned by classroom teachers provide the only information on student performance, because these can obviously vary with teacher's standards. By signalling the achievement of students relative to an external standard, the information provided by external-exam systems makes students' performance comparable to the performance of students in other classes and schools. As students receive marks relative to the national average, their educational achievement is made observable and transparent, facilitating the monitoring of the performance of students, teachers and schools.

A number of studies analysing the effects of external exit exams at the end of secondary school show that students perform substantially and statistically significantly better in countries that have external exit-exam systems than in countries without external exit-exam systems (*e.g.* Bishop 1997, Bishop 1999; Woessmann 2001, 2003a, Woessmann 2003b, Bishop 2006; Fuchs and Woessmann 2007). Likewise, positive effects of external exit exams have been found in countries where some regions have those exams in place and others do not, for example for Canadian provinces (Bishop 1997, 1999), US states (*e.g.* Bishop, Moriarty and Mane 2000) and German states (Juerges, Schneider and Buechel 2005; Woessmann 2007c).

Accountability policies must be implemented carefully

There is also evidence that school-focused accountability systems can lead to strategic responses from teachers and schools, for example by increasing placements of low-performing students in special-education programs which are outside the accountability system or by pre-emptively retaining students (Jacob 2005). Yet, recent experimental evidence from the US suggests that providing lower-income families with direct information on school test scores as well as the choice between good schools can in fact lead to better student achievement for disadvantaged students (Hastings and Weinstein 2008). High-stakes testing may also introduce incentives for cheating (Jacob and Levitt 2003).

Critics of accountability systems argue that the latter measures do not always take differences in resources between schools into account (Ladd and Walsh 2002) or that problems of measurement error undermine their objective of raising student achievement (Kane and Staiger, 2002). It is crucial to develop good value-added indicators as to counter this criticism at least to some extent,[2] and the specific implementation of accountability systems seems crucial.

School autonomy

Economic models of school governance often suggest that greater autonomy can lead to increased efficiency of public schools (*e.g.* Hoxby 1999; Nechyba 2003). School autonomy in process and personnel decisions has been found to be associated with significantly better student achievement (Woessmann 2001, 2003a; Fuchs and Woessmann 2007). These decisions include such areas as the purchase of supplies and budget allocations within schools, hiring and rewarding teachers (for a given budget) and choosing textbooks and instructional methods. The positive effects of these various forms of school autonomy are also found in international tests in primary school (Fuchs and Woessmann 2005). The existing cross-country evidence also reveals that there are important interaction effects between school autonomy and the accountability introduced by external exams (*cf.* Woessmann 2007b for a survey). The results show that school autonomy is more beneficial in systems with external exit exams (Woessmann 2005; Fuchs and Woessmann 2007). In several additional school decision-making areas there is evidence that school autonomy is better for student achievement when external exit exams are in place (Woessmann 2005). Woessmann (2003a; 2005) emphasises the role of different aspects of school autonomy for student achievement. Recent evidence suggests that a combination of school autonomy and accountability by means of central exams is beneficial for student performance (*e.g.* Fuchs and Woessmann 2007; Woessmann *et al.* 2009).

Educational resources: level and composition

A lot of research has been devoted to the analysis of the impact of the level of education spending on schools' performance and students' results, generally finding that resources do not have substantial effects on student achievement (*cf.* Gundlach, Woessmann, and Gmelin, 2001; Hanushek, 2003); most of the studies on class size do not find a substantial effect on student outcomes (*e.g.* Hanushek, 1994; Hoxby 2000; Woessmann, 2002, 2007a; Woessmann and West, 2006).

Less attention has been paid to the composition of spending. Teachers' quality is thought to be the most important determinant of students' results. The quality of teaching depends on the interaction of three groups of factors (OECD 2006): teachers' competencies, teaching and learning practice and school environment. It is not however easy to assess the importance of each of these factors, not least because most of them are imperfectly measured. Quality of teachers for instance may depend on some observables (initial qualification, experience, training, *etc.*) but also on latent dimensions (communication skills, motivation and commitment, *etc*). Overall quality of teachers is very hard to measure and it remains, to a large extent, unclear what are its main determinants (*e.g.* Rivkin *et al.*, 2005).

Eide *et al.* (2004) have deplored the declining academic proficiency of US teachers over the last decade, and have related this phenomenon to the compensation structures in the US labour market for teachers. Following this explanation, we argue that country average teacher wages in lower secondary education serve as an – admittedly imperfect – proxy of teacher quality in a country.

In our analysis we thus look at the association between the student-teacher ratio and student performance, as well as at the association between an aggregate measure of teacher wages in lower secondary education and student performance, when controlling for the aggregate level of spending.

The PISA dataset does not allow us to directly assess the role of teacher quality for student achievement, essentially for two reasons: PISA does not contain any information on teachers' profile (*e.g.* qualifications, teaching methods and so on), and students are not paired with their respective teachers in the dataset. Even if we had this information, valid conclusions with regard to teacher quality would be hampered by the cross-sectional nature of the data set which makes value added estimations impossible.

Empirical strategy and data

Following a relatively standard approach in the literature (*e.g.* Hanushek, 1994; Woessmann, 2003a, b; Woessmann *et al.* 2009), we analyze students' performance in international assessment surveys such PISA by estimating an educational production function. In this approach, educational outcomes are measured by test results, and inputs include a wide range of potential determinants of students' achievements, comprising policy and non-policy variables. This work follows closely the empirical exercise carried out in the PISA 2006 report, making use of the same statistical analysis (multi-level data estimation) and, to a large extent, of same explanatory variables. Differently from PISA 2006, estimations are carried out on a sample of OECD countries only.

PISA measures 15 year olds' achievements in three domains (namely science, mathematics and reading), and at the same time collects many pieces of information on socio-economic characteristics of students and their learning environment. Given that the focus of the PISA 2006 cycle is on measuring science literacy, with two thirds of the testing time devoted to science, we choose science performance as the outcome variable when estimating educational production functions. One of the characteristics of the PISA dataset is that data are sampled at two levels: students and schools. The structure of the dataset requires an appropriate statistical tool for empirical analysis (as discussed in OECD 2009, for example). As a result of the sampling strategy used,[3] the residuals are generally not independent,[4] violating a conventional assumption of Ordinary Least Squares estimators. Concretely, this means that achievement of students within a school is likely to be more similar than would be the case in a simple random sample because students in the same school share peer characteristics, teaching and learning conditions and common neighbourhood. As a result, the calculated standard errors of coefficient estimates are biased downwards (though the coefficients estimates themselves are consistent).

As in PISA 2006, student performance is analyzed as the result of sets of variables at three levels: students, schools and countries. The specification of multi-level modelling retained in this exercise is the "random intercept model". This means that, at all levels, slopes of coefficients are fixed (*i.e.* homogeneous across individuals, schools and countries) and only the intercept is randomised (which allows predicting the PISA scores for an individual *i* in the school *j* in the country *k,* as a function of deviations of school j and country k from the respective country and international average effects). The annex provides technical details about the statistical models used here.

The sample contains data from 246 562 students and 8 911 schools in 29 OECD countries. All OECD countries (except France) were used for the multi-level analysis. France was excluded because there, no data were collected on school-level variables from school principals.

Data on teacher per student ratio, teachers' wages and compensation per teachers are based on our calculations on the OECD Education dataset. Table 1 contains descriptive statistics for all explanatory variables used in our estimation while Table 2 decomposes variance into between countries and between schools variance, for variables measured at school level.

Results

An unconditional (*i.e.* no predictors at the student, school and country levels) Hierarchical Linear Model (HLM) analysis decomposed the total variance in our outcome variable science performance into variability among students within schools (60.4%), variability between schools within countries (31.0%) and variability between countries (8.6%).

Simple model without policy variables

The results of the background model (which excludes policy variables) are presented in the first column of Table 3. In line with previous literature, there is a strong effect of socio-economic background on science achievements, both at individual and at school level. The former result may be driven by a number of different (possibly concomitant) factors: socio-economically advantaged parents may be better able to promote their children's education, *e.g.* in early childhood or later on, providing better support with homework and encouraging learning effort. Moreover, higher income may help meet students' needs, for example by making possible private tutoring. It is also likely that parents' socio-economic status reflects their (unobserved) ability, which is genetically

Table 1. **Descriptive statistics of variables used in the model**

Variables	Observations	Minimum	Maximum	Mean	Standard	% missing
System level						
Student-teacher ratio	29	8.88	27.06	13.38	3.70	–
Central exams in science	29	0.00	1.00	0.59	0.46	–
Teachers wages in lower secondary education	29	9.10	88.67	35.37	14.88	10.34
Teachers wages in lower secondary education, relative to GDP	29	0.67	2.51	1.31	0.47	10.34
GDP per capita (per 1000 US-$ ppp)	29	7.21	64.84	28.42	10.95	–
School level						
Index of autonomy in staffing	8 911	–1.16	1.61	0.01	1.00	1.79
Index of autonomy in course contents	8 911	–2.30	1.03	0.00	0.99	1.79
Index of autonomy in budgeting	8 911	–2.91	0.77	0.01	0.99	1.79
Autonomy in firing	8 911	–1.00	1.00	0.09	0.93	1.79
Autonomy in determining starting salaries	8 911	–1.00	1.00	–0.48	0.82	1.79
Autonomy in determining salary increases	8 911	–1.00	1.00	–0.47	0.81	1.79
Composite index of autonomy	8 911	–1.18	0.82	0.00	0.49	1.79
School posting achievement data publicly	8 911	0.00	1.00	0.37	0.48	3.21
School size (per 100 students)	8 911	0.05	100.00	7.43	5.50	3.42
School average index of economic, cultural and social status	8 911	–3.13	1.75	0.00	0.64	0.00
School in city (100000 or more people)	8 911	0.00	1.00	0.32	0.47	2.11
School in a small town or village (15000 or less people)	8 911	0.00	1.00	0.32	0.47	2.11
Percentage of native students at school	8 911	0.00	1.00	0.91	0.16	0.00
Student level						
Learning time for self-study or homework (science), hours per week	246 562	0.00	7.00	1.42	1.45	3.09
Learning time for regular lessons (science), hours per week	246 562	0.00	7.00	2.98	1.96	3.62
Student's language spoken at home is the same as the test language	246 562	0.00	1.00	0.91	0.28	3.84
Student has no immigration background	246 562	0.00	1.00	0.89	0.31	2.11
Student index of economic, cultural and social status	246 562	–5.67	3.35	0.00	1.00	1.09
Student attends grade 7	246 562	0.00	1.00	0.01	0.08	0.23
Student attends grade 8	246 562	0.00	1.00	0.05	0.21	0.23
Student attends grade 9	246 562	0.00	1.00	0.36	0.48	0.23
Student attends grade 10	246 562	0.00	1.00	0.49	0.50	0.23
Student attends grade 11	246 562	0.00	1.00	0.09	0.29	0.23
Student attends grade 12	246 562	0.00	1.00	0.00	0.05	0.23
Student attends grade 13	246 562	0.00	1.00	0.00	0.00	0.23
Student is a female	246 562	0.00	1.00	0.49	0.50	0.23

Source: Our calculations on PISA 2006 dataset.

Table 2. **Decomposition of variance between countries and schools of the main institutional variables tested in the analysis**

School level explanatory variable	% variance between countries	% variance between schools (and within countries)
School posting achievement data publicly	22.6	77.4
Autonomy in firing	45.1	54.9
Autonomy in determining starting salaries	32.5	67.5
Autonomy in determining salary increases	34.4	65.6
Index of autonomy in budgeting	20.9	79.1
Index of autonomy in course contents	49.5	50.5
Composite index of autonomy	44.1	55.9

Note: Table 2 shows the variance decomposition for all institutional variables measured at school level. For example, 22.6% of the total variance in the accountability variable "school posting achievement data publicly" corresponds to variation between countries.
Source: Our calculations on PISA 2006 dataset.

transmitted to their offspring (see OECD 2007a for a discussion). At school level, the importance of socio-economic background concerns peer effects and social segregation in schools, for example. These effects are also well-established in the literature, and may vary considerably from one country to another (as documented in PISA 2006). Immigration background of students has bearing on students' achievements at individual level only: native students score on average 18 points higher than non-native students, students speaking the same language at home and school also do significantly better than those who do not. Schools in rural areas do better than schools in small cities, which do better than schools in large cities (this latter effect is however not statistically significant). Students attending bigger schools also display a better performance than students attending smaller schools. Finally, regressions find that girls have a lower performance in science than boys and that the performance is positively (negatively) related to the degree of student's grade anticipation (delay) in the educational system.

Educational resources

The association of educational resources with student performance is tested here in several ways. Following a standard approach in the literature, educational resources have been first expressed as levels of spending (at lower-secondary level, or as the cumulative spending for a 15 year old from primary to lower-secondary level) in final consumption PPPs, or as a share of GDPs per capita. Consistently with previous studies, none of these measures reveal a relationship between average students' achievements and aggregate financial resources spent on in education (regressions are not shown).

In order to test for the effects of the composition of spending, educational expenditure per student has been decomposed into two terms: teachers' wages and the ratio of teachers per student. The former can be measured in various ways, and Table 3 presents the results of regressions with various proxies.

In a first regression, teachers' wages (excluding employers' social security contributions) are associated with higher PISA performance (column 2). This effect is not large in absolute terms (the elasticity is around 0.1), but it is of comparable size to that of other PISA determinants in the regressions. When teachers' wages are normalised by GDP per capita, or/and when they are measured through compensation per teacher (*i.e.* including employers' social security contributions) they do not show any significant

Table 3. **Three-level regressions on PISA 2006 scores, selected specifications**

	1	2	3	4	5	6	7	8	9	10	11
Constant	499.26	502.23	501.67	501.00	498.78	499.61	499.93	499.86	499.77	499.38	498.70
System level:											
GDP per capita	–0.71	–2.13			–0.56	–0.75	–0.76	–0.74	–0.75	–2.35	
Student per teacher ratio		–2.18	–1.15	0.41						–2.79	–2.17
Teachers' wages		1.37								1.38	
Teachers' wages/GDP per capita			15.39								20.47
Teachers' compensation/GDP per capita				–4.52							
Accountability: central exam in science					21.72					17.88	24.40
School level:											
School average index of economic, social and cultural status	47.60	41.91	41.85	41.82	47.26	48.73	48.82	48.62	48.28	47.37	47.29
School in a city (100,000 or more inhabitants)	–4.01	15.20	15.32	15.35	14.30	14.19	14.69	14.92	14.92	14.16	14.37
School in small town or village (15,000 or less people)	7.96	6.98	6.95	6.94	7.84	8.02	8.03	7.91	7.79	7.92	7.88
School with high percentage of native	14.41	–3.65	–3.61	–3.60	–3.86	–3.79	–3.80	–3.80	–3.60	–3.98	–3.92
School size	0.82	0.79	0.79	0.80	0.78	0.79	0.79	0.80	0.76	0.78	0.78
Accountability, publicly publishing students' achievements					5.41				5.65	5.31	5.28
Autonomy, dismissal of teachers						–1.69	–1.21				
Autonomy, starting salary of teachers						–0.58	–0.25				
Autonomy, teachers' salary increase						–3.39	–3.02				
Autonomy, budgeting						1.01					
Autonomy, curricular content							–2.34				
Autonomy (composite index)								–8.16	–8.36		
Interaction Accountability (publicly publishing results) x autonomy (composite)									1.20		
Student level:											
Index of economic, social and cultural status	17.75	16.02	16.02	16.02	17.75	17.75	17.75	17.75	17.75	17.75	17.75
Native students	17.86	17.28	17.28	17.27	17.86	17.87	17.88	17.88	17.88	17.86	17.86
Student's language spoken at home is the same as the test language	20.16	20.14	20.14	20.14	20.16	20.13	20.12	20.12	20.12	20.16	20.16
Student is a female	–8.30	–10.33	–10.33	–10.33	–8.30	–8.30	–8.30	–8.30	–8.30	–8.30	–8.30
Student attends grade 7	–105.17	–96.46	–96.46	–96.47	–105.09	–105.31	–105.23	–105.17	–105.11	–105.15	–105.15
Student attends grade 8	–77.88	–71.03	–71.04	–71.04	–77.82	–77.93	–77.88	–77.84	–77.79	–77.84	–77.85
Student attends grade 9	–34.46	–31.72	–31.73	–31.74	–34.43	–34.48	–34.46	–34.47	–34.44	–34.42	–34.43
Student attends grade 11	17.46	17.15	17.16	17.15	17.41	17.49	17.50	17.50	17.47	17.41	17.42
Student attends grade 12	47.02	49.23	49.24	49.24	46.98	46.95	46.97	47.00	46.99	46.98	46.99
Student attends grade 13	125.27	97.48	97.50	97.49	124.89	124.92	125.12	125.28	124.89	124.90	124.92
Learning time in sciences for regular lessons (hours per week)		9.77	9.77	9.77							
Learning time in sciences for self–study or homework (hours per week)		–0.33	–0.33	–0.33							
Variance explained (expressed as % of variance at each level)											
Student level:	62.0%	68.7%	68.8%	68.8%	62.2%	62.2%	62.3%	62.2%	62.4%	62.1%	62.2%
School level:	19.6%	35.2%	18.3%	14.4%	28.6%	20.5%	17.4%	18.4%	16.7%	47.7%	32.0%
System level:	27.2%	33.5%	32.0%	31.7%	28.1%	27.4%	27.1%	27.2%	27.1%	29.7%	28.4%

Note: Coefficients which are statistically different from zero are in shade.
Source: OECD calculations.

association with students' achievements (columns 3-4). The teacher per student ratio is not significantly related to student performance in all three specifications.

The compositional effects of spending have also been tested by controlling for a given level of spending (with the idea of measuring the potential trade-off between number of teachers and their wages under a given budget constraint). Results presented above are robust to the inclusion of this additional control.

Finally, learning time in science is found to increase students' achievements by 10 points for each additional hour per week. Homework, however, does not have a significant influence, unlike in the PISA 2006 analysis.

Selected aspects of accountability

Various measures of accountability are available in the PISA dataset. We present and discuss here two measures which are found to be highly relevant for countries that need to increase their performance. One is benchmarking, as measured by the public posting of students' achievements at school level. As shown in column 5 of Table 3, publishing performance information at school level is associated with higher performance in the PISA sample: students attending schools which make use of benchmarking perform 4 points higher on the PISA score. A caveat is that this is a national-level variable, which may capture country-fixed effects correlated with external exams. For instance the "culture of evaluation" or the "value attributed to education" might be picked up by this variable in some countries (*e.g.* Korea or Japan).

The effect of publishing results at school level is relatively small compared with the importance of external standardised national exams in sciences at the end of lower-secondary level. In countries where such national external exams do exist, students display a higher PISA performance (22 points). These two accountability variables are of a different nature: the benchmarking variable is based on self-reporting by schools in the PISA sample; the national exam variable is defined at the country level.

Other accountability measures have been tested in a similar specification, without finding any significant relationship with students' results. This could reflect however some measurement errors in the variables, to the extent that they are measured through principal's subjective assessment. Information on the existence of central exams and publication of results is likely to be collected more objectively.

Autonomy

Previous research from the field has shown that managerial autonomy might have an ambiguous influence on students' achievements (see, for instance, Bishop and Woessmann, 2004; Woessmann *et al.,* 2009). It has been found that some aspects of autonomy that are positively related to student achievement (for instance autonomy of staffing) while others show a negative association with student performance (*e.g.* autonomy of budgeting). Moreover, research shows that autonomy has to go hand in hand with accountability for its effects on students' achievements to fully show up. In this chapter autonomy has been measured in different ways, by looking at specific dimensions and sub-dimensions of managerial freedom and also by testing a composite index, which encompasses various dimensions. Columns 6 to 8 of Table 3 show that there is no evidence that autonomous governance *per se* improves students' achievements. This is in line with PISA 2006 which finds no impact of autonomy at school level. The PISA report shows, however, that autonomy in setting educational content of learning and in budgeting is good at country-level, meaning that in countries with a large number of schools reporting autonomy as regards curriculum and budget, students' performance is higher, for given overall resource use.[5] We don't find similar evidence.

Autonomy and selected aspects of accountability

As highlighted above, autonomy may deliver good outcomes if accountability devices are also in place. This idea finds support in previous research (Woessmann *et al.,* 2009). The effect of combining autonomy and accountability on students' achievements is also tested in this analysis. Column 9 shows that there is a positive interaction effect between the two which is not, however, statistically significant. Various other specifications with different measures of autonomy and selected aspects of accountability have been tested: in none of them did the positive interaction between accountability and autonomy turn out to be statistically conclusive.

A combined model

Following Raudenbush and Bryk (2002) and Snijders and Bosker (1999), a build-up strategy was pursued to combine the different policy variables discussed above. This results in a combined model shown in the last two columns of Table 3. In line with specifications mentioned above, students' achievements are found to be positively related to selected aspects of accountability (national and school level). The aggregate teacher per student ratio is found to have no influence on educational output, while teachers' wages (relative to GDP per capita or in absolute terms) are associated with higher PISA scores.

Our combined model explains 32% of the between school variance, which corresponds to an increment of 12% variance explained compared with our background model (column 1).

Caveats on the causal interpretation of results

While offering many interesting insights on the relationship between educational outcomes and individual and school characteristics, this approach has some limitations that have to be borne in mind when interpreting results. First, estimates rely on a number of parametric assumptions for which there are few priors; hierarchical linear models are however quite robust to parametric assumptions. Second, the PISA dataset has some limitations. Notably:

- Given the cross-sectional nature of the data, policy variable effects cannot necessarily be interpreted in terms of causality. PISA is a single cross-section of observations, it is thus not possible to control for individual ability and previous student achievement, that is, to estimate value-added or panel data models. The latter, however, would be more appropriate if explanatory variables changed over time or if they were not exogenous to school or student performance. The coefficients shown in Table 3 are unbiased only to the extent that the institutional and resource variables of interest are uncorrelated with other unobserved characteristics that in turn have an impact on student performance (the grade variable, used in this way in most research, may not meet this requirement). Finally, we try to reduce the risk of omitted variables bias by including a large set of regression controls at the country, school and student level. However, to the extent that a country's institutions and resource endowments are related to unobserved, *e.g.* cultural, factors which in turn may be related to student performance, we cannot interpret the association between institutional and resource factors and student performance as causal effects. Thus, caution should prevail when drawing policy conclusions from the presented results. It should be noticed, however, that many of our findings are in line with literature using other empirical identification strategies such as natural experiments (see previous section).

- 15 year olds' performance is certainly the result of a cumulative learning process which starts much earlier in life; however, the impact of the learning environment from earlier ages is not taken into account.
- While weights are used to control for potential sampling variance bias, the PISA survey covers only a limited number of schools in each country (300 schools on average).
- Since the majority of explanatory variables rely on survey data, there are potential measurement errors.
- Interpretation of country-level variables must be done with caution because of potential omitted variables (*e.g.* the culture of learning, the attitude to the efforts, *etc.*) at country level.
- The meaning of policy variables when measured by self-reporting at school level may be different from when it is based on "objective" national policy settings.
- Assumptions made on educational resources variable deserve a specific comment. Differently from other institutional variables, the impact of resources variables is tested only at country level. The disadvantage of this is that only the aggregate effect is captured in the estimation, and one cannot strictly infer the effect of having richer or poorer schools. An advantage is instead that, measuring educational resources at national level makes it possible to avoid endogeneity problems (Woessmann 2003).

Discussion and conclusion

In line with previous literature, we see no evidence that the level of resources matters for learning outcomes. However, the composition of spending may make some difference. In particular, high number of teachers per student is not systematically associated to higher outcomes, while higher teachers' wages are. This suggests that the trade-off between the number of teachers and the wages bill should be reconsidered with the view of increasing the efficiency of educational spending. This conclusion comes with the caveat that the impact of teachers' wages is not statistically very robust and thus further evidence might be needed to corroborate this finding.

Consistently with PISA 2006, we find that selected accountability policies at school and national level, respectively consisting of publicly posting students' achievements and having central exams in sciences, are associated with higher students' scores. Autonomy was not found to have an association with students' results, not even when it goes hand in hand with the accountability policies tested here. This result should be however put in perspective with PISA 2006, which finds that autonomy at country level (still based on schools self-reporting, however) has a positive impact on students' performance. In a robustness exercise (results not shown but available upon request), similar findings were obtained when using the maths score as the dependent variable, instead of the science score.

Important policy conclusions emerge from the analysis of the role of institutional features of education system for learning outcomes. Increasing the accountability of education system is critical to improve the performance of teachers, principals and education authorities across the board. Making schools accountable is impossible without adequate information, as information is needed to provide feedback to both students and teachers on how well students are learning. Accountability is also essential to set the right incentives for schools to pursue the goals set by the education system. Feedback can be provided under various forms (see, for instance, OECD 2008a); this paper suggests that publishing school results is an effective tool, so are central examinations. This analysis comes with the caveat that the role of educational policies should not be interpreted in causality terms: this means that higher teachers' wages, publication of results and existence of centralised exams standardised at national level are all necessary to higher achievement of the education system. They are not however sufficient conditions and additional policy efforts in other areas may be needed. Accountability policies should be designed and implemented as to take into account the country-specific context. This is for instance crucial to overcome political resistance to some of these reforms (OECD 2008b).

The focus of this study is on analysing quality and effectiveness of school systems. Thus, in this paper, possible determinants of average outcomes are investigated, but we do not look at the relationship between these determinants and equality of opportunities within school systems. Previous research has however found little evidence of a trade-off between average quality and equity. Yet, policymakers should be aware that high-stakes accountability policies have been subject to the criticism of ignoring equity issues, in design and implementation (*e.g.* Diamond and Spillane 2004). If school accountability policies are based on student performance levels rather than the value added by schools, they may give undue advantages to schools serving students from high socioeconomic backgrounds. Furthermore, teachers may respond strategically to accountability measures by retaining disadvantaged students or excluding them from the test-taking population. Depending on their design, they may also induce teachers to concentrate primarily on achieving high average performance, and neglect the performance of weaker students, whose performance may be more difficult to improve. Also value-added accountability measures are not immune from critique. It has been noticed that differences in resources between schools are not always taken into account, or that problems of measurement error undermine their objective of raising student achievement. Yet, recent experimental evidence from the US suggests that providing lower-income families with information on school test scores in a public school choice plan can in fact increase performance of disadvantaged students. Based on these findings, it can be concluded that, if properly designed, accountability policies may safely improve average performance without impairing equity within school systems.

Notes

1. The impact of family's background is small in Italy when measured at individual student's level (*i.e.* within a given school). It is however sizeable when assessed at school's level (between schools variance).
2. The PISA questionnaire does not include a good question to measure the impact of reward measures for teachers and principals on students' achievements. For this reason, the current exercise does not provide a direct test of these measures, while other measures of accountability are considered.
3. This is a two-stage stratified sample design, with the first stage consisting of schools with 15 years old students, and the second stage consisting of students within the sampled schools.
4. The intra-class correlation is a measure that can be used to assess whether the assumption of independence of errors holds. On PISA dataset this assumption is rejected.
5. This model could not be estimated due to multicollinearity problems induced by a smaller sample than in PISA 2006.

Bibliography

Betts, J.R. (1998), “The Impact of Educational Standards on the Level and Distribution of Earnings”, *American Economic Review*, 88(1), pp. 266-275.

Bishop, J.H. (1992), “The Impact of Academic Competencies on Wages, Unemployment, and Job Performance”, *Carnegie-Rochester Conference Series on Public Policy*, 37, pp. 127-194.

Bishop, J.H. (1997), “The Effect of National Standards and Curriculum-Based Exams on Achievement”, *American Economic Review*, 87(2), pp. 260-264.

Bishop, J.H. (1999), “Are National Exit Examinations Important for Educational Efficiency?”, *Swedish Economic Policy Review*, 6(2), pp. 349-398.

Bishop, J.H. (2006), “Drinking from the Fountain of Knowledge: Student Incentive to Study and Learn – Externalities, Information Problems and Peer Pressure”, in E.A. Hanushek and F. Welch (eds.), *Handbook of the Economics of Education*, Vol. 2, Amsterdam: North-Holland, pp. 909-944.

Bishop, J.H., J.Y. Moriarty and F. Mane (2000), “Diplomas for Learning, Not Seat Time: The Impact of New York Regents Examinations”, *Economics of Education Review*, 19(4), pp. 333-349.

Bishop, J.H. and L. Woessmann (2004), “Institutional Effects in a Simple Model of Educational Production”, *Education Economics*, 12, pp. 17-38.

Cohen, J. and P. Cohen (1985), *Applied Multiple Regression and Correlation Analysis for the Behavioural Sciences* (2nd ed.), Hillsdale: Erlbaum.

Costrell, R.M. (1994), “A Simple Model of Educational Standards”, *American Economic Review*, 84(4), pp. 956-971.

Dee, Thomas S. (2004), “Are There Civic Returns to Education?”, *Journal of Public Economics*, 88(9), pp. 1697-1720.

Diamond, J.B. and J. Spillane (2004), “High-Stakes Accountability in Urban Elementary Schools: Challenging or Reproducing Inequality?”, *Teachers College Record*, 106(6), pp. 1145-1176.

Fuchs, T. and L. Woessmann (2005), “The Determinants of Differences in Primary-School Learning across Countries”, mimeo, Ifo Institute for Economic Research at the University of Munich.

Fuchs, T. and L. Woessmann (2007), “What Accounts for International Differences in Student Performance? A Re-examination using PISA Data”, *Empirical Economics*, 32(2-3), pp. 433-464.

Gundlach, E., L. Woessmann and J. Gmelin (2001), “The Decline of Schooling Productivity in OECD Countries”. *Economic Journal*, 111, C135-C147.

Hanushek, E.A. (1994), “Education Production Functions”, in T. Husén and T.N. Postlethwaite (eds.), *International Encyclopaedia of Education*, Second Edition, Vol. 3, Oxford, Pergamon, pp. 1756-1762.

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

Hanushek, E.A. (2006), “School Resources”, in E. Hanushek and F. Welch (eds.), *Handbooks of Education Economics*, pp. 865-908.

Hanushek, E.A. and D.D. Kimko (2000), “Schooling, Labor-Force Quality, and the Growth of Nations”, *American Economic Review*, 90(5), pp. 1184-1208.

Hanushek, E.A. and L. Woessmann (2007), *Education Quality and Economic Growth*. Washington DC, World Bank.

Hanushek, E.A. and L. Woessmann (2008), “The Role of Cognitive Skills in Economic Development”, *Journal of Economic Literature*, 46(3), pp. 607-668.

Hanushek, E.A. and Margaret E. Raymond (2004), “The Effect of School Accountability Systems on the Level and Distribution of Student Achievement”, *Journal of the European Economic Association*, 2(2-3), pp. 406-415.

Hastings, J.S. and J.M. Weinstein (2008), “Information, School Choice, and Academic Achiement: Evidence From Two Experiments”, *Quarterly Journal of Economics*, 123(4), pp. 1373-1414.

Hoxby, C.M. (1999), “The Productivity of Schools and Other Local Public Goods Producers”, *Journal of Public Economics*, 74(1), pp. 1-30.

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

Jacob, B.A. (2005), “Accountability, Incentives and Behavior: The Impact of High-stakes Testing in the Chicago Public Schools”, *Journal of Public Economics*, 89(5-6), pp. 761-796.

Jacob, B.A. and S.D. Levitt (2003), "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating", *Quarterly Journal of Economics*, 118(3), pp. 843-877.

Jones, M.P. (1996), "Indicator and Stratification Methods for Missing Explanatory Variables in Multiple Linear Regression", *Journal of the American Statistical Association*, 91(433), pp. 222-230.

Juerges, H., K. Schneider and F. Buechel (2005), "The Effect of Central Exit Examinations on Student Achievement: Quasi-Experimental Evidence from TIMSS Germany", *Journal of the European Economic Association*, 3(5), pp. 1134-1155.

Juhn, C., K.M. Murphy and B. Pierce (1993), "Wage Inequality and the Rise in Returns to Skill", *Journal of Political Economy*, 101(3), pp. 410-442.

Kane, Thomas J. and Douglas O. Staiger (2002), "The Promise and Pitfalls of Using Imprecise School Accountability Measures", *Journal of Economic Perspectives*, 16(4), pp. 91-114.

Kenkel, D., D. Lillard and A. Mathios (2006), "The Roles of High School Completion and GED Receipt in Smoking and Obesity", *Journal of Labor Economics*, 24(3), pp. 635-659.

Ladd, H.F. and R.P. Walsh (2002), Implementing Value-Added Measures of School Effectiveness: Getting the Incentives Right, *Economics of Education Review,* 21, pp. 1-17.

Lazear, E.P. (2003), Teacher Incentives, *Swedish Economic Policy Review*, 10(3), pp. 179-214.

Lochner, Lance and Enrico Moretti (2004), The Effect of Education on Crime. Evidence from Prison Inmates, Arrests, and Self-Reports, *American Economic Review*, 94(1), pp. 155-189.

McIntosh, S. and A. Vignoles (2001), Measuring and Assessing the Impact of Basic Skills on Labour Market Outcomes, *Oxford Economic Papers*, 53(3), pp. 453-481.

Milligan, K., E. Moretti and P. Oreopoulos (2004), Does education improve citizenship? Evidence from the United States and the United Kingdom, *Journal of Public Economics*, 88, pp. 1667-1695.

Mulligan, C.B. (1999), Galton *versus* the Human Capital Approach to Inheritance, *Journal of Political Economy*, 107(6, pt. 2), pp. 184-224.

Murnane, R.J., J.B. Willett, Y. Duhaldeborde and J.H. Tyler (2000), How Important Are the Cognitive Skills of Teenagers in Predicting Subsequent Earnings?, *Journal of Policy Analysis and Management*, 19(4), pp. 547-568.

Nechyba, T.J. (2003), Centralization, Fiscal Federalism, and Private School Attendance, *International Economic Review*, 44(1), pp. 179-204.

Nickell, S. (2004), Poverty and Worklessness in Britain, *Economic Journal*, 114(494), C1-C25.

Organisation for Economic Co-operation and Development (OECD) (2000), *Literacy in the Information Age: Final Report of the International Adult Literacy Survey*, Paris, OECD.

Organisation for Economic Co-operation and Development (2006), *Teachers matter,* Paris, OECD.

Organisation for Economic Co-operation and Development (2007a), *PISA 2006 Science Competencies for Tomorrow's World. Volume 1: Analysis,* Paris, OECD.

Organisation for Economic Co-operation and Development (2007b), *PISA 2006 Science Competencies for Tomorrow's World. Volume 2: Data,* Paris, OECD.

Organisation for Economic Co-operation and Development (2008a), *How do OECD countries Take Stock of Progress and Performance in Education Systems?,* unpublished document.

Organisation for Economic Co-operation and Development (2008b), *Assessing the challenges of Policy Implementation in Education,* unpublished document.

Organisation for Economic Co-operation and Development (2009), *PISA Data Analysis Manual SPSS, Second Edition,* Paris, OECD.

Raudenbush, S.W. and A.S. Bryk (2002), Hierarchical Linear Models: Applications and Data Analysis Methods, London, Sage.

Snijders, T. and R. Bosker (1999), Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling, London, Sage.

Woessmann, L. (2001), Why Students in Some Countries Do Better: International Evidence on the Importance of Education Policy, *Education Matters*, 1(2), pp. 67-74.

Woessmann, L. (2002), *Schooling and the Quality of Human Capital*, Berlin, Springer.

Woessmann, L. (2003a), Schooling Resources, Educational Institutions and Student Performance: the International Evidence, *Oxford Bulletin of Economics and Statistics*, 65(2), pp. 117-170.

Woessmann, L. (2003b), Central Exit Exams and Student Achievement: International Evidence, in P.E. Peterson and M.R. West (eds.), *No Child Left Behind? The Politics and Practice of School Accountability*, Washington DC, Brookings Institution Press, pp. 292-323.

Woessmann, L. (2005), The Effect Heterogeneity of Central Exams: Evidence from TIMSS, TIMSS-Repeat and PISA, *Education Economics,* 13(2), pp. 143-169.

Woessmann, L. (2007a), International Evidence on Expenditures and Class Size: A Review, *Brookings Papers on Education Policy 2006/2007*, pp. 245-272.

Woessmann, L. (2007b), International Evidence on School Competition, Autonomy and Accountability: A Review, *Peabody Journal of Education*, 82(2-3), pp. 473-497.

Woessmann, L. (2007c), Fundamental Determinants of School Efficiency and Equity: German States as a Microcosm for OECD Countries, Harvard University, *Program on Education Policy and Governance Research Paper PEPG 07–02*, Cambridge, MA: Harvard University.

Woessmann, L., E. Lüdemann, G. Schuetz and M.R. West (2009), *School Accountability, Autonomy, and Choice around the World*, Cheltenham, Edward Elgar.

Woessmann, L. and M.R. West (2006), Class-size Effects in School Systems around the World: Evidence from Between-grade Variation in TIMSS, *European Economic Review*, 50, pp. 695-736.

ANNEX A1

Statistical model

We estimate a three-level-random-intercept model, with predictors at student (a), school (X), and country level (W). See Table 1 for a list of all predictor variables as well as descriptive statistics.

Level-1 model: students

$$Y_{ijk} = \pi_{ojk} + \sum_{p=1}^{P} \pi_{pjk} \cdot a_{pijk} + e_{ijk}$$

Level-2 model: schools

$$\pi_{ojk} = \beta_{00k} + \sum_{q=1}^{Q_p} \beta_{pqk} \cdot X_{qjk} + r_{0jk}$$

$$\pi_{1jk} = \beta_{10k} + r_{1jk}$$

Level-3 model: countries

$$\beta_{00k} = \gamma_{000} + \sum_{s=1}^{S_{pq}} \gamma_{pqs} W_{sk} + u_{00k}$$

$$\beta_{10k} = \gamma_{100} + u_{10k}$$

This three-level hierarchical linear model was estimated using the commercial software HLM 6.08 (developed by Raudenbush, Bryk and Congdon). The model coefficients and statistics were estimated using a full maximum likelihood procedure.

Fixed/random effects and centring

In the models for examining the association of selected system and school-level variables on science performance, all slopes were fixed and only the intercept was randomised at all three levels. All variables including both background and explanatory variables were centred on the grand mean. Thus, in all models, the intercept is to be interpreted as the achievement score in science for a student who has the international mean in all variables included in the model.

Significance Tests

Throughout the multilevel analysis, an effect was considered statistically significant if the p-value was below 0.1 at country level and below 0.005 at school level. Different criterion values were chosen for the two levels to balance between significance and statistical power.

Student weights

For the multilevel analysis, weights were computed based on the student final weights in the PISA 2006 dataset. The student final weights contained in the PISA dataset (W_FSTUWT) were normalised at the international level including 29 participating countries to:

i) make the sum of the weights across the 29 countries equal to the number of students across the 29 countries in the dataset;

ii) maintain the same proportion of weights as in the student final weights (W_FSTUWT) within each country; and

iii) ensure that each individual country's contribution to the analysis is equal by introducing a country factor (*i.e.* the sum of the weights within each country is the same for all 29 countries).

Treatment of missing data

The proportion of missing values for the variables considered in the analysis is presented in Table 1, column 8. Even though the missing rate was less than 5% for most of the variables, list-wise deletion of all observations that have a missing value for at least one variable would have reduced the sample size by 28.21%, since more than 30 variables were included in the models. Therefore, missing values were imputed in order to include the maximum number of cases in the analysis.

Since the missing rates were not high for most of the variables, a simple imputation approach was used to circumvent the problem of missing data: predictors at the individual and school level were imputed using a dummy variable adjustment (Cohen and Cohen, 1985). Missing values for system-level variables were replaced by imputed values.

It is known that this imputation method generally produces biased estimates of coefficients (Jones, 1996), and that standard errors are underestimated since they do not account for the uncertainty introduced through imputation. However, given the fact that for only 1 out of 33 variables, more than 5% of the data were missing (Table 1, column 8), this bias was considered negligible.

A missing dummy variable was created for all variables with missing values regardless of whether a variable was continuous, categorical or dichotomous, which was set to 1 if the data were missing on that variable and it was set to 0 if the data were not missing.

Then, for continuous variables, missing values were replaced by the weighted school average of the variable. If all data on the respective variable were missing in one school such that the weighted school mean could not be computed, the weighted country mean was imputed. If all data on the respective variable were missing in a country, the weighted international mean was imputed. When a missing value was replaced by the country or school mean, the weights were proportional to the sampling probability (weighting factor W_FSTUWT from the PISA 2006 dataset). When a missing value was replaced by the international mean, each country was given an equal weight. As for categorical variables, missing values in dummy and variables were replaced by 0.