The Effect of Teacher Characteristics on Student Science Achievement

Pietro Sancassani[a]

[a] Ifo Institute, Poschingerstr. 5, 81679 Munich, Germany. E-mail address: sancassani@ifo.de

August 4, 2023

Abstract

Using TIMSS 2015, an international large-scale assessment of student skills, I investigate the effect of teacher characteristics on student science achievement. My identification strategy exploits the feature that in many education systems different science subjects–physics, biology, chemistry, and earth science–are taught by different teachers. The availability of student test scores as well as teachers' questionnaires for each of these subjects allows me to implement a within-student approach which controls for unobserved student heterogeneity. I find a positive and weakly significant effect of teacher subject-specific qualifications on student science test scores, equivalent to 1.7% of a standard deviation. Whether teachers hold a Master's degree, a major in education, or their experience have no significant effect on student science test scores.

Keywords: teachers, student achievement, teacher characteristics, education production function, TIMSS

JEL Code: I21, I29, C21, J24

1. Introduction

There is ample evidence that teachers have a large impact both on student 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). A vast literature has analyzed the effect of various teacher qualifications, such as their degrees, pedagogical or subject-specific qualifications and experience on student test scores. With the exception of teacher experience, many studies have shown that these characteristics are not consistently associated with student performance. However, the large majority of high-quality and causal studies in this literature have focused on single countries, particularly on the United States (for reviews of the literature, see Burroughs et al. 2019; Coenen et al. 2018; Wayne and Youngs 2003). It is therefore an open question whether results from this literature generalize to other and potentially very different countries. Considering the importance of teachers for student outcomes, this question has important implications for education systems worldwide.

In most settings it is often difficult to credibly estimate the impact of teacher characteristics on student 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, especially in international contexts. In this paper, I use international data to investigate the impact of four teacher characteristics, namely whether teachers hold a Master's degree, a subject-specific qualification, a major in education or their level of experience on student science performance. I exploit the availability of student test scores and teacher questionnaires for four scientific subjects– physics, chemistry, biology, and earth science–available for each 8th grade student participating in *Trends in International Mathematics and Science Study* 2015 (TIMSS 2015). I focus on countries in which these science subjects are taught by at least two different teachers. This is a unique setting that allows me to implement a within-student across-teacher approach by linking teachers' characteristics in one specific science subject to student outcomes in the same subject. Using student fixed effects, I eliminate any source of unobserved student heterogeneity, such as innate abilities or general motivation, that is not subject-specific. Unobserved sources of student heterogeneity which are subject-specific, such as student preferences or abilities, might still bias the estimates if they are systematically associated with teacher characteristics and student outcomes. However, this is less of a concern in this study since the subjects analyzed belong to the same field, as opposed to studies using a similar approach but with subjects that belong to different fields, such as math and reading (e.g. Metzler and Woessmann 2012; Bietenbeck, Piopiunik, and Wiederhold 2018; Hanushek, Piopiunik, and Wiederhold 2019) and I provide suggestive evidence that student heterogenous preferences are unlikely to bias my estimates.

The main result of this analysis is that teacher subject-specific qualifications have a positive effect of 1.7% of a standard deviation (SD) on student science test scores. This effect is equivalent to a one-hour increase in weekly instruction time, as estimated in Bietenbeck and Collins (2023). I do not find evidence of any effect of the other teacher characteristics–whether teachers hold a Master's degree, a major in education or years of experience–on student science test scores.

Additional analyses reveal that the average effect of subject-specific teachers' qualifications masks substantial heterogeneities. It is positive and largely significant for biology and chemistry (5.6% of SD) and virtually zero for physics and earth science. Similarly, the effect is driven by Western countries, as opposed to former Soviet or Yugoslavia countries for which the impact is seemingly zero. Further, the effect of subject-specific teachers' qualifications is stronger for female students and for students coming from more affluent backgrounds. It is also robust to the addition of student indicators aiming at capturing remaining subject-specific student heterogeneity, namely the extent to which students enjoy learning the subject or find the teaching engaging. Conversely, the other teacher characteristics considered do not consistently affect student test scores.

The main contribution of this paper is to provide evidence on the impact of teacher qualifications on student science achievement in an international context. The literature investigating the effect of teacher characteristics on student test scores abounds, yet it focuses primarily on the United States and on math or reading test scores. Findings from this paper are therefore more likely to apply to a broader context than what the literature currently suggests. Further, this paper focuses on student science outcomes, which have received less attention in the literature compared to those in math and reading, thereby providing evidence on topical subjects given the widespread attempts to promote Science, Technology, Engineering, and Math (STEM) skills in many countries (e.g., Carnevale, Smith, and Melton 2011; European Commission 2020; Mahboubi 2022).

Concerning the impact of subject-specific teacher qualifications on student test scores, most studies find positive effects, especially in math.[1] For example, using a within-student model akin to that used in this study, Clotfelter, Ladd, and Vigdor (2010) finds an impact of teachers' certification in math of 11% SD on student math test scores and that of a certification in English of 10% SD. Conversely, it does not find any impact of teachers' certification in biology. In a similar vein, previous studies also found positive albeit smaller impacts of subject-specific teacher qualifications on math test scores, but little to no impact on science test scores (Monk and King 1994; Rowan, Chiang, and Miller 1997; Goldhaber and Brewer 1997, 2000). Other studies have not found any impact of subject-specific teacher qualifications on student test scores (Aaronson, Barrow, and Sander 2007; Harris and Sass 2011). These studies use US data, and I enrich this literature with international evidence. Closest to this paper are two recent studies that investigate the impact of teacher subject-specific qualifications on student science test scores using TIMSS data and find a positive effect of 3.5-4% SD (Inoue and Tanaka 2022; Sancassani 2023). Compared to these studies, I focus on a different set of countries where different teachers teach the four science subjects, whereas they focus on countries where the same teacher teaches the four science subjects. In the context of this study teachers are more constrained since, for example, they cannot allocate more instruction time to the subjects they prefer or in which they have a subject-specific qualification, which could bias the estimates. I therefore complement these studies by confirming that teacher subject-specific qualifications positively affect student science test scores also in this context, albeit to a lesser extent and with substantial heterogeneity across science subjects.

Many studies have not found any impact of teachers holding a Master's or advanced degrees on student test scores (e.g., Murnane and Phillips 1981; Hanushek 1992; Goldhaber and Brewer 2000; Clotfelter, Ladd, and Vigdor 2010, 2007; Harris and Sass

[1] A distinct strand of the literature has analyzed direct measures of teacher subject knowledge rather than qualifications and consistently finds that these positively affect student test scores, particularly in math (e.g., Rowan, Chiang, and Miller 1997; Boyd et al. 2008; Clotfelter, Ladd, and Vigdor 2010; Metzler and Woessmann 2012) and also in international contexts (Bietenbeck, Piopiunik, and Wiederhold 2018; Hanushek, Piopiunik, and Wiederhold 2019; Bietenbeck, Irmert, and Sepahvand 2023).

2011; Ladd and Sorensen 2015). In line with results from this literature, I do not find any impact of whether teachers hold a Masters' degree on student science test scores also in the international context I analyze. Similarly, most of the literature has found no or little impact of whether teachers hold a major in education on student test scores (e.g., Goldhaber and Brewer 2000; Croninger et al. 2007; Harris and Sass 2011). Again, results from this analysis are in line with this literature although with more external validity.

Most studies have found that teachers' experience positively affects student test scores (e.g., Boyd et al. 2008; Clotfelter, Ladd, and Vigdor 2010; Harris and Sass 2011; Wiswall 2013; Papay and Kraft 2015; Ladd and Sorensen 2017), although some studies have not found any impact (e.g., Aaronson, Barrow, and Sander 2007; Croninger et al. 2007). I do not find any impact of teacher experience on student science test scores. A possible explanation for this finding is that the greatest gains in teachers' performance with respect to experience occur in the early years of their careers and then quickly flatten, as several studies using US data have shown (e.g., Rockoff 2004; Rivkin, Hanushek, and Kain 2005; Clotfelter, Ladd, and Vigdor 2006; Boyd et al. 2008; Harris and Sass 2011). Teachers in the sample I analyze tend to be relatively old–only 5% of the teachers have less than three years of experience–potentially hindering the detection of any impact of teacher experience on student test scores.[2]

Overall, results from this analysis tend to be in line with the literature, thereby confirming that teacher qualifications and experience–the traditional determinants of teacher compensation–are at best a weak predictor of teacher quality. Consistent with the literature investigating the impact of teacher subject knowledge on student achievement, subject-specific teacher qualifications seem to matter the most for student science test scores among the analyzed characteristics.

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 main results, heterogeneities and robustness checks are discussed in Section 4. Section 5 concludes.

[2] According to TIMSS 2015 data, US students are taught by teachers with an average of 12.5 years of experience, while the students in the sample I analyze are taught by teachers with an average of 19.9 years of experience.

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 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 the science test that TIMSS administers to them is made up of four science subjects: biology, chemistry, physics, and earth science.[3] Tests scores are available for each student and subject, thus yielding four observations at most for each student in science.[4] Depending on countries' curricula, some exceptions are possible; students in Sweden, for instance, are not tested in earth science as this subject does not belong to their 8th grade curriculum, which yields only three observations per student

I focus on ten countries in which the four science subjects are taught by different teachers. This setting allows me to implement a within-student across-subjects model in an international context, where the deviation of test score in one subject from the average science performance of each student is associated with the deviation of teacher characteristics in the same subject from the average science teacher characteristics of each student. Due to the design of international large-scale assessments like TIMSS, this

---

[3] With 35% of the questions asked in the science assessment, biology takes up the large share of the assessment, followed by physics (25%), chemistry, and earth science (both 20%). 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 subject throughout the analysis as previous analyses have shown results are very robust to using multiple plausible values with large samples (e.g., Jerrim et al. 2017; Bietenbeck and Collins 2023; Sancassani 2023).

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 identification strategy used in this study. Contrary to the *Programme for International Student Assessment* (PISA) considered in Jerrim et al. (2017), each booklet in TIMSS 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.[5] Hence, each student is tested in all the science subjects of the TIMSS science assessment.

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 a Master's degree or higher.[6] The subject-specific qualifications of teachers are determined by whether teachers report holding a major in the subject that they teach.[7] This allows me to identify whether teachers have a major in one of the four specific science subjects that are tested in TIMSS. Teachers can also report whether they have a major in education.[8] 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 holding subject-specific qualifications and holding a major in education capture the content and pedagogical qualifications of a teacher, respectively. Finally, I use different specifications of teachers' experience, which is reported in the teacher questionnaire as

[5] In TIMSS, biology, chemistry, physics and earth science are referred to as "domains" to distinguish them from the "subject", science, to which they belong. For simplicity, I refer to these domains as subjects. Each block in the TIMSS booklet contains between 12 and 18 items. For more information concerning the assessment design, see Mullis and Martin (2013).

[6] 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.

[7] 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 subjects of interest. I therefore consider a teacher as holding a subject-specific qualification only if she holds a major in the instruction subject.

[8] Teachers can report whether they have a major in education, education-science or education-mathematics. Results are robust to using an extended definition of major in education that comprises the major in education-science, and also in education-mathematics.

the number of years of experience. These variables provide a common metric to describe teacher preparation in an international context, although the actual quality of teacher preparation can be very different across countries. However, I only exploit variation within students across different teachers, thus accounting for cross-country heterogeneities in average teacher preparation.

Other variables of interest are the extent to which students like learning a subject, henceforth SLL, or find the teaching engaging, henceforth FTE. I use these indicators to ensure that results are not driven by non-random selection of students across science subjects, which might still bias my estimates. These variables are provided by TIMSS 2015 and are based on the student questionnaires. The *Student Likes Learning Biology* indicator, for instance, is based on student agreement with nine statements such as "I enjoy learning biology" or "Biology teaches me how things in the world work". Similarly, the *Student 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 within subjects, so that they have a mean of 0 and a standard deviation of 1 in each subject.

I standardize student test scores within subject in order to facilitate the interpretation of the coefficients. I impute missing values for teacher characteristics and control variables using mean imputation at the school-subject level.[9] The percentage of missing values is between 4.8 and 6.1% for all the variables in the analysis. As some countries in the sample have many more observations than others, I rescale individual weights provided by TIMSS so that each country carries the same weight in the analysis. This ensures that results are not driven by countries with a large number of observations and does not affect the weights within countries. 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 subjects of interest. This tends to be the exception across

[9] Whenever school mean is unavailable, I impute missing values at the country-subject level. All regressions include an imputation dummy for the four teacher characteristics analyzed in this study. Table A8 in the Appendix shows that dropping teachers for which there are missing values in any the four teacher characteristics analyzed (7.9% of the sample) does not affect the main results.

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 that can be used in the subsequent analysis. I also exclude six additional countries for which I am unable to link different teachers to the science subject(s) they teach.[10] In the remaining sample, I also exclude 4% of the students that are taught science by only one teacher or where I am unable to link teachers to a specific subject. 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

I report the descriptive statistics of the sample in Table 1. The large majority of students are taught by teachers with tertiary education: 43% of the students are taught by teachers whose highest level of education is a Bachelor's degree, whereas 48% are taught by teachers with a Master's degree. Students in this sample are also very likely to be taught science subjects by teachers with a subject-specific qualification, namely a major in the subject taught: 83% of the students are taught science subjects by such teachers. Almost half of the students are taught by teachers who have a major in education. On average, teachers in this sample are relatively old and report having 19.9 years of experience. For comparison, teachers in OECD countries report having on average 17 years of experience (OECD 2019). Descriptive statistics of teacher characteristics at the country level reported in Table A2 reveal that this is mostly due to former Soviet and Yugoslavia countries, where teachers tend to have many years of experience.[11] The majority of students, 80%, are taught by female teachers. The prevalence of female teachers is a well-documented phenomenon (e.g., OECD 2018, 2019), although it appears to be even more pronounced in this context, as the average share of female teachers in

[10] The excluded countries are: Dubai, United Arab Emirates, Israel, Japan, Korea and the US. This occurs whenever the variable provided by TIMSS indicating the "Subject Code" of the teacher, which indicates which subject the teacher teachers, does not refer to a particular subject but is coded as "Integrated Science".
[11] As noted in TALIS, a representative and international survey of teachers covering 48 countries, patterns of teacher experience across countries vary with respect to the degree to which teachers work in other roles throughout their careers (OECD 2019). In fact, the teachers' average years of work experience in other non-education roles in the countries that are part of my sample except for Armenia, that did not participate in TALIS, is 2.6, as opposed to 3.5 in OECD countries, and 7 in the United States (OECD 2019). This suggests that teachers in the sample that I analyze tend to work exclusively as teachers.

OECD countries is 68%. Again, figures in Table A2 show that the prevalence of female teachers is driven by former Soviet and Yugoslavia countries, where the share of female teachers is above 90% in some cases. Finally, students are taught each science subject for 1.58 hours (1 hour and 34 minutes) on average per week.

In Table A1, I report the main teacher characteristics separately for each science subject. Teachers' education level is rather homogenous across subjects, although it is slightly lower for physics teachers. At least 80% of the students in each subject are taught by teachers with a subject-specific qualification, with a share that varies between 80% in physics and 87% in earth science. Hence teachers without a major in the science subject that they teach are rather rare in this sample, and almost all students are taught by teachers with such major in some countries (see Table A2). Student earth science teachers are the most likely to hold a subject-specific qualification and less likely to hold a major in education. The main difference of teacher characteristics across subjects concerns teacher gender: the share of students taught by female teachers in physics is 68%, while it is above 80% in the other three subjects. 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.

The prevalence of former Soviet and Yugoslavia countries in the sample that I analyze might limit the external validity of this study, especially if student science performance would be relatively similar across these countries. However, this does not seem to be the case. As shown in Table A3, the student science performance of countries that are part of the analysis (in bold) is quite widespread. The large heterogeneity of student performance therefore speaks in favor of external validity of this study.

3. Empirical Strategy

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

$$A_{icsk} = \beta' T_{csk} + \gamma' X_{ick} + \delta' C_{csk} + \tau' S_{ck} + \theta_k + \sigma_s + \varepsilon_{icsk} \tag{1}$$

where A_{icsk} is the achievement of student i in class c in subject s in country k, T_{csk} is the vector of student i's teacher characteristics of interest, X_{ick} is a vector of student subject-invariant variables that control for student and family background, C_{csk} is a vector of teacher and subject-specific variables such as teacher gender, instruction time, whether the student likes the subject or find the teaching engaging, S_{ck} is a vector of class-specific variables, such as the number of students in the classroom, or the school location, θ_k is a vector of country fixed effects that accounts for country-specific heterogeneity. σ_s is a vector of subject fixed effects, that accounts for differences across subjects. For example, they account for the fact that the test might be more difficult on average in one subject or that teachers in one subject might be, on average, better prepared. ε_{icsk} 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. For example, the literature suggests that the allocation of teachers is unlikely to be random with respect to student socioeconomic status (SES). On the one hand, more wealthy parents try to secure better resources for their children by choosing better schools (Clotfelter, Ladd, and Vigdor 2006).[12] On the other hand, some countries try to improve the conditions in disadvantaged schools through smaller classes or lower student-teacher ratios.[13] In this case, controlling for observable student, teacher, class, and school characteristics can mitigate some concerns. Yet, estimates will be biased if there is non-random sorting of student and teacher characteristics based on unobservable characteristics, such as student innate ability or motivation. For instance, teachers with subject-specific qualifications might be systematically assigned to classes with more motivated and better performing students. Therefore, teacher characteristics might still not be allocated randomly

[12] Among the countries that are part of this study, 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).

[13] 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).

conditional on observable student characteristics, which could bias the OLS estimates of the teacher characteristics in Eq. (1).

As I observe the results of each student in at least three different science subjects, I can eliminate bias due to unobservable student characteristics that do not vary across science subjects. Multiple observations for each student allow me to implement a within-student across-subjects model which controls for unobserved and subject-invariant student traits. The only variation that is left to capture the effect of teacher characteristics is the within-student across-subjects variation. I implement this by estimating the linear model described in Eq. (1) including student fixed effects model with an OLS estimator:

$$A_{icsk} = \beta' T_{csk} + \delta' C_{csk} + \mu_i + \sigma_s + \varepsilon_{icsk} \qquad (2)$$

Adding the student fixed effects, μ_i, to the model makes the inclusion of all student, teacher, classroom, and school variables that do not vary across subjects redundant, and are therefore dropped from the model in Eq. (2). The vector β therefore captures the impact of teacher characteristics net of, for example, unobserved student characteristics that might be correlated with student characteristics and student science test scores.

The main assumption to obtain unbiased estimates of β is that teacher characteristics are randomly assigned to students conditional on student and subject fixed effects and the remaining subject-specific controls, such as teacher gender and instruction time. This is a reasonable assumption since the allocation of teachers and students that might bias the estimates depends primarily on student and school average SES (see footnote 12 and 13), which are controlled for in Eq. (2). However, estimates could still be biased if unobserved subject-specific characteristics were consistently associated with teacher characteristics. For example, if students of higher ability in one science subject were consistently taught by more experienced or higher ability teachers in that subject, estimates might still be biased. While this cannot be ruled out entirely, it seems unlikely given that the subjects analyzed belong to the same field. Hence, the main assumption of the model is more likely to hold in this study as compared to studies using similar models but with different subjects, such as math and reading (e.g., Clotfelter, Ladd, and Vigdor 2010; Lavy 2015; Hanushek, Piopiunik, and Wiederhold 2019).

To corroborate the validity of the main identifying assumption, I provide evidence that controlling for subject-specific variables aimed at capturing any remaining subject-specific heterogeneity, such as whether students like learning the subjects or find the teaching engaging, does not alter the main results appreciably.

Further, I focus on a subsample of students whose teachers teach more than one science subject, thereby including teacher fixed effects in Eq. (2). This identification strategy is the same used in Inoue and Tanaka (2022) and Sancassani (2023) and allows me to control for all teacher characteristics that do not vary across subjects, including all the teacher characteristics analyzed in this study except for the teacher subject-specific qualifications. Results suggest the impact of teacher subject-specific qualifications is robust to this specification and, if anything, larger in magnitude.

A possible drawback of the model estimated in Eq. (2) with subjects that belong to the same field is that the impact of a teacher might spill over into adjacent subjects, thus downward biasing the estimates of β. Further, the amount of variation in outcomes that can be exploited should be a priori smaller as performances in related subjects should not be too different. Therefore, this analysis will likely yield conservative estimates of the impact of teacher characteristics on student outcomes.

4. Results

4.1. Main Results

I first report the results of a model described in Eq. (1) that includes a large set of control variables, subject and country fixed effects in Table A4. In the pooled regression that includes all science subjects in column 1, only the coefficient of the major in education is positive and statistically significant at the 10% level. This association is equivalent to a 3.1% of a SD increase in student test scores. The magnitude of the teacher subject-specific qualifications coefficient is similar in magnitude, but it does not reach any conventional level of statistical significance. I also run the model in Eq. (1) separately for each science subject and report results in columns 2-5. All coefficients do not reach any conventional level of statistical significance, except for the major in education teacher coefficient for biology (column 3), which is positively and statistically significant.

Estimates reported in this table are likely biased as they do not account for unobserved and subject-invariant student heterogeneity. Nonetheless, they provide a useful benchmark for subsequent estimations of the model with student fixed effects described in Eq. (2). Further, they suggest that there is some heterogeneity with respect to the association between teacher characteristics and student science achievement across subjects. I will provide evidence that this heterogeneity persists also in the linear model with student fixed effects in Section 4.2.

I report the results obtained by estimating the within-student across-subjects model of Eq. (2) in Table 2. In columns 1-4, I present the impact of teacher characteristics and student science test scores net of student and subject fixed effects, teacher gender and instruction time separately for each characteristic. In column 5, I include all teacher characteristics simultaneously. Only teacher subject-specific qualifications have a positive and significant effect on student science achievement, while the other teacher characteristics do not have any significant effect on student science test scores.

The effect of teacher subject-specific qualifications on student achievement is equivalent to 1.7%-1.8% SD and is statistically significant only at the 10% level when all teacher characteristics are included. Compared to the results reported in studies similar to this, the magnitude of this effect is roughly half the size reported in Inoue and Tanaka (2022) and Sancassani (2023), who use a more restrictive within-student within-teacher across-subjects model with TIMSS data. The most likely explanation for such a difference is the prevalence of former Soviet and Yugoslavia countries in the sample that I analyze, which are not part of the analysis by Inoue and Tanaka (2022) and Sancassani (2023). In fact, results in section 4.2 show that the effect of teacher subject-specific qualifications in these countries is close to 0. Conversely, the different identification strategy is unlikely to be the reason for the different results found in Inoue and Tanaka (2022) and Sancassani (2023), and I provide evidence of this in Section 4.3.

Results from the model with student fixed effects confirm that neither holding a Master's degree nor years of experience significantly affect student test scores in this sample. The results for teachers' Master's degree are in line with the literature, which has repeatedly shown that teachers with a Master's degree are not more effective (e.g., Hanushek 1992; Clotfelter, Ladd, and Vigdor 2007; Ladd and Sorensen 2015). It is nevertheless interesting to note that this finding also holds outside of the US and with science test scores, a context that has been rarely investigated in the literature.

The impact of holding a major in education becomes insignificant and close to zero when student fixed effects are included in the model. Other studies have also found that holding a major in education does not significantly affect student test scores (e.g., Harris and Sass 2011). These results suggest that the estimate of the major in education is substantially upward biased in the model without student fixed effects. This might be due to unobservable student characteristics that affect student test scores and are associated with the likelihood of being taught by teachers with a major in education, such as student motivation or unobserved ability.

Results for teachers' experience are more puzzling, as most of the literature finds a positive impact on student test scores. As I show in Section 4.3, results for teacher experience are also robust to different specifications of teachers' experience that take into account that the largest gains typically occur at the beginning of teachers' careers. A possible reason for these results is that teachers in this sample are relatively old. Given the cross-sectional nature of the data, this model cannot disentangle cohort effects from experience effects, which might also explain these results. The average impact of teachers' experience might therefore be zero due to opposing cohort and experience effects. This could be the case if, for example, the average teachers' training or skills improved in the last decades and at the same time teachers become more effective with experience.

Finally, it is worth noting that including student fixed effects in the model leads to a considerable increase in the R-Squared. Hence, student fixed effects account for a very large portion of the variation in student test scores. This suggests that any remaining bias due to unobserved factors should be small, and I provide more evidence on this in Section 4.3.

4.2. Heterogeneities

I first explore whether the impact of teacher characteristics varies by subject. As already hinted by the model without student fixed effects, which allows to estimate the impact of teacher characteristics separately for each subject, there appears to be some heterogeneity across subjects. I explore this in the within-student model, where I estimate the impact of teacher characteristics using each possible combination of two science subjects, as this model cannot be estimated using only one subject.

I report results of this exercise in Table 3. First, it can be noted that holding a Master's degree does not seem to affect student science test scores, regardless of the subject.

Concerning the teacher subject-specific qualifications, the effect on student test scores seems to be driven by biology and chemistry, as shown in column 6. In these subjects, the impact on student test scores is equivalent to 5.6% SD. The effect is also positive and statistically significant for physics and biology (column 1), although it is considerably smaller in magnitude (2.4% SD). While this model does not allow to estimate the impact of teacher characteristics separately for each subject, this table suggests that teacher subject subject-specific qualifications in biology and chemistry positively affect student test scores in the same subjects, whereas the impact in physics and earth science is likely zero.

As already noted in Section 2.2, the share of female teachers in physics is much lower than in the other subjects (see Table A1), which might explain why the effect of teacher subject-specific qualifications is likely zero in this subject. Although to a lesser extent, a similar argument also holds for earth science. Previous studies have shown that female students benefit the most from having teachers with subject-specific qualifications, especially when teachers are also females (e.g., Inoue and Tanaka 2022; Sancassani 2023).[14] I will further this finding in the next analysis. Concerning the impact of teachers holding a major in education, the impact is seemingly positive for physics and biology (column 1) and negative for chemistry and earth science (column 2), and physics and chemistry (column 3). Overall, it is hard to draw any conclusive evidence from the pattern emerging from the impact of teachers holding a major in education from this exercise. Finally, teachers' year of experience do not seem to affect student test scores also in these specifications.

I explore heterogeneities by student characteristics in Table 4.[15] In columns 1-2, I explore heterogeneities in the impact of teacher characteristics according to student gender. As mentioned previously, several studies find evidence of a role-model effect of female teachers, whereby having a female teacher improves female student educational outcomes (e.g., Dee 2005, 2007; Winters et al. 2013; Gong, Lu, and Song 2018),

[14] In fact, if I restrict the analysis reported in Table 3 to female teachers only, the impact of teacher subject-specific qualifications is even larger for biology and chemistry (0.0611, p-value < 0.01), where the share of female teachers is relatively larger, and close to zero for physics and earth science (0.0007).

[15] I only report the specifications including all the explanatory variables of interest. Results obtained by including only one explanatory variable at the time are qualitatively the same.

especially in contexts where female role models are relatively scarce (e.g., Card et al. 2022; de Gendre et al. 2023). In line with these results, I also find that the impact of teacher subject-specific qualifications on female student test scores is positively significant and is equivalent to 2.2% SD, 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.

In line with the role-model effect of female teachers for female students, the higher impact of teacher subject-specific qualifications on female students might be due to positive classroom interactions between female teachers and female students.[16] As already mentioned, similar results have also been found in Inoue and Tanaka (2022) and Sancassani (2023), which point to a role model of female teachers for female students.

In columns 3-4, I divide the sample between low- and high-SES students, i.e., students whose SES is below or above the median in their respective country. Teacher subject-specific qualifications have a positive and significant effect only on students coming from relatively more affluent backgrounds, with an estimated impact of 2.8% SD. For teachers with subject-specific qualifications, 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, teachers with subject-specific qualifications might be able to work more efficiently with students who have more subject knowledge from the beginning.[17] This is captured to a large extent by student 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.

[16] In fact, estimating the same model with female teachers only yields an even higher coefficient of the impact of teacher subject-specific qualifications on female student science test scores equivalent to 2.5% SD.

[17] To substantiate this hypothesis, I also divide the sample between low- and high-achievers, i.e., students whose average math 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 between low- and high-SES students. For high-achieving students, the effect of teacher subject-specific qualifications 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 potentially endogenous to the treatments. I therefore stick to splitting the sample between low- and high-SES students as the preferred specification.

Finally, I explore heterogeneities with respect to two distinct groups of countries in the analysis, namely countries that have been part of the Soviet Union or Yugoslavia (Armenia, Georgia, Hungary, Kazakhstan, Lithuania, Russia, and Slovenia) and Western countries (England, Malta, and Sweden). Given the prevalence of countries in the former group, this analysis is useful to better characterize the findings of this study.

I report the results of this exercise in Table 5. Results reveal quite a clear pattern: the impact of teacher subject-specific qualifications seems to be entirely driven by Western countries, where the magnitude of the effect is equivalent to 5.5% SD, whereas it is close to zero for the group of former soviet or Yugoslavia countries. A possible reason for this result is that the share of teachers with a subject-specific qualifications is considerably lower in the group of Western countries (76%) than in former Soviet or Yugoslavia countries (86%), where it is even above 95% in some countries (Armenia, Georgia, Kazakhstan, and Russia, see Table A2). Hence, teachers without such qualifications in the latter group of countries are rather exceptional, which might affect the validity of the estimates for such countries. Findings from this exercise can also help rationalizing why the impact of teacher subject-specific qualifications that I find is considerably smaller than that reported in Inoue and Tanaka (2022) and Sancassani (2023), which use a similar identification strategy and TIMSS data, although they focus on a different and broader set of countries.

Finally, results from Table 5 also suggest that both holding a Master's degree and, especially, a major in education negatively affect student test scores in Western countries. One can only speculate about reasons why teachers with such qualifications might be less effective in these countries, as the data at hand do not allow a thorough investigation of these aspects. For example, teachers with these qualifications in Western countries appear to have fewer years of experience (11.5 for both for teachers with a Master's degree or major in education) than those who do not (13 years of experience, on average), with these differences being statistically significant.[18] This might signal that teachers with such qualifications also differ with respect to some other unobservable characteristics that negatively affect student science test scores. Finally, results from this exercise confirm

[18] Conversely, teachers with and without subject-specific qualifications are not statistically different from each other with respect to years of experience.

that teacher experience does not seem to affect student science test scores in either group of countries.

4.3. Robustness Checks

As discussed in Section 3, the main threat to the identification strategy comes from unobserved and subject-specific factors that affect student test scores and are associated with teacher characteristics. For example, if students that are more passionate about a specific science subject are systematically assigned to teachers that have a major in that subject, the estimate of the impact of teacher subject-specific qualifications will be biased. I previously argued that this seems unlikely given the proximity of the subjects analyzed. Nevertheless, I also provide empirical evidence to back this claim. First, I include in Eq. (2) two subject-specific controls discussed in Section 2, namely the extent to which students like learning a specific subject (SLL) or find the teaching in one subject engaging (FTE). These indicators should capture remaining subject-specific heterogeneities that might still bias the estimates, for example those related to student being more interested in a specific subject and being assigned to better qualified teachers in that subject. I report results from these specifications in Table A5. In column 1, I report the baseline results of Eq. (2), I then include the SLL indicator in column 2, the FTE indicator in column 3, and both indicators simultaneously in column 4. Unsurprisingly, both indicators are positively related to the student test scores, and the main results are robust to the inclusion of these controls.[19] Hence, this table confirms that results are unlikely to be biased by unobserved student or teacher subject-specific factors.

I provide further evidence of the robustness of the impact of teacher subject-specific qualifications on student test scores by restricting the analysis to those teachers that teach more than one science subject. This allows me to include teacher fixed effects in Eq. (2) thereby controlling for subject-invariant teacher characteristics, such as teacher general ability or motivation. This specification is the same as the model estimated in Inoue and Tanaka (2022) and Sancassani (2023) and in this context comes at the cost of a substantial

[19] These indicators are not included in the main specification as controls since they are themselves possible outcomes of teacher characteristics. Hence, they are likely bad controls in this analysis. In fact, in the working paper version of this study, I show that teacher experience has a negative and statistically significant impact on both indicators (Sancassani 2021).

loss in terms of sample size, as only 17% of the students in my sample are taught by such teachers. Further, it only allows the estimation of teacher characteristics that vary across subjects, namely the teacher subject-specific qualifications, whereas the characteristics that do not vary across subjects are absorbed by the teacher fixed effects. For these reasons, a direct comparison of the results from this analysis with the main results cannot be made, but it is nonetheless informative of the type of bias might still be present.

I report results for teacher subject-specific qualifications for this sample in Table A6. In column 1, I use the identification strategy described in Eq. (2), thus including student fixed effects only, and I include teacher fixed effects in column 2. Results in both columns are positive and statistically significant, and considerably larger than those reported for the main sample. It is important to note that the inclusion of teacher fixed effects in column 2 leads to an increase in magnitude of the impact of teacher subject-specific qualifications. Hence, this confirms that teacher unobserved characteristics are unlikely to substantially bias the estimate of the teacher subject-specific qualifications.

Next, I turn to the robustness of the results obtained for teacher experience. Many studies have shown that teachers improve most rapidly at the beginning of their careers (e.g., Rockoff 2004; Boyd et al. 2008), which implies a non-linear relationship between student test scores and teacher years of experience. I therefore also implement several non-linear specifications of experience in the within-student model of Eq. (2) and report results in Table A7. First, I report results of the impact of the years of teacher experience as reported in Table 2, column 4, in column 1. I then include teacher years of experience squared in column 2. I estimate a model with teachers' experience specified using three bins (namely 0-2 years of experience, 3-5, 6 ore more) in column 3, a specification that follows Harris and Sass (2011)'s binning of years of experience in column 4, and a specification with balanced bins, namely where each bin except for the first one contains roughly 20% of the sample, in column 5. Results clearly indicate that experience does not have any impact of student science test scores, regardless of the specification used.

Finally, Table A8 reports the main results obtained by dropping from the sample all teachers for which there are missing values in the teachers' characteristics analyzed in this study (7.9% of the sample). Results are robust to this specification, indicating that results are not driven by the imputation of the missing values.

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 8th graders are taught sciences by different teachers.

The main result of the analysis is that science teachers who hold a subject-specific qualification in the subject that they teach have a positive and weakly significant impact on student 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 student 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 policy implication is that subject knowledge should play a key role in the recruitment and compensation of teachers in lower secondary schools. Furthermore, the benefit of teacher subject-specific qualifications could be reaped at no additional cost by allocating science teachers according to their qualifications.

Acknowledgments
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. The OCCAM ETN project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant number 765400.

References

Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. "Teachers and Student Achievement in the Chicago Public High Schools." *Journal of Labor Economics* 25 (1): 95–135. https://doi.org/10.1086/508733.

Bietenbeck, Jan, and Matthew Collins. 2023. "New Evidence on the Importance of Instruction Time for Student Achievement on International Assessments." *Journal of Applied Econometrics* 38 (3): 423–31. https://doi.org/10.1002/jae.2957.

Bietenbeck, Jan, Natalie Irmert, and Mohammad H. Sepahvand. 2023. "Teacher Quality and Cross-Country Differences in Learning in Francophone Sub-Saharan Africa." *Economics of Education Review* 96 (3): 102437. https://doi.org/10.1016/j.econedurev.2023.102437.

Bietenbeck, Jan, Marc Piopiunik, and Simon Wiederhold. 2018. "Africa's Skill Tragedy." *Journal of 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." *Journal of Policy Analysis and Management* 27 (4): 793–818. https://doi.org/10.1002/pam.20377.

Burroughs, Nathan, Jacqueline Gardner, Youngjun Lee, Siwen Guo, Israel Touitou, Kimberly Jansen, and William Schmidt. 2019. "A Review of the Literature on Teacher Effectiveness and Student Outcomes." In *Teaching for Excellence and Equity*. Vol. 6, edited by Nathan Burroughs, Jacqueline Gardner, Youngjun Lee, Siwen Guo, Israel Touitou, Kimberly Jansen, and William Schmidt, 7–17. IEA Research for Education. Cham: Springer International Publishing.

Card, David, Ciprian Domnisoru, Seth Sanders, Lowell Taylor, and Victoria Udalova. 2022. "The Impact of Female Teachers on Female Students' Lifetime Well-Being." NBER Working Papers 30430.

Carnevale, Anthony P., Nicole Smith, and Michelle Melton. 2011. "STEM: Science Technology Engineering Mathematics. State-Level Analysis." *Tech. rep.* Georgetown University Center on Education and the Workforce. http://cew.georgetown.edu/stem/.

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.” *Journal of 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.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2010. “Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects.” *Journal of Human Resources* 45 (3): 655–81. http://www.jstor.org/stable/25703472.

Coenen, Johan, Ilja Cornelisz, Wim Groot, Henriette van den Maassen Brink, and Chris van Klaveren. 2018. “Teacher Characteristics and Their Effects on Student Test Scores: A Systematic Review.” *Journal of Economic Surveys* 32 (3): 848–77. https://doi.org/10.1111/joes.12210.

Croninger, Robert G., Jennifer King Rice, Amy Rathbun, and Masako Nishio. 2007. “Teacher Qualifications and Early Learning: Effects of Certification, Degree, and Experience on First-Grade Student Achievement.” *Economics of Education Review* 26 (3): 312–24. https://doi.org/10.1016/j.econedurev.2005.05.008.

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.” *Journal of 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.

European Commission. 2020. “European Skills Agenda for Sustainable Competitiveness, Social Fairness and Resilience.” News release. 2020. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2020:274:FIN.

Gendre, Alexandra de, Jan Feld, Nicolás Salamanca, and Ulf Zoelitz. 2023. “Same-Sex Role Model Effects in Education.” University of Zurich Working Paper 438.

Goldhaber, Dan D., and Dominic J. Brewer. 1997. "Why Don't Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity." *Journal of Human Resources* 32 (3): 505. https://doi.org/10.2307/146181.

Goldhaber, Dan D., and Dominic J. Brewer. 2000. "Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement." *Educational Evaluation and Policy Analysis* 22 (2): 129. https://doi.org/10.2307/1164392.

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. 1992. "The Trade-Off Between Child Quantity and Quality." *Journal of Political Economy* 100 (1): 84–117. https://doi.org/10.1086/261808.

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

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.

Inoue, Atsushi, and Ryuichi Tanaka. 2022. "Do Teachers' College Majors Affect Students' Academic Achievement in the Sciences? A Cross-Subfields Analysis with Student-Teacher Fixed Effects." *Education Economics*, 1–15. https://doi.org/10.1080/09645292.2022.2119549.

Jepsen, Christopher, and Steven G. Rivkin. 2009. "Class Size Reduction and Student Achievement." *Journal of 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.

Ladd, Helen F., and Lucy C. Sorensen. 2015. "Do Master's Degrees Matter? Advanced Degrees, Career Paths, and the Effectiveness of Teachers." *National Center for*

Analysis of Longitudinal Data in Education Research (CALDER) (Working Paper 136).

Ladd, Helen F., and Lucy C. Sorensen. 2017. "Returns to Teacher Experience: Student Achievement and Motivation in Middle School." *Education Finance and Policy* 12 (2): 241–79. https://doi.org/10.1162/EDFP_a_00194.

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.

Mahboubi, Parisa. 2022. "The Knowledge Gap: Canada Faces a Shortage in Digital and STEM Skills." https://www.wsj.com/articles/states-sue-.

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.

Monk, David, and Jennifer King. 1994. "Multi-Level Teacher Resource Effects on Pupil Performance in Secondary Mathematics and Science: The Role of Teacher Subject Matter Preparation." In *Contemporary Policy Issues: Choices and Consequences in Education*, edited by Ronald Ehrenberg, 29–58. Ithaca, NY: ILR.

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.

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

Murnane, Richard J., and Barbara R. Phillips. 1981. "What Do Effective Teachers of Inner-City Children Have in Common?" *Social Science Research* 10 (1): 83–100. https://doi.org/10.1016/0049-089X(81)90007-7.

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

OECD. 2019. *TALIS 2018 Results (Volume I): Teachers and School Leaders as Lifelong Learners.* TALIS, OECD Publishing, Paris.

Papay, John P., and Matthew A. Kraft. 2015. "Productivity Returns to Experience in the Teacher Labor Market: Methodological Challenges and New Evidence on Long-Term Career Improvement." *Journal of Public Economics* 130 (4): 105–19. https://doi.org/10.1016/j.jpubeco.2015.02.008.

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.

Rowan, Brian, Fang-Shen Chiang, and Robert J. Miller. 1997. “Using Research on Employees' Performance to Study the Effects of Teachers on Students' Achievement.” *Sociology of Education* 70 (4): 256. https://doi.org/10.2307/2673267.

Sancassani, Pietro. 2021. “The Effect of Teacher Characteristics on Students’ Science Achievement.” *ifo Working Paper* (No. 348).

Sancassani, Pietro. 2023. “The Effect of Teacher Subject-Specific Qualifications on Student Science Achievement.” *Labour Economics* 80 (1): 102309. https://doi.org/10.1016/j.labeco.2022.102309.

Wayne, Andrew J., and Peter Youngs. 2003. “Teacher Characteristics and Student Achievement Gains: A Review.” *Review of Educational Research* 73 (1): 89–122. https://doi.org/10.3102/00346543073001089.

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.

Wiswall, Matthew. 2013. “The Dynamics of Teacher Quality.” *Journal of Public Economics* 100 (1): 61–78. https://doi.org/10.1016/j.jpubeco.2013.01.006.

List of Tables

Table 1: Descriptive Statistics

	Mean (1)	SD (2)	Min/Max (3)
Teacher Variables			
Bachelors' Teacher	0.43	(0.49)	0.0-1.0
Masters' Teacher	0.48	(0.49)	0.0-1.0
Subject-Specific Qualification Teacher	0.83	(0.36)	0.0-1.0
Major in Education Teacher	0.49	(0.49)	0.0-1.0
Experience (y)	19.90	(11.18)	0.0-45.0
Female Teacher	0.80	(0.39)	0.0-1.0
Instruction Time per Week (h)	1.58	(0.71)	0.0-10.0
Student Variables			
Female Student	0.49	(0.50)	0.0-1.0
Speak Language of Test at Home	0.73	(0.44)	0.0-1.0
Born in Country	0.95	(0.22)	0.0-1.0
Student SES Indicator	10.73	(1.54)	4.2-13.9
Student Likes Learning Subject	-0.00	(0.99)	-3.3-2.0
Student Finds Teaching Engaging	-0.02	(0.98)	-3.2-1.4
# Observations		148,751	
# Students		39,827	
# Teachers		5,709	
# Countries		10	

Note: The table reports weighted descriptive statistics for the main variables of interest. The unit of observation is the student-subject combination. In the *Teacher Variables* panel, I report the share of students taught by teachers with a Bachelor's degree, a Master's degree, a subject-specific qualification, a major in education; the average number of years of experience; the share of students taught by female teachers; the average number of instruction time (in hours) per week. In the *Student Variables* panel, I report the share of female students, the share of students who speak the language of the test often at home, the share of students who are born in the country where the test took place, the average value of the student socio-economic status indicator, the average value of the "Student Likes Learning the Subject" indicator and the "Student Finds the Teaching Engaging" indicator. I also report the total number of observations, the number of distinct students, teachers, and countries.

Table 2: The Effect of Teacher Characteristics on Student Science Test Scores

Independent Variables	(1)	(2)	(3)	(4)	(5)
Masters' Teacher	0.0009				0.0006
	(0.0057)				(0.0057)
Subject-Specific Qual. Teacher		0.0178**			0.0169*
		(0.0088)			(0.0089)
Major in Education Teacher			-0.0087		-0.0074
			(0.0054)		(0.0054)
Experience (y)				0.0003	0.0003
				(0.0002)	(0.0002)
Students, Subject FE	YES	YES	YES	YES	YES
Observations	148,751	148,751	148,751	148,751	148,751
R-Squared	0.965	0.965	0.965	0.965	0.965

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

Table 3: Heterogenous Effect of Teacher Characteristics on Student Science Test Scores

	Phy. & Bio. (1)	Chem. & Earth Sc. (2)	Phy. & Chem. (3)	Bio. & Earth Sc. (4)	Phy. & Eath Sc. (5)	Bio. & Chem. (6)
Masters' Teacher	0.0004	0.0001	-0.0042	-0.0007	0.0077	0.0014
	(0.0086)	*(0.0132)*	*(0.0083)*	*(0.0098)*	*(0.0103)*	*(0.0102)*
Subject-Specific Qual. Teacher	0.0241**	0.0166	-0.0159	-0.0169	0.0104	0.0561***
	(0.0120)	*(0.0205)*	*(0.0112)*	*(0.0190)*	*(0.0187)*	*(0.0158)*
Major in Education Teacher	0.0210**	-0.0285**	-0.0174**	-0.0120	-0.0014	-0.0001
	(0.0088)	*(0.0121)*	*(0.0084)*	*(0.0095)*	*(0.0100)*	*(0.0107)*
Experience (y)	0.0000	0.0005	0.0004	-0.0001	0.0003	0.0005
	(0.0004)	*(0.0005)*	*(0.0003)*	*(0.0004)*	*(0.0004)*	*(0.0005)*
Students, Subject FE	YES	YES	YES	YES	YES	YES
Observations	74,858	63,758	73,882	64,434	66,774	74,628
R-squared	0.978	0.965	0.979	0.980	0.975	0.975

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

Table 4: The Effect of Teacher Characteristics on Student Science Test Scores by Gender and SES

	Student Gender		SES	
	Male	Female	Low-SES	High-SES
	(1)	(2)	(3)	(4)
Masters' Teacher	-0.0039	0.0049	-0.0011	0.0043
	(0.0060)	*(0.0062)*	*(0.0061)*	*(0.0066)*
Subject-Specific Qual. Teacher	0.0104	0.0220**	0.0112	0.0277**
	(0.0087)	*(0.0104)*	*(0.0086)*	*(0.0119)*
Major in Education Teacher	-0.0095	-0.0047	-0.0075	-0.0061
	(0.0058)	*(0.0058)*	*(0.0060)*	*(0.0061)*
Experience (y)	0.0001	0.0005*	0.0002	0.0003
	(0.0002)	*(0.0003)*	*(0.0003)*	*(0.0003)*
Students, Subject FE	YES	YES	YES	YES
Observations	76,350	72,401	85,538	63,213

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

Table 5: The Effect of Teacher Characteristics on Student Science Test Scores by Group of Countries

	Western Countries (1)	Former Soviet / Yugoslavia Countries (2)
Masters' Teacher	-0.0216*	0.0011
	(0.0125)	*(0.0064)*
Subject-Specific Qual. Teacher	0.0546***	-0.0006
	(0.0134)	*(0.0106)*
Major in Education Teacher	-0.0399***	-0.0031
	(0.0128)	*(0.0062)*
Experience (y)	0.0003	0.0002
	(0.0007)	*(0.0002)*
Students, Subject FE	YES	YES
Observations	21,348	127,403

*Note:*The table reports the results for the within-student across-subjects model that includes four science subjects (physics, biology, chemistry, earth science). The number of observations is given by all the student-subject combinations. All specifications control for instruction time and teacher gender and include student and subject fixed effects and imputation dummies of the reported teacher characteristics. In column 1, I report results for three Western countries (England, Malta, Sweden), whereas in column 2, I report results for former Soviet (Armenia, Georgia, Hungary, Kazakhstan, Lithuania, Russia) or Yugoslavia (Slovenia) countries. Test scores have been standardized within subjects and aggregated at the classroom-subject level to reduce measurement error. Standard errors are clustered at the classroom level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.h

Appendix

Table A1: Teacher Descriptive Statistics by Subject

	Physics		Biology		Chemistry		Earth Science	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Bachelors' Teacher	0.45	(0.50)	0.43	(0.50)	0.42	(0.49)	0.41	(0.49)
Masters' Teacher	0.45	(0.49)	0.48	(0.49)	0.49	(0.49)	0.51	(0.49)
Subject-Specific Qualification Teacher	0.80	(0.39)	0.85	(0.35)	0.82	(0.38)	0.87	(0.33)
Major in Education Teacher	0.50	(0.48)	0.53	(0.48)	0.52	(0.49	0.39	(0.48)
Experience (y)	20.23	(11.56)	18.95	(11.11)	19.90	(10.90	20.59	(11.03)
Female Teacher	0.68	(0.45)	0.85	(0.35)	0.85	(0.34)	0.82	(0.38)
Instruction Time per Week (h)	1.73	(0.80)	1.52	(0.69)	1.60	0.63	1.46	(0.64)
# Students	39,169		38,069		37,487		33,896	
# Teachers	1,722		1,710		1,636		1,360	

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

Table A2: Descriptives by Country

	Armenia		England		Georgia		Hungary		Kazakhstan	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A										
Bachelors' Teacher	0.13	(0.34)	0.62	(0.49)	0.09	(0.29)	0.65	(0.48)	0.93	(0.25)
Masters' Teacher	0.79	(0.38)	0.25	(0.39)	0.89	(0.31)	0.33	(0.46)	0.03	(0.17)
Subject-Specific Qual. Teacher	0.96	(0.18)	0.78	(0.38)	0.96	(0.19)	0.26	(0.43)	0.97	(0.18)
Major in Education Teacher	0.29	(0.43)	0.53	(0.46)	0.39	(0.48)	0.86	(0.34)	0.25	(0.43)
Experience (y)	22.96	(10.51)	12.83	(9.37)	22.39	(11.29)	23.23	(10.20)	19.38	(11.22)
Female Teacher	0.95	(0.21)	0.54	(0.46)	0.94	(0.23)	0.71	(0.45)	0.83	(0.38)
Instruction Time per Week (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	

	Lithuania		Malta		Russia		Slovenia		Sweden	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Panel B										
Bachelors' Teacher	0.55	(0.5)	0.7	(0.46)	0.24	(0.43)	0	(0.06)	0.5	(0.5)
Masters' Teacher	0.41	(0.48)	0.22	(0.4)	0.74	(0.43)	0.61	(0.48)	0.38	(0.47)
Subject-Specific Qual. Teacher	0.95	(0.22)	0.91	(0.27)	0.97	(0.16)	0.93	(0.25)	0.63	(0.46)
Major in Education Teacher	0.55	(0.48)	0.52	(0.48)	0.53	(0.5)	0.22	(0.4)	0.77	(0.39)
Experience (y)	24.38	(10.19)	10.99	(7.98)	22.95	(11.05)	21.98	(10.17)	12.57	(8.37)
Female Teacher	0.85	(0.35)	0.71	(0.44)	0.90	(0.30)	0.81	(0.39)	0.58	(0.48)
Instruction Time per Week (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 A3: 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 A4: OLS Regressions

	All Subjects	Physics	Biology	Chemistry	Earth Science
	(1)	(2)	(3)	(4)	(5)
Masters' Teacher	0.0142	0.0163	-0.00587	0.0161	0.0238
	(0.0172)	*(0.0253)*	*(0.0246)*	*(0.0266)*	*(0.0313)*
Subject-Specific Qual. Teacher	0.0263	-5.73e-05	0.0382	-0.0131	0.0663
	(0.0241)	*(0.0362)*	*(0.0337)*	*(0.0434)*	*(0.0521)*
Major in Education Teacher	0.0311*	-0.0170	0.0585**	0.0316	0.0532
	(0.0183)	*(0.0254)*	*(0.0268)*	*(0.0290)*	*(0.0325)*
Experience (y)	0.000939	-0.00189	0.00114	0.00189	-0.000109
	(0.000709)	*(0.00121)*	*(0.00110)*	*(0.00122)*	*(0.00122)*
Subject FE	YES	NO	NO	NO	NO
Country FE	YES	YES	YES	YES	YES
Student, Class, School Controls	YES	YES	YES	YES	YES
Observations	148,751	39,193	38,070	37,555	33,933
R-squared	0.451	0.478	0.481	0.455	0.514

Notes: Each column includes an OLS regression for the specified subjects. Column 1 includes all subjects. All specifications include country fixed effects, student, subject-specific, class and school controls and imputation dummies for the explanatory variables of interest. In column 1, subject fixed effects are included. Student controls are student SES, gender, language spoken at home, whether parents have foreign origins and expectations in educational achievement. Subject-specific controls are teacher gender, whether students enjoy learning the subject, 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 within subjects and aggregated at the class-subject level to reduce measurement error. Standard errors are clustered at the classroom level.*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A5: Additional Subject-Specific Controls

	(1)	(2)	(3)	(4)
Masters' Teacher	0.0006	0.0004	0.0006	0.0005
	(0.0057)	*(0.0057)*	*(0.0057)*	*(0.0057)*
Subject-Specific Qual. Teacher	0.0169*	0.0165*	0.0166*	0.0165*
	(0.0089)	*(0.0088)*	*(0.0088)*	*(0.0088)*
Major in Education Teacher	-0.0074	-0.0076	-0.0077	-0.0076
	(0.0054)	*(0.0054)*	*(0.0054)*	*(0.0054)*
Experience (y)	0.0003	0.0003	0.0003	0.0003
	(0.0002)	*(0.0002)*	*(0.0002)*	*(0.0002)*
Student Likes Subject Indicator (SLL)		0.0139***		0.0105***
		(0.0013)		*(0.0013)*
Student Finds the Teaching Engaging Indicator (FTE)			0.0135***	0.0059***
			(0.0016)	*(0.0016)*
Student, Subject FE	YES	YES	YES	YES
Observations	148,751	148,751	148,751	148,751

Note: The table reports the results for the within-student across-subjects model that includes four science subjects (physics, biology, chemistry, earth science). The number of observations is given by all the student-subject combinations. All specifications control for instruction time and teacher gender and include student and subject fixed effects and imputation dummies of the rewported teacher characteristics. In column 2, I also control for the *Student Likes the Subject* indicator, in column 3 for the *Student Finds the Teaching Engaging* indicator, and for both indicators in column 4. Test scores have been standardized within subjects and aggregated at the classroom-subject level to reduce measurement error. Standard errors are clustered at the classroom level. *** p<0.01, ** p<0.05, * p<0.1.

Table A6: The Impact of Teacher Subject-Specific Qualifications with Student and Teacher Fixed Effects

	(1)	(2)
Subject-Specific Qual. Teacher	0.0655***	0.0700***
	(0.0166)	*(0.0170)*
Students, Subject FE	YES	YES
Teacher FE	NO	YES
Observations	23,544	23,544

Note: The table reports the results for the within-student within-teacher across-subjects model that includes four science subjects (physics, biology, chemistry, earth science). The number of observations is given by all the student-subject combinations. All specifications control for instruction time and teacher gender and include student and subject fixed effects and imputation dummies of the teacher characteristics. In column 2, I also include for teacher fixed effects. The sample only includes students whose teachers teach more than one science subject. Test scores have been standardized within subjects and aggregated at the classroom-subject level to reduce measurement error. Standard errors are clustered at the classroom level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A7: Different Specifications of Teacher Experience

	Baseline	Exp. Squared	Three bins	Harris and Sass (2011)	Balanced bins
	(1)	(2)	(3)	(4)	(5)
Experience (y)	0.0003	0.0003			
	(0.0002)	*(0.0009)*			
Exp. Squared (/100)		-0.0001			
		(0.0021)			
Exp. (3-5 years)			-0.0077		
			(0.0136)		
Exp. (> 5 years)			0.0007		
			(0.0113)		
Exp. (3-4 years)				-0.0119	
				(0.0149)	
Exp. (5-9 years)				0.0031	
				(0.0127)	
Exp. (10-14 years)				-0.0032	
				(0.0126)	
Exp. (15-24 years)				-0.0056	
				(0.0122)	
Exp. (> 24 years)				0.0061	
				(0.0117)	
Exp. (2-9 years)					0.0085
					(0.0165)
Exp. (10-17 years)					0.0059
					(0.0167)
Exp. (18-24 years)					0.0014
					(0.0173)
Exp. (25-30 years)					0.0133
					(0.0169)
Exp. (>30 years)					0.0155
					(0.0170)
Students, Subject FE	YES	YES	YES	YES	YES
Teacher Characteristics	YES	YES	YES	YES	YES
Observations	148,751	148,751	148,751	148,751	148,751
R-squared	0.965	0.965	0.965	0.965	0.965

Note: The table reports the results for the within-student across-subjects model that includes four science subjects (physics, biology, chemistry, earth science). The number of observations is given by all the student-subject combinations. All specifications control for whether teachers hold a Master's degree, a subject-specific qualification, a major in education, instruction time, and teacher gender and include student and subject fixed effects and imputation dummies of the explanatory variables. In column 1, only years of teachers' experience is included. I then also include years of experience squared (divided by 100) in column 2. I use bins of teacher years of experience in columns 2-5. In column 3, I use three bins, namely <3 years, 3-5, and > 5 years. The omitted category is the "< 3 years". In column 4, I use Harris and Sass (2011)'s binning scheme. The omitted category is "< 3 years". In column 5, I use balanced bins, where each bin except for the first one contains roughly 20% of the observations. The omitted category is "<2 years". Test scores have been standardized within subjects and aggregated at the classroom-subject level to reduce measurement error. Standard errors are clustered at the classroom level. *** p<0.01, ** p<0.05, * p<0.1.

Table A8: The Effect of Teacher Characteristics on Student Science Test Scores without Missing Values

Independent Variables	(1)	(2)	(3)	(4)	(5)
Masters' Teacher	-0.0007				-0.0003
	(0.0059)				*(0.0059)*
Subject-Specific Qual. Teacher		0.0165*			0.0155*
		(0.0091)			*(0.0092)*
Major in Education Teacher			-0.0084		-0.0072
			(0.0055)		*(0.0056)*
Experience (y)				0.0003	0.0002
				(0.0002)	*(0.0002)*
Students, Subject FE	YES	YES	YES	YES	YES
Observations	136,991	136,991	136,991	136,991	136,991
R-Squared	0.965	0.965	0.965	0.965	0.965

Note: The table reports the results for the within-student across-subjects model that includes four science subjects (physics, biology, chemistry, earth science) only for teachers without missing values in any of the reported teacher characteristics. The number of observations is given by all the student-subject combinations. All specifications control for instruction time and teacher gender and include student and subject fixed. Test scores have been standardized within subjects and aggregated at the classroom-subject level to reduce measurement error. Standard errors are clustered at the classroom level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.