ifo WORKING PAPERS

348
2021

February 2021

The Effect of Teacher Characteristics on Students' Science Achievement

Pietro Sancassani

ifo INSTITUTE
Leibniz Institute for Economic Research
at the University of Munich

Imprint:

ifo Working Papers
Publisher and distributor: ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49(0)89 9224 0, Telefax +49(0)89 985369, email ifo@ifo.de
www.ifo.de

An electronic version of the paper may be downloaded from the ifo website:
www.ifo.de

ifo Working Paper No. 348

The Effect of Teacher Characteristics on Students' Science Achievement*

Abstract

Using data from TIMSS 2015, an international large-scale assessment of student skills, I investigate the effect of teacher characteristics on students' science achievement. My identification strategy exploits the feature that in many education systems different science domains (physics, biology, chemistry, and earth science) are taught by different teachers. The availability of students' test scores as well as teachers' questionnaires for each of these domains allows me to implement a within-student approach which controls for unobserved student heterogeneity. I find a positive and significant effect of teacher specialization in the specific science domain on students' results, equivalent to 1.7% of a standard deviation. Holding a Master's degree, pedagogical preparation and teaching experience have no significant effect. Teachers' experience has a negative impact on the extent to which students like to study a subject or find teaching engaging.

JEL Code: I21, I29, C21, J24
Keywords: Teachers, student achievement, teacher characteristics, education production function, TIMSS

Pietro Sancassani
ifo Institute – Leibniz Institute for
Economic Research
at the University of Munich
Poschingerstr. 5
81679 Munich, Germany
sancassani@ifo.de

* I gratefully acknowledge comments from seminar participants at the ifo Center for the Economics of Education and at the HIVA - Research Institute for Work and Society at KU Leuven. Pietro Sancassani is part of the European Training Network OCCAM. This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant number 765400.

1. Introduction

There is ample evidence that teachers have a large impact both on students' performance at school (e.g. Hanushek 1971; Murnane 1975; Rockoff 2004) as well as on a variety of outcomes later in life (Chetty, Friedman, and Rockoff 2014). However, little is known about what characteristics and teaching methods make a good teacher. The literature repeatedly demonstrates that observable teacher characteristics, especially those related to education and experience, do not tend to be good indicators of teacher quality (Hanushek 1986; Rivkin, Hanushek, and Kain 2005; Clotfelter, Ladd, and Vigdor 2007; Staiger and Rockoff 2010, among others). On the other hand, in most settings it is often difficult to credibly estimate the impact of teacher characteristics on students' performance. Unobserved student and teacher characteristics as well as sorting of students and teachers into classes and schools are only some of the most obvious threats to identification in this area.

In this paper, I investigate in an international context the impact of four teacher characteristics, namely teachers' education level, scope of experience, specialization, and pedagogical preparation, on students' performance. These are important characteristics as education and experience are the traditional determinants of teacher recruitment and compensation. I exploit the availability of test scores from four scientific domains (physics, chemistry, biology and earth science) available for each 8th grade student participating in TIMSS 2015 (Trends in International Mathematics and Science Study). Furthermore, I exploit the availability of teachers' questionnaires for each science teacher that teaches at least one science domain. My sample only includes countries in which these science domains are taught by at least two different teachers. This is a unique setting that allows me to implement a within-student across-teachers approach by linking teachers' characteristics in one specific science domain to students' outcomes in the same domain. Using student fixed effects, I eliminate any source of unobserved student heterogeneity, such as innate abilities or effort, that is not domain-specific. To uncover some possible mechanisms through which teacher characteristics affect student performance, I also explore their impact on the extent to which students enjoy learning a subject or find teaching engaging.

In the within-student approach, other unobserved sources of student heterogeneity which are domain-specific, such as student preferences or abilities, might still bias the estimates if they are consistently associated with the mechanism through which teachers are allocated. However, this is less of a concern when the multiple outcomes belong to the same field, as in this case. A further advantage of using closely related outcomes in a within-student across-teachers approach is that this model relies on the assumption that the impact of teachers is the same across subjects. In studies using a similar approach (e.g. Metzler and Woessmann 2012; Bietenbeck, Piopiunik, and Wiederhold 2018; Hanushek, Piopiunik, and Wiederhold 2019), multiple outcomes for a single student belong to different fields (math and reading, for instance). This study uses outcomes which are more alike and, therefore, more likely to require similar skills, thereby relying on weaker assumptions.

The main result of my analysis is that teacher specialization has a positive and significant effect on students' science test scores. This effect is equivalent to 1.7-1.8% of a standard deviation of the students' test scores. Evidence from the US links an increase in teacher value-added by one standard deviation to an increase in student achievement by 10-20% of a standard deviation.[1] From this perspective, teacher specialization would explain between 9-18% of the variation in teacher effectiveness.

This effect is relatively small if compared to teacher interventions reported in other studies. For example, Taylor and Tyler (2012) report an impact of 5-11% of a formal peer evaluation program for teachers on student performance. Jackson and Makarin (2018) find an impact of 6-9% of a standard deviation of providing teachers with high-quality lesson plans on student outcomes. With respect to other instructional inputs, Lavy (2015) finds an effect of 6% of a standard deviation for an additional hour of instruction time per week. On this basis, the effect of being taught by a specialized teacher corresponds to about 18 additional minutes of instruction time per week. Nevertheless, it should be kept in mind that the effect of teacher specialization stems from teachers teaching a science domain in which they are already specialized. Differently from the other teacher interventions mentioned previously, this effect could be achieved at virtually no cost by allocating science teachers according to their specializations.

[1] The figure for the US is reported in Jackson, Rockoff, and Staiger (2014). The lower- and upper-bound of the estimates refer to English and math teachers, respectively. Thus, teachers seem to have a larger impact in math, which, unlike English, is mostly learned in school. In this sense, science is more similar to math.

I find a larger effect for female students and for students coming from more affluent backgrounds. I do not find a significant impact of the other teacher characteristics (education level, experience, and pedagogical preparation) on students' achievements. The impact of specialization is robust to the addition of student indicators aiming to capture remaining domain-specific within-student heterogeneity, namely the extent to which students enjoy learning the subject or find the teaching engaging. As such indicators are also a potential channel through which teachers can affect students' test scores, I also perform a mediation analysis. The results of this analysis show that teacher experience has a significant negative impact on the extent to which students enjoy learning a subject or find the teaching engaging. This result is robust across all domains and model specifications. Other teacher characteristics do not have a significant impact on these indicators.

The effect of specialization is in line with the recent literature on the effects of subject-specific teacher skills. Bietenbeck, Piopiunik, and Wiederhold (2018), for example, find an effect of 3% of a standard deviation of teacher subject knowledge on 6^{th}-grade students' reading and math scores in Sub-Saharan Africa. Using a Peruvian 6^{th}-grade dataset, Metzler and Woessmann (2012) find that one standard deviation in subject-specific teacher achievement increases student achievement in math by about 9% of a standard deviation, although the effects on reading are mostly insignificant. Hanushek, Piopiunik, and Wiederhold (2019) find a significant effect, equivalent to 11% of a standard deviation in students' test scores, of teachers' numeracy and literacy skills in 31 developed countries.

I do not find an effect of teacher experience on students' test scores. The literature seems to suggest that the greatest gains in teacher performance from experience occur in the early years of their careers and then quickly flatten (e.g. Rivkin, Hanushek, and Kain 2005; Clotfelter, Ladd, and Vigdor 2006; Boyd et al. 2008; Harris and Sass 2011). This might not be reflected in this analysis as the average teaching experience in my sample is relatively high. Only 5% of the teachers have less than 3 years of experience.

It has been observed in several studies that holding a Master's degree is generally not a strong predictor of teacher performance, as summarized by Hanushek and Rivkin (2004), among others. I also do not find a significant effect. There is no conclusive evidence regarding the impact of pedagogical preparation. This aspect, however, has received little attention in the literature so far. In line with my results, Harris and Sass (2011), for example, report no impact of teachers having majored in education on their performance as measured by student outcomes.

This paper contributes to the literature by investigating the impact of teacher characteristics on student achievement in four closely related science domains in a unique setting. To the best of my knowledge, this is the first study that focuses on the performance of students in the natural sciences using a within-student across-teachers approach. In fact, the impact of teacher characteristics on student test scores may vary between subjects (e.g., Metzler and Woessmann 2012; Kane, Rockoff, and Staiger 2008). It is therefore important to increase our knowledge of the potentially different effects of teacher characteristics on different subjects. Furthermore, I provide additional insights into the possible mechanisms by which teacher characteristics affect student performance. Overall, the results tend to be in line with the literature and confirm that observable teacher characteristics only explain a limited amount of variation in student test scores. This can have important implications for the mechanisms by which teachers are selected and compensated, as other aspects might be more relevant.[2]

The remainder of the paper is structured as follows: Section 2 describes the data and provides some descriptive characteristics. Section 3 presents the estimation strategy. The results, mediation analysis and robustness checks are discussed in Section 4. Section 5 concludes.

2. Data and Descriptive Statistics

2.1. TIMSS 2015 and Sample Selection

I use data from TIMSS 2015, an international large-scale assessment which tests 4th and 8th grade students worldwide in math and science. TIMSS employs a two-stage clustered sampling design to draw a representative national sample from each participating country. It includes tests of entire classes within randomly selected schools

[2] A growing body of literature considers different forms of teachers' cognitive skills, such as teachers' scores on licensure tests (Clotfelter, Ladd, and Vigdor 2006; Goldhaber and Anthony 2007; Harris and Sass 2011), tests of teachers' subject knowledge (Metzler and Woessmann 2012; Bietenbeck, Piopiunik, and Wiederhold 2018) or country-level teachers' cognitive skills (Hanushek, Piopiunik, and Wiederhold 2019). These tend to be more consistent predictors of teacher effectiveness, but they are rarely observed.

in a country with sampling probabilities proportional to school size as well as background questionnaires for students, teachers, and schools. The TIMSS achievement scale was established in 1995 with a scale center point of 500 located at the mean of the combined distribution of the participating countries and a standard deviation of 100.

I focus on the achievements of 8th graders in science as this is the most suitable setting for my identification strategy. 8th graders are usually around 14 years old and their science test score is made up of four domains: biology (35%), chemistry (20%), physics (25%) and earth science (20%).[3] Tests scores are available for each student and domain,[4] thus yielding 4 observations at most for each student in science.[5] Furthermore, there are countries in which specific science domains are taught by at least two different teachers, which constitutes the type of variation I exploit in this analysis. This clear distinction between closely related domains is rather special as it typically does not occur at such an early stage of education.

In this setting, I implement a within-student across-teacher model in an international context, where the deviation of test score in one domain from the average science performance of each student is associated with the deviation of teacher characteristics in the same domain from the average science teacher characteristics of each student. Due to the design of international large-scale assessments like TIMSS, this approach is not immune to criticism (e.g. Jerrim et al. 2017). In fact, these tests typically use a matrix-sampling approach in which students complete different booklets that contain a subset of questions from a common pool. If a student's booklet does not contain any questions regarding a specific subject or domain, the score in the missing subject or domain would be derived from her performance in other subjects using item response theory. The resulting within-student variation would therefore only capture the noise caused by the imputation technique, which may be a problem for the kind of identification

---

[3] In a typical 8th grade science curriculum, biology includes topics such as the characteristics, systems and processes of living things. Physics and chemistry topics include the study of the matter and energy, electricity and magnetism. Earth Science topics are, e.g., the earth's physical features and the solar system. More information can be found in Mullis and Martin (2013).

[4] TIMSS provides 5 plausible values for each student test score. I use the first plausible valuable for each domain.

[5] Depending on countries' curricula, some exceptions are possible; students in Sweden, for instance, are not tested in earth Science as this domain does not belong to their 8th grade curriculum.

I use. However, each booklet of the TIMSS 2015 contains two science blocks and two math blocks and each science block replicates the proportion of domains that constitute a subject as indicated in TIMSS guidelines.[6] Thus, the scores available for each student reflect the actual performance in each domain. These features make this setting suitable for my analysis.

I obtain the main variables of interest from the teacher questionnaire. I consider teachers to hold a Master's degree if they report having completed at a Master's degree or higher.[7] The specialization of teachers in a specific domain is determined by the choice of their major in their instruction domain during their post-secondary education.[8] It is important to highlight that this allows me to identify whether teachers have a major in one of the four specific science domains that are tested in TIMSS. Pedagogical preparation is captured by a variable indicating whether teachers have a major in general education or in science education.[9] These variables are all binary indicators and constitute the main features of teacher preparation. Holding a Master's degree indicates that a teacher has an advanced education level, while being specialized and holding a major in education capture the content and pedagogical knowledge of a teacher, respectively. Years of experience constitute an important teacher characteristic, as more experience tends to be associated with more effectiveness in the job.

These variables provide a common metric to describe teacher preparation in an international context. Nevertheless, the actual quality of teacher preparation can be very different across countries regardless of teacher qualifications, thus making cross-country comparisons potentially misleading. However, cross-country differences are accounted for in a within-student across-teachers model which uses only the variation arising from

[6] Each block contains between 12 and 18 items. The examination time for each student is 90 minutes. For more information concerning the assessment design, see Mullis and Martin (2013).

[7] Therefore, this category also includes teachers who have a doctoral degree or an equivalent degree, who only represent 1.5% of the sample. Excluding them does not have an impact on the results.

[8] The question is formulated as: "During your post-secondary education, what was your major or main area(s) of study?". Among other options, teachers can indicate whether they have a major in biology, physics, chemistry, and earth science, which are the domains of interest. I will therefore consider a teacher as specialized only if she holds a major in the instruction subject.

[9] Teachers can report whether they have a major in education-science and education-general. Using only one of the two majors in the estimations has very little impact on the estimates.

the teacher preparation relative to the average preparation of teachers teaching in the same class.

Other variables of interest are the extent to which students like learning a domain, henceforth SLL, or find the teaching engaging, henceforth FTE. TIMSS 2015 provides these domain-specific indicators that are derived from the student questionnaire. The *Student Likes Learning Biology* indicator, for instance, is based on students' agreement with nine statements such as "I enjoy learning biology" or "Biology teaches me how things in the world work". Similarly, the *Students' Views on Engaging Teaching in Biology* indicator is based on ten questions, such as "I know what my teacher expects me to do" or "My teacher does a variety of things to help us learn". I standardize both indicators across domains, so that they have a mean of 0 and a standard deviation of 1 in each domain. I also standardize student test scores across domains in order to facilitate the interpretation of the coefficients. To reduce measurement errors due to the limited number of items in each domain,[10] I aggregate the normalized test scores at the class-domain level.

I impute missing values for control variables using mean imputation at the country-domain level.[11] The percentage of missing values is between 4.8 and 6.1% for all the variables in the analysis. There are no missing values for student test scores. I rescale individual weights provided by TIMSS so that each country has the same weight in the analysis. Weights within countries are therefore not affected. Throughout the analysis, I cluster standard errors at the class level as this is the level of the treatment.

In 2015, 40 countries and 285,119 students participated in the science-8th grade assessment. I select countries where a sizable part of the students is taught by at least two different teachers in the domains of interest. This tends to be the exception in most countries: in 24 out of 40 countries less than 8% of the students are taught science by at least two teachers. I drop all these countries as they contain too few (if any) observations

[10] For example, the individual student test score for physics, which constitutes 25% of the science test, is based on 6 to 9 items.

[11] I only use complete cases with respect to the main teacher variables of interest. Whenever school mean is unavailable, I impute missing values by country mean. Although not reported, the main results are robust to the exclusion of imputed values.

that can be used in the subsequent analysis. I also exclude 6 additional countries[12] for which I am unable to link different teachers to the domain(s) they teach.[13] In the remaining 10 countries, I exclude cases where students are taught science by only one teacher, where the teacher's characteristics of interest are missing or where I am unable to link teachers to a specific domain.[14] The final sample consists of 39,827 students and 5,709 teachers in 10 countries: Armenia, England, Georgia, Hungary, Kazakhstan, Lithuania, Malta, Russia, Slovenia and Sweden.

2.2. Descriptive Statistics

All countries participating in TIMSS 2015 are reported in Table 1 in descending order of performance. Countries that are part of the analysis are in bold. Countries on the left side of the table are above the international median, while those on the right side are below the international median. A large variation in the average score of the considered countries can be observed. The top performer, Slovenia, has an average score of 551 while the average score of Georgia, the lowest in the sample, is 443. This means that the difference between the country with the highest and the country with the lowest test score is larger than one standard deviation. Many of the countries in which science domains are taught separately are former soviet countries, while this is not the case for most of the other countries participating in TIMSS 2015. Nevertheless, the large variation in average test scores of the countries that are part of the analysis speaks in favor of the external validity of this study.

It is important to keep in mind that TIMSS selects representative samples of the students within countries, which does not necessarily yield a representative sample of teachers. Nevertheless, evidence from TALIS (OECD 2014), an international survey of the teacher population, does not indicate large discrepancies between the teachers

[12] Dubai, United Arab Emirates, Israel, Japan, Korea and the US.

[13] This occurs whenever the variable provided by TIMSS indicating the "Subject Code" of the teacher does not refer to a particular domain but is coded as "Integrated Science".

[14] These cases account for 4% of the sample in the 10 countries.

included in the descriptive statistics of TIMSS and the population of teachers in a country.[15]

Descriptive statistics of the sample are reported in Table 2. The total number of observations (148,751) is given by the student-domain combination. It can be noted that, on average, each student is observed 3.74 times. Students' teachers are highly educated: 91% of the students are taught by teachers who have at least a Bachelor's degree. The share of students taught by teachers who have a Master's degree is 48%. In their report covering 20 years of TIMSS, Mullis, Martin, and Loveless (2016) acknowledge that since 1995, the first year in which TIMSS was conducted, countries have increased the requirements for becoming a teacher.

With an average experience of almost 20 years, the teachers in the sample are considerably older than the average teacher in TIMSS who has around 15 years of experience.[16] It can also be noted that most teachers are female.

The *Home Resources* indicator is a comprehensive measure of the socioeconomic status (SES) of the students. It is based on questions regarding parents' education, number of books at home and number of home study supports available for students (such as an own room or internet connection).

The descriptive statistics by domain for the main teacher variables of interest are presented in Table 3. Physics teachers have, on average, a slightly lower level of education and specialization, while earth science teachers are less likely to have majored in pedagogy. Biology teachers are, on average, less experienced and earth science teachers are less likely to have majored in pedagogy. It can also be noted that there are fewer observations for chemistry and earth science. This is because students are not tested in subjects that are not taught in the current school year. For example, Swedish students did not take the earth science test. Therefore, only 3 test scores are available for Swedish students. Further descriptive statistics at the country level can be found in Table A1 in the Appendix. Overall, the descriptive statistics by domain do not reveal great differences. It

[15] To verify this, I compare the descriptive statistics of interest for the 13 countries that participated both in in the TIMSS 2015 (8th grade) and in TALIS 2013.

[16] Such a difference is due to the prevalence of countries in which teachers typically work as teachers throughout their entire career. The high average experience might make it harder to capture the effect of experience on students' achievements if it is concentrated in the first years of teachers' careers, as the literature suggests.

is important to highlight that, while substantial differences of teacher characteristics across domains do not represent a concern for the identification strategy *per se,* they might signal different selection mechanisms for teachers in different science domains. However, this does not seem to be supported by the data as descriptive statistics by domain do not reveal great differences.

A major threat to the identification strategy arises from domain-specific non-random allocation of teachers and students. With respect to students' socioeconomic status (SES), the literature suggests that the allocation of teachers is unlikely to be random. On the one hand, more wealthy parents try to secure better resources for their children by choosing better schools (Clotfelter, Ladd, and Vigdor 2006).[17] On the other hand, countries try to improve the conditions in disadvantaged schools through smaller classes or lower student-teacher ratios.[18] While all student background characteristics are held constant in a within-student model, domain-specific non-random allocation of teachers and students might still bias the estimates. However, there is no clear indication that such patterns apply to specific subjects or domains. To uncover possible non-random patterns of domain-specific allocation of teachers, I present the relevant average teacher characteristic by domain and the socioeconomic background of the students in Table 4. I also provide test statistics for differences in average teacher characteristics between high- and low-SES students. High-SES students are those who are above the median of the *Home Resources* indicator in their respective country. The figures highlight two important patterns in the sample. First, the hypothesis that teachers are not allocated randomly with respect to students' SES is confirmed. In all domains, low-SES students are on average less likely to be taught by teachers with a Masters' degree but more likely to be taught by teachers who majored in education. Similarly, low-SES students are more likely to be

[17] There is evidence that in Malta, Russia, Slovenia and the United Kingdom disadvantaged schools are significantly worse off than advantaged schools in terms of the proportion of teachers with a major in science; the same applies to Georgia with respect to the proportion of fully certified teachers (OECD 2018).

[18] In Georgia, for example, classes in the most disadvantaged schools have, on average, 10 students less than the classes in the most advantaged schools. In Hungary, Malta, Russia and Sweden the classes in disadvantaged schools are also significantly smaller than in advantaged schools. Furthermore, in Georgia, Hungary, Malta and Russia, the student-teacher ratio in the most disadvantaged schools is more than 30% lower than in the most advantaged schools (OECD 2018). However, it has also been shown that increasing the number teachers often comes at the expense of the quality of the teaching staff (Jepsen and Rivkin 2009; Dieterle 2015; OECD 2018).

taught by more experienced teachers. All these within-subject differences are highly statistically significant. As for specialized teachers, this is only true for biology and earth science.

The second important pattern is that the differences between the characteristics of teachers of high- and low-SES students always point to the same direction. This suggests that, despite the allocation of teachers with respect to student background characteristics being non-random, it is consistent across domains. This is relevant since a major threat to identification in a within-student across-teachers model lies in systematic differences in teacher allocation across domains, a pattern that is not supported by the data.

3. Empirical Strategy

As a first step, I estimate the following OLS model including a rich set of controls:

$$A_{icdk} = \beta' T_{cdk} + \gamma' X_{ick} + \delta' C_{cdk} + \tau' S_{ck} + \theta_k + \varepsilon_{icdk} \qquad 1$$

where A_{icdk} is the achievement of student i in class c in domain d in country k, T_{cdk} is the vector of student i's teacher characteristics of interest, X_{ick} is a vector of student domain-invariant variables that control for student and family background, C_{cdk} is a vector of domain-specific variables related to student preferences, instruction time and other teacher traits, S_{ck} is a vector of class-specific variables, such as the number of students or the school location, θ_k is a vector of country fixed effects that accounts for country-specific heterogeneity, and ε_{icdk} is the idiosyncratic error term.

The vector of interest, β, captures the association between teacher characteristics and student achievement. However, unobservable characteristics that are both correlated with student achievement and teacher characteristics might bias the estimates. In the previous section I provide evidence of non-random allocation of teacher characteristics with respect to students' SES. However, such non-random allocation might also occur along other unobserved student dimensions which cannot be accounted for in this model. For instance, specialized teachers might be systematically assigned to classes with more motivated and better performing students. Therefore, teacher characteristics might still not be allocated randomly conditional on observable student characteristics, which would bias the OLS estimates of the teacher characteristics.

As I observe the results of each student in at least three different domains, I can eliminate bias due to unobservable student characteristics that do not vary across science domains. Multiple observations for each student allow me to implement a within-student across-teacher model which controls for unobserved and domain-invariant student traits. The only variation that is left in order to capture the effect of teacher characteristics is the within-student and across-domains variation. This can be achieved empirically by estimating the following student fixed effects model:

$$A_{icdk} = \beta' T_{cdk} + \delta' C_{cdk} + \mu_i + \mu_d + \varepsilon_{icdk} \qquad 2$$

where A_{icdk} is the achievement of student i in class c in domain d and country k, T_{cdk} is the vector of student i's teacher characteristics of interest, namely whether a teacher holds a Master's degree, the years of experience , whether a teacher is specialized in the domain being taught and whether a teacher majored in education . The vector β captures the parameter of interest. C_{cdk} are domain-specific controls, such as teacher gender and instruction time, which account for the remaining domain-specific heterogeneity. Finally, μ_i and μ_d are student and domain fixed effects, respectively, so that all coefficients are estimated using only within-student variation, thus controlling for every variable that does not vary across domains. ε_{icdk} is the idiosyncratic error.

Student fixed effects control for a variety of characteristics that are known to largely affect student achievement, such as socioeconomic status and domain unspecific innate abilities. They also control for all domain-invariant school and class features, such as class size or the school environment. Domain fixed effects eliminate domain-specific test score heterogeneities as well as other unobserved factors that are specific to one domain. For example, they account for the fact that the test might be more difficult on average in one domain or that teachers in one domain might be, on average, more educated.

Estimates could still be biased if the association between unobservable student and teacher characteristics differs between domains. This might be the case if specialized physics teachers were more likely to be placed in a class with more motivated students but the same would not apply to biology teachers. Although this cannot be ruled out entirely, Table 4 in the previous section does not indicate different patterns of student-teacher matching across domains.

The model relies on the assumption that the impact of teacher characteristics is homogenous across domains. Compared to studies examining different subjects, this analysis relies on a weaker assumption as the multiple outcomes belong to the same field. Furthermore, I provide suggestive evidence that this does not seem to be the case. The OLS analysis in the following section demonstrates that the relation between teacher characteristics and student achievement is not substantially different across domains. On the other hand, the fact that the multiple outcomes are so closely related to each other makes it difficult to pin down the actual impact of a single teacher in the taught domain. There is indeed a potential for the impact of a teacher to spill over into adjacent domains. Furthermore, the amount of variation in outcomes that can be exploited should be a priori smaller as performances in related domains should not be too different. Therefore, it is likely that this analysis yields conservative estimates of the impact of teacher characteristics on student outcomes.

Student fixed effects also account for general science knowledge and therefore for the impact of characteristics of previous teachers. In fact, it should be kept in mind that students' performance in science is the result of several years of schooling during which students were potentially taught by many different teachers. Furthermore, it is likely that the allocation mechanisms between teachers and students remain in place throughout all years of schooling, which could exacerbate pre-existing differences. For these reasons, an excessive portion of the variation in student achievement might be attributed to the characteristics of current teachers, leading to a bias in the estimates. By capturing each student's stock of knowledge in the sciences, student fixed effects limit the amount of variation that can be falsely attributed to the current teacher. This might come at the cost of increasing the attenuation bias that is due to the fact that the binary indicators I use are a rough measure of teacher preparation. For all of these reasons, the estimated coefficients should be considered as a lower bound.

4. Results

4.1. Main Results

OLS results of a model that includes a large set of control variables and country fixed effects to account for country heterogeneity are reported in Table A2 in the Appendix. In the pooled regression that includes all science domains in Column 1, only the major in education is positive and marginally significant. This association is equivalent to 3% of a standard deviation in student achievement. The magnitude of the specialized teacher coefficient is virtually identical but due to a larger standard error, it is not significant. The results in Columns 2 to 5 are not significant, except for the result for the major in education in Column 3, which is positively and statistically significant. Figures in this table do not show substantial heterogeneity across domains. However, due to the possible correlation between teacher characteristics and unobservable student traits that might affect students' test scores, OLS estimates are likely to yield biased estimates.

To circumvent such possible bias, I implement the within-student across-teachers model of Equation 2. Results are reported in Table 5. In Columns 1 to 4, I present the relationship between teacher characteristics and student science test scores controlling for teacher gender and instruction time once student and domain fixed effects have been accounted for, separately for each characteristic. In Column 5, I include all the teacher characteristics of interest simultaneously. Results underline a positive and significant effect of specialized teachers on student achievement, equivalent to between 1.7%-1.8% of a standard deviation. The magnitude of this coefficient is considerably smaller than the one observed in the OLS model, although the parameter is estimated more precisely. All other characteristics considered do not seem to have a significant impact.

The impact of having majored in education is virtually zero, which suggests that the parameter estimated with the OLS model was substantially biased upwards even after controlling for student background characteristics. The small magnitude of the observed coefficients might also be due to the fraction of total variation that remains in the students' test scores. Once student and domain fixed effects are accounted for, the within-student standard deviation in the test scores is 0.33, or one-third of the standard deviation of the full sample. This can be considered as the amount of variation that can realistically be influenced by teachers, as it already takes into account the impact of important factors such as the socioeconomic status, gender or innate abilities. From this perspective, the observed impact of specialized teachers amounts to 5.1%-5.6% of the within-student standard deviation.[19]

I explore heterogeneities by students' characteristics in Table 6.[20] In Columns 1-2, I explore heterogeneities in the impact of teachers according to students' gender. The impact of specialized teachers on female students' test scores is positively significant and is equivalent to 2.2% of a standard deviation, while it is positive but insignificant for male students. Such a difference is sizable but not statistically significant. The impact of experience is positively significant for female students, although the magnitude is rather small and only marginally significant. Similarly, the impact of teachers who have majored in education is marginally significant and negative for males, with a magnitude smaller than 1% of a standard deviation.

As most teachers are female, the higher impact of specialized teachers on female students might be due to positive classroom interactions between female teachers and female students. This is not new to the literature and several studies find that having a female teacher improves female students' educational outcomes (e.g. Dee 2005, 2007; Winters et al. 2013; Gong, Lu, and Song 2018). However, including an interaction term between teacher specialization and teacher gender in Equation 2 with female students does not support this interpretation. In fact, the coefficient of the interaction between the teacher specialization and teacher being a female is negative but not significant (not shown).

In Columns 3-4, I divide the sample in low- and high-SES students, i.e. students whose SES is below or above the median in their respective country. Specialized teachers have a positive and significant effect only on students coming from more affluent backgrounds, with an estimated impact of 2.8% of a standard deviation. For specialized teachers, the difference between the coefficients of the two samples is significant. It is plausible to assume that teachers find an environment better suited for learning in schools attended by high-SES students and can therefore deploy their knowledge more effectively. Furthermore, specialized teachers might be able to work more efficiently with

[19] For consistency with the existing literature, I only consider effects relative to the full standard deviation of the model in the remainder of the paper.

[20] I only report the specifications including all the explanatory variables of interest as there is very little additional value in presenting the bivariate specifications as in Table 5.

students who have more subject knowledge from the beginning.[21] This is captured to a large extent by students' SES, with a difference in the average test scores of high- and low-SES students equivalent to 45% of a standard deviation. Although this difference includes current school input, a large part of it is probably due to knowledge accrued before the current school year.

4.2. Mediation Analysis

In this section, I explore potential channels through which teacher characteristics affect student achievement. There are two student indicators described in Section 2, which capture the extent to which students like learning the subject (SLL) and find the teaching engaging (FTE). As a first step, I include these indicators as additional domain-specific controls in the within-student model with student test scores as the dependent variable.

While including potential channels of the treatments in the regressions might come at the cost of over-controlling, this step ensures that the potential channels are relevant and that there is no omitted variable bias left from remaining domain-specific endogeneity.[22] Results in Table 7 do not seem to provide evidence of bias due to the omission of domain-specific controls. The impact of teacher characteristics is in fact robust to the inclusion of these domain-specific indicators. In particular, the impact of specialized teachers remains significant in all specifications but slightly decreases in its magnitude. Both indicators are positively associated with student test scores, but the results should not be interpreted causally.[23] The magnitude of their coefficients is virtually identical when they are included separately (Columns 2 and 3), but the SLL indicator

[21] To substantiate this hypothesis, I also divide the sample in low- and high-achievers, i.e. students whose average science test score is below or above the median science test score in their respective country. The results (not shown) are virtually identical to those obtained when I divide the sample in low- and high-SES students. For high-achieving students, the effect of specialized teachers is positive and significant, while for low-achieving students is positive but not significant. However, dividing the sample between low- and high-achieving students is likely to be endogenous to the treatments. I therefore stick to the previous specification (dividing the sample between low- and high-SES students) as the preferred one.

[22] In fact, one possible remaining concern is that students will perform better in one specific domain simply because they have a preference for it, and, therefore, will enjoy learning it and find the teaching more engaging. Thus, omitting the SLL and FTE indicators might cause an omitted variable bias if, for example, specialized science teachers tend to be assigned to classes where students have a preference for their subject.

[23] Reverse causation is likely to be an issue for these controls, as students who perform better in one subject are probably more likely to enjoy the subject and the teaching more. However, a causal analysis of these controls lies outside the scope of this paper.

seems to be more relevant for student achievement when both indicators are included (Column 4) in a horse-race regression. While the two indicators are strongly correlated to each other (0.70, p<0.01), the SLL indicator is a clearer indicator of student preferences and thus more suitable to account for this aspect. Conversely, the FTE indicator seems to be better suited as a mediator, as it is more likely to be affected by teacher characteristics.

I explore the role of the SLL and FTE indicators as potential mechanisms in Table 8 and 9, where I use them as outcomes of teacher characteristics using the same models of Equation 1 and 2. I report results from the pooled OLS model with various sets of controls and fixed effects (Columns 1, 2 and 3) as well as from the within-student model (Column 4). The OLS model should not include major biases in this context. In fact, there is no obvious way in which a non-random sorting of teachers and students might be based on students' appreciation for a subject or how engaging they find the teaching. The results seem to support such hypotheses, as the coefficients are virtually identical regardless of the model used. Only the parameter associated to teachers' Master's degrees becomes positive in the within-student specification, but it remains statistically insignificant in all specifications.

The main result illustrated in these tables is that teachers' experience has a clear and consistent significantly negative impact on whether students like the subject or find the teaching engaging. The results are robust to the inclusion of student, teacher, and school controls as well to the inclusion of student fixed effects.[24] The impact of an additional year of experience leads to a decrease in both indicators equivalent to roughly 0.4% of a standard deviation. While this analysis does not provide consistent estimates of the impacts of both the SLL and FTE indicators on student achievement, it is reasonable to assume that students will learn more if they are more engaged or enjoy a subject. These are also desirable outcomes per se.

The negative impact of teacher experience on the SLL and FTE indicators might help explain a pattern that is frequently discussed in the literature, namely that the largest

[24] A separate OLS regression for each domain (not shown) also confirms these results.

gain in experience is concentrated in the very first years of teachers' careers.[25] In fact, it is possible that net impact of teacher experience is a combination of factors that improve with increasing experience, such as classroom management or subject knowledge, and other factors that worsen with increasing experience, such as enthusiasm for the subject or for teaching in general. The marginal benefit of an additional year of experience might therefore fade out as the latter factors offset the former ones.

As final remark, it can be observed that the coefficients of the specialized teachers are quite large in both tables, especially in the within-student model, although they never reach statistical significance. Compared to the other teacher characteristics, the specialized teacher coefficients are estimated much less precisely, with standard errors that are almost 50% larger than those of the major in education. As a result, although the point estimates consistently point in the positive direction, I cannot reject the null hypotheses that students do not enjoy learning a subject more nor find the teaching more engaging when taught by specialized teachers.

4.3. Robustness

To ensure that results are not driven by single countries, where, for example, specialized teachers are particularly effective compared to non-specialized teachers, I repeat the analysis excluding one country at a time.[26] The results are reported in Table A3. While the effect remains largely positive in all columns, it does not reach any conventional level of statistical significance when Malta, Slovenia or Sweden are

[25] This suggests a non-linear relationship between student test scores and teacher experience. To explore this aspect, I implement several non-linear specifications of experience in the within-student model of Equation 2 with science test scores as outcome, namely experience squared, logarithm of experience and a piecewise specification, (i.e. having 2, 3-5 or 6 or more years of experience). However, the impact of teacher experience is not significant in any these specifications. As a further step, I restrict the sample to the youngest cohort of teachers (25 years or less) or teachers with less than 4 years of experience. Although the resulting samples are too small to draw reliable conclusions (1,024 and 4,028 observations, respectively), the impact of teacher experience is positively significant in this context, with a magnitude between 1.9-4.2% of a standard deviation for one additional year of experience. The positive impact disappears when the second-to-youngest cohort is included (25 to 30 years old teachers) or teachers with less than 5 years of experience are included, thus suggesting diminishing marginal returns to experience. This is also shown in Boyd et al. (2008), who report gains between 5-7% of a standard deviation during the first year of experience, with these gains accounting for more than half of the cumulative experience effect.

[26] In principle, it is possible to run a separate regression for each single country. However, some countries contribute very little to the identification due to very large or small shares of the variables of interest (e.g. only 3% of the teachers in Kazakhstan have a Master's degree). Thus, single-country regressions might not be particularly informative.

excluded (Column 7,9 and 10). On the other hand, excluding Hungary yields the largest estimate of the impact of specialized teachers, suggesting that they are not particularly effective in this country.[27] Overall, results suggest that results for specialized teachers are not driven by single countries.

When Armenia, Hungary or Lithuania are excluded (Columns 1, 4 and 6), the results for the major in education become marginally significant and negative, with a point estimate of around 1.1% of a standard deviation. Overall, the coefficient for the major in education always points to the negative direction in the within-student models with student test scores as outcome variable. A possible interpretation for this is that pedagogical and subject-specific knowledge are substitutes in the preparation of teachers. In fact, the correlation between being a specialized teacher and having majored in education is significantly negative (-0.29, $p<0.01$). Therefore, the major in education might also be capturing the effect of a lower level of subject knowledge.

I also perform a further robustness check in which I omit one domain at a time. Table A4 shows that the impact of specialization is stronger when earth science and especially physics are dropped. This suggests that specialized teachers are less effective in these domains. Conversely, the impact of specialization fades when biology is excluded from the analysis, indicating that the effect is driven by biology teachers. A possible explanation for this comes from the design of the test. As described in Section 2, biology constitutes the largest part of the science test (35%). Therefore, test scores in this domain should be considered more reliable and less noisy than test scores in other domains. Omitting biology from the within-student model might therefore leave only test scores that are too noisy to detect a relatively small effect.

As also observed when omitting some countries, the coefficient for the major in education becomes significantly negative when physics and, in particular, biology are dropped. Again, this might be due to the fact that a major in education might capture part of the effect of lower subject knowledge.

---

[27] Hungary is the country with the lowest share of specialized teachers (26% of the teachers are specialized, see Table A1 in the Appendix). This might suggest that expertise is not a priority in the training of lower secondary science teachers in this country. Nevertheless, the overall performance of Hungarian students in science is well above the international TIMSS average, suggesting that other factors contribute to a country's students achieving good results.

5. Conclusion

It is widely acknowledged that teachers play a fundamental part in student education and that education systems worldwide should strive to ensure teacher quality. Nevertheless, what constitutes teacher quality remains relatively unresolved. Available teacher characteristics such as education and experience tend to be weak predictors of teachers' effectiveness. This paper complements previous studies using within-student across-subject analyses in that it focuses exclusively on science achievement in a group of countries in which 8^{th} graders are taught sciences by different teachers.

The main result of the analysis is that science teachers who are specialized in the domain that they teach have a positive and significant impact on students' science performance, while neither having a Master's degree nor holding a major in education or the number of years of experience has a significant impact on students' performance. This result confirms that subject knowledge tends to be a stronger predictor of teacher effectiveness than, for example, the general education level or experience. A related policy implication might be that subject knowledge should play a key role in the recruitment and compensation of teachers in lower secondary schools. Furthermore, the benefit of teacher specialization could be reaped at no additional cost by allocating science teachers according to their specialization.

In the mediation analysis, I find that teacher experience negatively affects the indicators that measure how much students like a subject and find the teaching engaging. This result might help to explain a pattern which has often been observed in the literature, namely that most of the gains from teaching experience in terms of student performance seem to be concentrated in the very first years of the teaching career. A possible implication of this result is that teachers should be incentivized to update their teaching methods throughout their career in order to keep their students engaged.

References

Bietenbeck, Jan, Marc Piopiunik, and Simon Wiederhold. 2018. "Africa's Skill Tragedy." *J. Human Resources* 53 (3): 553–78. https://doi.org/10.3368/jhr.53.3.0616-8002R1.

Boyd, Donald, Hamilton Lankford, Susanna Loeb, Jonah Rockoff, and James Wyckoff. 2008. "The Narrowing Gap in New York City Teacher Qualifications and Its Implications for Student Achievement in High-Poverty Schools." *J. Pol. Anal. Manage.* 27 (4): 793–818. https://doi.org/10.1002/pam.20377.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* 104 (9): 2633–79. https://doi.org/10.1257/aer.104.9.2633.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2006. "Teacher-Student Matching and the Assessment of Teacher Effectiveness." *J. Human Resources* XLI (4): 778–820. https://doi.org/10.3368/jhr.XLI.4.778.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2007. "Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects." *Economics of Education Review* 26 (6): 673–82. https://doi.org/10.1016/j.econedurev.2007.10.002.

Dee, Thomas S. 2005. "A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?" *American Economic Review* 95 (2): 158–65. https://doi.org/10.1257/000282805774670446.

Dee, Thomas S. 2007. "Teachers and the Gender Gaps in Student Achievement." *J. Human Resources* XLII (3): 528–54. https://doi.org/10.3368/jhr.XLII.3.528.

Dieterle, Steven G. 2015. "Class-Size Reduction Policies and the Quality of Entering Teachers." *Labour Economics* 36:35–47. https://doi.org/10.1016/j.labeco.2015.07.005.

Goldhaber, Dan, and Emily Anthony. 2007. "Can Teacher Quality Be Effectively Assessed? National Board Certification as a Signal of Effective Teaching." *Review of Economics and Statistics* 89 (1): 134–50. https://doi.org/10.1162/rest.89.1.134.

Gong, Jie, Yi Lu, and Hong Song. 2018. "The Effect of Teacher Gender on Students' Academic and Noncognitive Outcomes." *Journal of Labor Economics* 36 (3): 743–78. https://doi.org/10.1086/696203.

Hanushek, Eric A. 1971. "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data." *American Economic Review* 61 (2): 280–88.

Hanushek, Eric A. 1986. “The Economics of Schooling: Production and Efficiency in Public Schools.” *Journal of Economic Literature* 24 (3): 1141–77.

Hanushek, Eric A., Marc Piopiunik, and Simon Wiederhold. 2019. “The Value of Smarter Teachers.” *J. Human Resources* 54 (4): 857–99. https://doi.org/10.3368/jhr.54.4.0317.8619R1.

Hanushek, Eric A., and Steven G. Rivkin. 2004. “How to Improve the Supply of High-Quality Teachers.” *Brookings Papers on Education Policy*.

Harris, Douglas N., and Tim R. Sass. 2011. “Teacher Training, Teacher Quality and Student Achievement.” *Journal of Public Economics* 95 (7): 798–812. https://doi.org/10.1016/j.jpubeco.2010.11.009.

Jackson, C. Kirabo, Jonah E. Rockoff, and Douglas O. Staiger. 2014. “Teacher Effects and Teacher-Related Policies.” *Annu. Rev. Econ.* 6 (1): 801–25. https://doi.org/10.1146/annurev-economics-080213-040845.

Jackson, Kirabo, and Alexey Makarin. 2018. “Can Online Off-the-Shelf Lessons Improve Student Outcomes? Evidence from a Field Experiment.” *American Economic Journal: Economic Policy* 10 (3): 226–54. https://doi.org/10.1257/pol.20170211.

Jepsen, Christopher, and Steven G. Rivkin. 2009. “Class Size Reduction and Student Achievement.” *J. Human Resources* 44 (1): 223–50. https://doi.org/10.3368/jhr.44.1.223.

Jerrim, John, Luis Alejandro Lopez-Agudo, Oscar D. Marcenaro-Gutierrez, and Nikki Shure. 2017. “What Happens When Econometrics and Psychometrics Collide? An Example Using the PISA Data.” *Economics of Education Review* 61:51–58. https://doi.org/10.1016/j.econedurev.2017.09.007.

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 (6): 615–31. https://doi.org/10.1016/j.econedurev.2007.05.005.

Lavy, Victor. 2015. “Do Differences in Schools’ Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries.” *The Economic Journal* 125 (588): F397-F424. https://doi.org/10.1111/ecoj.12233.

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 (2): 486–96. https://doi.org/10.1016/j.jdeveco.2012.06.002.

Mullis, Ina V. S., and Michael O. Martin, eds. 2013. *TIMSS 2015 Assessment Frameworks.* Chestnut Hill MA: TIMSS & PIRLS International Study Center Lynch School of Education Boston College.

Mullis, Ina V. S., Michael O. Martin, and Tom Loveless, eds. 2016. *20 Years of TIMSS: International Trends in Mathematics and Science Achievement, Curriculum, and Instruction:* TIMSS & PIRLS International Study Center,Lynch School of Education, Boston CollegeandInternational Association for the Evaluation of Educational Achievement (IEA).

Murnane, Richard. 1975. *The Impact of School Resources on the Learning of Inner City Children.* Cambridge, MA: Ballinger.

OECD. 2014. *TALIS 2013 Results:* TALIS, OECD Publishing, Paris.

OECD. 2018. *Effective Teacher Policies.* PISA, OECD Publishing, Paris.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica* 73 (2): 417–58. https://doi.org/10.1111/j.1468-0262.2005.00584.x.

Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review* 94 (2): 247–52. https://doi.org/10.1257/0002828041302244.

Staiger, Douglas O., and Jonah E. Rockoff. 2010. "Searching for Effective Teachers with Imperfect Information." *Journal of Economic Perspectives* 24 (3): 97–118. https://doi.org/10.1257/jep.24.3.97.

Taylor, Eric S., and John H. Tyler. 2012. "The Effect of Evaluation on Teacher Performance." *American Economic Review* 102 (7): 3628–51. https://doi.org/10.1257/aer.102.7.3628.

Winters, Marcus A., Robert C. Haight, Thomas T. Swaim, and Katarzyna A. Pickering. 2013. "The Effect of Same-Gender Teacher Assignment on Student Achievement in the Elementary and Secondary Grades: Evidence from Panel Data." *Economics of Education Review* 34:69–75. https://doi.org/10.1016/j.econedurev.2013.01.007.

List of Tables

Table 1: Average Science Score in TIMSS 2015, Entire Sample

Country	Average Scale Score	(SE)	Country	Average Scale Score	(SE)
Singapore	597	(3.2)	Turkey	493	(4.0)
Japan	571	(1.8)	**Malta**	**481**	**(1.6)**
Chinese Taipei	569	(2.1)	United Arab Emirates	477	(2.3)
Korea, Rep. of	556	(2.2)	Malaysia	471	(4.1)
Slovenia	**551**	**(2.4)**	Bahrain	466	(2.2)
Hong Kong SAR	546	(3.9)	Qatar	457	(3.0)
Russian Federation	**544**	**(4.2)**	Iran, Islamic Rep. of	456	(4.0)
England	**537**	**(3.8)**	Thailand	456	(4.2)
Kazakhstan	**533**	**(4.4)**	Oman	455	(2.7)
Ireland	530	(2.8)	Chile	454	(3.1)
United States	530	(2.8)	**Armenia***	**452**	**(-)**
Hungary	**527**	**(3.4)**	**Georgia**	**443**	**(3.1)**
Canada	526	(2.2)	Jordan	426	(3.4)
Sweden	**522**	**(3.4)**	Kuwait	411	(5.2)
Lithuania	**519**	**(2.8)**	Lebanon	398	(5.3)
New Zealand	513	(3.1)	Saudi Arabia	396	(4.5)
Australia	512	(2.7)	Morocco	393	(2.5)
Norway (9)	509	(2.8)	Botswana (9)	392	(2.7)
Israel	507	(3.9)	Egypt	371	(4.3)
Italy	499	(2.4)	South Africa (9)	358	(5.6)

Note: The figure has been obtained from TIMSS 2015 8th grade Science Achievement. Standard errors of the average country science achievement are in parentheses. Countries that are part of the analyzed sample are in bold. *Armenia took the test one year later and was not included in the original figure. I added it manually.

Table 2: Descriptive Statistics

	Mean	SD	Min	Max
Bachelors' Teachers	0.43	0.49	0.0	1.0
Masters' Teachers	0.48	0.49	0.0	1.0
Experience (y)	19.90	11.18	0.0	45.0
Specialized Teachers	0.83	0.36	0.0	1.0
Major in Education	0.49	0.49	0.0	1.0
Female Teachers	0.80	0.39	0.0	1.0
Instruction Time (h)	1.58	0.71	0.0	10.0
Home Resources	10.73	1.54	4.2	13.9
# Observations		148,751		
# Students		39,827		
# Teachers		5,709		

Note: The unit of observation is given by the student-domain combination. The table reports weighted descriptive statistics of the main variables of interest. Bachelors' Teachers hold only a Bachelors' degree, while Masters' Teachers also hold a Masters' degree as well. Experience is measured in years. Specialized Teachers are those who have a major in their instruction domain. The Home Resources indicator provided by TIMSS captures the socioeconomic status of the students and is based on parents’ education, number of books at home and home study supports available for students.

Table 3: Descriptives by Domain

Variables	Physics Mean	Physics (SD)	Biology Mean	Biology (SD)
Bachelors' Teachers	0.45	(0.50)	0.43	(0.50)
Masters' Teachers	0.45	(0.49)	0.48	(0.49)
Experience (y)	20.23	(11.56)	18.95	(11.11)
Specialized Teachers	0.80	(0.39)	0.85	(0.35)
Major in Education	0.50	(0.48)	0.53	(0.48)
Instruction Time (h)	1.73	(0.80)	1.52	(0.69)
# Students	39,169		38,069	
# Teachers	1,722		1,710	

Variables	Chemistry Mean	Chemistry (SD)	Earth Science Mean	Earth Science (SD)
Bachelors' Teachers	0.42	(0.49)	0.41	(0.49)
Masters' Teachers	0.49	(0.49)	0.51	(0.49)
Experience (y)	19.90	(10.90)	20.59	(11.03)
Specialized Teachers	0.82	(0.38)	0.87	(0.33)
Major in Education	0.52	(0.49)	0.39	(0.48)
Instruction Time (h)	1.60	(0.63)	1.46	(0.64)
# Students	37,487		33,896	
# Teachers	1,636		1,360	

Note: The table reports weighted descriptive statistics by domain. For each domain, the number of distinct students and teachers observed is also reported.

Table 4: Teacher Characteristics by Domain and Student SES

	Physics		Biology		Chemistry		Earth Science	
	Low-SES	High-SES	Low-SES	High-SES	Low-SES	High-SES	Low-SES	High-SES
Masters' Teachers	0.44	0.47	0.47	0.49	0.49	0.50	0.49	0.55
t-test statistic	*(4.47)* ***		*(5.40)* ***		*(3.28)* ***		*(11.07)* ***	
Experience (y)	20.66	19.61	19.30	18.44	19.95	19.83	20.77	20.31
t-test statistic	*(-8.86)* ***		*(-7.46)* ***		*(-1.00)*		*(-3.78)* ***	
Specialized Teachers	0.80	0.80	0.84	0.87	0.81	0.82	0.86	0.89
t-test statistic	*(0.34)*		*(7.53)* ***		*(1.59)*		*(6.66)* ***	
Major in Education	0.50	0.49	0.54	0.53	0.53	0.50	0.39	0.39
t-test statistic	*(-2.16)* **		*(-2.49)* **		*(-6.61)* ***		*(-0.82)*	

Note: The table reports the weighted means of the main independent variables of interest by student SES and domain. High-SES students are students who fall above the median SES level within their country. For each variable, I report the t-statistic associated with the difference in the means between High- and Low-SES students. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Teacher Characteristics' Effect on Student Test Scores

Independent Variables	(1)	(2)	(3)	(4)	(5)
Masters' Teachers	0.0011				0.0015
	(0.0057)				*(0.0057)*
Experience		0.0003			0.0002
		(0.0002)			*(0.0002)*
Specialized Teachers			0.0182**		0.0172*
			(0.0088)		*(0.0089)*
Major in Education				-0.0088	-0.0076
				(0.0054)	*(0.0055)*
Observations	148,751	148,751	148,751	148,751	148,751
Students, Domain FE	YES	YES	YES	YES	YES

Note: The table reports the results for the within-student across-teachers model that includes four science domains (physics, biology, chemistry, earth science). The number of observations is given by all the student-domain combinations. All specifications control for instruction time and teacher gender and include student and domain fixed effects. Test scores have been standardized across domains and aggregated at the classroom-domain level to reduce measurement error. Standard errors are clustered at the classroom level. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Main Results by Gender and SES

Independent Variables	Student Gender		SES	
	(1)	(2)	(3)	(4)
	Male	Female	Low-SES	High-SES
Masters' Teachers	-0.0029	0.0056	-0.0004	0.0052
	(0.0059)	*(0.0061)*	*(0.0061)*	*(0.0065)*
Experience	0.0001	0.0005*	0.0002	0.0003
	(0.0002)	*(0.0003)*	*(0.0003)*	*(0.0003)*
Specialized Teachers	0.0106	0.0224**	0.0115	0.0280**
	(0.0087)	*(0.0104)*	*(0.0086)*	*(0.0118)*
Major in Education	-0.0097*	-0.0049	-0.0077	-0.0063
	(0.0058)	*(0.0058)*	*(0.0060)*	*(0.0061)*
Observations	76,350	72,401	85,538	63,213
Students, Domain FE	YES	YES	YES	YES

Note: The table reports the results for the within-student across-teachers model that includes four science domains (physics, biology, chemistry, earth science). The number of observations is given by all the student-domain combinations. All specifications control for instruction time and teacher gender and include student and domain fixed effects. Each column reports the estimated coefficient in the indicated sub-sample. High-SES students are those above the median SES level within their country. Test scores have been standardized across domains and aggregated at the classroom-domain level to reduce measurement error. Standard errors are clustered at the classroom level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 7: Additional Controls

Independent Variables	(1)	(2)	(3)	(4)
Masters' Teachers	0.0015	0.0014	0.0015	0.0014
	(0.0057)	*(0.0056)*	*(0.0056)*	*(0.0056)*
Experience	0.0002	0.0003	0.0003	0.0003
	(0.0002)	*(0.0002)*	*(0.0002)*	*(0.0002)*
Specialized Teachers	0.0172*	0.0168*	0.0168*	0.0167*
	(0.0089)	*(0.0088)*	*(0.0088)*	*(0.0088)*
Major in Education	-0.0076	-0.0078	-0.0079	-0.0078
	(0.0055)	*(0.0054)*	*(0.0054)*	*(0.0054)*
SLL		0.0139***		0.0105***
		(0.0013)		*(0.0013)*
FTE			0.0135***	0.0059***
			(0.0016)	*(0.0016)*
Observations	148,751	148,751	148,751	148,751
Students, Domain FE	YES	YES	YES	YES

Note: The table reports the results for the within-student across-teachers model that includes four science domains (physics, biology, chemistry, earth science). The number of observations is given by all the student-domain combinations. All specifications control for instruction time and teacher gender and include student and domain fixed effects. Test scores have been standardized across domains and aggregated at the classroom-domain level to reduce measurement error. Standard errors are clustered at the classroom level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 8: Impact on the *Student Likes Learning* indicator

Independent Variables	(1) OLS	(2) OLS	(3) OLS	(4) Within-Student
Masters' Teachers	-0.0011	-0.0155	-0.0076	0.0116
	(0.0152)	*(0.0148)*	*(0.0148)*	*(0.0172)*
Experience	-0.0034***	-0.0032***	-0.0031***	-0.0038***
	(0.0006)	*(0.0006)*	*(0.0006)*	*(0.0006)*
Specialized Teachers	0.0298	0.0253	0.0298	0.0262
	(0.0240)	*(0.0230)*	*(0.0221)*	*(0.0214)*
Major in Education	0.0140	0.0118	0.0130	0.0136
	(0.0144)	*(0.0139)*	*(0.0137)*	*(0.0151)*
Observations	148,751	148,751	148,751	148,751
R-squared	0.0887	0.1151	0.1185	0.5593
Country FE	YES	YES	YES	NO
Student Controls	NO	YES	YES	NO
Class, School Controls	NO	NO	YES	NO
Student FE	NO	NO	NO	YES

Note: The table reports the results for an OLS model (Column 1,2,3) and a within-student across-teachers model (Column 4) that include four science domains (physics, biology, chemistry, earth science). The number of observations is given by all the student-domain combinations. The dependent variable is the "Student Likes Learning the Subject" indicator standardized across domains. Student controls are student SES, gender, language spoken at home, whether parents have foreign origins and expectations in educational achievement. Class controls are class size, share of students with language difficulties, class SES and the share of native speakers. School controls are the school location, whether science instruction is hindered by shortage of resources, school discipline problems and school emphasis on academic success. Domain-specific controls are teacher gender and instruction time. Country fixed effects are included in columns 1-3, student fixed effects are included in column 4. Standard errors are clustered at the classroom level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 9: Impact on the *Student Finds the Teaching Engaging* Indicator

Independent Variables	(1) OLS	(2) OLS	(3) OLS	(4) Within-Student
Masters' Teachers	-0.0239	-0.0333**	-0.0247	0.0045
	(0.0165)	(0.0164)	(0.0161)	(0.0172)
Experience	-0.0040***	-0.0039***	-0.0039***	-0.0039***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Specialized Teachers	0.0026	0.0006	0.0054	0.0243
	(0.0247)	(0.0243)	(0.0233)	(0.0217)
Major in Education	0.0248*	0.0230	0.0249*	0.0227
	(0.0147)	(0.0146)	(0.0143)	(0.0153)
Observations	148,751	148,751	148,751	148,751
R-squared	0.1058	0.1179	0.1236	0.6388
Country FE	YES	YES	YES	NO
Student Controls	NO	YES	YES	NO
Class, School Controls	NO	NO	YES	NO
Student FE	NO	NO	NO	YES

Note: The table reports the results for an OLS model (Column 1,2,3) and a within-student across-teachers model (Column 4) that include four science domains (physics, biology, chemistry, earth science). The number of observations is given by all the student-domain combinations. The dependent variable is the "Student Finds the Teaching Engaging" indicator standardized across domains. Student controls are student SES, gender, language spoken at home, whether parents have foreign origins and expectations in educational achievement. Class controls are class size, share of students with language difficulties, class SES and the share of native speakers. School controls are the school location, whether science instruction is hindered by shortage of resources, school discipline problems and school emphasis on academic success. Domain-specific controls are teacher gender and instruction time. Country fixed effects are included in columns 1-3, student fixed effects are included in column 4. Standard errors are clustered at the classroom level. *** p<0.01, ** p<0.05, * p<0.1

Appendix

Table A1: Descriptives by Country

Variables	Armenia		England		Georgia		Hungary		Kazakhstan	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Students' Science Score	452.4	(104.43)	568.92	(85.64)	437.54	(96.75)	526.49	(95.27)	530.15	(106.29)
Bachelors' Teachers	0.13	(0.34)	0.62	(0.49)	0.09	(0.29)	0.65	(0.48)	0.93	(0.25)
Masters' Teachers	0.79	(0.38)	0.25	(0.39)	0.89	(0.31)	0.33	(0.46)	0.03	(0.17)
Experience (y)	22.96	(10.51)	12.83	(9.37)	22.39	(11.29)	23.23	(10.20)	19.38	(11.22)
Specialized Teachers	0.96	(0.18)	0.78	(0.38)	0.96	(0.19)	0.26	(0.43)	0.97	(0.18)
Major in Education	0.29	(0.43)	0.53	(0.46)	0.39	(0.48)	0.86	(0.34)	0.25	(0.43)
Instruction Time (h)	1.72	(0.44)	-		1.69	(0.65)	1.39	(0.61)	1.77	(0.7)
# Students	5,002		819		4,035		4,893		4,887	
# Teachers	588		224		645		599		791	

Variables	Lithuania		Malta		Russia		Slovenia		Sweden	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Students' Science Score	516.38	(84.19)	502.62	(112.92)	543.93	(87.73)	553.42	(82.95)	518.78	(91.4)
Bachelors' Teachers	0.55	(0.5)	0.7	(0.46)	0.24	(0.43)	0	(0.06)	0.5	(0.5)
Masters' Teachers	0.41	(0.48)	0.22	(0.4)	0.74	(0.43)	0.61	(0.48)	0.38	(0.47)
Experience (y)	24.38	(10.19)	10.99	(7.98)	22.95	(11.05)	21.98	(10.17)	12.57	(8.37)
Specialized Teachers	0.95	(0.22)	0.91	(0.27)	0.97	(0.16)	0.93	(0.25)	0.63	(0.46)
Major in Education	0.55	(0.48)	0.52	(0.48)	0.53	(0.5)	0.22	(0.4)	0.77	(0.39)
Instruction Time (h)	1.45	(0.65)	2.19	(1.25)	1.58	(0.43)	1.45	(0.53)	1.13	(0.45)
# Students	4,347		2,756		4,780		4,257		4,051	
# Teachers	904		335		749		572		302	

Note: Each column reports weighted descriptive statistics by country. The number of distinct students and teachers are also reported.

Table A2: OLS Regressions

Independent Variables	(1) All Domains	(2) Physics	(3) Biology	(4) Chemistry	(5) Earth Science
Masters' Teachers	0.0133	0.0163	-0.00587	0.0161	0.0238
	(0.0169)	*(0.0253)*	*(0.0246)*	*(0.0266)*	*(0.0313)*
Experience	0.000820	-0.00189	0.00114	0.00189	-0.000109
	(0.000705)	*(0.00121)*	*(0.00110)*	*(0.00121)*	*(0.00122)*
Specialized Teachers	0.0297	-5.73e-05	0.0382	-0.0131	0.0663
	(0.0239)	*(0.0362)*	*(0.0337)*	*(0.0434)*	*(0.0521)*
Major in Education	0.0304*	-0.0170	0.0585**	0.0313	0.0532
	(0.0182)	*(0.0254)*	*(0.0268)*	*(0.0290)*	*(0.0325)*
Constant	-1.838***	-1.816***	-1.590***	-1.820***	-1.603***
	(0.179)	*(0.184)*	*(0.199)*	*(0.211)*	*(0.209)*
Observations	148,751	39,193	38,070	37,555	33,933
R-squared	0.451	0.478	0.481	0.455	0.514
Country FE	YES	YES	YES	YES	YES
Student, Class, School Controls	YES	YES	YES	YES	YES

Notes: Each column includes an OLS regression for the specified domains. Column 1 includes all domains. All specifications include country fixed effects, student, domain-specific, class and school controls. Student controls are student SES, gender, language spoken at home, whether parents have foreign origins and expectations in educational achievement. Domain-specific controls are teacher gender, whether students enjoy learning the domain, find the teaching engaging and instruction time. Class controls are class size, share of students with language difficulties, class SES and the share of native speakers. School controls are the school location, whether science instruction is hindered by shortage of resources, school discipline problems and school emphasis on academic success. Test scores have been standardized across domains and aggregated at the class-domain level to reduce measurement error. Standard errors are clustered at the classroom level.*** p<0.01, ** p<0.05, * p<0.1

Table A3: Piecewise Deletion of Countries

Independent Variables	(1) Armenia	(2) England	(3) Georgia	(4) Hungary	(5) Kazakhstan	(6) Lithuania	(7) Malta	(8) Russia	(9) Slovenia	(10) Sweden
Masters' Teachers	-0.0012	-0.0011	0.0021	0.0015	0.0013	0.0026	0.0039	0.0039	-0.0009	0.0023
	(0.0056)	*(0.0058)*	*(0.0059)*	*(0.0064)*	*(0.0056)*	*(0.0060)*	*(0.0060)*	*(0.0061)*	*(0.0065)*	*(0.0059)*
Experience	0.0003	0.0002	0.0003	0.0003	0.0001	0.0003	0.0001	0.0002	0.0002	0.0003
	(0.0003)	*(0.0002)*	*(0.0003)*	*(0.0003)*	*(0.0002)*	*(0.0003)*	*(0.0002)*	*(0.0003)*	*(0.0003)*	*(0.0002)*
Specialized Teachers	0.0146*	0.0153*	0.0250***	0.0324***	0.0187**	0.0165*	0.0110	0.0165*	0.0113	0.0115
	(0.0087)	*(0.0092)*	*(0.0092)*	*(0.0112)*	*(0.0082)*	*(0.0095)*	*(0.0091)*	*(0.0091)*	*(0.0094)*	*(0.0099)*
Major in Education	-0.0090*	-0.0078	-0.0091	-0.0112**	-0.0074	-0.0117**	-0.0059	-0.0097	0.0002	-0.0043
	(0.0054)	*(0.0057)*	*(0.0060)*	*(0.0057)*	*(0.0055)*	*(0.0058)*	*(0.0059)*	*(0.0060)*	*(0.0059)*	*(0.0054)*
Observations	129,058	146,286	132,975	129,253	129,277	131,506	142,003	129,908	131,877	136,616
Students, Domain FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES

Note: The table reports the results for the within-student across-teachers model that includes four science domains (physics, biology, chemistry, earth Science). The number of observations is given by all the student-domain combinations. The country indicated in each column has been dropped for the estimation. All specifications control for instruction time and teacher gender and include student and domain fixed effects. Test scores have been standardized across domains and aggregated at the classroom-domain level to reduce measurement error. Standard errors are clustered at the classroom level. ***p<0.01, **p<0.05, *p<0.1

Table A4: Piecewise Domain Deletion

Independent Variables	(1) All	(2) Physics	(3) Biology	(4) Chemistry	(5) Earth Science
Masters' Teachers	0.0015	0.0031	0.0019	0.0028	-0.0007
	(0.0057)	*(0.0079)*	*(0.0072)*	*(0.0064)*	*(0.0064)*
Experience	0.0002	0.0004	0.0004	0.0000	0.0003
	(0.0002)	*(0.0003)*	*(0.0003)*	*(0.0003)*	*(0.0003)*
Specialized Teachers	0.0172*	0.0312**	-0.0004	0.0148	0.0206**
	(0.0089)	*(0.0131)*	*(0.0111)*	*(0.0107)*	*(0.0091)*
Major in Education	-0.0076	-0.0144*	-0.0195***	0.0029	0.0014
	(0.0055)	*(0.0076)*	*(0.0068)*	*(0.0065)*	*(0.0067)*
Observations	148,751	107,779	110,377	111,042	113,247
Students, Domain FE	YES	YES	YES	YES	YES

Note: The table reports the results for the within-student across-teachers model. The number of observations is given by all the student-domain combinations. In column 1, all the domains are included. In columns 2-5, the indicated domain has been dropped for the estimation. All specifications control for instruction time and teacher gender and include student and domain fixed effects. Test scores have been standardized across domains and aggregated at the classroom-domain level to reduce measurement error. Standard errors are clustered at the classroom level. ***$p<0.01$, **$p<0.05$, *$p<0.1$