Smart Teachers, Smart Students:

Teacher Skills and Student Performance across Developed Countries*

Eric A. Hanushek, Marc Piopiunik, Simon Wiederhold§

Preliminary version. Please do not cite or quote without permission.

Student performance differs greatly across countries, but little is known about the role of teacher quality in explaining these differences. New international data from the PIAAC survey of adult skills allow quantifying country-specific teacher skills in numeracy and literacy for the first time. Our main identification strategy exploits exogenous variation in teacher skills attributable to international differences in relative wages of non-teacher public-sector employees. Using student-level performance data, we find that teacher skills are an important determinant of international differences in student performance. Results are supported by fixed-effects estimations that exploit within-country between-subject variation in teacher skills.

Keywords: teacher skills, student performance, instrumental variable, PIAAC, PISA

JEL classification: I20, H40, H52

* We would like to thank William Thorn, Veronica Borg, and Vanessa Denis for access to and help with the international PIAAC data. We are also indebted to David Deming, Bernhard Enzi, Oliver Falck, Josh Goodman, Brian Jacob, Susanna Loeb, Paul Peterson, Jesse Rothstein, Guido Schwerdt, Marty West, Ludger Woessmann, and participants at the CESifo Area Conference on the Economics of Education, the annual conference of the Verein für Socialpolitik, the RWI Research Network Conference on the "Economics of Education", and the Ifo Center for the Economics of Education and Innovation seminar for providing valuable comments. Piopiunik is indebted to the Program on Education Policy and Governance (PEPG), Kennedy School of Governance, Harvard University, in particular to Paul Peterson and Antonio Wendland, for their support and hospitality during his research visit. Wiederhold is thankful for the hospitality provided by the Center for International Development at Harvard University, especially to Ricardo Hausmann, Ljubica Nedelkoska, and Frank Neffke. Wiederhold also gratefully acknowledges financial support from the Fritz Thyssen Foundation.

§ Hanushek: Hoover Institution, Stanford University, CESifo, and NBER, hanushek@stanford.edu; Piopiunik: Ifo Institute at the University of Munich and CESifo, piopiunik@ifo.de; Wiederhold: Ifo Institute at the University of Munich, wiederhold@ifo.de.

1. Overview

Numerous international assessment tests have shown that the cognitive skills of students differ greatly across countries, including across developed economies. These differences take on considerable significance because historically the cognitive skills of the population have been an important driver of a country's long-run economic growth (e.g., Hanushek and Woessmann (2012)). While previous studies stressed the importance of institutional features of the schooling systems in explaining these differences, the potential role of teacher quality has remained largely unexplored. This paper investigates whether differences in measured teacher skills across developed countries can explain the huge international differences in student performance.

Various public discussions have emphasized the importance of teacher skills for improving student achievement. For example, the widely-cited McKinsey report on international achievement concludes that "the quality of an educational system cannot exceed the quality of its teachers" and then goes on asserting that "the top-performing systems we studied recruit their teachers from the top third of each cohort graduate from their school system: the top 5 percent in South Korea, the top 10 percent in Finland, and the top 30 percent in Singapore and Hong Kong." (Barber and Mourshed (2007), p. 16)

Our analysis exploits new international data in order to test whether differences in teacher quality can explain the huge international differences in student performance.[1] Using data from the Programme for the International Assessment of Adult Competencies (PIAAC), we can for the first time quantify differences in teacher skills in numeracy and literacy. Descriptively, we find that these teacher skills differ widely across countries. For example, average numeracy and literacy skills of teachers in the worst-performing countries (Italy and Russia) are similar to the skills of employed adults with just a post-secondary, non-tertiary education in Canada.[2] In contrast, the skills of teachers in the best-performing countries (Japan and Finland) are higher than the skills of adults with a master's or PhD degree in Canada. These differences in teacher skills reflect, as we discuss below, both where teachers are drawn from in each country's skill distribution and the overall level of skills in each country's population.

Combining this information on teacher quality with student achievement, we find that differences in teacher skills are an important determinant of international differences in student performance. Specifically, we use country-level measures of subject-specific teacher skills along with rich student-level micro data from the Programme for International Student Assessment

[1] The validity of the comparisons of teacher skills has also been questioned (Schleicher (2013)) and is considered below.

[2] We use Canada for the skill comparison because the Canadian sample is by far the largest among all countries surveyed in PIAAC, allowing for a fine disaggregation of individuals by educational degree.

(PISA) to estimate the impact of teacher skills on student performance in math and reading across 23 developed countries.

We pursue three different strategies to investigate the impact of teacher skills. First, we estimate OLS models with extensive sets of control variables, including student and family background, general and subject-specific school inputs as well as institutional features of the school systems. Furthermore, the PIAAC data enable us to control coarsely for the impact of parent skills on their child's academic performance. We use the PIAAC micro data to compute the numeracy and literacy skills of different groups of adults, defined by gender, educational attainment, and number of books at home. We match the average skill levels of these adult groups to the actual parents of students tested in PISA. Controlling for parent skills allows us to control for the persistence of skills across generations. However, the OLS coefficients on teacher skills cannot be interpreted causally as the OLS models likely suffer from omitted-variable bias. For instance, the educational attitude in a country or teachers' pedagogical capabilities may be correlated with both teacher skills and student performance.

Second, we use a within-student across-subject approach which controls for unobserved student-specific characteristics that similarly affect math and reading performance (e.g., innate ability and family background). The advantage of this student fixed-effects model over the OLS model is that it controls for all differences across countries that are not subject-specific, e.g., general education preferences. However, we worry that these estimates are still biased because country differences may well be subject-specific; e.g., some countries may particularly emphasize math skills, while others may attach more importance to reading skills. Moreover, the student fixed-effects estimations likely amplify the attenuation bias as our observed teacher skills are measured with error.

Third, our main identification strategy addresses these concerns by exploiting quasi-experimental variation in teacher skills due to differences in wage distributions across countries. Specifically, we use the gross hourly wages provided in the PIAAC micro data to instrument teacher skills with the position (i.e., the percentile rank) of the mean wages of non-teacher public-sector employees in the wage distribution of non-teacher private-sector college graduates. The basic idea of the instrument is that countries with relatively high wages for public-sector employees are able to recruit individuals with higher skills as teachers (who are predominantly public-sector employees in most developed countries). By excluding all persons working in the education sector (teachers, university professors, etc.) when constructing the instrument, we ensure that the instrument does not reflect the education preferences in a country. Because countries with high wages for public-sector employees might also have more resources to spend on education, we control for the cumulative educational expenditure per student.

The instrumental-variable results indicate sizeable impacts of teacher skills on student performance. We find that a one-standard-deviation increase in teachers' numeracy skills raises student math performance by 20 percent of an international standard deviation. The effect of teacher skills on student performance is about half this magnitude in reading, but is also highly statistically significant. The point estimates are smaller in the student fixed-effects model, consistent with larger measurement error when teacher skills are differenced across subjects. We also find that parent skills are always positively associated with student performance in both math and reading; however, only the association between parent numeracy skills and student math performance captures statistical significance.

We provide several specification checks to show the robustness of the teacher-skill effects. For instance, we create coarse measures of teachers' subject-specific pedagogical skills by using student-level information in PISA about the quality of math and language class teachers. Adding these instruction-quality indicators as additional control variables does not change the teacher-skills coefficients. Instruction quality itself is positively associated with student performance in both subjects, but is statistically significant only in reading.

Furthermore, results are robust to controlling in different ways for the general skill level in a country. We also find some evidence for effect heterogeneity, as the impact of teacher skills is stronger for students with low socioeconomic background than for students with high socioeconomic background. Finally, country-level regressions suggest that policymakers can attract and retain higher-skilled individuals in the teaching profession by increasing teacher wages.

The paper proceeds as follows. Section 2 considers relevant prior research. In Section 3, we introduce the datasets and describe the computation of teacher and parent skills. Section 4 presents the estimation strategies. Section 5 reports results on the impact of teacher skills on student performance in math and reading. Furthermore, we provide robustness checks and heterogeneity analyses for various student subsamples. In Section 6, we analyze the role of teacher wages as a potential leverage for policymakers to raise teacher skills. Section 7 concludes.

2. Relevant Literature

Large numbers of studies investigate the determinants of student achievement within individual countries.[3] We build upon these studies in our investigation of the determinants of achievement. The clearest conclusion from this "educational production function" literature is that achievement reflects a combination of a wide variety of family background factors, school inputs, and institutional factors. But, while these studies give some guidance, they generally are better suited to within-country analysis and are not structured to explain the differences in achievement that we observe across countries. In particular, all of these studies consider the impacts of school characteristics within a country's overall institutional structure – such as the amount of local decision making authority at schools, the requirements for teacher certification, or the overall salary levels for teachers – and do not necessarily give an accurate picture of their impact under differing institutional structures.

There has developed a parallel literature on international differences in achievement that builds on the comparative outcome data in existing international assessments (see Hanushek and Woessmann (2011a)). Perhaps one of the clearest explanatory factors from these international studies has been the role of family background in explaining student achievement.[4] In contrast, specific conclusions about the impact of resources have been much more limited. There has, for example, been considerable research on overall educational expenditures and on identified resource inputs such as class size, but the existing research has not identified these as being strong drivers of international differences in achievement.[5] The lack of findings on resources has led to a different set of international studies that focuses on the effects of institutional features of the school systems. These include the degree of local decision making, the use of accountability systems, and direct rewards for personnel in the schools.[6]

At the same time, individual country studies have emphasized the role of teacher quality, and they suggest that the consideration of differences in teacher quality in existing international studies may be incorrect. The detailed study of teachers within countries has generally shown that the common measures of teacher differences – teacher education and teacher experience levels – are not consistently related to student achievement, raising questions about the reliance on these in

---

[3] See, for example, the reviews in Hanushek (2002) and Glewwe et al. (2013).

[4] For example, see the review in Björklund and Salvanes (2011) or the analysis in Woessmann et al. (2009).

[5] See Hanushek (2006) for a review of the effects of school resources and the international evidence in Hanushek and Woessmann (2011a).

[6] For example, positive impacts have been estimated for school autonomy (especially in developed countries; cf. Hanushek, Link, and Woessmann (2013)) and the share of privately operated schools potentially increasing school competition (West and Woessmann (2010)). See the range of institutional studies in Hanushek and Woessmann (2011a).

international studies.[7] In a closely related set of within-country and international studies, researchers have relied on different measures of teacher salaries as proxies for teacher quality, implicitly assuming that higher-paid teachers have higher skills or are more motivated. However, existing within-country evidence indicates that teacher salaries are a weak measure of teacher quality (see the overview by Hanushek and Rivkin (2006)).[8] Two kinds of international studies have expanded on the within-country analysis of teacher effectiveness. Dolton and Marcenaro-Gutierrez (2011) construct a country panel with international student assessment tests in the period 1995–2006, showing that teacher salaries – both measured in absolute terms and relative to the average wages in a country – are positively associated with student performance, even after controlling for country fixed effects. Related analysis has looked at the use of performance pay, and the international research has tended to find that pay incentives are effective in improving performance – but incentives are not a measure of differences among teachers.[9] Moreover, despite the fact that education policy often targets teachers, none of this work provides evidence that teacher quality differences across countries explain international student performance gaps.

Specifically, the quality of teachers as a determinant of international student performance differences has not yet been studied in a way that is consistent with the body of research that has developed.[10] There is abundant evidence that teachers' impact on students' reading and math performance differs greatly. Numerous within-country studies (mostly from the United States) have demonstrated that there is huge variation in teachers' value added.[11] Little variation in teacher quality is explained by observable teacher characteristics and in turn by teacher salaries that are largely driven within countries by these characteristics (teacher education and experience). In contrast, teachers' academic skills as measured by scores on achievement tests, while not entirely consistently having an impact, are more strongly associated with student performance (Eide, Goldhaber, and Brewer (2004); Hanushek and Rivkin (2006)). While omitted variables and non-random sorting of students and teachers often hamper a causal interpretation of these patterns, Metzler and Woessmann (2012) show the relevance of teacher skills for student performance when

---

[7] The exception in the results is the first years of teaching experience. For evidence in developed countries, see Hanushek (2003). For developing countries, see Hanushek (1995) and Glewwe et al. (2013). For cross-country evidence, see Hanushek and Woessmann (2011a).

[8] We explore the relationship between teacher skills and teacher wages in Section 6.

[9] For a review on teacher performance pay, see Leigh (2013). See also the international investigation of performance pay in Woessmann (2011).

[10] See reviews of within-country studies of teacher quality in Hanushek and Rivkin (2006, 2012).

[11] For a sample of the research into teacher effectiveness, see Rockoff (2004), Rivkin, Hanushek, and Kain (2005), Kane, Rockoff, and Staiger (2008), Chetty, Friedman, and Rockoff (2014), and the summary in Hanushek and Rivkin (2010). As an indication of the magnitudes involved, Rivkin, Hanushek, and Kain (2005) estimate that the effect of a costly ten student reduction in class size is smaller than the benefit of moving the teacher quality distribution one standard deviation upwards.

these problems are likely circumvented. Exploiting within-teacher within-student variation using data from 6th-grade students in Peru, they find a positive impact of teacher subject knowledge on student performance in math, although student reading scores are hardly affected by better teacher knowledge. Using several identification approaches to overcome endogeneity problems, we show the importance of subject-specific teacher skills across developed countries.

3. Data

This section describes the construction of teacher and parent skills based on the PIAAC data and presents the PISA data on student performance.

3.1 Teacher Skills

Teacher skills are derived from the Programme for the International Assessment of Adult Competencies (PIAAC) survey. Developed by the Organisation for Economic Co-operation and Development (OECD) and collected in 2011/2012, PIAAC tested various cognitive skills of more than 160,000 adults in 24 countries that represent almost 75 percent of the world economy.[12] The target population was the non-institutionalized population aged 16-65 years, with at least 5,000 participants in each country. The survey was administered by trained interviewers either in the respondent's home or in a location agreed upon between the respondent and interviewer. The standard survey mode was to answer questions on a computer, but respondents without computer experience could opt for a pencil-and-paper interview.[13] Respondents could take as much time as needed to complete the assessment.[14]

PIAAC has been designed to be valid cross-culturally and cross-nationally in order to provide internationally comparable adult skills. The survey measures key cognitive and workplace skills needed to advance in the job and to participate in society in three domains: numeracy, literacy, and problem solving in technology-rich environments.[15] Cognitive skill tasks are often framed as real-

[12] We use 23 countries in our analysis: Australia, Austria, Belgium (Flanders), Canada, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Ireland, Italy, Japan, Korea, the Netherlands, Norway, Poland, the Russian Federation, the Slovak Republic, Spain, Sweden, the United Kingdom (England and Northern Ireland), and the United States. Cyprus did not participate in PISA. According to OECD (2013), data for the Russian Federation are preliminary, may still be subject to change, and are not representative of the entire Russian population because they do not include the population of the Moscow municipal area. Our results are not sensitive to dropping the Russian Federation from the sample.

[13] On average across countries, 77.5 percent of assessment participants took the computer-based assessment and 22.5 percent took the paper-based assessment. A field test suggests no impact of assessment mode (OECD 2013).

[14] PIAAC tests were conducted in the official language of the country of residence. In some countries, the assessment was also conducted in widely spoken minority or regional languages.

[15] *Literacy* is defined as the "ability to understand, evaluate, use and engage with written texts to participate in society, to achieve one's goals, and to develop one's knowledge and potential," and *numeracy* is the "ability to access, use, interpret, and communicate mathematical information and ideas in order to engage in and manage the mathematical

world problems, such as maintaining a driver's logbook (numeracy domain). PIAAC measures each of the skill domains on a 500-point scale.[16] Inspection of sample items indicates that the skills tested in PIAAC reflect knowledge and competencies that should have been acquired by the end of compulsory schooling, but do not reflect more advanced competencies (e.g., solving differential equations) that are acquired only at college; still, skills tested in PIAAC can probably be improved by a high-quality college education.

Before the skill assessment, all participants answered a background questionnaire providing information about occupation, education, and demographic characteristics. In the Public Use File, information on occupation is available only at the two-digit code in some countries (Germany, Ireland, Sweden, and the United States), while a few other countries (Austria, Canada, Estonia, and Finland) do not report any occupational code. For this study, however, we gained access to the four-digit ISCO-08 (International Standard Classification of Occupations) codes for all countries through the OECD, which allows us to identify teachers in fine categories.[17]

We define teachers as all PIAAC respondents who report as current four-digit occupation code "primary school teacher", "secondary school teacher", or "other teacher" (which includes, for example, special education teachers and language teachers).[18] We exclude university professors and vocational school teachers since the vast majority of PISA students (15-year-olds) are still in secondary school, and have therefore not been taught by these types of teachers. We also exclude pre-kindergarten teachers as this teacher group is more involved with the emotional and social upbringing of children than with rigorously teaching students in reading and math.[19]

PIAAC does not allow us to identify the subject that a teacher is teaching, so we use the numeracy and literacy skills of all teachers tested in PIAAC. We use country-level median of the teacher skills because the median is more robust to outliers than the mean, which is particularly

demands of a range of situations in adult life" (see OECD (2013) for more details). As we want to explain cross-country differences in students' reading and math performance, we do not use the PIAAC skills in the domain "problem solving in technology-rich environments." Moreover, four countries surveyed in PIAAC (Cyprus, France, Italy, and Spain) did not administer tests in this optional skill domain.

[16] Throughout, we use the first plausible value of both PIAAC and PISA scores in each domain.

[17] Australia and Finland report only two-digit occupation codes in PIAAC.

[18] Results are very similar if we drop the category "other teachers." However, we prefer to keep these teachers in the sample to increase sample size.

[19] For Australia and Finland we are not able to exclude pre-kindergarten teachers and university professors/vocational school teachers from our teacher sample. However, based on the 21 countries where teachers are defined using the four-digit code, it turns out that teacher skills based on the four-digit code are very similar to those defined with the two-digit code: The correlation of both skill measures is 0.97 for numeracy and 0.95 for literacy. On average, numeracy (literacy) skills based on the two-digit code are only marginally higher (by 0.5 (0.1) PIAAC points) than the respective skills based on the four-digit codes. The average absolute difference in the 21 countries is 2.1 points in numeracy and 1.9 points in literacy. Moreover, simultaneously excluding Australia and Finland from the analysis does not qualitatively change our results.

relevant in smaller samples.[20] We weight individual-level observations with inverse sampling probabilities when computing country-specific teacher skills.

Table 1 reports summary statistics of the teacher skills in the 23 countries and in the pooled sample. The number of teachers in the national PIAAC samples ranges from 124 teachers in Italy to 834 teachers in Canada, with 231 teachers per country on average.[21] Teachers in Finland and Japan perform best in both numeracy and literacy, while teachers in Italy and Russia perform worst in both domains. The difference in numeracy is 44 points, which is about 85 percent of the international *individual*-level standard deviation (53 points). Teachers in the United States (284 points) perform worse than the average teacher in numeracy (295 points), but are slightly above the international mean in literacy. Interestingly, the country ranking and the cross-country variation in teacher skills are similar to those of all prime-aged workers with full-time employment (see Table 1 in Hanushek et al. (2013)).[22] Also note that teacher numeracy skills are better than teacher literacy skills in some countries, while the reverse is true in other countries. We will exploit this variation in subject-specific teacher skills in the fixed-effects model that uses only variation within countries across subjects (see Section 5.1). Furthermore, both numeracy and literacy skills of teachers are completely unrelated to the number of teachers in the national PIAAC samples. For the econometric analysis, we standardize the country-specific teacher skills across the 23 countries (at the country level) to have a mean of zero and a standard deviation of one.

To get a better understanding of how strongly teacher skills differ, we compare the differences in teacher skills across countries with the skill differences of adults across educational groups within a single country (Figure 1). Because it provides by far the largest sample, we use Canada for this skill comparison.[23] The literacy skills of the lowest-performing teachers (in Italy and Russia) are similar to the literacy skills of adults with only a vocational degree (278 points). Teachers in Canada, the Netherlands, Norway, and Sweden have similar skills than adults with a bachelor degree (306 points). The literacy skills of the best-performing teachers (in Japan and Finland) are even higher than the skills of adults with a master or doctoral degree (314 points). This comparison, which looks similar for numeracy skills, indicates that teacher skills differ greatly across developed countries.

---

[20] The country-level correlation between teacher median skills and mean skills is 0.97 for both numeracy and literacy. Moreover, all results are robust to using mean teacher skills instead of median teacher skills (see Table 5 for a robustness check of our main specification).

[21] The sample size for Canada is substantially larger than for any other country surveyed in PIAAC because Canada decided to oversample to obtain regionally representative adult skills.

[22] Younger teachers have higher skills than older teachers in almost all countries in our sample. Also, male teachers have higher skills than female teachers, especially in numeracy. These patterns, however, are not specific to teachers, but are very similar among all college graduates in a country. Detailed results are available on request.

[23] The sample used for this comparison includes all employed individuals aged 25-65 years.

These variations in teacher skills reflect both where teachers are drawn from the skill distribution of the population and where a country's overall skill level falls in the world distribution. As most teachers have obtained a college degree (88 percent on average across all countries), we expect that teacher skills fall at or above the median of the skill distribution of the entire adult population. Across all 23 countries, teacher skills fall at the 68th (70th) percentile of the numeracy (literacy) skill distribution of all adults, ranging from the 53rd to the 78th percentile (see Table 1). The position of teacher skills relative to the skills of all other adults is a first indication that teacher skills based on PIAAC are in a plausible range (see below for further evidence).

These descriptive statistics also indicate that the overall statements about where teachers fall in the skill distribution of different countries (e.g., Barber and Mourshed (2007)) are not accurate and likely do not adequately indicate the important dimensions of teacher skills across countries.[24] From Table 1, teachers in France and Spain are drawn highest up from the country distributions in numeracy and literacy, respectively. This is the case even though Finnish teachers have the highest measured skills, reflecting that the country average of skills is so high.

As most teachers are college graduates, it may also be illuminating to compare teacher skills with the skills of all college graduates in a country (see Figure 2). While teacher skills fall in the middle of the 25th-75th skill range of college graduates in most countries, teachers come from the upper part of the skill distribution in some countries (e.g., Finland and Japan) and from the lower part of the college graduate skill distribution in other countries (e.g., Poland and Slovak Republic). The position in the overall skill distribution from which countries recruit their teachers can potentially be influenced by policymakers. We address this issue in Section 6.

Because the PIAAC tests are new and have not been fully validated, it is useful to compare the PIAAC-based teacher skills with the numeracy and literacy skills of teachers in larger datasets. More precisely, we compare the position of teacher skills in the adult skill distribution in the PIAAC data with the respective position in several national datasets. We first look at the U.S. datasets National Longitudinal Survey of Youth 1979 (NLSY79) and the NLSY97. The NLSY79 is a nationally representative sample of 6,111 young men and women who were born between 1957 and 1964. The NLSY97 is a nationally representative sample of 6,748 individuals born between 1980 and 1984. (Note that these age cohorts partly overlap with the age range of the PIAAC participants.) We measure NLSY79 respondents' occupation (using four-digit Census codes) in 2010 (last available year) and NLSY97 respondents' occupation in 2011 to make this sample as comparable as possible to PIAAC (survey year is 2011). Teachers are defined as in PIAAC (i.e., excluding pre-kindergarten teachers and university professors/vocational education teachers). We

[24] This point about teacher skills was first made by Schleicher (2013).

weight individual-level observations with the cross-sectional weights taken from the year in which the occupation is measured, giving each NLSY survey the same total weight.

We take the mathematics and language skills tested in the four AFQT subtests which are part of the Armed Services Vocational Aptitude Battery (ASVAB). The ASVAB was administered to 94 percent of NLSY79 respondents in 1980 and to 81 percent of NLYS97 respondents in 1997. We combine the scores from the mathematical knowledge and arithmetic reasoning tests to a numeracy skills measure and the scores from the word knowledge and paragraph comprehension tests to a literacy skills measure.[25] Using the same computation procedure as for PIAAC, teacher skills fall at the 67th (64th) percentile in the adult skill distribution in numeracy (literacy). This is quite close to the position of teacher skills in the PIAAC data for the USA (see Table 1): 70th (71st) percentile in numeracy (literacy).

Given that the position of teacher skills in the adult skill distribution as measured in PIAAC is very similar to that in other nationally representative datasets with larger sample sizes, we are confident that our PIAAC measures are a good proxy for the true teacher skills in a country.

3.2 Parent Skills

Because the parents of the PISA students (henceforth “PISA parents”) are not tested themselves in any skill domain, we use the PIAAC data to compute proxies for the numeracy and literacy skills of PISA parents. The idea is to use a sample of adult PIAAC participants that could in principle be the parents of PISA students. We then match the numeracy and literacy skills of the PIAAC adults to the actual PISA parents based on several observable characteristics. Specifically, we apply the following procedure. We take all adults in PIAAC aged 35-59 with children. With respect to age, these individuals are potential parents of the 15-year-old PISA students since PIAAC adults were 17–44 years old when PISA students were born. For each country separately, we then regress the numeracy/literacy skills of these adults on three characteristics: gender[26], education (3 categories), and number of books at home (6 categories).[27] Finally, we multiply the estimated coefficients with the same three characteristics (i.e., gender, education, and books at home) of the *actual* PISA

---

[25] As respondents were born in different years, we take out age effects by regressing test scores on year of birth dummies first (separately for NLSY79 and NYS97). We control for age effects in the NLSY data because participants were still children or adolescents at the time of testing. In contrast, we do not take out age effects in the PIAAC data because PIAAC participants have already mostly completed their education when being tested.

[26] We compute skills separately for PISA mothers and fathers because numeracy/literacy skills of women and men might differ. By predicting gender-specific skills, PISA students with single mothers, for example, are assigned only the skill level of women and not the average skill level of men and women.

[27] We collapsed the original 8 categories of the PIAAC education variable into 3 categories so that the education categories in PIAAC and PISA would exactly match. The 6 categories of the *number of books at home* variable are identical in PIAAC and PISA, so this variable was not modified. Sample sizes range from 1,074 adults in the Russian Federation to 11,933 adults in Canada. The average sample size is 2,851 adults per country (see Table A-1).

parents to obtain predicted numeracy/literacy skills of all PISA parents.[28] In the student-level analysis, we use the average skills of mother and father as a proxy for parent skills.[29]

Although the PIAAC-based parent skills are only coarse proxies for the true skills of PISA parents, controlling for the skill level of parents allows us to tackle several issues. Most importantly, student performance is likely to be persistent across generations, for example, because the quality of the education system or the valuation of education changes only slowly over time. We intend to capture at least part of this intergenerational persistence in skills by including a proxy for parent skills. Second, parent skills might be a determinant of student performance over and above the student's general family background as measured by parents' education, parental occupation, and number of books at home.

Table A-1 presents summary statistics of parent skills in numeracy and literacy by country. Similar to teacher skills, parent skills differ greatly across countries, ranging from 258 points in Poland to 301 points in Belgium (in numeracy). Also, parent skills differ substantially *within* countries. On average, the difference between the minimum and maximum skill in a country is 88 points, or 1.7 times the international individual-level standard deviation. The large variation in parent skills suggests that these measures can capture differences in student performance both across and within countries.

3.3 Student Performance and Further Control Variables

International data on student performance stem from the Programme for International Student Assessment (PISA), conducted by the OECD.[30] PISA is a triennial survey that tests math and reading competencies of nationally representative samples of 15-year-old students, an age at which students in most countries are approaching the end of compulsory schooling.[31] The tests emphasize understanding as well as flexible and context-specific application of knowledge, and hence do not test curriculum-specific knowledge. PISA contains both multiple-choice and open-answer questions and is set up to provide internationally comparable test scores.

We use the two PISA cycles 2009 and 2012 because the student cohorts in these two test cycles have largely been taught by the teacher cohorts tested in 2011 and 2012 in PIAAC. Student cohorts of earlier PISA cycles (2000, 2003, and 2006) have partially been taught by some PIAAC teachers,

[28] We use *number of books at home* in addition to educational degree, since this variable has been shown to be the single strongest predictor of student test scores (Woessmann (2003)).

[29] Results are very similar if we use the maximum skills of mother and father instead.

[30] We prefer PISA over TIMSS because students participating in PISA were tested in both math and reading, while TIMSS only assessed math performance. Note that math scores from TIMSS are strongly correlated with math scores from PISA at the country level.

[31] Since teachers in PIAAC were only tested in the domains numeracy and literacy, we discard the science test scores in PISA.

but teacher turnover would introduce additional error in the teacher skill measures for students in these earlier cycles. Another reason for combining PISA 2009 and 2012 is that students provide information about the teaching practices of their teachers just for the subject that is the focus in each round of PISA testing: reading in 2009 and math in 2012. From the survey information, we can compute country-specific instruction quality indicators for reading (based on PISA 2009) and for math (based on PISA 2012). These instruction quality indicators capture subject-specific pedagogical skills of teachers, which might be a potentially important confound of teacher skills (see Section 5.3).

Student characteristics, such as gender and migration status, and information about parents, such as education, occupation, and number of books at home, come from student background questionnaires.[32] Table A-2 provides summary statistics of student performance and student characteristics.[33] As is well-known, student performance in math and reading differs significantly across countries. Given that the learning progress in one school year is about 40 PISA points, the difference between the USA and Korea is almost two school years in math and one school year in reading. For the regressions, we normalize test scores at the student level across the 23 countries with a mean of 0 and a standard deviation of 1, separately for each PISA cycle. As we are interested in differences across countries, each country receives the same total weight in each PISA cycle.

In addition to parent skills, we use number of books at home, parents' highest educational degree, and parental occupation to control for family background (see Table A-3).

Based on student information, we also construct measures of weekly instructional time for both language and math classes. As in Lavy (forthcoming), we aggregate this information across students to the school level. Furthermore, school principals provide information on the lack of qualified math teachers and language teachers, whether the school is public or private, city size, total number of students in the school, and about three different types of autonomy (see Table A-4).

Finally, country characteristics include variables that have been used in previous cross-country analysis such as cumulative educational expenditure per student between age 6 and 15, GDP per capita, and school starting age (see Table A-5). As discussed above, we also add indicators of instruction quality of math and language teachers.

---

[32] As with all such surveys, the dataset of all students with performance data has missing values for some background questions. Since we consider a large set of explanatory variables and since a portion of these variables is missing for some students, dropping all student observations with any missing value would result in substantial sample reduction. We therefore imputed values for missing control variables by using the country-by-wave means of each. To ensure that imputed data are not driving our results, all our regressions include an indicator for each variable with missing data that equals one for imputed values and zero otherwise.

[33] All statistics are averages across PISA 2009 and PISA 2012. Again, we weight individual-level observations with inverse sampling probabilities.

4. Estimation Strategy

In the baseline OLS model, we estimate an international education production function of the following form:

$$y_{iksc} = \alpha + \delta T_{kc} + X_{isc}\beta_1 + X_{sc}\beta_2 + X_c\beta_3 + Z_{iksc}\gamma_1 + Z_{ksc}\gamma_2 + \varepsilon_{iksc}, \quad (1)$$

where y_{iksc} is the test score of student *i* in subject *k* (math or reading) in school *s* in country *c*. T_{kc} represents the median teacher skills in subject *k* in country *c*. X_{isc} is a vector of student-level variables measuring student and family background, X_{sc} is a vector of school-level characteristics, and X_c is a vector of country-level control variables.[34] The ***Z*'s** are also control variables, but they vary across subjects. Z_{iksc} is a vector containing student-level variables of parents' numeracy and literacy skills, and Z_{ksc} is a vector containing school-level variables measuring the shortage of qualified teachers and weekly instruction time in math and language classes. ε_{iksc} is an error term with mean zero.

Interpreting the OLS estimates of δ as the causal effect of measured teacher skills on student performance is problematic, however, because there might be unobserved omitted variables correlated with both teacher skills and student performance. These omitted variables could include, for example, the educational attitude in a country: Societies that emphasize the importance of good education likely have both teachers with higher skills and parents who strongly support their child's education (not perfectly captured by our measure of parent skills). Similarly, the persistence of the quality of the education systems would lead to a positive correlation between student performance and skills of teachers (who went through the same education system one generation earlier) even if teacher skills have no real impact on student performance. Another omitted variable would be teachers' pedagogical skills. Subject-specific skills and pedagogical capabilities might be correlated either because a high-quality teacher education raises both types of skills or simply because of differential self-selection of individuals into the teaching profession. Note that self-selection or sorting of students and teachers across schools (within countries) and within schools is no concern in our study because teacher skills are measured at the country level. Finally, country-specific teacher skills are likely measured with error such that OLS estimates are biased toward zero. We will discuss measurement error in the country-level teacher skills in detail below.

We use two independent strategies to address these concerns. The first strategy exploits the unique feature of the PISA and PIAAC data that both teacher skills and student performance are

[34] See Tables A-2, A-3, A-4 and A-5 for a complete list of all control variables.

observed in two subjects. This allows us to exploit within-student variation in teacher skills across math and reading. Therefore, we investigate whether differences in student performance between math and reading are systematically associated with differences in teacher skills between math and reading.[35] While student characteristics, student ability, family background, and school environment are the same for both subjects, teacher skills can differ between math and reading. Within-student effects of teacher skills on student performance are estimated by adding student fixed effects in Equation (1).[36] The student fixed effects capture any performance differences between students that are not subject-specific, for instance, due to family background, innate ability, and motivation. Adding student fixed effects also controls for any non-subject-specific differences across schools (and hence across countries) and for international differences in general pedagogical skills of teachers. Note that all control variables contained in the X vectors are absorbed by the student fixed effects, whereas all subject-specific control variables contained in the Z vectors, such as parent skills in numeracy and literacy, control for differences within students across subjects. Importantly, we control for instruction time in math and language classes at the school level which has been shown to affect student performance (Lavy (forthcoming)). In contrast to OLS, the effect of teacher skills is "net" of teacher skill spillovers across subjects (for example, if teacher literacy skills affect student math performance) in the fixed-effects model.[37]

The student-fixed-effects approach, however, has the disadvantage that it cannot control for unobserved differences across countries that differ across subjects. For example, if societies have both teachers with high numeracy skills and a strong preference for advancing children in math (with parents supporting their children accordingly), then fixed-effects estimates of teacher skills will still be biased. Furthermore, the coefficient on teacher skills might still be attenuated in the fixed-effects model if teacher skills are measured with error. In fact, the attenuation bias is likely more severe in the fixed-effects model than in the OLS model (see below). To address these concerns, we employ an alternative identification strategy.

This strategy is an instrumental-variable approach that uses exogenous variation in teacher skills due to cross-country differences in public-sector wages. The basic idea is that countries paying teachers relatively higher wages are more likely to attract and retain individuals with high skills in the teaching profession as teaching becomes more attractive relative to other professions. Instrumenting teacher skills with relative teacher wages would likely be invalid, however, as high

---

35 Within-student across-subject variation has already been used in previous research (e.g., Dee (2005), Metzler and Woessmann (2012), Lavy (forthcoming)).

36 We estimate the model as a first-difference specification which yields numerically identical results.

37 Note that spillover effects are completely eliminated in the student fixed-effects model when cross-subject spillovers are identical in math and reading.

teacher wages might reflect a high preference for children's education. Therefore, our instrument excludes all teaching professionals. Specifically, it is constructed as the position (i.e., the percentile rank) of the mean wages of *non-teacher* public-sector employees in the wage distribution of non-teacher private-sector college graduates.

Wages of non-teacher public-sector employees should be substantially correlated with teacher wages because teachers are predominantly public-sector employees themselves (76 percent in our sample). In fact, the correlation between the instrument and the position of teacher wages in the wage distribution of non-teacher private-sector college graduates is strong (0.79), but far from perfect. At the same time, wages of public-sector employees – excluding all persons working in the education sector – are likely uncorrelated with education preferences in a country. One might still worry that non-teacher wages are influenced by teacher wages if teachers are a dominant group among all public-sector employees. This would violate instrument exogeneity if the level of teacher wages would reflect country-specific education preferences. However, teaching professionals represent only a minority among all public-sector employees. In the national PIAAC samples, the share of teaching professionals among all public-sector employees ranges from 14 to 27 percent, with an average of 18.7 percent. Besides these low shares of teaching professionals, it seems furthermore plausible that fiscal arguments – and not teacher wages – determine the wage bargaining for non-teacher public-sector employees.

Another worry is that our instrument just reflects a country's preference for public-sector provisions, which may be correlated with preferences for education. However, neither of these conjectures is supported by the data. First, the instrument does not seem to be a proxy for the importance of the public sector in a country; the correlation between the instrument and public expenditure as a percentage of GDP is actually negative (-0.40).[38] Second, the instrument appears to be unrelated to a country's education preferences. The correlations with both cumulative expenditure per student between the age of 6 and 15 normalized with GDP per capita (r=0.04) and with public expenditure on education as a share of total public expenditure (r=0.12) are basically zero.[39] Given these findings, we are confident that the instrument is not correlated with education preferences.

A remaining issue is that countries with high public-sector wages also spend more on education simply because they have more resources. Since our standard set of control variables includes

[38] Data on public-expenditure as a share of GDP and public expenditure on education as a share of total public expenditure come from OECD (2014a). Data refer to the year 2011.

[39] Alternative data sources for gauging the importance of education in a country could be the World Value Survey or the European Social Survey. However, there is no adequate question in these datasets that could capture educational preferences in a country.

cumulative educational expenditure per student between age of 6 and 15 in a country, we control for this potentially confounding factor.

The reason for using the wages of college graduates (in the private sector) when constructing the instrument is that the vast majority of teachers are college graduates themselves, implying that teachers are recruited mainly from the national pool of college graduates.[40] Because the instrument uses college graduates as a comparison group, we need to express the skill level of teachers as a relative measure as well. We do so by additionally including the country-specific median skills of college graduates in the regression. The instrument thus predicts the skill level of teachers *relative* to the skill level of college graduates in the country. Moreover, we use all non-teacher public-sector employees when constructing the instrument – instead of restricting ourselves to college graduates in the public sector – to ensure a reasonable sample size.[41]

Predicted values of teacher skills are obtained in the following first-stage model:

$$T_{kc} = \alpha + \varphi Wage_c + \mu CollSkills_{kc} + X_{isc}\beta_4 + X_{sc}\beta_5 + X_c\beta_6 + Z_{iksc}\gamma_4 + Z_{ksc}\gamma_5 + Z_{kc}\gamma_6 + \eta_{iksc}, \qquad (2)$$

where teacher skills in subject k in country c, T_{kc}, are regressed on the instrument, i.e., the relative wages of non-teacher public-sector employees, $Wage_c$, the median skills of college graduates in subject k in country c, $CollSkills_{kc}$, and all other control variables from Equation (1).

The instrumental-variable approach uses distributional information on *non-teacher* wages within countries. Therefore, the instrument is likely not correlated with subject-specific preferences across countries, solving the potential bias that might have plagued the within-student approach. Another advantage of the instrumental-variable strategy is that the instrument is likely to suffer less from measurement error than teacher skills, as the distributional wage data used to construct the instrument provides more accurate country-specific information than the subject-specific skills that are based on a smaller sample of teachers. Since the instrument is probably not correlated with the measurement error in the teacher skills variables, the instrumental-variable approach likely solves the bias due to measurement error that plagues the OLS and fixed-effects estimates (see the discussion on measurement error below).

Measurement Error

---

[40] In the PIAAC data, the share of teachers who are college graduates varies between 67 percent in Austria and 98 percent in Poland. Across the 23 countries in our sample, the mean share is 88 percent.

[41] In some countries, the number of non-teacher public-sector employees who graduated from college is well below 200. Results are similar if we only use all non-teacher public-sector college graduates for constructing the instrument.

Our country-level teacher skills are measured with error. First, we do not observe the skills of the individuals teachers who teach the students tested in PISA. Second, the observed country-level skills are a noisy measure of the true country-level teacher skills because we use the numeracy (literacy) skills not only of math (language) teachers, but also of all other teachers. Suppressing subject and school indices, one can write the population model we would like to estimate as follows:

$$y_{ic} = \alpha + \delta T_{ic}^{*} + X_{ic}\beta + \varepsilon_{ic}, \tag{3}$$

where T_{ic}^{*} represents the true skills of student *i*'s teacher (in country *c*).[42] To keep notation brief, the vector X_{ic} now contains all other control variables. The individual-level teacher skills, T_{ic}^{*}, can be expressed as follows: $T_{ic}^{*} = T_{c}^{*} + u_{ic}$, where T_{c}^{*} represents the true, but unobserved, median skills of teachers in country *c* (relevant for our PISA student population). The error term u_{ic} is uncorrelated with T_{c}^{*} as skills of individual teachers are scattered around the median skill in each country.[43] As T_{c}^{*} is unobserved, we rewrite the last equation as follows:

$$T_{ic}^{*} = T_{c} + \omega_{c} + u_{ic}, \tag{4}$$

where T_c is our observed measure of country-level teacher skills, and $\omega_c = T_c^{*} - T_c$ is the difference between true and observed country-level teacher skills. Plugging (4) in (3) yields a model we can estimate:

$$y_{ic} = \alpha + \delta T_{c} + X_{ic}\beta + (\delta\omega_{c} + \delta u_{ic} + \varepsilon_{ic}). \tag{5}$$

As discussed above, $\text{cov}(T_c, \varepsilon_{ic})$ is probably positive because of omitted variables such as country-specific education preferences. T_c and u_{ic} are likely uncorrelated since T_c^{*} and u_{ic} are uncorrelated as skills of individual teachers are scattered around the median skill in each country. Assuming that the observed country-level skills suffer from classical measurement error[44]; i.e.,

---

[42] Conceptually, T_{ic}^{*} does not represent the skills of a single teacher, but rather a skill average of all the teachers who have taught student *i* in the current and past years, with more recent teachers receiving more weight. To keep language simple, we will refer to T_{ic}^{*} as representing the skills of a single teacher.

[43] Note that using true *country*-level teacher skills, T_c^{*}, instead of true *individual*-level teacher skills, T_{ic}^{*}, would still yield consistent estimates of δ. The noisy macro-level measure would, of course, imply less precisely estimated coefficients.

[44] The large similarity between the teacher skill ranks of the national PIAAC samples and the other nationally representative datasets (for Germany and the U.S.) suggests that our observed teacher skills do not systematically over- or understate the true teacher skills.

$\text{cov}(T_c^*, \omega_c) = 0$; it follows that $\text{cov}(T_c, \omega_c) = \text{cov}(T_c^* - \omega_c, \omega_c) = -\sigma_{\omega_c} < 0$. To specify the bias in the coefficient of interest, δ, let us assume for the moment that teacher skills, T_c, are uncorrelated with the control variables in X_{ic}. In this case, δ will be overestimated if the positive omitted-variable bias is larger in magnitude than the attenuation bias due to the measurement error in the country-level teacher skills. In contrast, β will be biased downward if the measurement error is more severe than the omitted-variable bias.

The measurement error ω_c might differ between numeracy and literacy skills. On the one hand, numeracy skills might be tested more precisely than literacy skills; or numeracy skills might be more comparable across countries than literacy skills (see, e.g., Hanushek et al. (2013)). On the other hand, as our teacher skills are based on subject-specific test scores of all teachers in PIAAC, irrespective of the subject the teachers actually teach, it seems plausible that teachers' numeracy skills suffer from more measurement error than teachers' literacy skills as math/science teachers are probably more likely to answer literacy questions correctly than are language teachers in the numeracy test. In this case, the OLS teacher skills coefficient in math likely suffers from more attenuation bias than that in reading.

Finally, the attenuation bias due to measurement error in the country-level skills is probably aggravated in the first-differenced model because differencing is particularly problematic when the (true) numeracy and literacy skills are more strongly correlated than the measurement error in numeracy and literacy skills (Griliches and Hausman (1986), Pischke (2007)). Because true teacher skills in numeracy and literacy are certainly strongly correlated at the country level (the correlation of *observed* teacher skills across subjects is 0.77), differencing country-level teacher skills likely leads to a more severe attenuation bias.

5. Results

It is easiest to motivate the analysis with simple visual evidence showing that teacher skills and student performance are positively associated at the country level. The two upper graphs in Figure 3 show the unconditional correlations between teacher numeracy skills and student math performance (left panel) and between teacher literacy skills and student reading performance (right panel), respectively. Student test scores are aggregated to the country level. Both numeracy and literacy skills of teachers are positively associated with student performance, with a coefficient of 0.08 in math and 0.13 in reading. The two bottom graphs in Figure 3 show the association between teacher skills and student performance after controlling for country-specific skill levels of all adults aged 25-65 to net out the skill persistence across generations.[45] Although losing statistical significance, the coefficient on teacher numeracy skills is reduced only modestly, while the coefficient on teacher literacy skills even increases. When Korea, the most obvious outlier, is excluded, the coefficient on teacher numeracy skills becomes larger (0.074) and statistically significant at the 10 percent level.[46]

In the following empirical analysis, we begin with OLS and student fixed-effects estimates of the impact of teacher skills. We then present the instrumental-variable results, followed by robustness checks and heterogeneity analyses.

5.1 OLS and Student Fixed-Effects Estimates

OLS Results

Table 2 reports results from the least squares estimation of Equation (1), which serve as a benchmark for the fixed-effects and instrumental-variable estimates. The unconditional correlation between teacher numeracy skills and individual-level student math performance (Column 1) is identical to the country-level estimate presented in Figure 3. The coefficient on teacher numeracy skills remains statistically significant when adding a large set of background factors at the individual, family, school, and country level (Column 2) and when including the numeracy skills of parents of PISA students (Column 3).[47] In terms of magnitude, the coefficient estimate in Column (3) implies that a one-standard-deviation increase in teacher numeracy skills increases student math performance by almost 10 percent of a standard deviation. Even though various parent characteristics, such as education level and number of books at home, are included, parent numeracy skills are significantly related to student performance, but are rather modest in size compared to teacher skills.

Columns (4)-(6) report results for reading. In the specification with all controls (Column 6), the point estimate on teacher literacy skills is slightly below the coefficient on teacher numeracy skills. In contrast to math, parent literacy skills do not appear to matter for student performance in reading. The estimate is small, albeit positive, and statistically insignificant.

First-Differenced (Student Fixed-Effects) Results

[45] The country-level correlations between teacher skills and adult skills are 0.70 for numeracy and 0.77 for literacy. Skills of teachers and adults are substantially correlated since both have been educated in the same education system at about the same time. To some extent, skills are also correlated because teachers are included in the computation of adult skills.

[46] When omitting teacher skills, adult skills and student performance are strongly positively correlated in both math and reading. However, when conditioning on teacher skills, the estimates for adult skills substantially decrease in size and lose statistical significance.

[47] Table A-6 reports the estimated coefficients on all other control variables of specifications (3) and (6).

As discussed above, our naïve OLS estimations are prone to bias due to omitted variables. As many of the omitted variables we are concerned about vary at the country level, one strategy to overcome these problems is to use only within-country variation to identify the effect of teacher skills on student performance. Having test scores in two different subjects for students and teachers, as well as substantial variation in teacher skills across subjects,[48] we implement the fixed-effects model by regressing the difference in student performance (math minus reading test score) on the difference in teacher skills (numeracy minus literacy), thereby eliminating any non-subject-specific bias due to student, school, and country heterogeneity.

Table 3 presents the results of the student fixed-effects estimates. The specifications are the same as in Table 2, except that control variables that do not differ across subjects are dropped. With full controls, the fixed-effects estimate on teacher skills is about 40 percent smaller than the corresponding OLS estimate, but is still statistically significant (Column 3). This decrease in coefficient magnitude might occur for three distinct reasons. First, country-specific omitted variables that are similar across subjects, such as general education preferences – which likely bias the OLS coefficient upward – are controlled for in the fixed-effects model. Second, as discussed above, the attenuation bias becomes more severe as the measurement error in teacher skills very likely becomes larger when differencing numeracy and literacy skills. Third, the numeracy-literacy skill differences of teachers and parents are strongly correlated (r=0.77); unsurprisingly, the drop in the coefficient occurs when parent skills are included (Column 3).[49] Hence, the effect of teacher skills is identified only from the limited part of the skill variation that is independent of variation in parent skills.

As is standard in the literature that exploits within-student across-subject variation, the first-differenced model assumes that the effect of teacher numeracy skills on student math performance is identical to the effect of teacher literacy skills on student reading performance (e.g., Lavy (forthcoming)). To allow for differential effects of teacher numeracy and literacy skills, we also included them separately in the estimation equation (not shown). Even without imposing the uniformity of effects in the two subjects, we find very similar coefficients on teachers' numeracy (0.052) and literacy skills (0.058), both significant at the 10 percent level. In line with the OLS estimates, the within-student across-subject results suggest that the effect of teacher skills on student performance is similar across subjects.

A couple of other results are worth mentioning. The coefficient on parent skills in Column (3) is slightly larger than in the OLS model for math and almost statistically significant at the 10

[48] The country-level correlation between teachers' numeracy skills and literacy skills is 0.77, and thus far below 1.

[49] The levels of teacher and parent skills are much less correlated (0.34 in math and 0.41 in reading).

percent level (p=0.1002). Interestingly, the effect of instructional time on student performance is similar to the effect size in Lavy (forthcoming), who exploits within-student between-subject variation using PISA data from 2006.

While both OLS and student fixed-effects results suggest a positive impact of teacher skills on student performance, we are still concerned about a causal interpretation. Most importantly, if unobserved country-level determinants of student performance are subject specific, then the fixed-effects coefficients would still be biased. For example, the attitude toward education in a country may not be similar for both subjects, but knowledge and skills might be valued higher in one subject than in the other. Furthermore, the fixed-effects estimates are likely biased towards zero as we difference two variables that are measured with error. To address these concerns, we employ an instrumental-variables approach that exploits arguably exogenous variation in teacher skills across countries.

5.2 Instrumental-Variable Results

Given the limitations of the within-country between-subject identification strategy, we now provide quasi-experimental evidence on the impact of teacher skills. Specifically, we instrument the country-specific teacher skills with the relative wages of non-teacher public-sector employees in a country. The basic idea is that countries with high wages for public-sector employees are able to attract higher-skilled college graduates into the teaching profession (and retain them in the job). Controlling for the direct effect of well-endowed public sectors on student performance through higher education expenditure, the instrument exploits variation in teacher skills that is unlikely to be correlated with a country's (subject-specific) preference for education or other omitted variables simultaneously affecting teacher skills and student performance.

Table 4 reports results from the IV regressions. The first-stage results in the bottom panel show that the relative wage of non-teacher public-sector employees is a strong predictor of teacher skills. In the model with all controls (Column (3) for math and Column (6) for reading), we find that a one-standard-deviation increase in the wage position of public-sector employees (i.e., an increase by about 13 percentile ranks) is associated with an increase in teacher skills of 39 (43) percent of an international standard deviation in numeracy (literacy). The F-statistic of the instrument exceeds 10 by far in all models, suggesting that our estimations do not suffer from a weak-instrument problem.[50] As expected, IV standard errors are substantially larger than those in OLS.

[50] Weak instruments can lead to inconsistencies in the IV estimates (Bound, Jaeger, and Baker (1995)). Moreover, if instruments are weak, the conventional asymptotic approximations used for hypothesis tests and confidence intervals will usually be unreliable (Stock, Wright, and Yogo (2002).

Since identification relies on only 23 independent observations at the country level, one potential worry is that the positive association between the instrument and teacher skills is driven by a few outliers. An added-variable plot of the first-stage relationship that includes all control variables shows that this is not the case (see Figure 4). To construct this graph, we have aggregated the residuals of the student-level regressions to the country level, the level where instrument and teacher skills vary. We observe a clear positive relationship between the relative wages of non-teacher public-sector employees and teacher skills, as indicated by the solid regression lines. Excluding the three outliers Finland, Italy, and Japan leads to similar regression lines, with even slightly larger slopes. Therefore, the correlation between instrument and teacher skills in first-stage estimation is not driven by outliers.[51]

The second-stage results of the IV estimations are reported in the upper panel of Table 4. Higher teacher numeracy skills significantly increase student math performance (Columns 1-3). In the preferred specification in Column (3) which controls for the skills of parents of PISA students to net out the intergenerational persistence in skills, we find that a one-standard-deviation increase in teacher numeracy skills increases student math performance by 20 percent of an international standard deviation in test scores.

The IV coefficient on teacher numeracy skills thus indicates a sizable impact of a country's teacher skills on student performance. It is important to note that this estimate does not capture the effect of a single school year, but rather reflects the cumulative effect of teacher skills on student performance over all school years because teachers in the PIAAC sample have been teaching PISA students for several years.[52] Moreover, the IV estimate in the last specification is about twice as large as the corresponding OLS estimate. This coefficient increase is likely to reflect the elimination of the attenuation bias due to measurement error in the teacher-skill variable. As the instrument uses distributional information on all non-teacher public-sector employees, it provides more precise country-specific information than the observed teacher skills that helps to refine the picture of how teachers differ across countries. Furthermore, and most importantly, the instrument is likely uncorrelated with the measurement error in the teacher skills variable (see discussion in Section 4).

Our other main control variable, parent skills, also enters positively and significantly in the second stage for math. However, the coefficient on parent skills decreases somewhat in magnitude compared to the OLS specifications and captures less of the effect of teacher skills (the coefficient on teacher skills decreases by only 7 percent in the IV regressions between Columns 2 and 3, but

[51] In Section 5.3, we additionally provide a robustness check that excludes the three outlier countries from the sample.

[52] The average teacher age in our sample is 42.2 years.

decreases by 18 percent in the corresponding OLS models). This suggests that cross-country differences in teacher skills attributable to international differences in relative wages of public-sector employees are less correlated with parent skills.

A similar pattern holds for reading (Columns 4-6). Better teacher skills lead to improved student performance, irrespective of the included control variables. In the specification with all controls (Column 6), an increase in teacher literacy skills by one standard deviation improves student performance by about 10 percent of an international standard deviation. This effect size is only half of that in math, indicating that subject-specific teacher skills are more important for math than for reading. This finding is consistent with individual-level evidence provided in Metzler and Woessmann (2012) for Peruvian students.

In contrast to the numeracy results, the IV estimate of teachers' literacy skills is very close to the OLS estimate. A potential explanation for this finding is that attenuation bias due to measurement error is less severe for literacy skills than it is for numeracy skills of teachers. One source of such domain-specific measurement error could be that teacher skills in numeracy are more noisily measured than literacy skills because the numeracy questions are relatively hard to solve for non-math teachers in the national PIAAC samples, whereas math teachers have fewer difficulties answering literacy questions correctly.

5.3 Further Results

In this section, we show that our main results reported in Table 4 are robust to alternative specifications and samples. We furthermore investigate whether our teacher-skill measures just reflect pedagogical skills of teachers. Finally, we also explore whether the impact of teacher skills differs by gender, socioeconomic background, or migrant status of students.

Robustness Checks

Since teacher skills vary across countries, our first robustness check replaces individual-level parent skills with country-level parent skills, as measured by the median skills of all PIAAC respondents aged 35-59 with children (i.e., the same PIAAC respondents used to construct the individual-level parent skills). Using country-level parent skills increases the coefficients on teacher skills slightly (Columns (1) and (5) in Table 5). We obtain very similar results when we replace the country-specific parent skills with country-specific *adult* skills, as measured by the median skill level of all adults aged 25-65 (Columns 2 and 6).These findings show that the impact of teacher skills remains unchanged even if we control for the general skill level of the population at the country level (the level where teacher skills vary).

Any strategy that exploits international variation with limited degrees of freedom might suffer from the problem that the results are driven by a few outlier countries. Therefore, we replicate the main specifications, but now exclude the three countries that are outliers in the first-stage regressions (see Figure 4). Even without the outlier countries, teacher skills enter significantly in the second-stage regressions, and first-stage results still indicate that the instrument is strong (Columns 3 and 7). As we exclude these three countries from the sample, the impact of teacher skills gets even larger, especially in math. Due to large standard errors, however, the increase is not statistically significant. We also excluded each country individually from the sample (results available upon request). The estimated teacher-skill effects are always close to the baseline coefficients, confirming that the results are not driven by an individual country. As a final specification check, we use *average* teacher skills instead of median teacher skills. The coefficients, reported in Columns (4) and (8), are very close to the baseline estimates.[53]

Adding Teaching Practices

One worry is that our subject-specific teacher-skill measures just reflect pedagogical skills. To investigate whether pedagogical skills indeed confound the teacher-skill effects, we use information from the PISA students about their teachers' activities in language and math classes (e.g., how often does a teacher ask questions that make students reflect on a problem). We construct indicators of subject-specific instruction quality – as measures for teachers' pedagogical skills – in the following way: We first aggregate all teaching activities that are likely to promote student learning at the student level.[54] Then, we average the student-level answers for each school to obtain school-level instruction-quality indicators. As discussed in Section 3, teaching practices are unfortunately asked only regarding the subject that was the focus in the respective PISA cycle. For the PISA cycle when a subject (math or language) was not the focus, we "impute" the subject-specific instruction-quality indicator by using the country-level instruction quality from the other PISA survey, assuming that

[53] Although the first-stage F statistic in the reading regression decreases compared to the analogous result in Table 4, it is still sizeable, and the point estimate in the second stage is practically identical.

[54] PISA students are asked about the frequency of the activities their teachers do in language classes (PISA 2009) and in math classes (PISA 2012), respectively. For reading, we use the following items (each measured on a 4-point scale ranging from "never or hardly ever" to "in all lessons") to construct the instruction-quality indicator: asking students to explain the meaning of a text; asking questions that challenge students to get a better understanding of a text; giving students enough time to think about their answers; recommending books or author to read; encouraging students to express their opinion about a text; helping students relate the stories they read to their lives; and showing students how the information in texts builds on what they already know. For math, we use the following items (each measured on a very similar 4-point scale ranging from "never or rarely" to "almost or almost always"): asking questions that make students reflect on the problem; giving problems that require students to think for an extended time; presenting problems in different contexts so that students know whether they have understood the concepts; helping students to learn from mistakes they have made; asking students to explain how they have solved a problem; and presenting problems that require students to apply what they have learnt to new contexts.

the teaching practices in the same subject are highly correlated at the country level across the three-year period.

Table 6 reports the instrumental-variable results when we take into account the instruction quality in math and language classes. For comparison, we report the baseline results for math in Column (1) and for reading in Column (3). When instruction quality is added to the model, the coefficients on teacher skills change only little, suggesting that the subject-specific teacher skills have a strong independent impact on student performance. In fact, the coefficients on teacher skills even increase slightly when instruction quality is included since teacher skills and instruction quality are negatively correlated at the country level (r=-0.30 in math and r=-0.42 in reading). As expected, the instruction-quality indicators are positively related to student performance, although only the instruction quality in language classes captures statistical significance. The magnitude of the language instruction quality is sizeable; 0.036 SD improvement in student reading performance for a 1 SD increase in (country-level) instruction quality. One potential problem that these estimates suffer from is that the country-level instruction-quality indicators likely reflect cultural differences, partly just capturing how actively teachers communicate with their students. Therefore, it does not come as a surprise that the instruction quality is highest in Anglo-Saxon countries, but lowest in Asian countries.

To gain confidence that the negative correlation between subject-specific teacher skills and instruction quality is neither an artifact of the construction of this particular instruction-quality indicator nor driven by systematic misreporting by students, we have additionally looked at country-level information on teaching practices from TALIS 2013 (see OECD, 2014b for details). In contrast to PISA, TALIS asks teachers to report their own teaching practices.[55] In line with the PISA-based instruction-quality results, all teaching practices surveyed in TALIS are negatively correlated with teacher skills.[56] Thus, our results consistently indicate that the impact of the subject-specific teacher skills does not merely (or even mainly) reflect better pedagogical skills of teachers.

[55] Teaching practices assessed in TALIS include: present a summary of recently learned content; students work in small groups to come up with a joint solution to a problem or task; give different work to the students who have difficulties learning and/or to those who can advance faster; refer to a problem from everyday life or work to demonstrate why new knowledge is useful; let students practice similar tasks until teacher knows that every student has understood the subject matter; check students' exercise books or homework; students work on projects that require at least one week to complete; students use ICT for projects or class work.

[56] We do not use teaching practices from TALIS in the student-level regressions for three reasons. First, four of the 23 countries in our sample (Austria, Germany, Ireland, and the Russian Federation) did not participate in TALIS 2013, which would substantially reduce our sample. Second, at the time of writing, TALIS 2013 micro-data were not available, so we would have to rely on the aggregate data published by the OECD. However, the OECD does not provide sufficient information on how the country-level indicators of teaching practices have been constructed. Third, the OECD only provides teaching practices for all (lower secondary) teachers, which means that the teaching practices in TALIS are *not* subject-specific.

Effect Heterogeneity

Thus far, the effect of teacher skills were estimated for the entire student sample, thus showing the impact for the average student. In Table 7, we explore whether the impact of teacher skills differs across various student subgroups. Panel A reports results for math and Panel B for reading, while all specifications include the full set of control variables. First, we stratify the sample by student gender. We find identical effect sizes in reading, but a larger effect for girls in math. Due to the large standard errors, however, this gender difference is not statistically significant.

Next, we split the sample by students' socioeconomic background, as measured by the PISA index of economic, social, and cultural status (ESCS). This index captures a range of aspects of a student's family and home background that combines information on parents' education, occupations, and home possessions. Using the country-specific median ESCS scores to split the sample, we find that the effect of teacher skills on student performance is substantially larger for students with low socioeconomic background. The results furthermore suggest that higher teacher literacy skills (at least when measured at the country level) do not improve the reading performance of high-SES students (while the effect in math is sizeable). Interestingly, parent skills seem to be more important for high-SES students than for low-SES students. A one-standard-deviation increase in parent numeracy skills is associated with an increase in math performance of high-SES students of 4.5 percent of a standard deviation; the corresponding estimate for low-SES students is only about half the size. In reading, parent literacy skills are also significantly positive for high-SES students (zero for low-SES students). These results suggest that the benefits of teacher subject skills mainly accrue to students with low socioeconomic background, while parental skills are more important for students with high socioeconomic background.

Finally, we estimate teacher-skill effects separately for natives and migrants.[57] The pattern is less conclusive here. Teacher skills seem somewhat more important for migrants in math and for natives in reading. However, the differences of the point estimates are not statistically significant for either math or reading. Moreover, a cautious interpretation of the results for migrants is in order given the limited sample size.

[57] Because first-generation migrants might have migrated into the PISA test country just shortly before the PISA test, we can hardly ascribe their math and reading performance to the skill level of teachers in the test country. Therefore, we use only second-generation migrants here since these students were born in the PISA test country and have spent their school career in the education system of the test country.

6. Policy Implications

Our analysis consistently indicates that students living in the countries at the top of the PISA rankings perform better in math and reading because their teachers have higher numeracy and literacy skills. This raises the natural question what policymakers can do to improve teacher skills.

Before addressing this question, it is useful to understand what the estimates say about the impact of raising teacher skills. When we look across our 23 sampled countries, we see that Finland does in fact have the most skilled teachers by the PIAAC measures. Table 8 uses the estimated achievement models to simulate the improved student performance if each country brought its teachers up to the level of Finnish teachers. For some, such as Japan, this is not a huge change, but even Japanese schools would improve noticeably (0.10 s.d. in mathematics and 0.03 s.d. in reading). But for other countries, the improvements in student achievement would be dramatic (if they could improve their teachers). The U.S. would be expected to improve by roughly 0.55 s.d. in math. Russia and Italy would be expected to improve by almost 0.75 s.d. in student achievement. Of course, these are long-run impacts since they presume that the quality of students' teachers in the first ten grades would improve to the level of Finland – something that would take some time and effort to realize.

One approach to increase teacher skills, which has other advantages for a country, is to increase the overall achievement of its population. Of course, this is not easy, and considerable controversy surrounds the best way to do this. The clearest policy direction, however, appears to be improving the incentives for higher achievement (e.g., see Hanushek and Woessmann (2011b)). While beyond the scope of this paper, this approach would, by available evidence, rely on strong accountability of achievement, parental choice of schools, and rewards for performance.

Another option for policymakers is to try to attract better performers out of the existing skill distribution for the country. One way to do this may be to raise teacher wages to attract better-skilled individuals into the teaching profession. In fact, the argument that teacher pay is significantly related to teacher quality has been in the heart of the debate about educational policy for many years (see, e.g., Dolton and Marcenaro-Gutierrez (2011)). The idea is that countries that pay teachers relatively better recruit teachers from a higher part of the skill distribution and also manage to retain teachers in their profession.[58] If this link was present, there would be leverage for

[58] Raising pay might provide already-recruited teachers with more incentives to exert higher effort to improve the educational outcomes of the children they teach. The evidence on this is, however, not very encouraging. See Springer et al. (2010). For developing countries, however, the evidence is stronger; see Muralidharan and Sundararaman (2011).

policymakers to raise the skills of teachers in the country by paying them higher wages, with positive effects on student performance.[59]

In Table 9, we investigate whether teacher skills are indeed higher in countries that pay teachers (relatively) higher wages. Based on the PIAAC data, we run country-level regressions of teacher skills (separately for numeracy and literacy) on relative teacher wages, measured as the percentile rank of country-specific mean teacher wages in the wage distribution of all non-teacher college graduates. Importantly, estimates are conditioned on the skill level of all non-teacher college graduates to account for the differences in skills levels between countries. The results indicate that higher relative teacher pay is systematically related to higher teacher skills. For example, controlling for the wage level of college graduates (Columns 3 and 6), we find that a one-standard-deviation increase in relative teacher salary (that is, 15 percentile ranks) is associated with an increase in teacher skills in numeracy (literacy) of about 40 percent (30 percent) of an international standard deviation.[60] The coefficient on college graduates' skills is always close to unity, reflecting the fact that most teachers are college graduates themselves.

7. Conclusion

Student performance differs greatly across countries, but so far little is known about the importance of teacher quality in explaining these differences. In this paper, we use newly available data from the Programme for the International Assessment of Adult Competencies (PIAAC) to calculate country-level measures of teacher skills in numeracy and literacy in 23 developed economies. We first show that teacher skills differ substantially across countries. We then combine teacher skills with micro data on student performance from PISA to estimate international education production functions that extensively control for student, school, and country background factors, including coarse measures of the skills of PISA students' parents.

To overcome biases due to omitted country-level factors and measurement error in teacher skills, we exploit international differences in relative wages of non-teacher public-sector employees to obtain exogenous variation in teacher skills. The instrumental-variable results indicate that cross-country differences in teacher skills are an important determinant of international differences in student performance: A one-standard-deviation increase in teacher numeracy skills raises student

[59] Another channel through which a positive association between teacher pay and teacher skills may materialize (at least in the long run) is that higher salaries for teachers may improve the status of the teaching profession. As a result, more children might want to become teachers in the future, facilitating the recruitment of more able individuals.

[60] These estimates reflect both long-term incentives and teacher-sorting mechanisms of teacher pay (see Woessmann, 2011, for a discussion).

performance in math by 20 percent of an international standard deviation in test scores. The effect in reading is 10 percent of a standard deviation and also highly statistically significant.

Additional specifications that control for the general skill level in a country in various ways confirm that the teacher-skill effects do not just reflect the intergenerational persistence in skills. Results are also supported by an alternative identification strategy that relies on within-country variation across subjects, eliminating all factors that equally affect student performance in math and reading.

Further country-level regressions indicate that teacher skills (both in numeracy and literacy) are higher in countries where teachers are paid relatively well compared to other college graduates. This finding suggests that policymakers can attract and retain higher-ability individuals in the teaching profession by paying higher wages.

This paper provides evidence that teacher subject knowledge is one important determinant of differences in student performance across developed countries. However, there is reason to suspect that teacher quality plays a different role in developing countries. For instance, teacher skills might be less important in developing countries because of high absenteeism rates of teachers (Duflo, Rema, and Ryan (2012)). On the other hand, teacher skills could also be more relevant in developing countries since educational institutions are typically weaker there (e.g., Hanushek, Link, and Woessmann (2013)). Studying the impact of subject-specific teacher skills in developing countries is therefore an important avenue for future research that would complement the evidence provided in this paper.

References

Ashenfelter, Orley, and David J. Zimmerman. 1997. "Estimates of the returns to schooling from sibling data: Fathers, sons, and brothers." *Review of Economics and Statistics* 79, no. 1: 1-9.

Barber, Michael, and Mona Mourshed. 2007. *How the world's best-performing school systems come out on top*: McKinsey and Company.

Björklund, Anders, and Kjell G. Salvanes. 2011. "Education and family background: Mechanisms and policies." In *Handbook of the Economics of Education, Vol. 3*, edited by Stephen Machin Eric A. Hanushek and Ludger Woessmann. Amsterdam: North Holland: 201-247.

Bound, John, David A. Jaeger, and Regina M. Baker. 1995. "Problems with instrumental variables estiamtion when the correlation between the instruments and the endogenous explanatory variable is weak." *Journal of the American Statistical Association* 90, no. 430 (June): 443-450.

Chetty, Raj, John N. Friedman, and Jonah Rockoff. 2014. "Measuring the impacts of teachers II: Teacher value-added and the student ourcomes in adulthood." *American Economic Review*: forthcoming.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2013. "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood." Mimeo. Cambridge, MA: Harvard University.

Dee, Thomas S. 2005. "A teacher like me: Does race, ethnicity, or gender matter?" *American Economic Review* 95, no. 2: 158-165.

Dolton, Peter, and Oscar D. Marcenaro-Gutierrez. 2011. "If you pay peanuts do you get monkeys? A cross-country analysis of teacher pay and pupil performance." *Economic Policy* 26, no. 65 (January): 5-55.

Eide, Eric, Dan Goldhaber, and Dominic Brewer. 2004. "The teacher labour market and teacher quality." *Oxford Review of Economic Policy* 20, no. 2: 230-244.

Glewwe, Paul, Eric A. Hanushek, Sarah D. Humpage, and Renato Ravina. 2013. "School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010." In *Education Policy in Developing Countries*, edited by Paul Glewwe. Chicago: University of Chicago Press: 13-64.

Hanushek, Eric A. 1995. "Interpreting recent research on schooling in developing countries." *World Bank Research Observer* 10, no. 2 (August): 227-246.

Hanushek, Eric A. 2002. "Publicly provided education." In *Handbook of Public Economics, Vol. 4*, edited by Alan J. Auerbach and Martin Feldstein. Amsterdam: North Holland: 2045-2141.

Hanushek, Eric A. 2003. "The failure of input-based schooling policies." *Economic Journal* 113, no. 485 (February): F64-F98.

Hanushek, Eric A. 2006. "School resources." In *Handbook of the Economics of Education*, edited by Eric A. Hanushek and Finis Welch. Amsterdam: North Holland: 865-908.

Hanushek, Eric A., and Dennis D. Kimko. 2000. "Schooling, labor force quality, and the growth of nations." *American Economic Review* 90, no. 5 (December): 1184-1208.

Hanushek, Eric A., Susanne Link, and Ludger Woessmann. 2013. "Does school autonomy make sense everywhere? Panel estimates from PISA." *Journal of Development Economics* 104(September): 212-232.

Hanushek, Eric A., and Steven G. Rivkin. 2006. "Teacher quality." In *Handbook of the Economics of Education, Vol. 2*, edited by Eric A. Hanushek and Finis Welch. Amsterdam: North Holland: 1051-1078.

Hanushek, Eric A., and Steven G. Rivkin. 2010. "Generalizations about using value-added measures of teacher quality." *American Economic Review* 100, no. 2 (May): 267-271.

Hanushek, Eric A., and Steven G. Rivkin. 2012. "The distribution of teacher quality and implications for policy." *Annual Review of Economics* 4: 131-157.

Hanushek, Eric A., Guido Schwerdt, Simon Wiederhold, and Ludger Woessmann. 2013. "Returns to skills around the world." NBER Working Paper 19762. Cambridge, MA: National Bureau of Economic Research (December).

Hanushek, Eric A., and Ludger Woessmann. 2011a. "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: 89-200.

Hanushek, Eric A., and Ludger Woessmann. 2011b. "How much do educational outcomes matter in OECD countries?" *Economic Policy* 26, no. 67 (July): 427-491.

Hanushek, Eric A., and Ludger Woessmann. 2012. "Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation." *Journal of Economic Growth* 17, no. 4 (December): 267-321.

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. 2008. "What does certification tell us about teacher effectiveness? Evidence from New York City." *Economics of Education Review* 27, no. 6 (December): 615-631.

Lavy, Victor. forthcoming. "Do Differences in School's Instruction Time Explain International Achievement Gaps in Math, Science, and Reading? Evidence from Developed and Developing Countries." *Economic Journal*.

Leigh, Andrew. 2013. "The Economics and Politics of Teacher Merit Pay." *CESifo Economic Studies* 59, no. 1 (March): 1-33.

Metzler, Johannes, and Ludger Woessmann. 2012. "The impact of teacher subject knowledge on student achievement: Evidence from within-teacher within-student variation." *Journal of Development Economics* 99, no. 2 (November): 486-496.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India." *Journal of Political Economy* 119, no. 1: 39-77.

OECD. 2013. *OECD skills outlook 2013: First results from the survey of adult skills*. Paris: Organisation for Economic Co-operation and Development.

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.

Schleicher, Andreas. 2013. "What teachers know and how that compares with college graduates around the world." In *Education Today: Global perspectives on education and skills*. Paris: OECD.

Springer, Matthew G., Dale Ballou, Laura Hamilton, Vi-Nhuan Le, J.R. Lockwood, Daniel F. McCaffrey, Matthew Pepper, and Brian M. Stecher. 2010. *Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching*. Nashville, TN: National Center on Performance Incentives, Vanderbilt University.

Stock, James H., Jonathan H. Wright, and Motohiro Yogo. 2002. "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments." *Journal of Business & Economic Statistics* 20, no. 4 (October): 518-529.

West, Martin R., and Ludger Woessmann. 2010. "'Every Catholic child in a Catholic school': Historical resistance to state schooling, contemporary private competition and student achievement across countries." *Economic Journal* 120, no. 546: F229-F255.

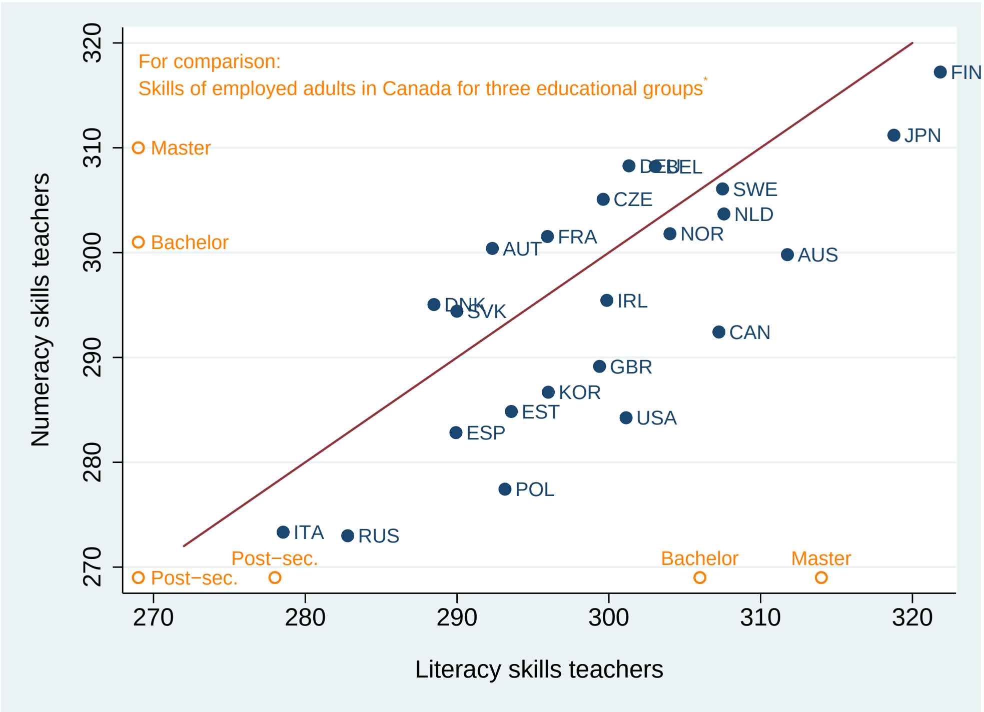
Woessmann, Ludger. 2003. "Schooling resources, educational institutions, and student performance: The international evidence." *Oxford Bulletin of Economics and Statistics* 65, no. 2: 117-170.

Woessmann, Ludger. 2011. "Cross-country evidence on teacher performance pay." *Economics of Education Review* 30, no. 3 (June): 404-418.

Woessmann, Ludger, Elke Luedemann, Gabriela Schuetz, and Martin R. West. 2009. *School accountability, autonomy, and choice around the world*. Cheltenham, UK: Edward Elgar.

Figures and Tables

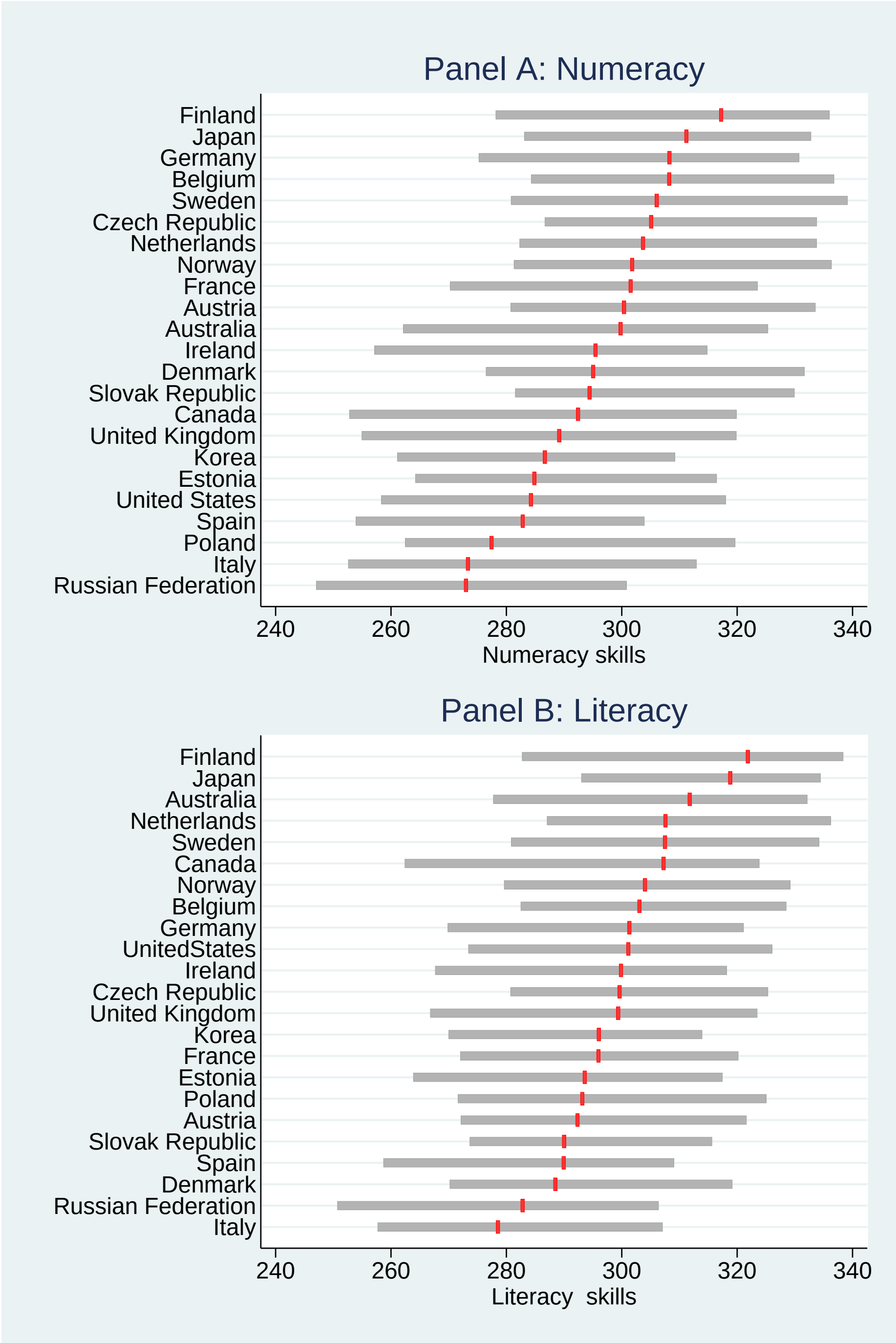
Figure 1: Teacher Skills



Note: The blue dots indicate country-specific teacher skills in numeracy and literacy (see text for construction of teacher skills). The orange circles indicate the median skills for three educational groups of employed adults in Canada aged 25-65 years. *Post-sec.* includes individuals with vocational education (post-secondary, non-tertiary) as highest degree (2,434 observations); *Bachelor* includes individuals with bachelor degree (3,671 observations); *Master* includes individuals with a master or doctoral degree (1,052 observations). *Data source*: PIAAC.

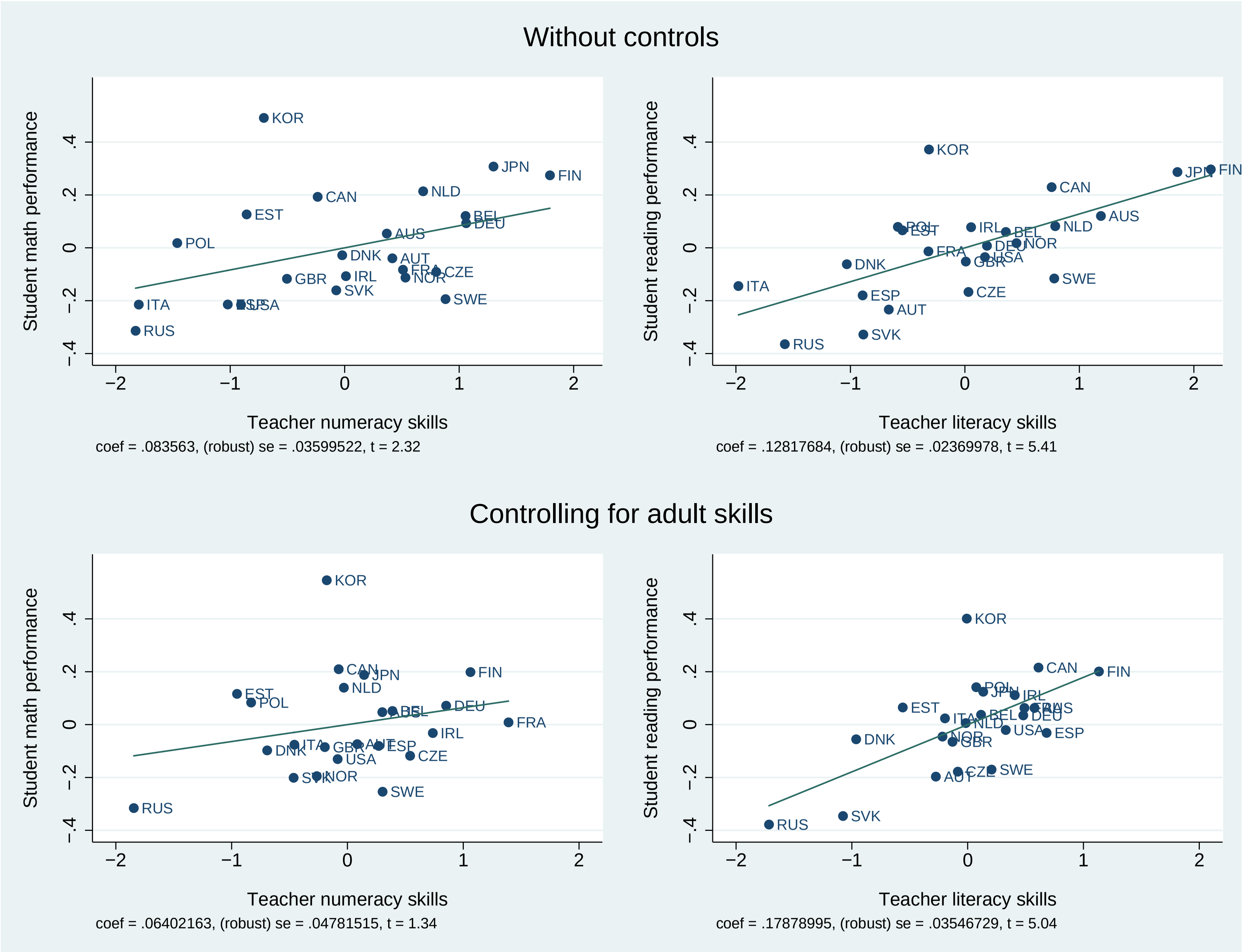
* Canada is used for the skill comparison as it provides by far the largest national sample in PIAAC.

Figure 2: Position of Teacher Skills in the Skill Distribution of College Graduates



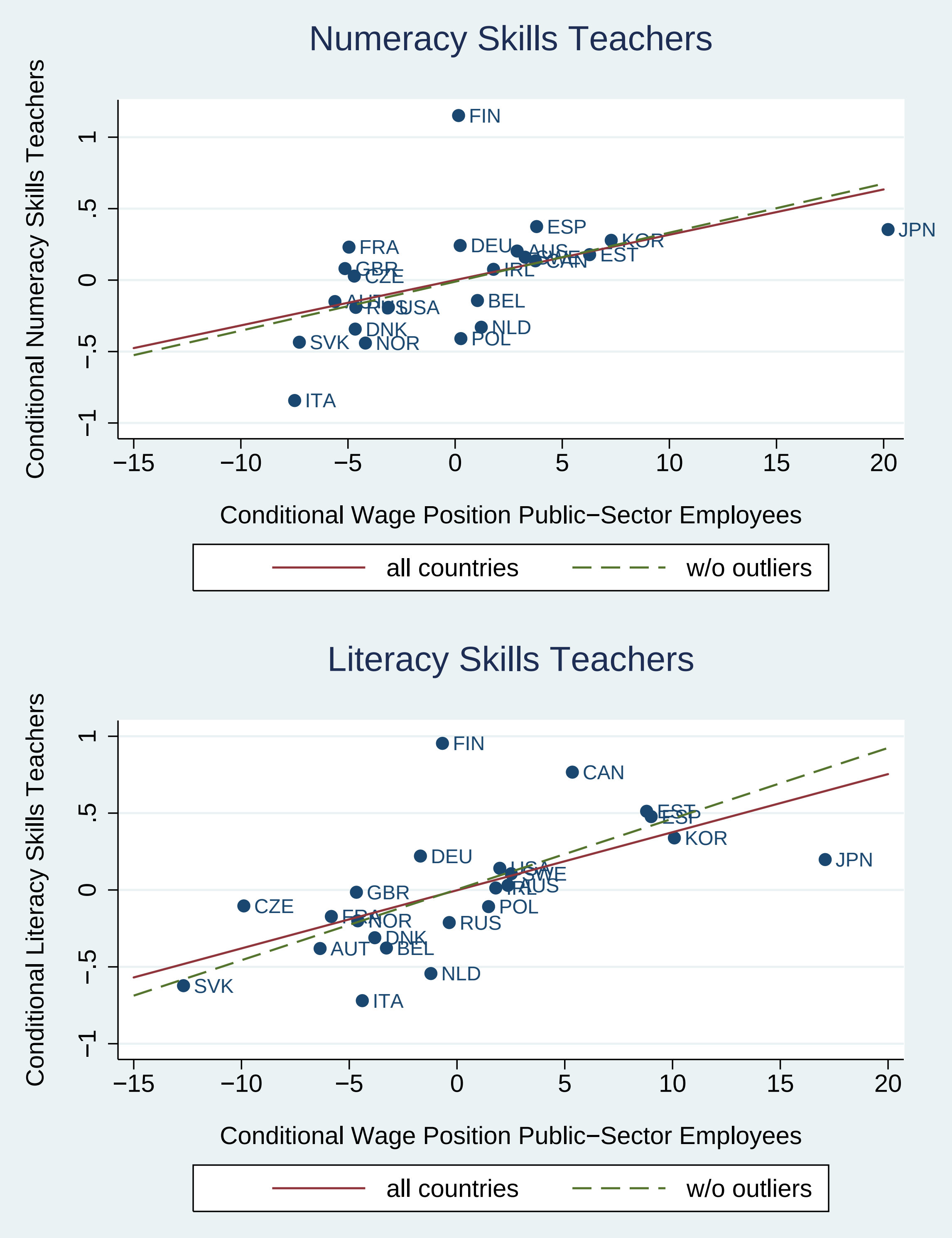
Note: Figure in accordance to Schleicher (2013). Vertical red bars indicate median teacher skills in a country. Horizontal grey bars show the interval of skills levels of all college graduates (including teachers) between the 25th and 75th percentile. Countries are ranked by the median teacher skills in numeracy and literacy, respectively. *Data source*: PIAAC.

Figure 3: Student Performance and Teacher Skills



Note: The two graphs in the first column do not include any controls. Two graphs in bottom row: Added-variable plot which controls for country-specific average literacy skills of all adults aged 25-65. *Data source*: PISA 2009 and 2012, PIAAC.

Figure 4: Relative Wages Public-Sector Employees and Teacher Skills (First Stage)



Note: Added-variable plot of a regression of teacher skills and wage position of public-sector employees (w/o teachers) in the distribution of all employees and all the control variable included in (2). Upper (lower) panel shows teacher numeracy (literacy) skills. Based on student-level regressions that are then aggregated to the country level. Solid line is fitted through all country-level observations; for fitting the dashed line, the outliers Finland, Italy, and Japan are excluded. *Data source*: PISA 2009 and 2012, PIAAC.

Table 1: Teacher Skills

	Pooled	Australia	Austria	Belgium	Canada	Czech R.	Denmark	Estonia	Finland	France	Germany	Ireland
Numeracy	295	300	300	308	292	305	295	285	317	302	308	295
Literacy	299	312	292	303	307	300	288	294	322	296	301	300
Difference	-4	-12	8	5	-15	5	7	-9	-5	6	7	-4
Rank numeracy	68	71	69	68	67	73	56	60	73	80	72	75
Rank literacy	70	75	70	71	72	77	60	69	74	77	74	74
Observations	5,322	248	188	215	834	141	413	188	221	163	127	180
	Italy	Japan	Korea	Netherl.	Norway	Poland	Russia	Slovak R.	Spain	Sweden	U.K.	U.S.
Numeracy	273	311	287	304	302	277	273	294	283	306	289	284
Literacy	279	319	296	308	304	293	283	290	290	307	299	301
Difference	-5	-8	-9	-4	-2	-16	-10	4	-7	-1	-10	-17
Rank numeracy	67	70	72	63	65	64	53	66	75	62	65	70
Rank literacy	73	67	74	67	68	73	54	60	80	65	67	71
Observations	124	147	217	197	279	199	137	133	183	147	310	132

Notes: Teacher skills are country-specific average skills of primary school teachers, secondary school teachers, and “other” teachers (including, e.g., special education teachers and language teachers). Because occupation in these countries is reported only at the two-digit level, teachers in Australia and Finland include all "teaching professionals" (ISCO-08 code 23), i.e. additionally include pre-kindergarten teachers and university professors. All skill measures are rounded to the nearest integer. Rank refers to the position of average teacher skills in the skill distribution of all adults aged 25-65 excluding teachers. Individuals are weighted with PIAAC final sample weights. Observations refer to the number of teachers used to construct country-specific teacher skills. *Data source:* PIAAC.

Table 2: Ordinary Least Squares Results

Dependent variable: student performance						
	Math			Reading		
	(1)	(2)	(3)	(4)	(5)	(6)
Teacher skills	0.084**	0.117***	0.096***	0.128***	0.086***	0.082***
	(0.035)	(0.021)	(0.021)	(0.023)	(0.020)	(0.022)
Parent skills			0.039***			0.007
			(0.011)			(0.010)
Student characteristics	No	Yes	Yes	No	Yes	Yes
Parent characteristics	No	Yes	Yes	No	Yes	Yes
School characteristics	No	Yes	Yes	No	Yes	Yes
Country characteristics	No	Yes	Yes	No	Yes	Yes
Students	406,564	406,564	406,564	406,564	406,564	406,564
Countries	23	23	23	23	23	23
Adj. R2	0.01	0.26	0.26	0.02	0.29	0.29

Notes: Least squares regressions weighted by students' inverse sampling probability, giving each country the same weight. Dependent variable: student PISA test score in math (columns (1–3) and in reading (columns 4–6). Student test scores are z-standardized at the individual level within each PISA cycle. Country-level teacher skills refer to numeracy in columns (1)–(3) and to literacy in columns (4)–(6). Teacher skills are z-standardized across countries. Parent skills are computed as the mean of mother's and father's skills in numeracy (columns 1–3) or literacy (columns 4–6). Parent skills are standardized using teacher skills as "numeraire" scale. Student characteristics are age, gender, migrant status (first-generation or second-generation), and language spoken at home. Parent characteristics include parents' educational degree, number of books at home, and type of occupation. School characteristics include school location, number of students per school, and three autonomy measures. Country characteristics are expenditures per student, GDP per capita, and school starting age (Table A1 reports results for all control variables). All regressions include controls for respective imputation dummies and a dummy indicating the PISA wave. Robust standard errors, adjusted for clustering at the country level, in parentheses. Significance levels: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. *Data sources:* PIAAC, PISA 2009 and 2012.

Table 3: First-Differenced (Student Fixed-Effects) Results

Dependent variable: student performance difference: math – reading			
	(1)	(2)	(3)
Teacher skills: numeracy – literacy	0.075***	0.090***	0.053*
	(0.024)	(0.021)	(0.027)
Parent skills: numeracy – literacy			0.053
			(0.031)
Instruction time: math – reading		0.066***	0.073***
		(0.017)	(0.015)
Shortage teachers: math – reading		–0.004	–0.003
		(0.007)	(0.007)
Students	406,564	406,564	406,564
Countries	23	23	23
Adj. R2	0.01	0.02	0.02

Notes: Dependent variable: student test score difference between math and reading. All regressions include controls for respective imputation dummies and for the PISA wave. Specifications give equal weight to each country. Robust standard errors, adjusted for clustering at country level, in parentheses. Significance levels: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. *Data sources:* PIAAC, PISA 2009 and 2012.

Table 4: Instrumental-Variable Results

Second stage (Dependent variable: student performance)						
	Math			Reading		
	(1)	(2)	(3)	(4)	(5)	(6)
Teacher skills	0.319***	0.217***	0.202***	0.326***	0.103**	0.099**
	(0.106)	(0.070)	(0.072)	(0.084)	(0.047)	(0.050)
Parent skills			0.029**			0.008
			(0.014)			(0.011)
Student characteristics	No	Yes	Yes	No	Yes	Yes
Parent characteristics	No	Yes	Yes	No	Yes	Yes
School characteristics	No	Yes	Yes	No	Yes	Yes
Country characteristics	No	Yes	Yes	No	Yes	Yes
First stage (Dependent variable: teacher skills)						
	Math			Reading		
Wage position public-sector employees	0.034***	0.031***	0.031***	0.029***	0.034***	0.034***
	(0.008)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Parent skills			0.038			0.006
			(0.055)			(0.049)
Instrument F statistic	18.0	19.0	17.9	22.6	30.3	29.4
Students	406,564	406,564	406,564	406,564	406,564	406,564
Countries	23	23	23	23	23	23

Notes: Dependent variable: student PISA test score in math (columns 1–3) and reading (columns 4–6), respectively. *Wage position public-sector employees* is the country-specific percentile rank of the mean wage of non-teacher public-sector employees in the wage distribution of non-teacher private-sector college graduates; the cross-country standard deviation is 12.7. All skill measures in columns (1)–(3) (4–6) refer to numeracy (literacy). Student, parent, school, and country characteristics are the same as in the OLS models (see Table 2). All regressions additionally control for median skills of university graduates in a country. All regressions include controls for respective imputation dummies and for the PISA wave. Specifications give equal weight to each country. Robust standard errors, adjusted for clustering at country level, in parentheses. Significance levels: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. *Data sources:* PIAAC, PISA 2009 and 2012.

Table 5: Robustness Checks

Second stage (Dependent variable: student performance)	Math				Reading			
	Parent skills	Adult skills	Excluding outliers	Mean teacher skills	Parent skills	Adult skills	Excluding outliers	Mean teacher skills
Teacher skills	0.224***	0.215**	0.394***	0.188**	0.136**	0.143*	0.131*	0.103**
	(0.075)	(0.083)	(0.138)	(0.091)	(0.065)	(0.074)	(0.077)	(0.052)
Parent skills (country level)	–0.007				–0.051			
	(0.039)				(0.040)			
Adult skills (country level)		0.002				–0.063		
		(0.048)				(0.053)		
Parent skills			0.038**	0.029**			0.007	0.011
			(0.017)	(0.012)			(0.009)	(0.010)
Student characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First stage (Dependent variable: teacher skills)	Math				Reading			
Wage position public-sector employees	0.031***	0.029***	0.028***	0.033***	0.032***	0.030***	0.033***	0.033***
	(0.006)	(0.007)	(0.005)	(0.007)	(0.006)	(0.007)	(0.005)	(0.010)
Instrument F statistic	23.7	19.6	25.9	22.4	28.9	20.8	45.4	10.4
Students	406,564	406,564	317,508	406,564	406,564	406,564	317,508	406,564
Countries	23	23	20	23	23	23	20	23

Notes: Robustness checks of the instrumental-variable estimations. Dependent variable: student PISA test score in math (columns 1–4) and reading (columns 5–8), respectively. *Wage position public-sector employees* is the country-specific percentile rank of the mean wage of non-teacher public-sector employees in the wage distribution of non-teacher private-sector college graduates. All skill measures in columns (1)–(4) (5–8) refer to numeracy (literacy). In columns (1) and (5), we replace individual-level parent skills by the country-specific median skill level of PIAAC respondents aged 35–59 with children. Analogously, in columns (2) and (6), we use median skill level of all PIAAC respondents aged 25–65 instead of individual parent skills. In columns (3) and (7), we drop Finland, Italy, and Japan, which appear as outliers in the first-stage regression (see Figure 3). In columns (4) and (8), we use teacher mean skills instead of median skills. Student, parent, school, and country characteristics are the same as in the baseline IV models (see Table 4). All regressions also control for median skills of university graduates in a country and include controls for imputation dummies and the PISA wave. Specifications give equal weight to each country. Robust standard errors, adjusted for clustering at country level, in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01. *Data sources:* PIAAC, PISA 2009 and 2012.

Table 6: Teaching Practices

Second stage (Dependent variable: student performance)	Math		Reading	
	Baseline	Instruction quality	Baseline	Instruction quality
Teacher skills	0.202***	0.218***	0.099**	0.104**
	(0.072)	(0.070)	(0.050)	(0.048)
Parent skills	0.029**	0.027*	0.008	0.004
	(0.014)	(0.015)	(0.011)	(0.012)
Instruction quality		0.120		0.350**
		(0.129)		(0.157)
Student characteristics	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes
School characteristics	Yes	Yes	Yes	Yes
Country characteristics	Yes	Yes	Yes	Yes
First stage (Dependent variable: teacher skills)				
	Math		Reading	
Wage position public-sector employees	0.031***	0.033***	0.034***	0.034***
	(0.007)	(0.008)	(0.006)	(0.006)
Instrument F statistic	17.9	17.8	29.4	29.7
Students	406,564	406,564	406,564	406,564
Countries	23	23	23	23

Notes: Dependent variable: student PISA test score in math (columns 1–2) and reading (columns 3–4), respectively. *Wage position public-sector employees* is the country-specific percentile rank of the mean wage of non-teacher public-sector employees in the wage distribution of non-teacher private-sector college graduates. All skill measures in columns (1)–(2) ((3–4)) refer to numeracy (literacy). We derive the indicator for teaching practices using the PISA data. See text for details on the construction of the teaching-practices indicator. All control variables are the same as in the baseline IV models (see Table 4). Specifications give equal weight to each country. Robust standard errors, adjusted for clustering at country level, in parentheses. Significance levels: * p<0.10, ** p<0.05, *** p<0.01. *Data sources:* PIAAC, PISA 2009 and 2012.

Table 7: Heterogeneity Analysis

Dependent variable: student performance						
	Panel A: Math					
	Gender		Parental background		Natives vs. Migrants	
	Boys	Girls	High SES	Low SES	Natives	Migrants
Teacher skills	0.162**	0.244***	0.158**	0.283***	0.188***	0.246*
	(0.073)	(0.072)	(0.078)	(0.078)	(0.071)	(0.140)
Parent skills	0.034**	0.023	0.045**	0.024*	0.037**	–0.005
	(0.014)	(0.015)	(0.018)	(0.014)	(0.015)	(0.018)
Instrument F statistic	17.9	18.0	20.7	16.7	17.8	14.9
	Panel B: Reading					
Teacher skills	0.099**	0.100*	0.026	0.199***	0.087*	0.059
	(0.048)	(0.053)	(0.053)	(0.053)	(0.051)	(0.110)
Parent skills	0.009	0.005	0.026*	0.000	0.014	–0.018
	(0.011)	(0.012)	(0.013)	(0.010)	(0.012)	(0.013)
Instrument F statistic	28.4	30.6	29.3	31.1	29.0	22.6
	Additional controls in Panels A + B					
Student characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Parent characteristics	Yes	Yes	Yes	Yes	Yes	Yes
School characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Country characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Students	204,424	202,140	207,914	198,650	350,912	20,433
Countries	23	23	23	23	23	22

Notes: Table reports estimates of the effect of teacher skills on student performance for the following subsamples: boys, girls, student with a high socioeconomic background, students with as low socioeconomic background, natives, and first-generation or second-generation immigrants. Socioeconomic background is measured by the PISA index of economic, social and cultural status (ESCS). This index captures a range of aspects of a student's family and home background that combines information on parents' education, occupations, and home possessions. Migrants are second-generation migrants. To account for the unequal distribution of migrants across countries, we re-weight regressions based on the sample of natives and migrants, respectively, giving equal weight to each country within each subsample. Korea has no second-generation migrants and thus drop out from the subsample of migrants. All skill measures in the upper (lower) part in the table refer to numeracy (literacy). Student, parent, school, and country characteristics are the same as in the OLS models (see Table 2). All regressions additionally control for median skills of university graduates in a country. All regressions include controls for respective imputation dummies and a dummy indicating the PISA wave. Specifications give equal weight to each country. Robust standard errors, adjusted for clustering at country level, in parentheses. Significance levels: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. *Data sources:* PIAAC, PISA 2009 and 2012.

Table 8: Simulation Analysis: Raising Teacher Skills to Finnish Level

	Numeracy		Literacy	
	Teacher skills difference to Finnish teachers (in PIAAC points)	Student perf. increase (in % of internat. SD)	Teacher skills difference to Finnish teachers (in PIAAC points)	Student perf. increase (in % of internat. SD)
Australia	17	28.7	10	9.5
Austria	17	27.8	30	27.9
Belgium	9	14.9	19	17.7
Canada	25	40.9	15	13.8
Czech R.	12	20.0	22	21.0
Denmark	22	36.6	33	31.5
Estonia	32	53.4	28	26.7
France	16	25.9	26	24.5
Germany	9	14.8	21	19.4
Ireland	22	35.9	22	20.8
Italy	44	72.4	43	40.9
Japan	6	9.9	3	2.9
Korea	31	50.4	26	24.4
Netherl.	14	22.3	14	13.5
Norway	15	25.4	18	16.8
Poland	40	65.6	29	27.1
Russia	44	73.0	39	36.9
Slovak R.	23	37.6	32	30.1
Spain	34	56.7	32	30.2
Sweden	11	18.4	14	13.6
U.K.	28	46.3	22	21.2
U.S.	33	54.4	21	19.6

Notes: This table shows by how much student performance would increase if teacher skills in numeracy and literacy, respectively, were at the levels in Finland (i.e., the country with highest teacher skills in both numeracy and literacy). Columns (1) and (3) show difference in teacher skills to Finland. *Data sources:* PIAAC, PISA 2009 and 2012.

Table 9: Teacher Pay and Teacher Quality

Dependent variable: teacher skills	Numeracy			Literacy		
	(1)	(2)	(3)	(4)	(5)	(6)
Teacher position in college grads wage distribution /10	0.28***	0.26***	0.27***	0.20***	0.19***	0.20***
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Numeracy skills college grads (w/o teachers)	1.05***	0.98***	1.00***			
	(0.10)	(0.12)	(0.12)			
Literacy skills college grads (w/o teachers)				1.03***	0.94***	1.00***
				(0.13)	(0.16)	(0.15)
GDP per capita		0.03*			0.02	
		(0.01)			(0.01)	
Wage college grads (w/o teachers)			0.02			0.01
			(0.02)			(0.02)
Constant	–1.67***	–2.48***	–1.96***	–1.01***	–1.65***	–1.17***
	(0.29)	(0.55)	(0.45)	(0.28)	(0.50)	(0.40)
Countries	23	23	23	23	23	23
Adj. R2	0.76	0.80	0.76	0.72	0.73	0.71

Notes: Dependent variable: teacher skills in numeracy (columns 1-3) and literacy (columns 4-6). *Teacher position in college grads wage distribution / 10* is the percentile rank of the country-specific mean teacher wage in the wage distribution of all non-teacher college graduates (divided by 10). Country-level regressions. Robust standard errors in parentheses. Significance levels: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. *Data source:* PIAAC.

Table A-1: Parent Skills

	Pooled	Australia	Austria	Belgium	Canada	Czech R.	Denmark	Estonia	Finland	France	Germany	Ireland
Numeracy												
Mean	279	287	280	301	282	267	293	264	299	275	289	275
Std. Dev.	23	21	15	22	20	17	21	11	18	26	21	22
Max – Min	88	128	50	108	120	51	141	40	102	132	126	96
Literacy												
Mean	277	293	268	289	284	261	278	262	297	272	279	280
Std. Dev.	20	19	15	20	18	12	20	12	17	21	19	18
Max – Min	80	113	47	96	116	37	148	39	101	106	109	86
Observations	65,576	3,137	2,231	2,251	11,933	2,105	3,352	3,463	2,252	3,086	2,293	2,371

	Italy	Japan	Korea	Netherl.	Norway	Poland	Russia	Slovak R.	Spain	Sweden	U.K.	U.S.
Numeracy												
Mean	267	295	276	295	277	258	266	274	265	275	281	267
Std. Dev.	19	7	17	22	16	12	7	19	22	20	20	32
Max – Min	104	26	85	120	62	43	26	61	94	78	109	135
Literacy												
Mean	264	294	281	293	273	259	276	270	266	272	285	277
Std. Dev.	16	6	15	21	15	10	9	15	21	19	18	27
Max – Min	86	22	76	109	51	36	34	48	87	71	95	122
Observations	1,789	2,103	3,361	2,276	2,228	1,793	1,074	2,442	2,614	1,864	3,578	1,980

Notes: Summary statistics of parents' skills (average skill of mother and father) based on actual parents of PISA students. See text for computation of parent skills. *Observations* refer to the number of adults in the PIAAC samples used for computing parents' skills. *Data sources:* PIAAC, PISA 2009 and 2012.

Table A-2: Summary Statistics of Student Performance and Student Characteristics

	Pooled	Australia	Austria	Belgium	Canada	Czech R.	Denmark	Estonia	Finland	France	Germany	Ireland
Math performance	504	509	500	515	522	496	502	516	530	496	513	494
	(93)	(95)	(94)	(103)	(88)	(94)	(84)	(81)	(85)	(100)	(97)	(86)
Reading performance	502	513	480	508	524	486	496	508	530	501	503	509
	(96)	(98)	(96)	(102)	(91)	(91)	(84)	(82)	(91)	(108)	(93)	(92)
Age (in years)	15.8	15.8	15.8	15.8	15.8	15.8	15.7	15.8	15.7	15.9	15.8	15.7
Female	0.49	0.50	0.51	0.49	0.50	0.48	0.50	0.49	0.49	0.51	0.49	0.49
First-gen. migrant	0.05	0.12	0.06	0.09	0.13	0.02	0.04	0.01	0.02	0.05	0.05	0.12
Second-gen. migrant	0.06	0.12	0.11	0.08	0.15	0.01	0.06	0.07	0.01	0.10	0.11	0.02
Other language	0.08	0.09	0.11	0.22	0.16	0.02	0.05	0.04	0.04	0.08	0.09	0.05
Observations	406,564	28,732	11,345	17,098	44,751	11,391	13,405	9,506	14,639	8,911	9,980	8,953
	Italy	Japan	Korea	Netherl.	Norway	Poland	Russia	Slovak R.	Spain	Sweden	U.K.	U.S.
Math performance	484	533	550	524	494	506	475	489	484	486	493	484
	(93)	(94)	(94)	(90)	(88)	(90)	(86)	(99)	(89)	(93)	(91)	(90)
Reading performance	488	529	537	510	503	509	467	470	485	491	497	498
	(96)	(100)	(83)	(91)	(96)	(89)	(90)	(98)	(90)	(103)	(96)	(94)
Age (in years)	15.7	15.8	15.7	15.7	15.8	15.7	15.8	15.8	15.9	15.7	15.7	15.8
Female	0.48	0.48	0.47	0.50	0.49	0.51	0.50	0.49	0.49	0.49	0.51	0.49
First-gen. migrant	0.06	0.00	0.00	0.04	0.05	0.00	0.05	0.01	0.10	0.06	0.07	0.07
Second-gen. migrant	0.02	0.00	0.00	0.08	0.04	0.00	0.07	0.00	0.01	0.08	0.05	0.13
Other language	0.14	0.00	0.00	0.06	0.07	0.01	0.09	0.06	0.18	0.09	0.07	0.14
Observations	61,978	12,439	10,022	9,220	9,346	9,524	10,539	9,233	51,200	9,303	24,838	10,211

Notes: Means and standard deviations (in parentheses) reported. *Other language* indicates a student who speaks a foreign language at home. *Observations* refer to the number of students in both PISA cycles. Statistics are based on student-level observations weighted with inverse sampling probabilities, giving each PISA cycle the same total weight. *Data sources:* PISA 2009 and 2012.

Table A-3: Summary Statistics of Parent Characteristics

	Pooled	Australia	Austria	Belgium	Canada	Czech R.	Denmark	Estonia	Finland	France	Germany	Ireland
Number of books at home												
0-10 books	0.12	0.09	0.13	0.16	0.10	0.10	0.13	0.07	0.07	0.16	0.11	0.14
11-25 books	0.15	0.12	0.16	0.17	0.14	0.14	0.16	0.14	0.12	0.17	0.13	0.15
26-100 books	0.31	0.30	0.31	0.29	0.31	0.35	0.32	0.31	0.34	0.29	0.29	0.30
101-200 books	0.19	0.21	0.17	0.17	0.21	0.19	0.18	0.21	0.22	0.17	0.20	0.19
201-500 books	0.15	0.18	0.14	0.13	0.16	0.15	0.14	0.17	0.18	0.13	0.17	0.15
More than 500 books	0.08	0.10	0.09	0.08	0.08	0.07	0.07	0.09	0.06	0.07	0.10	0.07
Highest education parents												
HH ISCED 0	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.00
HH ISCED 1	0.01	0.01	0.01	0.02	0.01	0.00	0.01	0.00	0.01	0.01	0.00	0.02
HH ISCED 2	0.06	0.05	0.04	0.03	0.02	0.01	0.05	0.03	0.02	0.09	0.15	0.07
HH ISCED 3B,C	0.09	0.07	0.29	0.05	0.00	0.18	0.13	0.02	0.08	0.19	0.12	0.02
HH ISCED 3A,4	0.28	0.32	0.18	0.28	0.25	0.49	0.15	0.38	0.09	0.19	0.23	0.35
HH ISCED 5B	0.21	0.13	0.28	0.22	0.24	0.09	0.41	0.22	0.27	0.22	0.18	0.18
HH ISCED 5A,6	0.35	0.42	0.20	0.40	0.48	0.23	0.24	0.35	0.53	0.30	0.30	0.35
Occupational status parents												
Blue collar-low skilled	0.06	0.05	0.05	0.09	0.06	0.07	0.05	0.06	0.03	0.07	0.06	0.05
Blue collar-high skilled	0.10	0.08	0.14	0.10	0.07	0.13	0.07	0.14	0.07	0.11	0.10	0.09
White collar-low skilled	0.25	0.17	0.26	0.23	0.21	0.27	0.25	0.23	0.20	0.26	0.29	0.26
White collar-high skilled	0.57	0.68	0.53	0.56	0.64	0.52	0.62	0.55	0.69	0.54	0.53	0.58

Table A-3: Summary Statistics of Parent Characteristics (continued)

	Italy	Japan	Korea	Netherl.	Norway	Poland	Russia	Slovak R.	Spain	Sweden	U.K.	U.S.
Number of books at home												
0-10 books	0.12	0.09	0.05	0.16	0.08	0.11	0.09	0.15	0.09	0.09	0.14	0.21
11-25 books	0.19	0.13	0.09	0.18	0.11	0.20	0.19	0.17	0.15	0.11	0.16	0.18
26-100 books	0.08	0.09	0.10	0.07	0.11	0.07	0.08	0.05	0.09	0.11	0.08	0.05
101-200 books	0.30	0.35	0.29	0.30	0.30	0.34	0.34	0.37	0.32	0.30	0.29	0.29
201-500 books	0.18	0.19	0.23	0.15	0.22	0.17	0.17	0.17	0.21	0.20	0.18	0.15
More than 500 books	0.13	0.15	0.24	0.13	0.19	0.11	0.13	0.10	0.15	0.19	0.15	0.11
Highest education parents												
HH ISCED 0	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.01
HH ISCED 1	0.01	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.07	0.01	0.00	0.02
HH ISCED 2	0.21	0.02	0.04	0.04	0.02	0.04	0.01	0.02	0.18	0.04	0.03	0.05
HH ISCED 3B,C	0.06	0.06	0.07	0.00	0.03	0.39	0.01	0.14	0.02	0.07	0.20	0.00
HH ISCED 3A,4	0.37	0.30	0.34	0.32	0.25	0.33	0.08	0.54	0.25	0.18	0.18	0.34
HH ISCED 5B	0.07	0.15	0.06	0.39	0.39	0.00	0.44	0.06	0.14	0.21	0.23	0.15
HH ISCED 5A,6	0.28	0.47	0.48	0.21	0.30	0.24	0.46	0.23	0.33	0.48	0.36	0.43
Occupational status parents												
Blue collar-low skilled	0.07	0.07	0.04	0.04	0.03	0.07	0.06	0.11	0.09	0.05	0.05	0.07
Blue collar-high skilled	0.17	0.08	0.06	0.06	0.04	0.27	0.11	0.16	0.18	0.05	0.05	0.06
White collar-low skilled	0.28	0.36	0.29	0.20	0.16	0.23	0.26	0.31	0.29	0.24	0.26	0.21
White collar-high skilled	0.45	0.48	0.59	0.68	0.75	0.43	0.54	0.40	0.43	0.65	0.62	0.64

Notes: Mean shares reported. Statistics are based on student-level observations weighted with inverse sampling probabilities, giving each PISA cycle the same total weight. *Data sources:* PISA 2009 and 2012.

Table A-4: Summary Statistics of School Characteristics

	Pooled	Australia	Austria	Belgium	Canada	Czech R.	Denmark	Estonia	Finland	France	Germany	Ireland
Instructional time math	3.5	4.0	2.6	3.5	5.3	3.1	3.7	3.7	2.9	3.5	3.3	3.1
Instructional time reading	3.6	3.9	2.4	3.6	5.3	3.0	5.2	3.3	2.5	3.7	3.1	3.0
Shortage math teachers	1.47	1.89	1.33	1.92	1.44	1.25	1.23	1.45	1.16	1.35	1.78	1.40
Shortage language teachers	1.34	1.53	1.36	1.54	1.26	1.12	1.17	1.30	1.10	1.36	1.46	1.16
Private school	0.21	0.41	0.11	0.69	0.08	0.06	0.24	0.04	0.04	0.20	0.06	0.60
Students per school	706	981	558	718	1032	450	480	557	429	822	702	593
Content autonomy	0.68	0.71	0.58	0.56	0.37	0.88	0.68	0.77	0.64	0.64	0.63	0.69
Personnel autonomy	0.43	0.39	0.08	0.38	0.30	0.88	0.58	0.54	0.24	0.05	0.15	0.34
Budget autonomy	0.83	0.93	0.86	0.69	0.75	0.79	0.96	0.84	0.92	0.97	0.88	0.87
	Italy	Japan	Korea	Netherl.	Norway	Poland	Russia	Slovak R.	Spain	Sweden	U.K.	U.S.
Instructional time math	3.8	3.9	3.6	2.8	3.2	3.4	3.6	3.0	3.5	3.1	3.7	4.3
Instructional time reading	4.7	3.5	3.5	2.8	3.8	3.7	3.1	3.0	3.4	3.0	3.8	4.4
Shortage math teachers	1.69	1.27	1.57	2.10	1.73	1.03	1.71	1.13	1.09	1.35	1.64	1.37
Shortage language teachers	1.64	1.21	1.57	1.74	1.70	1.01	1.63	1.10	1.08	1.19	1.38	1.20
Private school	0.06	0.30	0.42	0.67	0.02	0.03	0.00	0.09	0.33	0.12	0.26	0.08
Students per school	752	750	1116	1023	340	324	566	480	701	420	1062	1381
Content autonomy	0.72	0.92	0.89	0.93	0.49	0.75	0.59	0.59	0.53	0.63	0.89	0.48
Personnel autonomy	0.05	0.32	0.23	0.89	0.42	0.46	0.65	0.70	0.18	0.72	0.75	0.66
Budget autonomy	0.84	0.90	0.85	0.99	0.88	0.26	0.58	0.72	0.94	0.93	0.96	0.76

Notes: Country means reported. *Shortage math/language teachers* is based on the following school principal question: "Is your school's capacity to provide instruction hindered by any of the following issues? A lack of qualified mathematics/test language teachers" Possible answer categories are: not at all (1), very little (2), to some extent (3), a lot (4). School autonomy measures are binary. *Data sources:* PISA 2009 and 2012.

Table A-5: Summary Statistics of Country Characteristics

	Pooled	Australia	Austria	Belgium	Canada	Czech R.	Denmark	Estonia	Finland	France	Germany	Ireland
Expenditure per student	77	85	107	89	80	50	99	49	79	79	72	85
GDP per capita	34	39	39	36	38	25	38	20	36	33	36	43
School starting age	6.06	5	6	6	5	6	7	7	7	6	6	4
Instruction quality math	0.60	0.66	0.57	0.56	0.70	0.62	0.64	0.59	0.58	0.59	0.64	0.69
Instruction quality reading	0.49	0.53	0.41	0.43	0.56	0.44	0.57	0.50	0.37	0.52	0.44	0.51
	Italy	Japan	Korea	Netherl.	Norway	Poland	Russia	Slovak R.	Spain	Sweden	U.K.	U.S.
Expenditure per student	81	84	65	88	112	49	12	43	78	89	91	111
GDP per capita	32	34	28	41	49	18	21	22	32	38	35	46
School starting age	6	6	6	6	6	7	7	6	6	7	5	6
Instruction quality math	0.59	0.46	0.38	0.57	0.52	0.60	0.69	0.54	0.64	0.51	0.73	0.72
Instruction quality reading	0.49	0.44	0.34	0.37	0.37	0.59	0.80	0.47	0.44	0.42	0.54	0.61

Notes: Only country-level characteristics reported. The *instruction quality* indicators are based on student information provided in PISA; in 2009 for language teachers and in 2012 for math teachers. See text for details on the construction of the instruction quality indicators. The remaining country characteristics come from OECD statistics. Expenditure per student and GDP per capita are expressed in PPP-US-\$. *Data sources:* PISA 2009 and 2012, OECD.

Table A-6: OLS Estimations: Results on All Other Covariates

Dependent variable: student performance	Math	Reading
Student characteristics		
Age	0.140***	0.140***
	(0.019)	(0.014)
Female	–0.154***	0.349***
	(0.012)	(0.016)
First-generation migrant	–0.144***	–0.124**
	(0.050)	(0.049)
Second-generation migrant	–0.092*	–0.030
	(0.050)	(0.043)
Other language at home	–0.090**	–0.177***
	(0.033)	(0.037)
Family background		
11-25 books	0.204***	0.253***
	(0.024)	(0.022)
26-100 books	0.431***	0.507***
	(0.037)	(0.036)
101-200 books	0.607***	0.699***
	(0.048)	(0.045)
201-500 books	0.805***	0.883***
	(0.053)	(0.053)
More than 500 books	0.830***	0.883***
	(0.057)	(0.056)
HH ISCED 1	0.100*	0.166**
	(0.053)	(0.077)
HH ISCED 2	0.091	0.221***
	(0.065)	(0.062)
HH ISCED 3B,C	0.188**	0.313***
	(0.075)	(0.071)
HH ISCED 3A, 4	0.234***	0.353***
	(0.072)	(0.069)
HH ISCED 5B	0.188**	0.352***
	(0.078)	(0.066)
HH ISCED 5A, 6	0.260***	0.417***
	(0.078)	(0.063)
Blue collar-high skilled	0.112***	0.094***
	(0.013)	(0.016)
White collar-low skilled	0.180***	0.177***
	(0.017)	(0.017)
White collar-high skilled	0.399***	0.400***
	(0.022)	(0.021)

(continued on next page)

Table A-6 (continued)

Dependent variable: student performance	Math	Reading
School characteristics		
Small Town	–0.005	0.022
	(0.025)	(0.024)
Town	0.003	0.051*
	(0.028)	(0.029)
City	–0.001	0.062*
	(0.031)	(0.032)
Large City	0.019	0.089*
	(0.040)	(0.044)
Private school	0.188***	0.164***
	(0.028)	(0.032)
No. students per school (in 1000)	0.294***	0.247***
	(0.062)	(0.055)
Content autonomy	0.056	0.018
	(0.038)	(0.030)
Personnel autonomy	–0.164***	–0.159***
	(0.042)	(0.034)
Budget autonomy	0.031	0.029
	(0.040)	(0.041)
Shortage math teacher	–0.034**	
	(0.013)	
Shortage language teacher		–0.046***
		(0.016)
Weekly hours math classes	0.060**	
	(0.028)	
Weekly hours language classes		0.005
		(0.022)
Country-level measures		
Educational expenditure per student	0.001	0.003
	(0.002)	(0.002)
GDP per capita	–0.012**	–0.010*
	(0.005)	(0.006)
School starting age	0.076*	0.030
	(0.042)	(0.046)
Students	406,564	406,564
Countries	23	23
Adj. R2	0.26	0.29

Notes: The table reports results on all further covariates of the ordinary least squares estimations with the full set of control variables, corresponding to Column 3 (math) and Column 6 (reading) in Table 2. Omitted categories of family background and school characteristics: *0-10 books*; *parents have no educational degree*; *blue collar-low skilled*; and *village*. Significance levels: * p<0.10, ** p<0.05, *** p<0.01. *Data sources:* PIAAC, PISA 2009 and 2012.