

Program on Education Policy and Governance Working Papers
Series **Cross-Country Evidence on Teacher
Performance Pay** Ludger Woessmann Ludwig-Maximilians Universität
München

Ifo Institute for Economic Research
PEPG 10-11 **Prepared for the PEPG Conference Merit Pay: Will It
Work? Is It Politically Viable?** Harvard Kennedy School Cambridge,
Massachusetts

**June 3-4,
2010**

*Cross-Country Evidence on Teacher Performance Pay

†Ludger Woessmann University of
Munich and Ifo Institute

Preliminary version – Do not cite or circulate without author's permission

Abstract The general-equilibrium effects of performance-related teacher pay include long-term incentive and teacher-sorting mechanisms that usually elude experimental studies but are captured in cross-country comparisons. Combining country-level performance-pay measures with rich PISA-2003 international achievement microdata, this paper estimates student-level international education production functions. The use of teacher salary adjustments for outstanding performance is significantly associated with math, science, and reading achievement across countries. Scores in countries with performance-related pay are about one quarter standard deviations higher. Results avoid bias from within-country selection and are robust to continental fixed effects and to controlling for non-performance-based forms of teacher salary adjustments.

Keywords: teacher performance pay, student achievement, international comparison, PISA

JEL classification: I20, J33

May 14, 2010

· Paper prepared for the Harvard PEPG Conference on Merit Pay in June 2010. Hospitality provided by the W. Glenn Campbell and Rita Ricardo-Campbell National Fellowship of the Hoover Institution, Stanford University, when working on this paper during the first half of 2010 is gratefully acknowledged.

† University of Munich and Ifo Institute for Economic Research, Poschingerstr. 5, 81679 Munich, Germany; CESifo and IZA; woessmann@ifo.de.

1. Introduction

¹Educational reform discussions around the globe regularly include the idea of paying teachers based on how they perform in advancing their students' educational achievement. From a theoretical viewpoint, performance-related pay may elicit both incentive effects – raising motivation and effort of current teachers who strive to increase their pay – and sorting effects – attracting graduates into the teaching profession who expect to do well under performance-based compensation schemes.² Given recent evidence that teacher quality is very important for student achievement but unrelated to most observable teacher characteristics, motivating teachers to perform well and attracting a pool of high-performing teacher applicants are promising policies to improve student outcomes. This paper provides international evidence whether variation in the use of performance-related pay for teachers is associated with student achievement across countries.

³The best available evidence on the effects of teacher performance pay stems from a set of recent experimental studies performed in Israel, Kenya, and India.⁴ However, there are obvious limits to what we can learn from such experimental studies. Because of their short-term character, experimental studies are bound to miss any sorting effect into the teaching profession, which is “perhaps the most important aspect of compensation effects” (Lazear (2003), p. 186). In addition, while experimental studies clearly identify the short-term incentive effect for current teachers, long-term incentive effects may differ. On the one hand, time required to accustom oneself to incentives contracts may render the long-term incentive effect larger than the identified short-term incentive effect. On the other hand, adjustments to the detriment of nonincentivized teaching tasks as well as regular experience of failure may render it smaller. Furthermore, it is unclear to what extent the available experimental evidence generalizes to

¹ See Lazear (2003). Besley and Ghatak (2006) model sorting when teachers are motivated by a mission. Falk and Dohmen (forthcoming) provide evidence that pay schemes may affect who sorts into the teaching profession.

² See Rockoff (2004) and Rivkin, Hanushek, and Kain (2005) for seminal contributions and Hanushek and Rivkin (2010) for an overview of recent value-added studies on the contribution of teachers to student achievement.

³ See Lavy (2002, 2009), Glewwe, Ilas, and Kremer (2008), and Muralidharan and Sundararaman (2009).

⁴ See Holmstrom and Milgrom (1991) and Dixit (2002) for theoretical expositions. Ballou (2001) presents evidence that performance-related pay is more common in private than in public schools, arguing that multitask problems are thus unlikely to be a feature that makes performance-related pay inherently futile in teaching.

developed countries. The available non-experimental evidence for developed countries⁵ has the additional interpretation problem that better schools may choose to adopt teacher performance pay and that teachers and students may select themselves into schools on the basis of whether there is performance-related pay or not, possibly introducing substantial selection bias.

⁶One way of capturing long-term effects of teacher performance pay, including generalequilibrium effects of attracting different pools of teachers, while at the same time getting around within-system selection bias, is to draw on cross-country variation to compare countries that do and do not use teacher performance pay. A number of countries have extensive experience with performance-related adjustments of teacher salaries, whereas other countries do not.⁷ We combine country-level indicators of the use of performance-related pay with the micro database of the 2003 Programme for International Student Assessment (PISA) to estimate international education production functions that account for a rich set of family, school, and institutional determinants of student achievement at the individual, school, and country levels. Based on data for 190,000 students who form a representative sample of 15-year-olds in 28 OECD countries, the analysis provides descriptive evidence on the association between the use of salary adjustments for teacher performance (at the national level) and student achievement on an internationally comparable test. While the list of topics analyzed using international education production functions is long – including the effects of family background, resources, teacher salary levels, accountability, school autonomy, choice, tracking, and more – to our knowledge this is the first cross-country econometric study to address the topic of teacher performance pay.

We find that students in countries that adjust teacher salaries for outstanding performance in teaching perform about 25 percent of a standard deviation higher on the international math test than students in countries without teacher performance pay, after controlling extensively for student, school, and country measures. Similar associations are found for reading achievement, and somewhat smaller, but still substantial and significant associations for science achievement.

⁵ See Figlio and Kenny (2007) and Atkinson et al. (2009) for recent examples from the United States and England. Murnane and Cohen (1986) provide an early and Podgursky and Springer (2007) a recent overview of the available U.S. evidence.

⁶ Our analyses employ the information collected by the Organisation for Economic Co-operation and Development (2006) in a survey of member countries. See also Sclafani and Tucker (2006) for a review of international experiences.

⁷ See Hanushek and Woessmann (forthcoming) for an extensive overview of international education production functions.

While there are only limited signs of effect heterogeneity overall, there is some indication that the association is stronger where parents have higher socioeconomic background, but also for students who speak a foreign language at home. Tertiary degrees in pedagogy of teachers are positively associated with student achievement in countries without teacher performance pay, but not in countries with teacher performance pay.

Using international variation for the identification of effects of teacher performance pay has several advantages over national studies, but also clear limitations. At the most basic level, cross-country evidence can exploit institutional variation that is usually not available, or only available at a much more limited scale, within single countries. A main additional advantage is that cross-country evidence is able to capture the long-run, general-equilibrium effects of teacher performance pay that often elude studies in a single country. A second additional advantage is that by using measures of teacher performance pay aggregated at the country level, cross-country evidence circumvents the selection issues that usually plague within-country identification.

Obvious limitations of cross-country identification relate to the limited number of independent observations available at the country level and the possibility of measurement error in national reported data on the use of teacher performance pay. More fundamentally, possible issues of omitted country variables and reverse causation make identification less clean than in randomized experiments. For example, countries with low student achievement may feel pressure to introduce teacher performance incentives, which would bias the estimated effect of teacher performance pay downward. More generally, the use of teacher performance pay may be related to other unobserved country factors, such as general cultural features or the use of pay adjustment mechanisms more generally, that are themselves related to student achievement.

We implement a series of specification tests to check whether some of these remaining concerns with cross-country identification are likely to be a major source of bias in our estimates. In particular, to address issues of divergent national cultures around the world, we show that our results are robust to – or even stronger in – specifications that include continental fixed effects or restrict the analysis to European countries only. To address issues of possible correlation with the general prevalence of more flexible teacher pay schemes, the international database allows us to control for a rich set of indicators of the availability of teacher salary adjustments for other, non-

⁸ Hanushek and Woessmann (forthcoming) provide a more general discussion of advantages of and concerns with the use of cross-country data on educational achievement.

performance-related criteria, such as based on teaching conditions and responsibilities, teachers' qualifications and training, and demographics. Hardly any of these other forms of teacher salary adjustments is significantly related to student achievement, and none of them affects the qualitative result on teacher performance pay. Apart from these falsification tests, all our models control for a rich set of about 50 family, input, and institutional factors measured at the student, school, and national level that are usually unavailable in national datasets. While these analyses rule out the most obvious sources of bias in the international identification, the limits to the interpretation of results based on an international cross-section of student achievement data mentioned above require a cautious interpretation of the reported results as descriptive associations indicative of areas of possible important effects in educational production. The remainder of the paper is structured as follows. The next section introduces the

international data on teacher performance pay and the micro database of the PISA international student achievement test. Section 3 describes and discusses the empirical model. Section 4 presents the results of our estimations together with robustness and specification tests. Section 5 concludes.

2. Data

2.1 Cross-Country Measures of Teacher Performance Pay In 2003, the Organisation for Economic Co-operation and Development (OECD) for the first

time executed a survey among member countries about whether the base salary for public-school teachers could be adjusted based on an extensive list of criteria (see Organisation for Economic Co-operation and Development (2006), pp. 374-394, for details). One of these criteria related to salary adjustments awarded to teachers with "outstanding performance in teaching". Among the total of 16 possible criteria for salary adjustments, the Organisation for Economic Co-operation and Development (2006), p. 382, referred to this payment as "the only additional payment that could be classified as a performance incentive."

⁹ To identify causal effects of such institutional features as private-school competition and tracking in international education production functions, West and Woessmann (forthcoming) and Hanushek and Woessmann (2006) employ historical instrumental variables and between-grade difference-in-differences, respectively. Obvious similar sources of identification are not available in the case of teacher performance pay.

¹⁰The data were collected in the framework of the OECD's International Indicators of Education Systems (INES) survey. INES is an ongoing project among member countries explicitly set up to develop and collect a system of indicators for cross-national comparisons in education, mounting own cross-national surveys. In an interactive process, member countries design indicators, negotiate the definitions for data collection, and supply data for their respective country. The data on teacher bonuses was collected within the system-level data collection of INES Network C, whose 2005 Survey on Teachers and the Curriculum included a battery of questions on teacher salary adjustments, referring to formal policies for public institutions in the school year 2003/04.¹¹ National reporting by each member country is based on a mixture of sources including national ministries and statistical offices, law and policy documents, communications with associations of teachers and of local authorities, and sometimes specific teacher questionnaires. Efforts by the OECD and the fact that each member country inspects and approves the collected indicators before publication ensure a high degree of international comparability of the collected data.

Among the 28 OECD countries with available data in our database, thirteen countries report having adjustments of teacher salaries based on outstanding performance in teaching (see Table 1). The form of incentive and the method for identifying outstanding performance varies across countries. For example, in Finland, according to the general agreement for teachers, local authorities and education providers have an opportunity to encourage individual teachers in their work by personal cash bonuses on the basis of individual professional proficiency and performance at work. Outstanding performance may also be measured based upon the assessment of the head teacher (Portugal), assessments performed by education administrations (Turkey), or the measured learning achievements of students (Mexico).

Using information from the 2003 INES survey on locus of decision-making, the OECD also differentiates performance pay by whether it is awarded at the school level (by the head teacher

¹⁰ The definition of salary adjustments is as follows (Organisation for Economic Co-operation and Development (2006), p. 383): "An adjustment to base salary is defined as any difference in salary between what a particular teacher actually receives for work performed at a school and the amount that he or she would be expected to receive on the basis of level of experience Adjustments may be temporary or permanent, and they can effectively move a teacher off the scale and onto a different salary scale or onto a higher step on the same salary scale."

¹¹ For example, the United States conducted a sample survey of self-administered questionnaires to a representative sample of public school teachers. See Annex 3 to Organisation for Economic Co-operation and Development (2006) for details on national interpretation, sources, and references.

or school principal), by local or regional government authorities, or by the national government authority. In seven of the thirteen countries with performance pay, the decision is made at the school level, in five countries at the level of local or regional authority, and in three countries at the national level (Table 1).

As the question on performance pay was part of a more general questionnaire on possibilities for salary adjustments, we have information on the availability of teacher salary adjustments on the basis of a wide battery of additional, non-performance-related criteria, which we will use in part of our analyses below. In particular, there is information on seven types of criteria based on teaching conditions and responsibilities (such as taking on management responsibilities, teaching additional classes, and teaching in particular areas or subjects), five types of criteria related to teachers' qualifications and training (such as holding initial qualifications or later certification above the minimum required level and completion of professional development activities), two types of criteria based on demography (family status and age), and a residual category. Table 1 reports the share of available adjustments in a country in each of the three broad areas of categories (additional detail is provided below). Additional payments are most often given for particular responsibilities or working conditions.

Any cross-national collection scheme of data on teacher compensation schedules is bound to be subject to measurement error. Despite the extensive efforts of the OECD to ensure comparability and the authorization of the national reported data by all member countries, interpretation of what constitutes a performance pay scheme may differ across countries. The concept of "outstanding performance in teaching" seems vague enough to allow for differing interpretations, and in some countries, performance pay may be available in some regions or schools but not in others. In addition, information was collected only on the availability of payment adjustments, but not on payment amounts, which may differ substantially across countries. In econometric applications, classical measurement error tends to introduce attenuation bias, making it harder to identify statistically significant associations. The fact that the information was collected on a survey whose primary objective was not to ask about performance related pay, but salary levels and adjustments more generally, makes the eliciting of political-economy type reactions of under- or over-reporting of performance pay less – and the assumption of classical measurement error more – likely. In robustness specifications, we will also ensure that results are not driven by any one particular country.

2.2 The Programme for International Student Assessment (PISA) Micro Database We combine the performance pay data with the extensive student-level database on student achievement and background from the 2003 Programme for International Student Assessment (PISA), also conducted by the OECD. The PISA study provides comparable information on student achievement in mathematics, science, and reading for all OECD member countries (see Organisation for Economic Co-operation and Development (2004) for details). The target population was the 15-year-old students in each country, regardless of the grade they currently attended. Thus, in most participating countries, the target population consists of individuals nearing the last stages of compulsory schooling. Most of our analyses look at math achievement, which with about 70 percent of testing time was the focus of the 2003 PISA test, which had less detailed testing in science and reading. In contrast to earlier international studies that follow a curriculum-based testing approach, the questions in the three PISA literacy domains aim to test how well students are prepared to meet the real-life challenges of modern societies. Using item response theory, PISA mapped performance in the three subjects on a scale with an international mean of 500 and a (student-level) standard deviation of 100 test-score points across the OECD countries.

¹²Importantly for use in cross-country comparisons, the PISA sampling procedure ensured that a representative sample of the target population was tested in each country. The student sample sizes in the OECD countries range from 3,350 students in 129 schools in Iceland to 29,983 students in 1,124 schools in Mexico, yielding an international dataset of nearly 200,000 OECD country students. We weight all regressions by sampling probabilities to obtain consistent student population estimates within each country. Across countries, each country carries the same aggregate weight. This weighting scheme ensures that each system-level observation of teacher performance pay contributes equally to the estimation.

The PISA 2003 database also contains a host of background information on the participating students and schools that will be used as control variables in our international education production functions. Separate background questionnaires completed by students and by school principals provide detailed information on students' demographic characteristics, their family backgrounds, and their home environments, as well as school characteristics such as resource

¹² Most countries employed a two-stage sampling design, drawing a stratified random sample of schools in a first stage and then randomly testing 35 students in each school in a second stage.

¹³endowments, location, and institutional characteristics such as aspects of accountability, autonomy, and choice. We supplement this rich student-level and school-level database from PISA 2003 with additional country-level data on GDP per capita in 2003, cumulative educational expenditure per student between age 6 and 15 in 2002, and the existence of curriculum-based external exit exams from several international statistical sources (see Woessmann, Luedemann, Schuetz, and West (2009) for details). In addition, we obtain data on teacher salary levels in 2004 from the Organisation for Economic Co-operation and Development (2006). Combining the available data, we constructed a dataset containing over 190,000 students in

¹⁴28 of the 30 OECD countries. The dataset combines students' test scores in math, science, and reading with students' characteristics, family-background data, and school characteristics. We imputed missing observations on the PISA questionnaire items using the conditional mean imputation method (see Little and Rubin (1987)) described in Woessmann, Luedemann, Schuetz, and West (2009), Appendix B.3. To ensure that our results are not biased by the inclusion of imputed observations, all regression specifications include separate dummy variables identifying observations with imputed data on each variable as well as interactions between these dummies and the relevant variable.

Table A.1 in the Appendix reports international descriptive statistics for the variables employed in our analyses. Table 1 presents country means of selected variables for each participating country. For additional details on the PISA test and the database used in this paper, see Woessmann, Luedemann, Schuetz, and West (2009).

3. Empirical Model and Cross-Country Identification

Our basic model is a standard education production function that expresses student achievement as a function of a set of input factors:

¹³² of 0.94. Salary data are missing for the Slovak Republic. Specifically, we use a measure of statutory public-school teacher salary after 15 years of experience and minimum training, averaged over primary and lower secondary education, expressed in purchasing power parity. Similar results are obtained when using just the measure for lower secondary education, or when using starting salaries or salaries at the top of the scale. Lower-secondary education salaries in Turkey, missing in the data source, are imputed from a regression of salaries in lower-secondary salaries on salaries in primary education, which has an R

¹⁴ France had to be dropped from the sample because no school-level background information was provided. Canada had to be dropped because it did not provide data in the OECD survey of teacher salaries and salary adjustments.

$$CRBPT e\beta\beta\beta\beta +++++=iscscisciscisc43210 \quad (1)$$

is the achievement test score of student i in school s in country c . The variable of interest in this paper is P_c , an indicator of whether country c has performance-related teacher pay. The production function holds constant potential influence factors in the area of student-level family background B , school-level resources R , and country-level measures of expenditure and institutions C . The zero-mean error term e_{ics} is adjusted to allow for the clustering of observations by country, to take account of the fact that despite the student-level model, our variable of interest varies only at the country level.

We take the specification of the control model directly from Woessmann, Luedemann, Schuetz, and West (2009), which was also employed in West and Woessmann (forthcoming). The rich set of control factors is enlisted in Table A.1 in the Appendix. Specifically, the vector of student and family background characteristics consists of 31 variables, including such indicators as the student's gender and age, attendance of institutions of pre-primary education, immigration status, family status, and parental occupation and work status. The vector of variables measuring school resources and location comprises nine variables such as class size, availability of materials, instruction time, teacher education, and size of community location. The vector of country-level variables includes per-capita GDP of the country, average expenditure per student, external exit exams, school autonomy in budget and in staffing decisions, the share of privately operated schools, and the share of government funding for schools. The latter are measures of the institutional structure of national school systems that have been shown in previous work to be significantly associated with student achievement (see Hanushek and Woessmann (forthcoming)).

Our models also include a control for the teacher salary level. While teacher salaries tend to be subsumed in the general measure of educational expenditure in most of the literature – given that teacher salaries are the largest single cost component in education – holding teacher salary levels constant seems particularly relevant when focusing on possible effects of bonuses. Positive associations of teacher salary levels with student achievement have been reported in international education production functions estimated both at the country level (Lee and Barro (2001) and Dolton and Marcenaro-Gutierrez (2010)) and at the student level (Woessmann (2005)).

Maybe the main concerns for identification in national education production functions – omitted student-level ability and non-random selection of students and teachers into schools and classrooms – are bypassed in international education production functions by measuring the variable of interest at the country level. Within-country variation in the use of teacher performance pay may be the outcome of choices of parents, administrators, and schools that are correlated with the error term of the production function, despite the inclusion of a rich set of observable factors (Figlio and Kenny (2007)). In cross-country analyses, we can aggregate the performance-pay variable to the country level, thereby using only between-country variation and circumventing the selection problem that better schools within a system may adopt teacher incentives. Such cross-country analysis cannot be biased by standard issues of selection at the individual level, as patterns of sorting cancel out at the system level. Under the assumption that ability is uniformly distributed across OECD countries, the same is true for student ability.

The main concerns with identification in international education production functions stem from different kinds of omitted-variable and selection bias: system-level features that are correlated with both the adoption of performance pay and student achievement (see Hanushek and Woessmann (forthcoming) for a general discussion). In particular, performance-related pay may be associated with inherent cultural difference across nations that might be related to attitudes and performance in learning, and it may be more likely in countries that use flexible salary scales more generally also in other dimensions than teacher performance, such as teaching conditions and teacher qualifications.

One way to assess the severity of the first concern is to restrict the identifying variation to groups of countries that are culturally more homogeneous. We will thus estimate a robustness specification that includes continental fixed effects, as well as a specification that restricts the country sample to European countries only. Also note that any cultural features that could give rise to bias would have to be unrelated to the country-level controls in the model (as well as the rich set of student-level and school-level controls). That is, the cultural variation would have to be unrelated to the country's level of economic development and educational expenditure and to the adoption of other achievement-enhancing institutional measures such as external exams, school autonomy, and private-sector competition.

A way to address the second concern – that countries that adopt performance-related pay may be more likely to have more flexible salary schedules also in other dimensions, which may

in turn be related to student outcomes – is to take into account direct information on alternative forms of salary adjustments. We can draw on the rich information on non-performance-related pay adjustments outlined above to test whether any association of performance-related pay with student achievement just captures its relation with other forms of adjustments in teacher pay. Below, we will thus run a set of specification tests that control for several indicators of salary adjustments in such other areas as teaching conditions and responsibilities, teacher qualifications and training, and demography. Note that such specifications will also ensure that results are not driven by country-specific measurement error in the reported salary-adjustment data that are common to the different adjustment questions in the same survey.

While none of these specification checks can ensure strict exogeneity of our performancepay measure beyond doubt, they can rule out the most obvious sources of possible bias in the cross-country identification. A remaining concern of reverse causation, in that poor achievement could lead some countries to introduce performance-related pay as a possibly remedy, would lead to downward bias in our estimates.

4. Results

4.1 Basic Results Table 2 reports our basic results on country-level institutional measures in international

education production function. The dependent variable is student-level achievement on the PISA math test. As a benchmark comparison to the existing literature, column (1) replicates the basic model of Woessmann, Luedemann, Schuetz, and West (2009), Table 2.1, only eliminating Canada (which did not provide teacher salary information in the OECD survey) from the analysis. The estimation yields the same qualitative results on the institutional measures of external exams, school autonomy, and private vs. public operation and funding of schools.

Column (2) adds our measure of performance-related pay to this international education production function. The indicator of availability of teacher performance pay in a country enters statistically significantly positive in the model. Column (3) adds the teacher salary level as an additional control variable, which also enters significantly positive. The coefficient on teacher

¹⁵performance pay increases in size and statistical significance when salary levels are controlled for. The full control model of this specification is reported in Table A.1 in the Appendix.¹⁶ The basic result suggests that student achievement is significantly higher in countries that

¹⁷make use of teacher performance pay than in countries that do not use it, after extensively controlling of student, school, and country background measures. On average, students in countries with performance-related pay score 24.8 percent of a standard deviation higher on the PISA math test. Compared to the existing experimental studies, this is very similar to the effect size found by Muralidharan and Sundararaman (2009) in India and about twice as large as the one found by Lavy (2009) in Israel. Figure 1 depicts our main result graphically. It shows an added-variable plot of test scores

against teacher performance pay, after controlling for our extensive set of controls and then collapsing to the country level. It shows a clear positive association between the variation in country-average test scores that remains unexplained by the rich set of control variables and the variation in teacher performance pay that cannot be attributed to variation in the other observed determinants.

The figure also indicates that the main association is not driven by any single country or outlying observation. In fact, as a sensitivity test of the international identification, we also performed a robustness check that dropped one country at a time from the analysis. The qualitative result is robust in all estimations with reduced samples, confirming that the observed positive association between student achievement and performance-related pay is not driven by any particular country.

Columns (4) to (6) of Table 2 enter the performance pay measure separately by locus of decision-making on the salary adjustment. In the full sample of countries, the effect is relatively similar and independent of whether decisions are made at the school level, at the level of local or regional authorities, or at the national level. Even though – with the exception of Finland – the

¹⁵ The increase in the coefficient estimate on teacher performance pay is not driven by the change in sample due to missing data on teacher salaries in the Slovak Republic. Estimating the model of column (2) without the Slovak Republic – that is, on the 27-country sample of column (3) – yields a coefficient estimate on teacher performance pay of 16.831.

¹⁶ The general results on the other institutional measures are robust to the inclusion of the performance pay measure, with the exception of the coefficient on autonomy in staffing decisions, which becomes smaller and loses statistical significance.

¹⁷ In Kenya, Glewwe, Ilas, and Kremer (2008) find only temporary test-score effects on exams linked to the incentives, but not on exams that were not linked to the incentives.

three groups of countries are non-overlapping, the estimates across columns (4) to (6) are surprisingly close to each other. Note, however, that country observations with performance pay get relatively small (five and three, respectively) in the specifications of columns (5) and (6). Robustness specifications (not shown) reveal that significance in these latter two categories is sensitive to having Finland in the model, which is not the case in the specifications of columns (3) and (4). Thus, while the limited degrees of freedom in the cross-country identification do not allow for a conclusive evaluation of the locus of decision-making, they lend more confidence in a positive association with teacher performance pay when decisions are made at the school level than when they are made at a higher, administrative level.

4.2 Checks on the Specification As argued above, some of the main threats to identification in the cross-country setting come

from omitted country-level factors of culture and of other, non-performance-based salary adjustments. With respect to concerns of unobserved cultural differences, note that these would have to be unrelated to the included control variables, some of which – such as levels of teacher salaries and educational spending and measures of performance-conducive institutional structures – are obvious correlates of possible cultural factors that may be correlated with economic and educational performance. As a first attempt to additionally reduce heterogeneity, column (1) of Table 3 drops two countries – Mexico and Turkey – with particularly low levels of GDP per capita, which also stand out by having an average socioeconomic status that is a full standard deviation below the OECD average (as measured by the PISA index of economic, social, and cultural status). Dropping these countries hardly affects results.

One way to account for unobserved broad cultural heterogeneity is to fully exclude any variation that exists between continents by including continental fixed effects. Column (2) thus includes fixed effects for the four continents with OECD countries – Europe, North America, Oceania, and East Asia. The results show that drawing only on variation within continents actually increases the estimated coefficient on teacher performance pay. The basic result therefore does not just capture cultural differences between major world regions. Note also that the explained variance hardly increases when adding the continental fixed effects (relative to model (3) in Table 2), suggesting that the control variables contained in the basic model capture the cross-continental variation very well.

Another possibility to address issues of major cultural differences is to restrict the sample to European countries only, as Europe is the one world region with enough OECD country observations to allow for identification. The estimated coefficient on performance-related pay again increases when identification is restricted to the sample of 21 European countries (column (3) of Table 3).

To test whether the association between performance-related pay and student achievement only captures compensation schemes that are more flexible in general, Table 4 reports results of a set of specifications that control for the availability of other, non-performance-related salary adjustment mechanisms. In this table, each row represents the results of a separate regression. The first column reports the coefficient on teacher performance pay, and the second column reports the coefficient on the other type of salary adjustment that is controlled for in the respective row. The first three rows control for aggregate measures of salary adjustments in three alternative categories – teaching conditions and responsibilities, teacher qualifications and training, and demography. The aggregates are simple averages of the sub-categories available in each category, as reported in the bottom part of the table.

None of the three measures of salary adjustments for alternative, non-performance-related criteria enters the model significantly. The coefficient estimate on performance-related pay remains statistically significant and at the same order of magnitude. The positive association between performance-related pay and student achievement thus does not just capture effects of countries that are more prone to salary adjustments more generally, be it for teaching conditions, teacher qualifications, or demographic features.

The additional rows of Table 4 report results of specifications that control for indicators of each alternative salary adjustment mechanism separately. The estimated coefficient on performance-related pay remains significantly positive in every single specification. Furthermore, only the existence of salary adjustments for taking on management responsibilities in addition to teaching duties (and the residual “other” category of salary adjustments) enters the model statistically significantly. None of the other types of salary adjustment is statistically significantly associated with student achievement, including salary adjustments based on such criteria as teaching additional classes or hours, teaching in specific locations, taking on special activities, teaching students with special educational needs, teaching specific subjects, holding a

higher educational qualification or certification than required, completion of professional development activities, family status, and age.

4.3 Effect Heterogeneity and Additional Subjects The results so far estimate the average association of teacher performance pay with student

achievement in the whole population. Table 5 reports results of a specification that allows for the association to differ depending on teacher characteristics and basic student characteristics. All coefficients in the table stem from one joint regression that contains interaction terms between the country-level indicator on performance-related pay and the variables reported in the lead column. The first column reports estimates on the main effect of each variable, which represents the association of student achievement with the respective variable in countries that do not have performance-related pay. The second column reports estimates on the interaction term between performance-related pay and each variable, which represents the difference in the association of student achievement with the respective variable between countries without and with performance-related pay.

¹⁸The association between student achievement and teacher salary levels is slightly larger in countries with performance-related pay, but the difference to countries without performance-related pay is not significant. Likewise, the positive association between student achievement and the share of fully certified teachers in the school does not vary significantly by performance pay status. However, while the association between student achievement and the share of teachers in that school who hold a tertiary degree in pedagogy (the only measure of teachers' initial education available in the PISA database) is significantly positive in countries without performance-related pay, it turns to basically zero in countries with performance-related pay. The difference between the two groups of countries is highly significant. That is, teachers' educational degrees seem to matter when there are no performance incentives, but with performance incentives, teachers with lower educational degrees perform just as well as teachers with higher educational degrees on average.

¹⁸ Unfortunately, the PISA database does not provide teacher-level information, but only school-level information, and does not provide information on additional teacher characteristics. Woessmann (2003) uses the individual teacher data on education, experience, age, and gender available in the Trends in International Mathematics and Science Study (TIMSS) in international education production functions.

With respect to basic measures of student characteristics, there are small but statistically significant interaction effects of teacher performance pay with a general index of socioeconomic background and an indicator of books in the student's home, suggesting that the effect of performance-related pay may be slightly larger where parents have a higher socioeconomic background. At the same time, the negative association of student achievement with students' speaking a foreign language at home is significantly lower (in absolute terms) in countries that have performance-related pay. No similar effect heterogeneity is found for the majority of other student and family background measures contained in the model (as shown in Appendix Table A.1).

While all results so far refer to math achievement, Table 6 reports results for student achievement in science and reading. Teacher performance pay is significantly positively related to achievement in all three subjects. At 24.3 percent of a standard deviation, the coefficient estimate in reading is very similar to math. The coefficient estimate in science is somewhat smaller, at 15.4 percent of a standard deviation. The estimates increase to 29.9 and 24.1 percent of a standard deviation in reading and science, respectively, when including controls for continental fixed effects (columns (2) and (4)).

5. Conclusion

This paper provides first evidence on the association between the use of performance-related teacher pay and student achievement across countries. The results of cross-country education production functions that extensively control for student, school, and country background factors suggest that students in countries that make use of teacher performance pay perform significantly better in math, science, and reading than students in countries that do not use teacher performance pay. The size of the association between use of performance-related pay and student achievement is about one quarter of a standard deviation on the math and reading tests and about 15 percent of a standard deviation on the science test. Given the well-established nature of teacher bonuses in many countries, these associations are likely to capture general-equilibrium effects of performance-related pay which are likely to combine long-term incentive effects of motivating current teachers with sorting effects of changing the pool of applicants for the teaching profession.

Several specification tests rule out some of the most obvious threats to a causal interpretation of the association. The result is robust in specifications with continental fixed effects and within the sample of European countries. In both specifications, point estimates are actually larger, suggesting that omitted broad cultural patterns do not drive the results. Specifications that control for a large set of alternative salary adjustment mechanisms, not based on performance but based on such criteria as teaching conditions, teacher qualifications, and teacher demographics, confirm that the main result does not just capture effects of more flexible pay scales in general, but is specific to performance-based salary adjustments. In fact, apart from salary adjustments for taking on management responsibilities, no other specific adjustment scheme is significantly associated with student achievement.

The provided cross-country evidence is complementary to existing national, especially experimental estimates. While the cross-country results include long-term general-equilibrium effects that usually escape experimental studies, the latter provide more confidence in the exogeneity of treatment. Furthermore, cross-country estimation only allows for identification of general patterns, pointing towards the stylized fact of a basic association between teacher performance pay and student achievement. Given the limited degrees of freedom at the country level, the identification strategy does not have much to say about the specific design of performance-pay schedules or about issues of implementation. Many of these issues will have country-specific answers. Combining analyses based on different – national and international – approaches to identification helps to deepen our understanding of the effects of teacher performance pay.

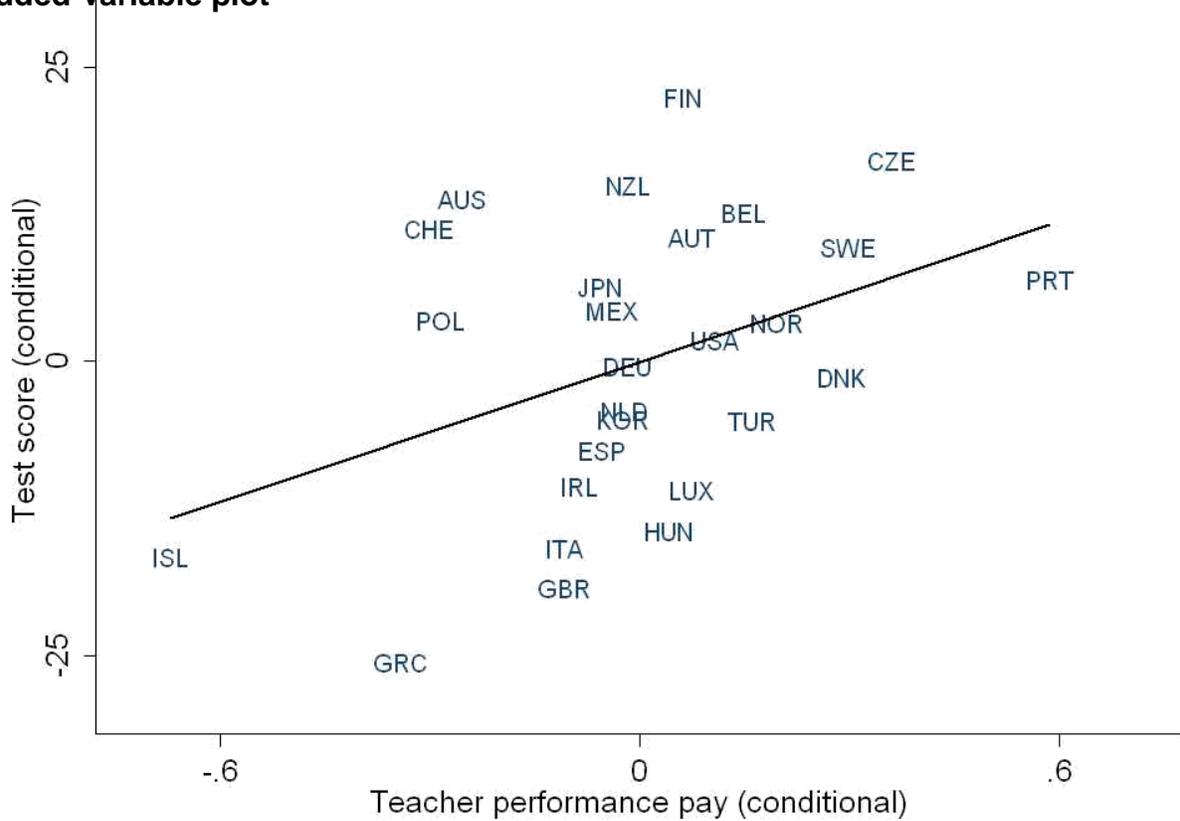
Many aspects of the cross-country association between student achievement and teacher performance pay remain open for future research. More work is needed to identify the causal relationships underlying the associations reported in this paper. The relevance of specific aspects of the performance pay scheme, such as the role of measurement of teacher performance, the size of the bonus, and the locus of decision-making, deserve further attention. Maybe most demanding of all, the current work does not speak to the relative importance of incentive and sorting effects in the overall effect of teacher performance pay, an issue crucial to understand the mechanisms of the long-run effects of performance-related teacher pay.

References

- Atkinson, Adele, Simon Burgess, Bronwyn Croxson, Paul Gregg, Carol Propper, Helen Slater, and Deborah Wilson. 2009. "Evaluating the impact of performance-related pay for teachers in England." *Labour Economics* 16, no. 3: 251-261.
- Ballou, Dale. 2001. "Pay for performance in public and private schools." *Economics of Education Review* 20, no. 1 (February): 51-61.
- Besley, Timothy, and Maitreesh Ghatak. 2006. "Sorting with motivated agents: Implications for school competition and teacher incentives." *Journal of the European Economic Association* 4, no. 2-3: 404-414.
- Dixit, Avinash. 2002. "Incentives and organizations in the public sector: An interpretative review." *Journal of Human Resources* 37, no. 4 (Fall): 696-727.
- Dolton, Peter, and Oscar D. Marcenaro-Gutierrez. 2010. "If you pay peanuts do you get monkeys? A cross country analysis of teacher pay and pupil performance." Mimeo. London: Royal Holloway College, University of London.
- Falk, Armin, and Thomas Dohmen. forthcoming. "You get what you pay for: Incentives and selection in the education system." *Economic Journal*.
- Figlio, David N., and Lawrence W. Kenny. 2007. "Individual teacher incentives and student performance." *Journal of Public Economics* 91, no. 5-6 (June): 901-914.
- Glewwe, Paul, Nauman Iltis, and Michael Kremer. 2008. "Teacher incentives." Revised version of NBER Working Paper 9671. Cambridge, MA: Harvard University.
- Hanushek, Eric A., and Steven G. Rivkin. 2010. "Generalizations about using value-added measures of teacher quality." *American Economic Review* 100, no. 2 (May): forthcoming.
- Hanushek, Eric A., and Ludger Woessmann. 2006. "Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries." *Economic Journal* 116, no. 510 (March): C63-C76.
- . forthcoming. "The economics of international differences in educational achievement." In *Handbook of the Economics of Education, Vol. 3*, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Amsterdam: North Holland.
- Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design." *Journal of Law, Economics and Organization* 7, Spec. Iss.: 24-52.
- Lavy, Victor. 2002. "Evaluating the effect of teachers' group performance incentives on pupil achievement." *Journal of Political Economy* 110, no. 6 (December): 1286-1317.
- . 2009. "Performance pay and teachers' effort, productivity, and grading ethics." *American Economic Review* 99, no. 5 (December): 1979-2011.

- Lazear, Edward P. 2003. "Teacher incentives." *Swedish Economic Policy Review* 10, no. 3: 179-214.
- Lee, Jong-Wha, and Robert J. Barro. 2001. "Schooling quality in a cross-section of countries." *Economica* 68, no. 272: 465-488.
- Little, Roderick J.A., and Donald B. Rubin. 1987. *Statistical analysis with missing data*. New York: Wiley.
- Muralidharan, Karthik, and Venkatesh Sundararaman. 2009. "Teacher performance pay: Experimental evidence from India." NBER Working Paper 15323. Cambridge, MA: National Bureau of Economic Research.
- Murnane, Richard J., and David K. Cohen. 1986. "Merit pay and the evaluation problem: Why most merit pay plans fail and a few survive." *Harvard Educational Review* 56, no. 1 (February): 1-17.
- Organisation for Economic Co-operation and Development. 2004. *Learning for tomorrow's world: First results from PISA 2003*. Paris: OECD.
- . 2006. *Education at a Glance: OECD Indicators 2006*. Paris: Organisation for Economic Co-operation and Development.
- Podgursky, Michael J., and Matthew G. Springer. 2007. "Teacher performance pay: A review." *Journal of Policy Analysis and Management* 26, no. 4: 909-949.
- Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, schools, and academic achievement." *Econometrica* 73, no. 2 (March): 417-458.
- Rockoff, Jonah E. 2004. "The impact of individual teachers on student achievement: Evidence from panel data." *American Economic Review* 94, no. 2 (May): 247-252.
- Sclafani, Susan, and Marc S. Tucker. 2006. *Teacher and principal compensation: An international review*. Washington, DC: National Center on Education and the Economy.
- West, Martin R., and Ludger Woessmann. forthcoming. "'Every Catholic child in a Catholic school': Historical resistance to state schooling, contemporary school competition, and student achievement." *Economic Journal*.
- Woessmann, Ludger. 2003. "Schooling resources, educational institutions, and student performance: The international evidence." *Oxford Bulletin of Economics and Statistics* 65, no. 2: 117-170.
- . 2005. "Educational production in Europe." *Economic Policy* 20, no. 43: 446-504.
- Woessmann, Ludger, Elke Luedemann, Gabriela Schuetz, and Martin R. West. 2009. *School accountability, autonomy, and choice around the world*. Cheltenham, UK: Edward Elgar.

**Figure 1: Performance pay and student achievement across countries:
Added-variable plot**



Notes: Added-variable plot of a regression of PISA 2003 math test scores on teacher performance pay and an extensive list of control variables for student characteristics, family background, school resources and location, and country-level features, as shown in Appendix Table A.1. Based on the student-level regression reported in column (3) of Table 2 that is then aggregated to the country level.

Table 1: Selected descriptive statistics by country

P o u t s t a n d i n g performance
 made by Adjustments of teacher salaries based on

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	Australia	524.1	27,872	44,065	0	0	0	0	0	0.714	0.4	0.5
Austria	505.1	27,567	34,822	0	0	0	0	0.571	0.0	1.0	Belgium	529.1	25,262	38,157	0	0	0	0	0.143	0.1	0.0
Czech Rep.	516.1	14,642	19,994	1	1	0	0	0.429	0.0	0.5	Denmark	513.7	27,970	37,925	1	1	0	0	0.714	0.8	0.0
Finland	544.2	23,786	35,430	1	1	1	1	0.714	0.2	0.0	Germany	503.1	25,189	47,551	0	0	0	0	0.286	0.0	1.0
Greece	444.5	15,787	28,646	0	0	0	0	0.429	0.4	0.5	Hungary	490.3	13,016	14,512	1	1	0	0	1.000	0.8	0.5
Iceland	514.7	26,352	22,396	0	0	0	0	0.857	0.6	0.5	Ireland	503.5	28,247	44,185	0	0	0	0	0.286	0.6	0.0
Italy	465.8	22,924	30,011	0	0	0	0	0.714	0.0	0.5	Japan	533.6	24,036	45,753	0	0	0	0	0.714	0.0	0.5
Korea	541.6	17,595	48,815	0	0	0	0	0.571	0.0	0.5	Luxembourg	493.3	49,261	73,580	0	0	0	0	0.571	0.4	1.0
Mexico	384.9	7,939	18,931	1	0	0	1	0.714	0.8	0.0	Netherlands	538.1	26,154	42,628	0	0	0	0	0.143	0.0	0.0
New Zealand	524.1	22,197	36,063	1	1	0	0	0.857	0.6	0.0	Norway	495.4	34,013	35,420	1	0	1	0	0.714	0.4	0.0
Poland	490.1	9,217	10,263	0	0	0	0	0.571	0.6	0.5	Portugal	466.1	17,333	31,635	1	0	1	0	0.714	0.6	0.5
Slovak Rep.	498.6	10,943	–	1	0	1	0	0.143	0.0	0.0	Spain	485.6	20,642	38,503	0	0	0	0	0.429	0.2	0.5
Sweden	509.6	26,138	29,971	1	1	0	0	0.143	0.0	0.0	Switzerland	526.1	28,792	53,536	0	0	0	0	0.714	0.0	0.5
Turkey	423.8	5,634	18,188	1	0	0	1	0.571	0.4	0.5	Unit. Kingdom	508.0	26,044	43,831	0.5	0.5	0	0	0.357	0.1	0.0
United States	483.5	34,875	39,914	1	0	1	0	0.714	0.4	0.0											

Mean 498.4 22,837 35,730 0.446 0.232 0.179 0.107 0.554 0.300 0.339 Country means. See text for definitions of variables.

PISA math score	GDP per capita	Schoo	Local	National	Condi	Qualifi	Demo
	Teacher salaries	l	autho	authority	tions	cations	graph
	Any	princip	rity				y
		al					

(
1
5
.
8
2
3
)
(
2
6
.
9
7
3
)

Ob
(cc

0.389 0.392 0.401 0.400 0.399 0.398 Dependent variable: PISA 2003 math score. Least squares regressions weighted by students' sampling probability.

** Robust standard errors adjusted for clustering at the country level in parentheses. Reported variables are measured at the country level. All regressions control for: 15 student characteristics, 16 family background measures, 9 measures of school resources and location, GDP per capita, expenditure per student, imputation dummies, and interaction terms between imputation dummies and the variables. Significance level: 1 percent,

Table 3: Restricted samples and continental fixed effects

Teacher performance pay 26.305 Sample: GDP > \$9,000 Full Europe (1) (2) (3) *** 32.619*** 35.889*** 1.492** 1.312
 (6.915) (6.509) (6.832) Teacher salaries 1.455 (0.459) (0.495) (0.615) Continental fixed effects yes
 2 Observations (students) 149,657 184,495 121,843 Clustering units (countries) 25 27 21 R 0.362 0.408 0.394 Dependent
 variable: PISA 2003 math score. Least squares regressions weighted by students' sampling probability.
 *** Robust standard errors adjusted for clustering at the country level in parentheses. Reported variables are
 measured at the country level. All regressions control for: 5 system-level institutional measures, 15 student
 characteristics, 16 family background measures, 9 measures of school resources and location, GDP per capita,
 expenditure per student, imputation dummies, and interaction terms between imputation dummies and the
 variables. Significance level: * 1 percent, 5 percent, 10 percent.

Table 4: Performance pay vs. other types of salary adjustment

Teacher performance pay	Other adjustment to teacher salary
Type of other adjustment: (1) (2)	Teaching conditions and responsibilities 27.024 18.531 (7.253) (18.395)
Teachers' qualifications and training 24.311*** 14.310 (5.885) (9.876)	Demography 22.737 12.039 (7.368) (13.661)
Criteria based on teaching conditions and responsibilities Management responsibilities in addition to teaching duties 24.493 18.350 (6.129) (5.670) Teaching more classes or hours than required by full-time contract 25.440*** 6.303 (6.887) (7.752) Special tasks (career guidance or counseling) 24.883 0.394 (6.790) (6.657) Teaching in a disadvantaged, remote, or high-cost area (location allowance) 16.227*** -14.201 (9.348) (10.136) Special activities (e.g., sports and drama clubs, summer school) 25.063*** 9.346 (7.315) (8.101) Teaching students with special educational needs (in regular schools) 26.670*** 7.001 (6.618) (5.669) Teaching courses in a particular field 27.643 -10.686 (7.514) (9.609)	
Criteria related to teachers' qualifications and training Holding an initial educational qualification higher than the minimum 23.786 -7.013 qualification required to enter the teaching profession (5.990) (6.249) Holding a higher than minimum level of teacher certification or training 24.054*** -4.650 obtained during professional life (6.202) (6.042) Successful completion of professional development activities 25.201*** -8.998 (6.164) (5.635) Reaching high scores in the qualification examination 24.573*** -4.093 (6.762) (9.050) Holding an educational qualification in multiple subjects 25.856 -6.434 (6.952) (10.683)	
Criteria based on demography Family status (married, number of children) 23.062 -13.161 (6.167) (11.330) Age (independent of years of teaching experience) 24.686*** -0.704 (7.437) (7.627) Other 41.178 28.422*** (5.146) (3.333)	

*** Each row reports the results of a separate regression, with column (1) showing the coefficient on teacher performance pay and column (2) showing the coefficient on the type of salary adjustment indicated in the lead column. Dependent variable: PISA 2003 math score. Least squares regressions weighted by students' sampling probability. Robust standard errors adjusted for clustering at the country level in parentheses. Reported variables are measured at the country level. All regressions control for: teacher salaries, 5 system-level institutional measures, 15 student characteristics, 16 family background measures, 9 measures of school resources and location, GDP per capita, expenditure per student, imputation dummies, and interaction terms between imputation dummies and the variables. Significance level: 1 percent, *5 percent, 10 percent.

Table 5: Heterogeneous effects: Interactions of performance pay with teacher and student characteristics

Main effect Interaction with
teacher performance pay (1) (2)

Students (observations) 184,495 Schools 6,877 Countries 27 R

the lead column and column (2) showing the coefficient on an interaction term of the variable indicated in the lead column with teacher performance pay. Dependent variable: PISA 2003 math score. Least squares regressions weighted by students' sampling probability. Robust standard errors adjusted for clustering at the country level for country-level variables and at the school level for student-level and school-level variables in parentheses. All regressions control for: teacher salaries, 5 system-level institutional measures, 15 student characteristics, 13 additional family background measures, 7 additional measures of school resources and location, GDP per capita, expenditure per student, imputation dummies, and interaction terms between imputation dummies and the variables. Significance level: 1 percent, * 5 percent, 10 percent.

0.394 All coefficients stem from one joint regression, with column (1) showing the coefficient on the variable indicated in

T e a c h e r c h a r a c t e r i s t i c s T e a c h e r
s a l a r i e s (c o u n t r y - l e v e l) 1 . 4 5 5

*** 0
11.
(s
(E
hor

T
a
b
l
e
6:
A
c
h
i
e
v
e
m
e
n
t
i
n
s
c
i
e
n
c
e
a
n
d
r
e
a
d
i
n
g

.....
.
(7.
41
)
(5.
12
5)
(6.
95
1)
(5.
15
9)
Te
ac
he
r
sa
lar
ie



s
1.
48
8
1.
60
9**
.
1.
28
7**
.
1.
36
7

(0.325) (0.308) (0.344) (0.398) Continental fixed effects no yes no yes

0.349 0.354 0.369 0.373 Dependent variable: PISA 2003 score. Least squares regressions weighted by students' sampling probability. Robust

standard errors adjusted for clustering at the country level in parentheses. Reported variables are measured at the country level. All regressions control for: 5 system-level institutional measures, 15 student characteristics, 16 family background measures, 9 measures of school resources and location, GDP per capita, expenditure per student, imputation dummies, and interaction terms between imputation dummies and the variables. Significance level: ** * 5 percent, 10 percent. 1 percent,

Descriptive statistics Basic model Mean Std. dev. Coeff. Std. err.

Parents' job Blue collar low skilled 0.096 0.289
Blue collar high skilled 0.141 0.341 0.561 (1.011) White collar
low skilled 0.233 0.414 1.960*** (0.969) White collar high skilled 0.530 0.490
7.812*** (1.035) **Books at home** 1 - 10 books 0.093
0.290 11 - 25 books 0.142 0.349 5.102*** (0.993)
26-100 books 0.314 0.464 22.502*** (1.038) 101-200 books 0.201 0.400 32.280*** (1.150)
201-500 books 0.156 0.363 49.756*** (1.256) More than 500 books 0.095 0.293 51.663*** (1.444)
Index of economic, social, and cultural status (ESCS) -0.016 1.003 19.318
(0 . 5 4 9) **S c h o o l**
r e s o u r c e s a n d l o c a t i o n
Class size (mathematics) 23.180 7.410 1.378 (0.069) Shortage of
i n s t r u c t i o n a l m a t e r i a l s Not at all 0.070
0.252 - 10.289*** (2.723) Strong 0.386 0.479 7.536*** (1.308) Instruction time
(mathematics, minutes per week) 197.034 86.807 0.049*** (0.005) **T e a c h e r**
e d u c a t i o n (s h a r e a t s c h o o l)
Fully certified teachers 0.905 0.177
19.624*** (3.896) Tertiary degree in pedagogy 0.646 0.318 6.234*** (2.059)
S c h o o l ' s c o m m u n i t y l o c a t i o n Village or rural area (<3,000)
0.108 0.307 Town (3,000-100,000) 0.571 0.488 -1.570 (1.550)
City (100,000-1,000,000) 0.209 0.401 5.764** (1.933) Large city with > 1
million people 0.111 0.310 5.536 (2 . 4 5 8)
C o u n t r y - l e v e l
m e a s u r e s ***GDP per capita (1,000
\$) 22.837 9.036 -5.070*** (0.923) Educational expenditure per student (1,000 \$) 56.867 25.955 1.565
(0.293) Students 191,841 184,495 Schools 7,158 6,877