The Effect of Teacher Subject-Specific Qualifications on Student Science Achievement*

Pietro Sancassani[a]

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

October 7, 2022

Abstract

I investigate the effect of teacher subject-specific qualifications on student science achievement using data from TIMSS 2015, a large-scale assessment of student skills. I exploit the availability of student test scores in four different science subjects—biology, chemistry, physics, and earth science—to test whether teachers holding a subject-specific qualification raise student test scores. Using a within-student within-teacher approach, which controls for student and teacher heterogeneity, I find that teacher subject-specific qualification in one subject increases student test scores by 3.5% of a standard deviation in the same subject. The effect is stronger for female students, especially when they are taught by female teachers, for disadvantaged students, and in lower-performing countries. The mediation analysis reveals that 20% of the effect is explained by teachers feeling more confident to teach topics in subjects in which they hold subject-specific qualifications.

Keywords: teachers, teacher qualification, student achievement, teacher characteristics, TIMSS

JEL Code: I21, I29, C21, J24

---

*Together with what is now a separate paper (Sancassani 2021), a previous version of this paper circulated under the title "The Effect of Teacher Characteristics on Student Science Achievement". For valuable comments and discussion, I would like to thank David Figlio, Sandra McNally, Lukas Mergele, Steven Hemelt, Ludger Woessmann, Ulf Zoelitz, two anonymous referees and seminar participants at ifo Institute, the 6th International NEPS Conference, and OCCAM meetings. Pietro Sancassani is part of the European Training Network OCCAM. which received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant number 765400.

1. Introduction

What makes a good teacher? This question has been at the center of a large literature spanning several decades. Although a definitive answer remains elusive, a consensus has seemingly emerged on some facts. Many studies have shown that generic teacher qualifications, such as teacher degree level, advanced degrees, or certification status are not good predictors of teacher quality (Hanushek 1986; Rivkin, Hanushek, and Kain 2005; Clotfelter, Ladd, and Vigdor 2007; Buddin and Zamarro 2009; Staiger and Rockoff 2010; Ladd and Sorensen 2015). Conversely, subject-specific teacher qualifications tend to better predict teacher quality (Monk and King 1994; Goldhaber and Brewer 1997, 2000; Croninger et al. 2007; Clotfelter, Ladd, and Vigdor 2010), although findings in this field are more mixed and less abundant. Yet, a striking feature of this literature calls for caution when interpreting results: the vast majority of studies uses US data. A recent survey of high-quality studies from 2003 to 2018 investigating the effect of any teacher characteristics on student scores features no studies investigating teacher subject-specific qualifications outside the US (Coenen et al. 2018).[1] A concurrent survey of the literature on teacher effectiveness and student outcomes highlights the same issue, thus questioning the extent to which existing evidence applies to other contexts (Burroughs et al. 2019). As teacher education programs vary greatly from country to country (Blömeke, Kaiser, and Lehmann 2010; Tatto et al. 2012), policymakers worldwide should be wary of existing evidence when devising policies concerning teachers. This deficit of evidence is even more critical for developing countries, which likely benefit the most from improving student achievement (Hanushek and Woessmann 2015).

In this paper, I investigate the impact of subject-specific teacher qualifications—as captured by teachers holding a major in science subjects—on student science test scores in an international setting. In most contexts, estimating the impact of teacher characteristics on student outcomes is challenging. Non-random assignment of teachers to students as well as unobservable student and teacher characteristics are the most

[1]Among the reviewed studies, only 9 out of the 58 studies considered do not use US data. Further, a previous review of this literature for the period until 2003 only covered US studies. The rationale for doing so was that the authors were aware of only one study not conducted in the US (Wayne and Youngs 2003).

obvious concerns from an econometric standpoint. I tackle these issues in a novel way by using within-student within-teacher across-subjects variation. I exploit the availability of test scores and teacher qualifications in four science subjects—biology, chemistry, physics, and earth science—available for each 8th-grade student participating in the *Trends in Mathematics and Science Study 2015* (TIMSS 2015). I focus on 30 countries where science is taught as an integrated subject, namely where all science subjects are taught by the same teacher, which constitute most of the countries in TIMSS 2015. Estimates obtained using the within-student within-teacher variation are not biased by unobserved student or teacher characteristics that do not vary across subjects, thus mitigating the most serious sources of bias.

I find that teacher subject-specific qualifications have a positive and statistically significant impact on student test scores. The magnitude of the impact is equivalent to 3.5% of a standard deviation (SD). Putting this figure into perspective, evidence from the US links an increase in teacher effectiveness by one SD to an increase in student math achievement by 20% SD (Jackson, Rockoff, and Staiger 2014). If the variation in teacher effectiveness in the international sample that I use is similar to that in the US, teacher subject-specific qualifications would explain approximately 17.5% of the variation in teacher effectiveness. Similarly, a student would gain approximately $6,825 on average in cumulative lifetime income from being taught by a teacher with subject-specific qualifications in a single grade.[2] Compared to other educational inputs, the effect of teacher subject-specific qualifications is equivalent to an increase of 2 hours and 10 minutes of weekly classroom instruction.[3]

Heterogeneity analyses reveal that the impact is stronger for female students, especially when they are taught by female teachers, and for students with a lower socio-economic status (SES). Concerning teacher characteristics, the impact of teacher subject-specific qualifications is stronger for teachers who also hold a major in

[2]I obtain this figure by multiplying the average gain in cumulative lifetime income from a one SD improvement in teacher value-added in a single grade ($39,000) calculated in Chetty, Friedman, and Rockoff (2014) in the US, by the share of teacher value-added "explained" by teacher subject-specific qualifications (17.5%).

[3]This estimate is obtained by dividing the estimated effect of teacher subject-specific qualifications (3.5% SD) by the average of the impact of a one-hour increase in weekly instruction time on student test scores (1.6% SD) computed in Bietenbeck and Collins (forthcoming) using six waves of TIMSS and PISA data, weighted by the number of countries in each wave.

education and follows a concave path with respect to years of teacher experience. The analysis of cross-country heterogeneities suggests that students in lower-achieving countries benefit more from being taught by teachers with subject-specific qualifications. These findings, together with the previously described larger impact of teacher with subject-specific qualifications for students with lower SES, suggest that students in more disadvantaged contexts might benefit the most from having such teachers. To shed light on the possible mechanisms through which teachers with subject-specific qualifications affect student achievement, I conduct a mediation analysis. I find that up to 20% of the impact of subject-specific qualifications is explained by teachers being more confident to teach subjects in which they hold a major.

This paper contributes to the literature of the impact of teacher characteristics on student test scores in three ways. First, it contributes to the existing evidence on subject-specific teacher qualifications as a determinant of student achievement in an international setting. Previous studies have generally found positive effects of subject-specific teacher qualifications on student test scores, especially for math (Monk and King 1994; Goldhaber and Brewer 1997, 2000; Clotfelter, Ladd, and Vigdor 2010), although other studies have not found any effect (Aaronson, Barrow, and Sander 2007; Harris and Sass 2011).[4] However, all the evidence in this field comes from studies conducted in the US. I enrich this literature by providing first evidence that teacher subject-specific qualifications positively affect student test scores in an international setting. Further, I find a stronger effect in developing and low-performing countries. This result suggests that the current consensus of the literature on teacher qualifications may underestimate the benefits that teacher qualifications bring to students around the world. Nudging teachers to acquire subject-specific qualifications is therefore likely to be beneficial for countries worldwide, especially in developing countries.

Second, I analyze the impact of subject-specific teacher qualifications in a novel way by using a within-student within-teacher across-subjects approach, with subjects

[4]A related strand of this literature has focused on teacher subject knowledge measured with subject-specific test scores rather than qualifications, showing that these are a consistent determinant of student test scores, especially in math (Clotfelter, Ladd, and Vigdor 2007; Boyd et al. 2008; Kukla-Acevedo 2009; Metzler and Woessmann 2012), and also in international settings (Bietenbeck, Piopiunik, and Wiederhold 2018; Hanushek, Piopiunik, and Wiederhold 2019).

belonging to the same field. Much like the more commonly used within-student across-subjects approach[5], it accounts for subject-invariant student characteristics that are known to affect student achievement, such as student ability or socioeconomic background. However, it has the additional advantage of holding constant any teacher characteristics that do not differ across subjects.[6] Further, to the best of my knowledge, this is the first study that applies this approach in a context where the subjects belong to the same field, i.e., science, as opposed to different fields, such as math and reading. A key assumption of all the approaches that exploit within-student across-subjects variation is that unobserved sources of subject-specific student or teacher heterogeneity do not bias the estimates. Given the relatedness of the subjects, this assumption is more likely to hold in this case.

The third contribution of this paper is that I focus exclusively on an important yet understudied subject: science. In the recent survey of the effect of teacher characteristics on student test scores by Coenen et al. (2018), science was among the subjects analyzed in only 11 of the 58 reviewed studies, while the majority of studies focused on math and/or reading. The lack of interest in science is at odds with the current educational and political debate. Calls to nurture science skills in school to address the need for employees with a STEM background and for a scientifically literate public have been pervasive in the last decade (Carnevale, Smith, and Melton 2011; PCAST 2012; OECD 2016; European Commission: DG Employment Social Affairs and Inclusion 2020). The literature has also shown that the impact of teacher qualifications on student test scores varies across subjects. For example, the US study by Clotfelter, Ladd, and Vigdor (2010) finds that the effects of teacher subject-specific certifications are, on average, positive, but very heterogeneous. Test scores of students taught by teachers with math or English certification are 11% SD and 10% SD higher, respectively, while it finds no effect for biology. Similarly, Monk and King (1994) and

[5]The within-student across-subjects approach has been used extensively in the literature to study the impact of teacher characteristics (Clotfelter, Ladd, and Vigdor 2010; Bietenbeck, Piopiunik, and Wiederhold 2018; Hanushek, Piopiunik, and Wiederhold 2019; Sancassani 2021) as well as other educational inputs, such as instruction time (Lavy 2015; Wedel 2021; Bietenbeck and Collins forthcoming) or teaching practices (Bietenbeck 2014) on student outcomes.

[6]Among the studies investigating the impact of teacher qualifications on student test scores, only Harris and Sass (2011) includes one specification with teacher fixed effects. However, they exploit within-teacher variation over time rather than over subjects.

Goldhaber and Brewer (2000) find that teacher subject-specific qualifications have a positive impact on student math test scores, but little or no effect in science.[7] Harris and Sass (2011) does not find evidence of the impact of teacher subject knowledge in math and reading acquired through undergraduate coursework on students' math or reading test scores, but it speculates that in other areas, such as science in secondary school, teacher subject knowledge may be a determinant of student test scores. I provide evidence in favor of this hypothesis.

The remainder of the paper is structured as follows: Section 2 describes the data and provides some descriptive statistics. Section 3 presents the estimation strategy. The main results, heterogeneities, international evidence, and robustness checks are discussed in Section 4. The mediation analysis is discussed in Section 5. Section 6 concludes.

2. Data and Descriptive Statistics

2.1. TIMSS 2015 and Sample Construction

I use data from TIMSS 2015, an international large-scale assessment of math and science skills of 4th- and 8th-grade students, which was the latest wave available at the start of this project. I replicate my main results also using data from the previous TIMSS wave, namely TIMSS 2011. TIMSS includes mathematics and science questions aimed at measuring students' grade-specific curriculum knowledge and a rich set of background questionnaires about students, teachers and schools that gather information about the educational and social contexts of students. The grade-specific focus of the TIMSS assessment makes it more suitable to study the impact of teacher subject-specific qualifications, as these are more likely to affect students' knowledge in a specific grade.[8] TIMSS employs a two-stage random sample design. In the first stage, a random sample of schools is drawn from each participating country with sampling probabilities proportional to school size. In the second stage, one or more entire classes

[7]Using teacher math and reading test scores, Metzler and Woessmann (2012) finds that an increase of one SD in teacher math test scores raises 6th-grade students' math test scores by 9% SD, but has no effect on reading test scores in Peru.

[8]Conversely, the better-known Programme for International Student Assessment (PISA) tests 15-year-old students' general problem-solving ability in math, science, and reading, regardless of students' curriculum and school grade.

of students are randomly selected from each school.[9] By sampling entire classes, TIMSS offers the ideal setting to study the relationship between teacher characteristics and student outcomes. The TIMSS achievement scale was established in 1995 by setting the mean of the average score of all participating countries in TIMSS 1995 to 500 and the standard deviation to 100.

I focus on 8th graders in science. I exclude 4th graders since teachers in primary school are typically trained as generalist teachers (Tatto et al. 2012), which raises important questions regarding the representativeness of the minority of teachers with subject-specific qualifications.[10] In 2015, 40 countries took part in the TIMSS 8th-grade study. 8th graders are around 14 years old and their TIMSS science assessment is made up of the following four subjects, with the share of questions concerning each domain reported in parentheses: biology (35%), chemistry (20%), physics (25%) and earth science (20%).[11] On top of the students' overall science test scores, TIMSS provides test scores for the above-mentioned four science subjects.[12] This is crucial for my identification strategy, which exploits within-student across-subjects variation. I consider the 30 countries where a single teacher teaches all four science subjects, i.e., countries where science is taught as an integrated subject, which allows me to exploit the within-teacher variation.

While compelling from an econometric standpoint, the within-student approach has recently received some criticism due to the design of international large-scale assessments (Jerrim et al. 2017). These tests typically use a matrix-sampling approach that involves splitting the entire pool of test questions into achievement booklets. Students are then randomly assigned to complete only one booklet. This approach

---

[9]In the sample of my analysis, in 76% of the cases only one class per school was sampled in each school.

[10]In TIMSS 2015, 79% of 8th graders have science teachers with subject-specific qualifications in science, while only 38% of 4th graders have such teachers (Martin et al. 2016).

[11]TIMSS distinguishes between "subjects", i.e., math and science, and the "domains" that constitute each subject, such as biology, chemistry, physics, and earth science for science. To ease exposition, I refer to biology, chemistry, physics, and earth science as subjects. For 4th graders, the TIMSS assessment does not make a distinction for biology and chemistry, which are grouped together under the name "life science". This distinction does not map directly into teachers' majors and is a further reason to exclude 4th graders from the analysis.

[12]For each test score, TIMSS provides five plausible values. Throughout the analysis, I use the first plausible value for each subject. As a robustness check, I replicate the main result of the analysis using all five plausible values for all science subjects. Results are robust to this specification (see Table A10).

ensures a comprehensive picture of the achievement of the student population while keeping the length of the test for each student manageable. Focusing on the Programme for International Student Assessment (PISA), Jerrim et al. (2017) highlights that if a student's booklet does not contain any questions regarding a specific subject or domain, multiple imputation is used to create the missing test scores. The resulting within-student variation would then mostly capture the noise induced by the imputation technique. However, unlike PISA, each TIMSS 2015 booklet contains both math and science questions, and, importantly for my application, each science block replicates the proportion of science subjects that constitute science (see Mullis and Martin (2013) for further details about the TIMSS 2015 assessment design), thus limiting this concern.

The explanatory variable of interest comes from the teacher questionnaire, where teachers are asked to indicate their major(s) during their post-secondary education in a pre-specified set of subjects.[13] I construct the sample of interest so that each observation consists of a student-subject combination, which yields four observations for each student. Teacher subject-specific qualifications, the explanatory variable, is a dummy variable that takes value one if the science teacher reports holding a major in the corresponding subject and zero otherwise. For example, if a student's teacher reports holding a major in biology but not in other science subjects, the teacher-subject-specific-qualifications variable will take value one for that student-biology observation and zero for the other student's observations (student-physics, student-chemistry, student-earth science). This constitutes the source of variation in the explanatory variable that I exploit in the within-student within-teacher across-subjects approach.

Teacher subject-specific qualifications might affect student achievement through different channels. For example, teachers might be more prepared to teach subjects in which they have a major. Using the teacher questionnaire, I construct a variable to substantiate this hypothesis.[14] For each science subject, the teacher questionnaire

[13]The original wording of the question is: *"During your <post-secondary> education, what was your major or main area(s) of study?"*. And the possible subjects are: Mathematics, Biology, Physics, Earth Science, Education-Mathematics, Education-Science, Education-General, Other. Teachers can indicate as many majors are they see fit.

[14]Potentially, other channels might also be relevant, such as subject knowledge, motivation, or teaching methods. Unfortunately, the data at hand do not allow any investigation of these or other channels.

includes a list of topics (5.5 on average) addressed by the TIMSS science test.[15] For each of these topics, teachers can indicate whether they feel *very well prepared, somewhat prepared,* or *not well prepared* to teach it.[16] If a teacher holds a subject-specific qualification in a subject, she might feel more prepared and, therefore, confident to teach topics that belong to that subject, and this might raise student test scores. I define a variable that captures the level of preparedness of teachers as the share of topics in each subject that a teacher feels *very well prepared to teach* and test whether it is a mediator of the effect of teacher subject-specific qualifications in the mediation analysis.

In 2015, 40 countries and 285,119 8th graders participated in the TIMSS science assessment. While in most countries science at the 8th grade is taught as an integrated subject, with a single teacher teaching all science subjects, this is not the case in 10 of the countries participating in TIMSS 2015 (namely Armenia, Georgia, Hungary, Kazakhstan, Lebanon, Lithuania, Malta, Russia, Slovenia, and Sweden). In countries where science is taught as separate subjects, a teacher only teaches one of the four science subjects in a classroom. I exclude the 10 countries where science is taught as separate subjects, thus excluding 47,292 students (16.6% of the original sample), as they are not suitable for the within-student within-teacher approach. In the remaining countries, I also exclude 13,383 students (4.7% of the original sample) who are taught science by more than one teacher.[17] The resulting sample consists of 224,454 students, 11,243 teachers and 30 countries. As each student is observed in four science subjects and the unit of observation is the student-subject combination, the total number of observations is 897,760. Throughout the analysis, I use the student sampling weights.

[15]For the full list of topics and the exact wording of the question, see Table A1 in the Appendix. An example of a topic for biology is *"Cells, their structure and functions, including respiration and photosynthesis as cellular processes"*. For chemistry, *"Physical and chemical properties of matters"*. For physics *"Energy forms, transformations, heat, and temperature"*. For earth science *"Earth's structure and physical feature […]"*.

[16]Teachers can also select the option "*not applicable*" if the topic is not in the 8th grade curriculum or they are not responsible for teaching that topic. For the same list of topics, teachers are also asked whether they taught the topic this year, before this year or not (see Table A1, Panel B). However, the topics taught might reflect differences in curricula rather than being an outcome of teacher qualifications and are therefore not included in the mediation analysis.

[17]The only exception is Morocco, where students are taught physics and chemistry by one teacher and biology and earth science by another teacher. This framework also yields within-teacher variation as I observe each teacher in two subjects.

I standardize all test scores within-subject so that the average test score has mean zero and standard deviation one in each subject. Regression coefficients can be therefore interpreted in terms of percentage of a standard deviation. Missing values in the explanatory variable of interest as well as in the controls are imputed using country-level mean imputation. For the main explanatory variable of interest, teacher subject-specific qualifications, I include an imputation dummy in all the regressions. 11.8% of values in the teacher subject-specific qualifications variable are missing. All regression results are robust to the exclusion of observations where teacher subject-specific qualifications are missing.

2.2. Descriptive Statistics

I report the main descriptive statistics for the sample of interest in Table 1. Concerning the main explanatory variable, biology is the most common teacher subject-specific qualification, with 42% of the students taught by a teacher with a major in biology, followed by chemistry (36% of the students), physics (31%), and earth science (20%). It is important to remind that teachers can report more than one subject-specific qualification; in fact, students are taught on average by teachers with 1.24 subject-specific qualifications in science. The modal student is taught by a science teacher with one subject-specific qualification.[18] This figure varies substantially across countries, with the highest average number of teacher subject-specific qualifications in Israel and the lowest in Ontario (Canada) (see column 1 in Table A3 in the Appendix). On average, 73% of the students are taught by teachers who hold at least one subject-specific qualification. Again, this figure masks important cross-country heterogeneities, with the highest share of such students being in England and Morocco and the lowest in Iran (see column 2 in Table A3 in the Appendix). Overall, these data suggest that most 8th-grade science teachers have acquired university-level content knowledge in at least

[18]For the distribution of the number of subject-specific qualifications, see Table A2 in the Appendix (column 3). Along with subject-specific qualifications, teachers can also indicate whether they have majors in other subjects, including education, education-science, or education-math. I also report the distribution of the number of subject-specific qualifications by whether teachers also hold any major in education in Table A2 (column 1 and 2).

one of the science subjects that they teach.[19] Even teachers without a major in a certain subject likely received some form of training in the content of the subject that they teach. In fact, according to the international teacher survey TALIS 2018 led by the OECD, 92% of a representative sample of lower secondary education teachers in 48 countries report having received training in the content of the subject that they teach (OECD 2019). The source of variation in the explanatory variable that I exploit for the preferred identification strategy stems from students being taught by teachers having at least one and less than four science subject-specific qualifications. It is therefore important that a considerable number of students are taught by teachers who satisfy this requirement. This is in fact the case, as 66% of the students are taught by such teachers (see column 3 in Table A3 in the Appendix).

Apart from the subject-specific qualifications, the TIMSS background questionnaires provide a wealth of information on teachers' and students' backgrounds, which I now briefly describe. On average, science teachers of 8^{th}-grade students report high levels of education. 62% of the students are taught by teachers with a Bachelor's degree and 22% by teachers with a Master's degree. These figures are in line but slightly smaller than those reported for the entire TIMSS 2015 8^{th}-grade science sample, in which 92% of students are taught by teachers with at least a Bachelor's degree. Teachers report having, on average, 14.54 of experience, in line with the figure reported for the entire TIMSS 2015 sample of 15 years of experience. The share students taught by teachers who report having a major in education is 61%[20] and having a major in education is negatively correlated with also having a subject-specific qualification in science (-.28, p-value < .001). The share of female teachers is 58%. The average weekly

[19]The observed cross-country heterogeneities might be due to how teachers are trained and selected in different countries. Another explanation is that the concept of majoring in one subject differs across countries. Thus, the subject knowledge acquired by majoring in one subject might also vary accordingly, affecting the independent variable's cross-country comparability. Nonetheless, this concern is not an issue for my estimates as I do not exploit variation stemming from cross-country variation in the independent variable.

[20]This figure includes teachers that report having either a major in education, education-science or education-mathematics. The figure for teachers who report having a major in education-science is 51%, for teachers who report having a major in education is 27%, and for education-mathematics is 9%. According to the TALIS 2018 survey, 92% of teachers across OECD countries and all the countries participating in TALIS received training in general pedagogy and in the pedagogy of the subjects that they teach (OECD 2019). It is therefore unlikely that teachers in my sample do not have any pedagogical preparation, regardless of whether they report holding any major in education.

instruction time for students in science is 5.65 hours. On average, students are taught by teachers who feel confident to teach 54% of the topics tested in TIMSS.

To explore country heterogeneities, I include country-level data from a variety of sources. For the distinction between developed and developing countries, I use the World Economic Situation and Prospects (WESP) 2014 classification of the United Nations (United Nations 2014). For GNI per-capita measures of countries in 2015, I use the World Bank data (World Bank 2021). The large variation in average science performance of the considered countries as well as other factors such as geographical location or economic development speaks in favor of the external validity of this study.

3. Empirical Strategy

To causally estimate the effect of teacher subject-specific qualifications on test scores, one would need to assume that teachers are randomly assigned to students and subject-specific qualifications to teachers. In practice, however, this is unlikely to be the case. First, the allocation of teachers is typically non-random, as, for example, wealthy parents tend to secure better resources for their children by choosing better schools (Clotfelter, Ladd, and Vigdor 2006). Second, teachers' decision to obtain subject-specific qualifications might depend on preferences or ability. If the teacher subject-specific qualifications are correlated with determinants of student test scores, the estimated effect of teacher subject-specific qualifications will be biased.

To address these concerns, I first implement a standard OLS approach estimating an education production function with a rich set of controls, which account for observable heterogeneities. I then implement a within-student within-teacher approach which also accounts for unobserved student and teacher heterogeneity that are subject invariant.

I first estimate the following linear model using OLS with a rich set of controls:

$$A_{istk} = \alpha S_{istk} + \gamma' X_{tk} + \delta' Z_{itk} + \theta_k + \varphi_s + \varepsilon_{istk} \tag{1}$$

where A_{istk} denotes the test score of student i in subject $s \in$ (*biology, chemistry, physics, earth science*), taught by teacher t in country k. A_{istk} is determined by the teacher subject-specific qualifications of student i's teacher t in subject s, S_{istk}, a vector of teacher as well as class and school characteristics X_{tk}, a vector of students characteristics Z_{itk}, country fixed effects θ_k and subject fixed effects, φ_s, with ε_{istk} being the idiosyncratic error. This model accounts for several factors that are known to affect students' outcome, such as students' socioeconomic status or gender (included in the vector Z_{itk}), teachers' experience (included in the vector X_{tk}) as well as country and subject heterogeneities (captured by the fixed effects θ_k and φ_s, respectively).

The main identifying assumption to obtain an unbiased estimate of the parameter of interest, α, is that teacher subject-specific qualifications, S_{istk}, are uncorrelated with the error term conditional on the included regressors. While controlling for observable student, teacher and class characteristics alleviates some of the concerns mentioned previously, unobservable determinants of students' test scores that are correlated with teacher subject-specific qualifications might still lead to a violation of the identifying assumption. If, for example, higher ability students are systematically sorted into classes with teachers with subject-specific qualifications, the estimated α in Eq. (1) is potentially upward biased. Conversely, α could be downward biased if teachers with subject-specific qualifications tend to be assigned to classrooms with lower-ability students. Similarly, more motivated, or higher-ability teachers might be more likely to hold a subject-specific qualification. Thus, both student and teacher unobserved characteristics can potentially bias the estimate of α and can do so independently from each other. It is therefore important to develop an identification strategy that can tackle both sources of bias.

To this purpose, I estimate a within-student within-teacher across-subjects model. As I observe the results of each student in four distinct science subjects, I can eliminate the heterogeneity due to unobservable student characteristics that do not vary across science subjects by including student fixed effects in Eq. (1). Further, I also observe every teacher in the same four subjects. I therefore include teacher fixed effects in Eq. (1), which control for all unobserved teacher characteristics that do not vary across subjects.[21] Essentially, student and teacher fixed effects account for all the observable and unobservable characteristics at the student, teacher, class, and school

[21]This represents the main difference with respect to the identification strategy in Sancassani (2021), where teachers are observed in only one science subject, thus preventing the exploitation of the within-teacher variation.

level that do not vary across subjects. Empirically, I estimate the following linear model:

$$A_{ist} = \beta S_{ist} + \sigma_i + \tau_t + \varphi_s + \varepsilon_{ist} \tag{2}$$

Where the subject-specific test score A_{ist} is determined by the teacher subject-specific qualifications S_{ist} and the student, teacher, and subject fixed effects (σ_i, τ_t, and φ_s respectively). Student and teacher fixed effects make the inclusion of all the subject-invariant student (Z_{itk}), teacher and classroom (X_{tk}) variables as well as country fixed effects (θ_k) redundant and are therefore omitted from Eq. (2).

Student fixed effects control for many subject-invariant characteristics that are known to affect student achievement, such as the socioeconomic status, general motivation, innate abilities, as well as classroom and school characteristics. Similarly, teacher fixed effects control for the subject-invariant components of observables teacher characteristics, such as teacher experience, education level or gender, as well as the subject-invariant components of unobserved teacher characteristics, such as motivation or ability. Finally, subject fixed effects eliminate subject-specific test score heterogeneities and other subject-specific unobserved factors, such as different curriculum coverage in different subjects. The estimation of β in Eq. (2) is therefore unlikely to be biased by the two main sources of bias mentioned: the unobserved subject-invariant student and teacher characteristics.

The main threat to the identification strategy consists of unobserved subject-specific heterogeneities. In fact, the estimated β might still be biased if unobserved subject-specific determinants of student outcomes, such as subject-specific instruction time, student or teacher ability or passion for the subject are correlated with the teacher subject-specific qualifications. To alleviate such concerns, I show that the results are robust to the inclusion of subject-specific instruction time and to restricting the sample to schools where student sorting is unlikely. Furthermore, following Oster's bounding exercise (Oster 2019), I show that any remaining bias due to unobserved factors should be negligible. Another concern for my identification strategy is that the estimated β might capture the effect of being taught by a teacher with subject-specific qualifications in the 8th grade and in previous years. Unfortunately, the data at hand do not allow to control for the qualifications of teachers in previous years. Nonetheless, the focus on the grade-specific knowledge of the curriculum of the TIMSS assessment ensures that any bias through this channel is most likely small. Finally, it is worth reminding that the more likely sorting of student and teachers based on student SES, general ability or interest for science is accounted for by student and teacher fixed effects.

A further assumption of this model is that the impact of teacher subject-specific qualifications is homogenous across subjects. Compared to studies that use a similar within-student identification strategy but using different subjects, it is a far weaker assumption in this setting, as the student test scores belong to the same field. Other things being equal, it is unlikely that science subject-specific qualifications might have a larger or smaller impact in different science fields, and I provide evidence of this.[22] Further, I show that the effect of teacher subject-specific qualifications is robust and stable with respect to the individual exclusion of each science subject in the robustness checks, which alleviates this concern.

A potential downside of using closely related outcomes is that the effect of teacher subject-specific qualifications in one science subject might spill over into other subjects. Relatedly, being the subjects so closely related to each other, the amount of variation that can be exploited should not be too large, as they probably require a similar set of student innate abilities. Considering these points, my estimates likely reflect a lower bound of the true effect.

4. Results

4.1. Main Results

Table 2 presents the main results of the impact of teacher subject-specific qualifications on student test scores. I first report results of the linear model described in Eq. (1) pooling the student test scores in the four science subjects —biology, chemistry, physics, and earth science—with an increasingly rich set of control variables (columns

[22]I directly test this by estimating the linear model in Eq. (1) including an interaction term between teacher subject-specific qualifications and subjects. I then perform a Wald test of equality of all the coefficients of the interaction terms, which I cannot reject (*p*-value = .77, F-statistic = .26); pairwise tests of equality of the coefficients also rule out heterogeneity in the coefficients.

1-3). I then report the result using the within-student within-teacher across-subjects approach described in Eq. (2) (column 4). The impact of teacher subject-specific qualifications on student test scores is positive and statistically significant and varies between 3.3% SD to 3.6% SD. The preferred estimate, the one obtained with the within-student within-teacher across-subjects approach (column 4), lies between the coefficients of the pooled linear models. It is positive and statistically significant at the 1% level and implies that teacher subject-specific qualifications raise student test scores in the subject in which a teacher holds a subject-specific qualification by 3.5% SD. The estimated coefficient in column 1 changes very little when including controls and fixed effects in the regressions, despite a substantial increase in the R-squared. This suggests that the remaining bias due to unobserved subject-specific factors is likely small. I substantiate this claim formally in Section 4.4, where I perform an analysis of unobservable selection and coefficient stability following Oster (2019).

Results show that teacher subject-specific qualifications matter for student science test scores. The magnitude of the effect, equivalent to 3.5% SD, is relatively small for a single school year but can become substantial if considered over a school cycle of six years, the average duration of secondary education worldwide (UNESCO 2021).

4.2. Heterogeneity – Student and Teacher Characteristics

I explore heterogeneities of the impact of teacher subject-specific qualifications in Table 3 using the within-student within-teacher across-subjects approach in Eq. (2). Several studies have found that student and teacher gender matters for educational achievement, especially for female students (Dee 2005; Paredes 2014; Lim and Meer 2017; Sansone 2017). This is even more important in science and, more in general, STEM subjects, where females have been historically underrepresented. I interact the teacher subject-specific qualifications separately with student and teacher gender (column 1 and 2, respectively) to tease out heterogeneities in the effect of teacher subject-specific qualifications with respect to student and teacher gender. Estimates suggest that female students benefit more from being taught by a teacher with subject-specific qualifications (column 1), whereas teacher gender alone does not seem to play a role for the effectiveness of teacher subject-specific qualifications (column 2). As a further step, I explore whether female students, who already benefit more from being taught by teachers with subject-specific qualifications, benefit even more when these teachers are also females. The rationale for this analysis follows the role-model effect of teachers observed in the literature (Dee 2005; Paredes 2014), according to which girls benefit from being assigned to female teachers without negative effects for boys. Such effect is possibly because female students might be more confident in learning science if the role-model to which they are exposed is a female teacher. I therefore test whether the interaction term between the teacher subject-specific qualifications and student gender varies by teacher gender.[23] I find that female students taught by teachers with subject-specific qualifications perform significantly better when their teachers are also females (table not shown), in line with the teacher role-model effect mentioned previously.

Teacher subject-specific qualifications may have a different impact on students with different SES, which, to a large extent, also captures student prior achievement. Theoretically, the marginal increase in teacher subject knowledge induced by teachers acquiring subject-specific qualifications might have different returns based on students' prior knowledge. Differences in the impact of teacher's subject-specific qualifications with respect to student SES might therefore reveal different functional forms that characterize the relationship between teacher subject knowledge and students' achievement. I explore such heterogeneity in column 3, where I interact teacher subject-specific qualifications with an indicator for student SES. I find that the effect of teacher subject-specific qualifications decreases as student SES increases. This finding suggests a steeper relationship between teacher subject knowledge and student achievement for

[23]Empirically, I include an interaction between teacher gender and the interaction between teacher subject-specific qualifications and student gender to the model estimated in column 1, but without including the main effects for the triple interaction. This is equivalent to estimating the interaction term between teacher subject-specific qualifications and student gender separately for the sample of female and male teachers. The coefficient associated with the triple interaction, which captures the effect for female students taught by female teachers with subject-specific qualifications, is positive and statistically significant (.018, p-value < .10). Similarly, the effect of teacher subject-specific qualifications for female students is larger when estimated for the sample of female teachers as opposed to the sample of male teachers (.072 and .053, respectively; with p-value < .01 for both terms).

lower SES students.[24] It also has important equity implications, as students from more disadvantaged contexts benefit the most from having teachers with subject-specific qualifications.

A similar theoretical argument can be made for other teacher qualifications. Teacher subject-specific qualifications could affect students' test scores differently based on teachers' general educational attainment or pedagogical knowledge. A steeper relationship between teacher subject knowledge and student test scores for teachers who also have a Master's degree or a major in education might indicate a complementarity between these additional qualifications. I explore these hypotheses in column 4 and 5. I do not find a statistically significant interaction between teacher subject-specific qualifications and teacher holding a Master's degree (column 4). I therefore do not find supporting evidence for the complementarity between such qualifications. Conversely, the interaction between teacher subject-specific qualifications and whether the teacher holds a major in education is positive and statistically significant (column 5), which implies that the effect of teacher subject-specific qualifications is larger for teachers who also have a major in education. This result suggests that that teacher pedagogical knowledge, captured by the major in education, and teacher subject knowledge, captured by the teacher subject-specific qualifications, are complementary ingredients for effective teaching.

Finally, I explore the role that teacher experience plays in the effectiveness of teacher subject-specific qualifications (column 6). I include both a linear and quadratic term for teacher experience[25] to tease out the largely documented non-linear relationship between teacher experience and student test scores (Rivkin, Hanushek, and Kain 2005; Boyd et al. 2008; Clotfelter, Ladd, and Vigdor 2010). The coefficients suggest a concave relationship between the effect of teacher subject-specific qualifications interacted with teacher experience and students' achievement. I provide a graphical

[24]I find similar results by interacting teacher subject-specific qualifications with a more direct measure of student prior achievement, student math test scores (not shown). Students in the lower part of the distribution of math test scores benefit the most from teachers with subject-specific qualifications. The student SES indicator correlates highly with the math test scores, but due to the potential endogeneity of the math test scores; I stick to the interaction with student SES as the main specification for this analysis.

[25]Following the existing literature on the (non-linear) effect of teacher experience on student test scores, I also define teacher experience in bins (namely 0-1 year, 2-5 years, 6-9 years, 10-12 years, 13-16, 17-23, 24+). Results from this specification (not shown) are qualitatively the same.

representation of this result in Figure 1, which shows that the effect of subject-specific qualifications reaches its peak around the midpoint of teacher experience (at roughly 18 years of experience), after which it declines. It is important to remind that teacher experience is collinear to teacher age. It is possible that the observed pattern is due to an experience effect, meaning that teachers improve their effectiveness in the first part of their career by, for example, learning by doing. Alternatively, this pattern could also be due to a cohort effect, meaning that the ability of teachers differs by cohort.[26] Given the cross-sectional nature of the data, I cannot disentangle these two components, but the pattern observed in this analysis is more in line with the vast literature reporting diminishing returns to teacher experience.

4.3. Heterogeneity – Country Subsamples

The wide heterogeneity of the countries considered is advantageous for the external validity of the results, although it brings additional challenges. If teacher training differs markedly across countries, holding subject-specific qualifications might mean different things. I therefore focus on the sub-group of OECD countries in the sample, for two main reasons. First, teachers in OECD countries report, on average, fewer subject-specific qualifications despite a higher level of education.[27] This likely indicates that subject-specific qualifications represent teachers' main field of study in OECD countries. Second, OECD surveys provides a wealth of information regarding teacher training. This allows me to provide a clearer picture about the framework in which teachers are selected and trained in these countries. According to the TALIS 2018 survey, in the OECD countries included in my sample except for Canada and Ireland, which are not covered in TALIS 2018, 92.7% of teachers report to have received training in the content of some or all subjects taught, 90% have received training in pedagogy of some or all subjects taught, and 92% in general pedagogy. These figures suggest that teachers in OECD countries likely received some training in both pedagogy and the content of the subjects they teach, regardless of their subject-specific

[26]For example, Nagler, Piopiunik, and West (2020) show that teachers who enter the profession during economic downturns are significantly more effective in raising student test scores.

[27]34% of students in non-OECD countries are taught by teachers who report two or more subject-specific qualifications, while only 26% of students in OECD countries are taught by such teachers.

qualifications. Further, the educational requirements for entry into initial teacher training differ little across OECD countries, where the minimum requirement is usually an upper secondary qualification (OECD 2022).

I report the main results for this subgroup of countries in Table 4 with the same specifications used for Table 2. All estimated coefficients are positive and statistically significant, although they decrease as I include more controls in the model. Interestingly, the R-squared in column 1 is much smaller than the R-squared in the same specification in Table 2, which indicates that this group of countries is much more homogenous. In the preferred specification of column 4, the magnitude of the coefficient is 2.8% SD, which is slightly smaller than the coefficient estimated in Table 2 for the full sample, although not statistically significantly different from it, as I show in Table 5. This implies that, even in the context of OECD countries where teachers likely received extensive training, students perform better in those subjects where their teachers hold subject-specific qualifications. To test whether the impact of teacher-subject specific qualifications varies by country subsamples, I include interactions between teacher subject-specific qualifications and a series of country indicators[28] in Eq. (2) and report the results in Table 5. First, I explore whether the effect of teacher subject-specific qualifications varies in countries that belong to the OECD (column 1) or are developed countries[29] (column 2). A priori, it is unclear if teacher subject-specific qualifications could be more effective in OECD (developed) or non-OECD (developing) countries. This ultimately depends on a variety of factors, such as the already mentioned teacher preparation, the attractiveness of the teaching career and so on. While the interaction term in column 1 points to the negative area, it does not reach any conventional level of statistical significance. However, the interaction term in column 2 suggests that teachers with subject-specific qualifications are more effective in developing countries.

The effect of teacher subject-specific qualifications might also depend on countries' average science achievement. A priori it is unclear whether students in

[28]For the list of all countries and the country indicators, see Table A3 in the Appendix.

[29]For the developed vs. developing countries classification, I used the WESP classification (United Nations 2014). This classification includes a further category of countries "in transition". However, none of these countries is in the sample I analyze. Being the OECD a club of mostly rich countries, the developed countries group is a subset of the OECD group.

countries with high average achievement could benefit more from having teachers with subject-specific qualifications. I therefore split the sample in countries that perform above and below the median science test score in my sample. Results show that teachers with subject-specific qualifications are more effective in countries with average science performance below the median (column 3). A further distinction between countries that are above and below the median GNI per capita does not show significant heterogeneities between relatively rich and poor countries (column 4). A possible explanation for results from this table is that the counterfactual teacher effectiveness, i.e., the effectiveness of teachers in science subjects in which they do *not* have a major, is lower in developing or lower-performing countries. As previously argued, teachers in OECD countries seemingly received pedagogical and content training in the subjects that they teach. While the data at hand do not allow to make similar claims for developing and lower-performing countries, it is possible that teachers in these countries received, on average, less training. For this reason, subject-specific qualifications might have larger value-added for teachers in these countries.

4.4. Robustness Checks

As discussed in Section 3, the main threat to the identification strategy comes from unobserved subject-specific confounders, while subject-invariant confounders are accounted for by student and teacher fixed effects. I therefore perform a series of robustness checks to ensure that any remaining bias due to subject-specific heterogeneities should not invalidate my estimates. A possible concern comes from different instruction time devoted to science subjects. If schools or countries that emphasize one science subject over the others are also more likely to appoint teachers with subject-specific qualifications in that subject and devote more instruction time to the same subject, estimates might be upward biased.[30] To mitigate this concern, I replicate my main analysis using TIMSS 2011, which allows me to control for the share of instruction time that teachers report to dedicate to each science subject. Results are reported in Table A4. First, it is reassuring to see that I can essentially replicate the main

[30]However, instruction time can also be an outcome of teacher subject-specific qualifications if teachers systematically devote more instruction time to the subjects in which they have a major. In this case, controlling for instruction time would be problematic.

result of the paper also using TIMSS 2011. The within-student within-teacher across-subjects specification in column 4 is positive and statistically significant, albeit slightly smaller in magnitude than the main specification in column 4, Table 2. Second, the results are robust when I control for instruction time in column 3 and 5, although the coefficient in the preferred within-student within-teacher across-subjects specification in column 5 slightly decreases. Following Bietenbeck, Piopiunik, and Wiederhold (2018), I address the issue of the remaining subject-specific student and teacher sorting by restricting the sample of my main analysis with TIMSS 2015 to students living in areas with less than 30 thousands, 15 thousands people or in rural areas. Students in these areas likely have little choice between different schools, which makes the issue of sorting less worrying. I report the results from this analysis in Table A5. Results are robust to these specifications and, if anything, they are larger in magnitude. Finally, I conduct an analysis of unobservable selection and coefficient stability following Oster (2019). I compare the coefficient estimated through the within-student within-teacher across-subjects specification (column 4 of Table 2) to the specification including only country and subject fixed effects (column 1 of Table 2) and setting $R_{max} = 1$ and $\delta = 1$.[31] Results, reported in Table A6, indicate that the estimated bias-adjusted treatment effect β^* is .035, which is identical to the preferred estimate. The value of δ for which $\beta = 0$ is 19.51, which far exceeds the standard cutoff of 1 and implies that the selection on unobservable characteristics needs to be almost 20 times larger than the selection on observables characteristics to drive the effect of teacher subject-specific qualifications to zero.

To ensure that results are not driven by a specific subject where teachers might benefit particularly from holding a subject-specific qualification, I replicate the main result by excluding one subject at a time. Results in Table A7 show that the effects are robust to the exclusion of each science subject. These results also address a concern raised in Section 3 about the potential bias induced by heterogenous effects of teacher

[31]These values denote the R-squared from a hypothetical regression of the outcome on the treatment and both observed and unobserved controls, and the relative degree of selection on observed and unobserved variables (Oster 2019), respectively. In practice, Oster (2019) recommends an $R_{max} = 1.3\tilde{R}$, where $\tilde{R}$ denotes the R-squared obtained in the regression with all controls, which in my case is .94 (see column 4 of Table 2). I therefore set $R_{max} = 1$ since setting $R_{max} = 1.3\tilde{R}$ would imply an implausible $R_{max} > 1$.

subject-specific qualifications and confirm that the results are rather homogeneous across different science subjects.

Given the heterogeneity of the countries considered in my analysis, it is possible that the effect of teacher subject-specific qualifications is driven by some countries where teachers with such qualifications are particularly effective in raising student test scores. I address this concern by replicating the main result excluding one country at a time. Results from the leave-one-country-out exercise in Table A8 are robust to the exclusion of each country in the sample. It seems therefore unlikely that results a driven by some outliers in the sample of countries considered. The effect of teacher subject-specific qualifications varies between 2.3% and 3.9% of a SD, with the lower and upper bound obtained when Egypt and Japan are excluded, respectively. Japan and Egypt lie at the opposite extremes of the distribution of science performance, with Japan being among the highest performing countries and Egypt among the lowest performing countries in the sample. This finding corroborates the evidence that the effect of teacher subject-specific qualifications is stronger in lower-performing countries.

A further issue concerns the weight that each country has in the analysis. Due to different sample sizes across countries, different countries carry different weights in the analysis. Instead of using the sampling weights provided by TIMSS, I replicate the results using rescaled weights so that each country carries the same weight ("senate weights"). Results, shown in column 2 in Table A9, are robust to this specification, although slightly smaller in magnitude.[32]

I also address issues related to the complex design of international assessment in Table A10. First, to minimize manipulation of the test scores, I replicate the main results using the raw (i.e., non-standardized) first plausible value for each science subject as outcome (column 2). I find that the impact of being taught by a specialized teacher is equivalent to 4.37 points, which corresponds to 3.7% SD,[33] in line with the coefficient estimated in the main specification (3.5% SD). Second, to account for the

[32]Some studies using international assessments (Lavy 2015; Rivkin and Schiman 2015; Cattaneo, Oggenfuss, and Wolter 2017; Bietenbeck, Piopiunik, and Wiederhold 2018) do not apply weights. I also check that my results are robust to this specification (in Table A9, column 3) and similar to those obtained using "senate weights".

[33]This coefficient is obtained dividing the coefficient in column 2 (4.37) by the SD of the first plausible values of all science subjects (118.56).

uncertainty about the process through which student test scores are computed, I use all five plausible values for each science subject.[34] The results (column 3) show that the effect of teacher subject-specific qualifications is robust to using all five plausible values and virtually identical to those obtained using only the first plausible value, and the standard error is roughly 10% larger. Finally, I address the issue of sampling variance typical of large-scale assessment such as TIMSS. To estimate standard errors that consider its multistage cluster sampling design, TIMSS suggests using the Jackknife Repeated Replication (JRR) technique.[35] Again, results in column 4 are robust to this specification, with the JRR technique inflating the standard errors by a further 10% with respect to column 3. I also replicate the main results clustering standard errors at different levels, namely at the school, student, or teacher level. Results (not shown) are robust to these specifications.

Last, I check the robustness of the results by dropping all observations for which teacher subject-specific qualifications is missing (11.8% of the sample). Results are also robust to this specification and virtually identical to those obtained in the main specification (teacher subject-specific qualifications coefficient = .034, *p*-value < .01).

5. Mediation Analysis

Having shown that teacher subject-specific qualifications increase student science test scores, I now explore a possible mediator through which this effect materializes. I focus on the share of topics that teachers feel confident to teach described in Section 2. Thanks to the increased subject knowledge that teachers acquire through a subject-specific qualification, teachers might feel more confident to teach topics in subjects in which they hold such qualification. A more confident teacher could be more effective in teaching a certain subject. Thus, the increased confidence in teaching certain

[34] I touched upon this point in Section 2. It has been generally acknowledged that the use of single plausible values does not make a substantial difference in large samples (Jerrim et al. 2017). However, my study slightly deviates from the cases discussed in the literature as the test scores for each science subject that I use are based on a limited number of questions (between 12 and 18), thus making the issue potentially relevant.

[35] Interested readers may find more detail about this technique and its application to the TIMSS data in Mullis and Martin (2013). In a nutshell, the JRR technique consists of subdividing the sample into clusters of sampling units (e.g., schools) and repeatedly replicating the statistics of interest by modifying the weight given to the sampling units within the cluster.

topics is a possible channel through which teacher subject-specific qualifications affect student test scores. To substantiate this hypothesis, I perform a mediation analysis in the spirit of Heckman, Pinto, and Savelyev (2013) and Heckman and Pinto (2015), following recent empirical implementations (Kosse et al. 2020; Resnjanskij et al. 2021; Hermes et al. 2021).

Variables must satisfy two conditions to act as mediators: they must be significantly affected by the independent variable of interest (specifically, teacher subject-specific qualifications) and be related to the outcome (student test scores). To test the first condition, I estimate the model described in Eq. (2) with the mediator as the dependent variable instead of student test scores. Results in Table A11 (Panel B) suggest that teachers with subject-specific qualifications are significantly more confident to teach topics that belong to the subject in which they hold a major. The result confirms that the mediator is significantly affected by teacher subject-specific qualifications. Looking at the magnitude of the coefficient, teacher subject-specific qualifications seem to have a large impact on the share of topics that teacher feel confident to teach, equivalent to 14.2 percentage points (or 39% SD).

To test the second condition, I include the mediator on the right-hand side of the baseline model of Eq. (2). Results are reported in Table A11 (Panel A). First, I report the impact of teacher subject-specific qualifications excluding the mediator (column 1) and then with the mediator (column 2). The mediator is significantly related to the outcome. As expected, the magnitude of the impact of subject-specific qualifications on student test scores decreases when the mediator is included, as the mediator captures part of the impact.

Finally, I compute the share of the effect of teacher subject-specific qualifications that can be attributed to the mediator.[36] As graphically shown in Figure 2, 20% of the effect of teacher subject-specific qualifications on student test scores is explained by teachers being more confident to teach topics that belong to subject in which they hold a major, while the remaining part is due to unobserved factors. Such

[36]The share is obtained by multiplying the coefficient of the coefficient of the impact of the independent variable on the mediator (.142, reported in Table A11, Panel B) by the association between the mediator and the outcome of interest (.05, report in Table A11, column 2, Panel A) and dividing by the impact of the independent variable on the outcome (.035, reported in Table A11, column 1, Panel A).

factors might be, for example, increased subject or pedagogical knowledge acquired through teacher subject-specific qualifications.

6. Conclusion

In this paper, I explore the effect of teacher subject-specific qualifications on student science test scores. I find that teachers with subject-specific qualifications raise student science test scores in the subjects in which teachers hold a major by 3.5% SD. The effect is robust to a variety of specifications and across different groups. The effect is larger for female students, especially when they are taught by female teachers, and for students from more disadvantaged backgrounds. Further, I find that the effect of teacher subject-specific qualifications is stronger in lower-performing countries. The mediation analysis reveals that 20% of the effect can be explained by the fact that teachers with subject-specific qualifications feel more confident to teach topics that belong to the subject in which they hold a major.

These findings are important for three reasons. First, I provide evidence of the importance of teacher subject-specific qualifications for student test scores in a broad set of countries. This finding adds to the existing literature on teacher subject-specific qualifications, which has focused almost exclusively on the US. Second, I shed light on an understudied yet important subject, science, for which existing evidence is mixed. Third, I exploit the richness and international nature of the data to provide further insights into the contexts and countries where subject-specific qualifications may have the greatest impact.

In terms of policy implications, countries should promote the acquisition of subject-specific qualifications, especially for science teachers in secondary schools. For example, countries could raise the standards required to become science teachers. This appears to be even more important for female students, for disadvantaged students and for lower-performing countries. Such policies could therefore increase both equity and efficiency in education systems worldwide. It is unclear whether students would also benefit from a further division of labor where teachers would only teach subjects in which they hold a major. Previous findings on such division of labor in elementary schools for math and reading are not encouraging (Fryer 2018), although findings for science are more promising (Bastian and Fortner 2020), thus calling for more research on this topic.

References

Bastian, Kevin C., and C. Kevin Fortner. 2020. "Is Less More? Subject-Area Specialization and Outcomes in Elementary Schools." *Education Finance and Policy* 15 (2): 357–82. https://doi.org/10.1162/edfp_a_00278.

Bietenbeck, Jan. 2014. "Teaching Practices and Cognitive Skills." *Labour Economics* 30 (3): 143–53. https://doi.org/10.1016/j.labeco.2014.03.002.

Bietenbeck, Jan, and Matthew Collins. "New Evidence on the Importance of Instruction Time for Student Achievement on International Assessments." *Journal of Applied Econometrics* (forthcoming).

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.

Blömeke, Sigrid, Gabriele Kaiser, and Rainer Lehmann, eds. 2010. *TEDS–M 2008: Professionelle Kompetenz Und Lerngelegenheiten Angehender Mathematiklehrkräfte Für Die Sekundarstufe I Im Internationalen Vergleich [Cross-National Comparison of the Professional Competency of and Learning Opportunities for Future Secondary School Teachers of Mathematics].* Münster: Waxmann.

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.

Buddin, Richard, and Gema Zamarro. 2009. "Teacher Qualifications and Student Achievement in Urban Elementary Schools." *Journal of Urban Economics* 66 (2): 103–15. https://doi.org/10.1016/j.jue.2009.05.001.

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.

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/.

Cattaneo, Maria A., Chantal Oggenfuss, and Stefan C. Wolter. 2017. “The More, the Better? The Impact of Instructional Time on Student Performance.” *Education Economics* 25 (5): 433–45. https://doi.org/10.1080/09645292.2017.1315055.

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.

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.” *J. 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.

European Commission: DG Employment Social Affairs and Inclusion. 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.

Fryer, Roland G. 2018. "The "Pupil" Factory: Specialization and the Production of Human Capital in Schools." *American Economic Review* 108 (3): 616–56. https://doi.org/10.1257/aer.20161495.

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." *J. 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.

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 Ludger Woessmann. 2015. *Universal Basic Skills:* OECD.

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.

Heckman, James, and Rodrigo Pinto. 2015. "Econometric Mediation Analyses: Identifying the Sources of Treatment Effects from Experimentally Estimated Production Technologies with Unmeasured and Mismeasured Inputs." *Econometric reviews* 34 (1-2): 6–31. https://doi.org/10.1080/07474938.2014.944466.

Heckman, James, Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review* 103 (6): 2052–86. https://doi.org/10.1257/aer.103.6.2052.

Hermes, Henning, Philipp Lergetporer, Frauke Peter, and Simon Wiederhold. 2021. "Behavioral Barriers and the Socioeconomic Gap in Child Care Enrollment.".

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.

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.

Kosse, Fabian, Thomas Deckers, Pia Pinger, Hannah Schildberg-Hörisch, and Armin Falk. 2020. “The Formation of Prosociality: Causal Evidence on the Role of Social Environment.” *Journal of Political Economy* 128 (2): 434–67. https://doi.org/10.1086/704386.

Kukla-Acevedo, Sharon. 2009. “Do Teacher Characteristics Matter? New Results on the Effects of Teacher Preparation on Student Achievement.” *Economics of Education Review* 28 (1): 49–57. https://doi.org/10.1016/j.econedurev.2007.10.007.

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).

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.

Lim, Jaegeum, and Jonathan Meer. 2017. “The Impact of Teacher–Student Gender Matches.” *J. Human Resources* 52 (4): 979–97. https://doi.org/10.3368/jhr.52.4.1215-7585R1.

Martin, M. O., I. V. S. Mullis, P. Foy, and M. Hooper. 2016. *TIMSS 2015 International Results in Science:* Boston College, TIMSS & PIRLS International Study Center. http://timssandpirls.bc.edu/timss2015/international-results/.

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.

Nagler, Markus, Marc Piopiunik, and Martin R. West. 2020. "Weak Markets, Strong Teachers: Recession at Career Start and Teacher Effectiveness." *Journal of Labor Economics* 38 (2): 453–500. https://doi.org/10.1086/705883.

OECD. 2016. *PISA 2015 Results (Volume I):* OECD.

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

OECD. 2022. *Education at a Glance 2022.* PISA, OECD Publishing, Paris: OECD.

Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics* 37 (2): 187–204. https://doi.org/10.1080/07350015.2016.1227711.

Paredes, Valentina. 2014. "A Teacher Like Me or a Student Like Me? Role Model Versus Teacher Bias Effect." *Economics of Education Review* 39 (2): 38–49. https://doi.org/10.1016/j.econedurev.2013.12.001.

President's Council of the Advisors on Science and Technology. 2012. "Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics." *Tech. rep.*

Resnjanskij, Sven, Jens Ruhose, Simon Wiederhold, and Ludger Woessmann. 2021. "Can Mentoring Alleviate Family Disadvantage in Adolscence? A Field Experiment to Improve Labor-Market Prospects." *CESifo Working Paper*.

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.

Rivkin, Steven G., and Jeffrey C. Schiman. 2015. "Instruction Time, Classroom Quality, and Academic Achievement." *The Economic Journal* 125 (588): F425-F448. https://doi.org/10.1111/ecoj.12315.

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

Sansone, Dario. 2017. "Why Does Teacher Gender Matter?" *Economics of Education Review* 61 (6453): 9–18. https://doi.org/10.1016/j.econedurev.2017.09.004.

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.

Tatto, Maria Teresa, Ray Peck, John Schwille, Kiril Bankov, Sharon L. Senk, Michael Rodriguez, Lawrence Ingvarson, Mark Reckase, and Glenn Rowley. 2012. *Policy, Practice, and Readiness to Teach Primary and Secondary Mathematics in 17 Countries: Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-M).* Amsterdam.

UNESCO. 2021. “World Development Indicators.” Accessed November 02, 2021. https://databank.worldbank.org/reports.aspx?source=2&series=SE.SEC.DURS&country=.

United Nations. 2014. *World Economic Situation and Prospects 2014 - Country Classification.* New York. Accessed November 02, 2021. https://www.un.org/en/development/desa/policy/wesp/wesp_current/2014wesp_country_classification.pdf.

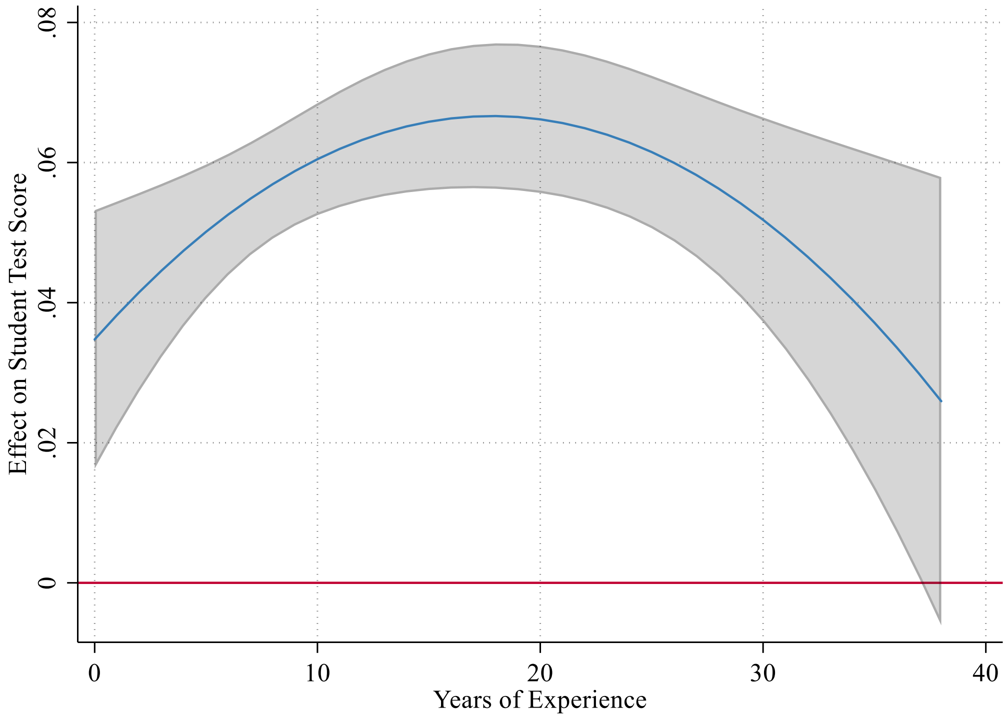
Wayne, Andrew J., and Peter Youngs. 2003. “Teacher Characteristics and Student Achievement Gains: A Review.” *Review of Educational Research* 73 (1): 89–122. https://doi.org/10.3102/00346543073001089.

Wedel, Katharina. 2021. “Instruction Time and Student Achievement: The Moderating Role of Teacher Qualifications.” *Economics of Education Review* 85 (27): 102183. https://doi.org/10.1016/j.econedurev.2021.102183.

World Bank. 2021. “World Development Indicators.” Accessed November 02, 2021. https://databank.worldbank.org/reports.aspx?source=2&series=NY.GNP.PCAP.KD&country=.

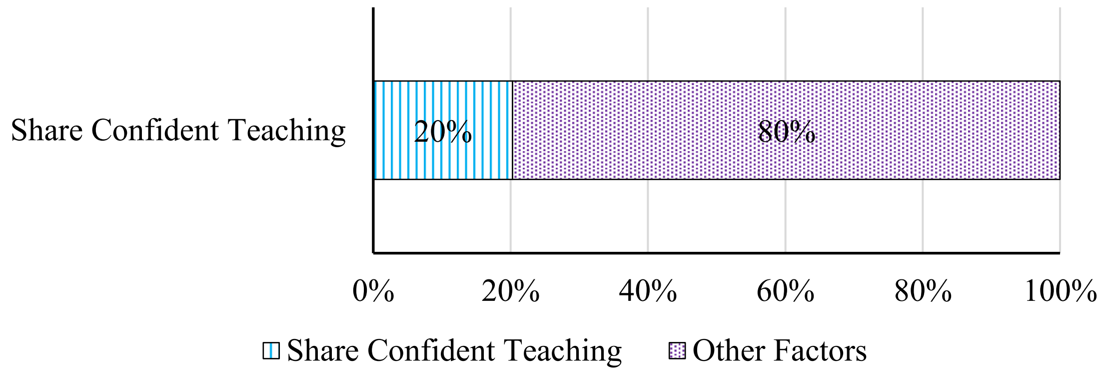
List of Figures

Figure 1: Effect of Teacher Subject-Specific Qualifications - Interaction with Teacher Experience



Note: The figure depicts the marginal effect of teacher subject-specific qualifications on student test scores along the domain of teacher experience with 95% confidence intervals. Estimates have been obtained by interacting teacher subject-specific qualifications with teacher experience in Eq. (2) and are reported in Table 3 in column 6.

Figure 2: Share of the Effect of Teacher Subject-Specific Qualifications Attributed to the Mediator



Note: The figure depicts the share of the effect of teacher subject-specific qualifications on student test scores that can be attributed to the mediator. The estimates to compute such share can be found in Table A11.

List of Tables

Table 1: Descriptive Statistics

	Mean (1)	SD (2)	Min-Max (3)
Teacher Subject-Specific Qualifications			
Biology	0.42	(0.47)	0.0-1.0
Chemistry	0.36	(0.46)	0.0-1.0
Physics	0.31	(0.44)	0.0-1.0
Earth Science	0.20	(0.37)	0.0-1.0
Teacher Variables			
N. of Subject-Specific Qualifications	1.24	(1.13)	0.0-4.0
At Least One Subject-Specific Qualification	0.73	(0.44)	0.0-1.0
Bachelors' Teachers	0.62	(0.49)	0.0-1.0
Masters' Teachers	0.22	(0.40)	0.0-1.0
Experience (y)	14.54	(9.26)	0.0-38.0
Any Major in Education	0.61	(0.47)	0.0-1.0
Female Teacher	0.58	(0.48)	0.0-1.0
Teaching time per week (hours)	5.65	(1.00)	3.0-10.0
Share Topics Confident to Teach	0.54	(0.37)	0.0-1.0
Student Variables			
Female Student	0.50	(0.50)	0.0-1.0
Student SES Indicator	10.04	(1.93)	4.2-13.9
Speak Language of Test at Home	0.79	(0.41)	0.0-1.0
Born in Country	0.95	(0.21)	0.0-1.0
# Observations		897,760	
# Students		224,454	
# Teachers		11,243	
# Countries		30	

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 Subject-Specific Qualifications* panel, I report the average number of students taught by teachers with a subject-specific qualification, separately for each science subject. In the *Teacher Variables* panel, I report the share of students taught by teachers who hold at least one subject-specific qualifications (i.e., at least one major in either biology, chemistry, physics, or earth science); the number of subject-specific qualifications refers to the number of science majors that teachers have. I also report the share of students taught by teachers who hold a Bachelors' degree, a Masters' degree, the years of experience of teachers, the share of teachers who hold any major in education (i.e., either in education, education-science or education math). The teaching time per week is the overall weekly instruction time in science reported by the teachers. The share of topics that teachers feel confident to teach is calculated within each subject as the share of topics that teachers feel very confident to teach. In the *Student Variables* panel, I report the student gender, the student SES indicator provided by TIMSS, which is a comprehensive measure of the socioeconomic status of the students, and 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). Speak language of test at home is a dummy variable that takes value "one" if a student speaks the language of the test always or almost always at home and "zero" otherwise. Born in country is a dummy variable that takes value "one" if a student is born in the country where the test is administered. I also report the total number of observations, the number of distinct students, teachers, and countries. As each student is observed four times (one for each subject), the total number of observations is equal to the number of distinct students multiplied by four.

Table 2: Effect of Teacher Subject-Specific Qualifications on Student Test Scores

	(1)	(2)	(3)	(4)
Teacher Subject-Specific Qualifications	0.033**	0.036***	0.035***	0.035***
	(0.016)	*(0.011)*	*(0.011)*	*(0.004)*
Subject FE	YES	YES	YES	YES
Country FE	YES	YES	YES	NO
Student, School Controls	NO	YES	YES	NO
Teacher Controls	NO	NO	YES	NO
Student, Teacher FE	NO	NO	NO	YES
Observations	897,760	897,760	897,760	897,760
R-squared	0.33	0.48	0.48	0.94

Note: The table reports OLS estimation using a set of controls (column 1,2,3) and student and teacher fixed effects (column 4). The outcome of interest is the standardized subject-specific (biology, chemistry, physics, and earth science) test score. Test scores have been standardized within each subject. The explanatory variable is teacher subject-specific qualifications. An observation corresponds to a student-subject combination. All regressions include weights, subject fixed effects, and an imputation dummy for teacher subject-specific qualifications. Student controls include: student SES, gender, language spoken at home, mother's immigrant status, father's immigrant status, student's immigrant status, student's education expectations. School and class controls include class size, share of students with language difficulties, share of economically disadvantaged students, indicator for shortage of resources for science instruction, school discipline problems, school location, school emphasis on academic success. Teacher controls include teacher experience, gender, level of education, major in education. Standard errors (in parentheses) have been clustered at the classroom level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3: Heterogenous Effect of Teacher Subject-Specific Qualifications on Student Test Scores – Student and Teacher Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
Teacher Subject-Specific Qualifications	0.005	0.031***	0.078***	0.034***	0.025***	0.013
	(0.005)	*(0.006)*	*(0.017)*	*(0.005)*	*(0.005)*	*(0.010)*
× F. Student	0.059***					
	(0.006)					
× F. Teacher		0.006				
		(0.008)				
× Student SES Indicator			-0.004**			
			(0.002)			
× Teacher holds Masters' Degree				0.004		
				(0.008)		
× Teacher holds Major in Ed.					0.020***	
					(0.008)	
× Teacher Experience						0.004***
						(0.001)
× Teacher Experience^2 (× 100)						-0.010***
						(0.004)
Subject, Student, Teacher FE	YES	YES	YES	YES	YES	YES
Observations	897,760	897,760	897,760	897,760	897,760	897,760

Note: The table reports OLS estimation using subject, student and teacher fixed effects. The outcome of interest is the standardized subject-specific (biology, chemistry, physics, and earth science) test score. Test scores have been standardized within each subject. The explanatory variable is teacher subject-specific qualifications. An observation corresponds to a student-subject combination. All regressions include weights and an imputation dummy for teacher subject-specific qualifications. I include an interaction between teacher subject-specific qualifications and student gender in column 1, and teacher gender in column 2. In column 3 I include an interaction with the student SES indicator. In column 4 and 5 I include an interaction for whether the teacher holds a Masters' degree or major in education, respectively. In column 6, I include an interaction with teacher years of experience and years of experience squared multiplied by 100. Standard errors (in parentheses) have been clustered at the classroom level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 4: Effect of Teacher Subject-Specific Qualifications on Student Test Scores – OECD Countries

	(1)	(2)	(3)	(4)
Teacher Subject-Specific Qualifications	0.052***	0.044***	0.043***	0.028***
	(0.020)	*(0.012)*	*(0.012)*	*(0.004)*
Subject FE	YES	YES	YES	YES
Country FE	YES	YES	YES	NO
Student, School Controls	NO	YES	YES	NO
Teacher Controls	NO	NO	YES	NO
Student, Teacher FE	NO	NO	NO	YES
Observations	349,244	349,244	349,244	349,244
R-squared	0.09	0.31	0.31	0.92

Note: The table reports OLS estimation using a set of controls (column 1,2,3) and student and teacher fixed effects (column 4) for OECD countries only (for the list of OECD countries, see Table A3). The outcome of interest is the standardized subject-specific (biology, chemistry, physics, and earth science) test score. Test scores have been standardized within each subject. The explanatory variable is teacher subject-specific qualifications. An observation corresponds to a student-subject combination. All regressions include weights, subject fixed effects, and an imputation dummy for teacher subject-specific qualifications. Student controls include: student SES, gender, language spoken at home, mother's immigrant status, father's immigrant status, student's immigrant status, student's education expectations. School and class controls include class size, share of students with language difficulties, share of economically disadvantaged students, indicator for shortage of resources for science instruction, school discipline problems, school location, school emphasis on academic success. Teacher controls include teacher experience, gender, level of education, major in education. Standard errors (in parentheses) have been clustered at the classroom level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5: Heterogenous Effect of Teacher Subject-Specific Qualifications on Student Test Scores – Country Characteristics

	(1)	(2)	(3)	(4)
Teacher Subject-Specific Qualifications	0.039***	0.041***	0.048***	0.033***
	(0.008)	*(0.006)*	*(0.008)*	*(0.005)*
× OECD Country	-0.007			
	(0.009)			
× Developed Country		-0.015*		
		(0.008)		
× High-Performing Country			-0.023***	
			(0.009)	
× High-GNI p.p. Country				0.007
				(0.008)
Subject, Student, Teacher FE	YES	YES	YES	YES
Observations	897,760	897,760	897,760	897,760

Note: The table reports OLS estimation using subject, student and teacher fixed effects. The outcome of interest is the standardized subject-specific (biology, chemistry, physics, and earth science) test score. Test scores have been standardized within each subject. The explanatory variable is teacher subject-specific qualifications. An observation corresponds to a student-subject combination. All regressions include weights and an imputation dummy for teacher subject-specific qualifications. I include an interaction between teacher subject-specific qualifications and an indicator for whether a country belongs to the OECD (column 1), whether a country is a developed country according to the WESP classification (column 2), whether a country average science score is above the median of the science test scores in the sample (column 3) and whether a country GNI per capita in 2015 is above the median GNI per capita of the countries in the sample (column 4). Standard errors (in parentheses) have been clustered at the classroom level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Appendix A – Additional Tables

Table A1: List of Science Topics Covered in TIMSS 2015

Panel A: Topics
Biology
a) Differences among major taxonomic groups of organisms (plants, animals, fungi, mammals, birds, reptiles, fish, amphibians)
b) Major organs and organ systems in humans and other organisms (structure/function, life processes that maintain stable bodily conditions)
c) Cells, their structure and functions, including respiration and photosynthesis as cellular processes
d) Life cycles, sexual reproduction, and heredity (passing on of traits, inherited versus acquired/learned characteristics)
e) Role of variation and adaptation in survival/extinction of species in a changing environment (including fossil evidence for changes in life on Earth over time)
f) Interdependence of populations of organisms in an ecosystem (e.g., energy flow, food webs, competition, predation) and factors affecting population size in an ecosystem
g) Human health (causes of infectious diseases, methods of infection, prevention, immunity) and the importance of diet and exercise in maintaining health
Chemistry
a) Classification, composition, and particulate structure of matter (elements, compounds, mixtures, molecules, atoms, protons, neutrons, electrons)
b) Physical and chemical properties of matter
c) Mixtures and solutions (solvent, solute, concentration/dilution, effect of temperature on solubility)
d) Properties and uses of common acids and bases
e) Chemical change (transformation of reactants, evidence of chemical change, conservation of matter, common oxidation reactions – combustion, rusting, tarnishing)
f) The role of electrons in chemical bonds
Physics
a) Physical states and changes in matter (explanations of properties in terms of movement and distance between particles; phase change, thermal expansion, and changes in volume and/or pressure)
b) Energy forms, transformations, heat, and temperature
c) Basic properties/behaviors of light (reflection, refraction, light and color, simple ray diagrams) and sound (transmission through media, loudness, pitch, amplitude, frequency)
d) Electric circuits (flow of current; types of circuits - parallel/series) and properties and uses of permanent magnets and electromagnets
e) Forces and motion (types of forces, basic description of motion, effects of density and pressure)
Earth Science
a) Earth's structure and physical features (Earth's crust, mantle, and core; composition and relative distribution of water, and composition of air)
b) Earth's processes, cycles, and history (rock cycle; water cycle; weather versus climate; major geological events; formation of fossils and fossil fuels)
c) Earth's resources, their use and conservation (e.g., renewable/nonrenewable resources, human use of land/soil, water resources)
d) Earth in the solar system and the universe (phenomena on Earth - day/night, tides, phases of moon, eclipses, seasons; physical features of Earth compared to other bodies)
(continues)

Table A1
(continued)

Panel B: Answer choices for each topic
Choose the response that best describes when the students in this class have been taught each topic
Mostly taught before this year
Mostly taught this year
Not yet taught or just introduced
How well prepared do you feel you are to teach the following science topics?
Not applicable
Very well prepared
Somewhat prepared
Not well prepared

Note: The list of topics comes from the TIMSS 2015 8th-grade science teacher questionnaire and comprises 7 topics in Biology, 6 in chemistry, 5 in physics and 4 in earth science (Panel A). For each topic, teachers are asked when students have been taught a topic and how well they feel prepared to teach it (Panel B).

Table A2: Descriptive Statistics – Number of Subject-Specific Qualifications by Major in Education

	Any Major in Education		
N. of Subject-Specific Qualifications	No (1)	Yes (2)	Total (3)
Zero	4.8	24.9	29.7
One	25.2	15.1	40.2
Two	6	8.4	14.4
Three	2.1	5.8	7.9
Four	1.1	6.7	7.8
Total	39.2	60.8	100

Note: The table reports the weighted share of students taught by teachers who hold zero, one, two, three or four subject-specific qualifications by whether they also hold any major in education (i.e., either major in education, education-science or education-mathematics).

Table A3: Descriptive Statistics by Country

	N. of Subject-Specific Qualifications	At Least One Subject-Specific Qualification	Within-Teacher Variation	OECD	Developed	High Perf.	High GNI	Science	# Observations
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Australia	1.60	0.86	0.82	Yes	Yes	Yes	Yes	511.6	39,404
Bahrain	1.78	0.95	0.86	No	No	No	Yes	460.3	18,512
Botswana	1.01	0.67	0.67	No	No	No	No	390.4	23,232
Canada	0.76	0.53	0.39	Yes	Yes	Yes	Yes	526.2	35,008
Canada (Ontario)	0.40	0.41	0.26	Yes	Yes	Yes	Yes	524.1	18,080
Canada (Quebec)	1.12	0.70	0.67	Yes	Yes	Yes	Yes	529.5	15,800
Chile	1.19	0.66	0.61	Yes	No	No	No	451.5	17,972
Chinese Taipei	1.17	0.93	0.90	No	No	Yes	No	567.4	21,832
Egypt	1.74	0.77	0.61	No	No	No	No	370.2	31,288
England	1.87	0.97	0.92	Yes	Yes	Yes	No	531.4	14,776
Hong Kong SAR	0.92	0.79	0.79	No	No	Yes	Yes	544.4	16,352
Iran	0.73	0.30	0.18	No	No	No	No	456.4	24,520
Ireland	1.54	0.94	0.93	Yes	Yes	Yes	Yes	529.4	18,808
Israel	2.14	0.92	0.82	Yes	No	Yes	No	505.0	16,716
Italy	1.95	0.95	0.78	Yes	Yes	Yes	No	498.1	17,924
Japan	1.15	0.85	0.78	Yes	Yes	Yes	No	567.6	16,240
Jordan	1.12	0.83	0.77	No	No	No	No	426.1	31,460
Kuwait	1.79	0.90	0.74	No	No	No	Yes	409.8	18,012
Malaysia	1.04	0.76	0.75	No	No	Yes	No	470.8	38,904
Morocco	1.36	0.97	0.97	No	No	No	No	392.8	51,840
New Zealand	1.39	0.92	0.91	Yes	Yes	Yes	No	512.8	32,568
Norway	0.77	0.48	0.38	Yes	Yes	Yes	Yes	509.3	18,364
Norway (8th Grade)	0.83	0.53	0.45	Yes	Yes	No	Yes	488.8	18,724

(continues)

Table A3
(continued)

	N. of Subject-Specific Qualifications	At Least One Subject-Specific Qualification	Within-Teacher Variation	OECD	Developed	High Perf.	High GNI	Science	# Observations
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Oman	1.82	0.96	0.83	No	No	No	No	454.1	35,532
Qatar	1.82	0.93	0.81	No	No	No	Yes	448.6	20,548
Saudi Arabia	1.14	0.83	0.77	No	No	No	Yes	396.2	15,036
Singapore	1.66	0.95	0.93	No	No	Yes	Yes	596.8	24,464
South Africa	1.58	0.83	0.80	No	No	No	No	356.6	50,056
South Korea	1.00	0.93	0.91	Yes	No	Yes	No	553.9	15,208
Thailand	0.98	0.61	0.52	No	No	No	No	455.8	25,928
Turkey	1.37	0.57	0.44	Yes	No	No	No	492.9	24,316
United Arab Emirates	1.19	0.86	0.83	No	No	No	Yes	470.4	62,716
United Arab Emirates (Abu Dhabi)	1.19	0.84	0.80	No	No	No	Yes	453.3	18,868
United Arab Emirates (Dubai)	1.31	0.89	0.86	No	No	No	Yes	517.4	19,416
United States	0.95	0.71	0.68	Yes	Yes	Yes	Yes	531.6	29,336
All Countries	1.24	0.73	0.66	16	12	17	17	478.3	897,760

Note: The table reports weighted statistics and indicators for each national entity included in the sample. In column 1, I report the average number of teachers subject-specific qualifications. In column 2, the share of students taught by teachers who hold at least one subject-specific qualification is reported (i.e., at least one major in either biology, chemistry, physics, or earth science). In column 3, I report the share of students whose teachers differ in their subject-specific qualifications across subjects (i.e., students who are taught by teachers who have one, two or three subject-specific qualifications). In columns 4-7, I report country indicators for whether a country belongs to the OECD (column 4), is a developed country according to the WESP classification (column 5), is above the median science test score of the countries in the sample (column 6), or is above the median GNI in 2015 of the countries in the sample (column 7). The average science test score is reported in column 8 and the number of observations in column 9. In the last row, the weighted average of column 1, 2, 3, and 8 is reported, while the sum of the indicators for column 4-7 and 9 is reported.

Table A4: TIMSS 2011 with Instruction Time

	(1)	(2)	(3)	(4)	(5)
Teacher Subject-Specific Qualifications	0.045***	0.045***	0.043***	0.026***	0.022***
	(0.014)	*(0.011)*	*(0.011)*	*(0.004)*	*(0.004)*
Subject FE	YES	YES	YES	YES	YES
Country FE	YES	YES	YES	NO	NO
Student, School Controls	NO	YES	YES	NO	NO
Teacher Controls	NO	NO	YES	NO	NO
Instruction Time	NO	NO	YES	NO	YES
Student and Teacher FE	NO	NO	NO	YES	YES
Observations	867,012	867,012	867,012	867,012	867,012
R-squared	0.37	0.49	0.49	0.94	0.94

Note: The table reports OLS estimation using a set of controls (column 1,2,3) and student and teacher fixed effects (column 4) using TIMSS 2011 data. The outcome of interest is the standardized subject-specific (biology, chemistry, physics, and earth science) test score. Test scores have been standardized within each subject. The explanatory variable is teacher subject-specific qualifications. An observation corresponds to a student-subject combination. All regressions include weights, subject fixed effects, and an imputation dummy for teacher subject-specific qualifications. Student controls include: student SES, gender, language spoken at home, mother's immigrant status, father's immigrant status, student's immigrant status, student's education expectations. School and class controls include class size, share of students with language difficulties, share of economically disadvantaged students, indicator for shortage of resources for science instruction, school discipline problems, school location, school emphasis on academic success. Teacher controls include teacher experience, gender, level of education, major in education. I include instruction time as a control in column 3 and 5. Standard errors (in parentheses) have been clustered at the classroom level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A5: Sample of Schools Located in Scarcely Populated Areas

	< 30k	< 15k	Small Town/Village
	(1)	(2)	(3)
Teacher Subject-Specific Qualifications	0.043***	0.055***	0.038***
	(0.008)	*(0.011)*	*(0.009)*
Subject, Student, Teacher FE	YES	YES	YES
Observations	320,556	210,072	227,956
R-squared	0.94	0.94	0.94

Note: The table reports OLS estimation using subject, student and teacher fixed effects. The outcome of interest is the standardized subject-specific (biology, chemistry, physics, and earth science) test score. Test scores have been standardized within each subject. The explanatory variable is teacher subject-specific qualifications. An observation corresponds to a student-subject combination. All regressions include weights and an imputation dummy for teacher subject-specific qualifications. In column 1, I report the result for schools located in areas with less than 30,000 inhabitants, in column 2 in areas with less than 15,000 inhabitants, and in column 3 for schools located in small towns, villages or rural areas. Standard errors (in parentheses) have been clustered at the classroom level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A6: Analysis of Unobservable Selection and Coefficient Stability following Oster (2019)

	(1)	(2)
Teacher Subject-Specific Qualifications	0.033**	0.035***
	(0.016)	*(0.004)*
Subject FE	YES	YES
Country FE	YES	NO
Student and Teacher FE	NO	YES
Observations	897,760	897,760
R-squared	0.33	0.94
Oster (2019) diagnostics		
Bound β^* for $\delta = 1$	0.035	
δ to match $\beta = 0$	19.51	

Note: The table reports OLS estimation using a country (column 1) and student and teacher fixed effects (column 2). The outcome of interest is the standardized subject-specific (biology, chemistry, physics, and earth science) test score. Test scores have been standardized within each subject. The explanatory variable is teacher subject-specific qualifications. An observation corresponds to a student-subject combination. All regressions include weights, subject fixed effects, and an imputation dummy for teacher subject-specific qualifications. The table also reports Oster (2019) diagnostics computed with $R_{max} = 1$ and $\delta = 1$ using TIMSS 2015. Standard errors (in parentheses) have been clustered at the classroom level. *** p<0.01, ** p<0.05, * p<0.1

Table A7: Leave One Subject Out

	Full Sample	Excluding Biology	Excluding Physics	Excluding Chemistry	Excluding Earth Science
	(1)	(2)	(3)	(4)	(5)
Teacher Subject-Specific Qualifications	0.035***	0.034***	0.038***	0.036***	0.033***
	(0.004)	*(0.005)*	*(0.005)*	*(0.004)*	*(0.005)*
Subject, Student, Teacher FE	YES	YES	YES	YES	YES
Observations	897,760	673,286	673,326	673,326	673,286

Note: The table reports OLS estimation using subject, student and teacher fixed effects. The outcome of interest is the standardized subject-specific (biology, chemistry, physics, and earth science) test score. Test scores have been standardized within each subject. The explanatory variable is teacher subject-specific qualifications. An observation corresponds to a student-subject combination. All regressions include weights and an imputation dummy for teacher subject-specific qualifications. In column 1, I report the result for the entire sample. I then replicate the results by excluding one science subject at a time, namely biology (column 1), physics (column 2), chemistry (column 3) and earth science (column 4). Standard errors (in parentheses) have been clustered at the classroom level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A8: Leave One Country Out

Excluded Country	Teacher Subject-Specific Qualifications	Std. Error	Observations
	(1)	(2)	(3)
Australia	0.035^{***}	*(0.004)*	858,356
Bahrain	0.035^{***}	*(0.004)*	879,248
Botswana	0.035^{***}	*(0.004)*	874,528
Canada	0.035^{***}	*(0.004)*	862,752
Canada (Ontario)	0.035^{***}	*(0.004)*	879,680
Canada (Quebec)	0.035^{***}	*(0.004)*	881,960
Chile	0.035^{***}	*(0.004)*	879,788
Chinese Taipei	0.035^{***}	*(0.004)*	875,928
Egypt	0.023^{***}	*(0.003)*	866,472
England	0.037^{***}	*(0.004)*	882,984
Hong Kong SAR	0.035^{***}	*(0.004)*	881,408
Iran	0.038^{***}	*(0.004)*	873,240
Ireland	0.035^{***}	*(0.004)*	878,952
Israel	0.035^{***}	*(0.004)*	881,044
Italy	0.035^{***}	*(0.004)*	879,836
Japan	0.039^{***}	*(0.005)*	881,520
Jordan	0.035^{***}	*(0.004)*	866,300
Kuwait	0.035^{***}	*(0.004)*	879,748
Malaysia	0.036^{***}	*(0.004)*	858,856
Morocco	0.037^{***}	*(0.004)*	845,920
New Zealand	0.035^{***}	*(0.004)*	865,192
Norway	0.035^{***}	*(0.004)*	879,396
Norway (8th Grade)	0.035^{***}	*(0.004)*	879,036
Oman	0.035^{***}	*(0.004)*	862,228
Qatar	0.035^{***}	*(0.004)*	877,212
Saudi Arabia	0.036^{***}	*(0.004)*	882,724
Singapore	0.034^{***}	*(0.004)*	873,296
South Africa	0.035^{***}	*(0.005)*	847,704
South Korea	0.037^{***}	*(0.004)*	882,552
Thailand	0.036^{***}	*(0.004)*	871,832
Turkey	0.032^{***}	*(0.004)*	873,444
United Arab Emirates	0.035^{***}	*(0.004)*	835,044
United Arab Emirates (Abu Dhabi)	0.035^{***}	*(0.004)*	878,892
United Arab Emirates (Dubai)	0.035^{***}	*(0.004)*	878,344
United States	0.024^{***}	*(0.004)*	868,424

Note: The table reports OLS estimation using subject, student and teacher fixed effects. The outcome of interest is the standardized subject-specific (biology, chemistry, physics, and earth science) test score. Test scores have been standardized within each subject. The explanatory variable of interest is teacher subject-specific qualifications. An observation corresponds to a student-subject combination. All regressions include weights and an imputation dummy for teacher subject-specific qualifications. In each row, I report the coefficient of teacher subject-specific qualifications obtained estimating Eq. (2) by dropping from the estimation sample the country indicated in each row; the corresponding estimated coefficient is reported in column 1, the standard error of the estimate in column 2 and the number of observations in column 3. Standard errors (in parentheses) have been clustered at the classroom level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A9: Different Weights

	Sampling Weights (1)	Senate Weights (2)	Without Weights (3)
Teacher Subject-Specific Qualifications	0.035***	0.022***	0.020***
	(0.004)	*(0.002)*	*(0.002)*
Subject, Student, Teacher FE	YES	YES	YES
Observations	897,760	897,760	897,760

Note: The table reports OLS estimation using subject, student and teacher fixed effects. The outcome of interest is the standardized subject-specific (biology, chemistry, physics, and earth science) test score. Test scores have been standardized within each subject. The explanatory variable is teacher subject-specific qualifications. An observation corresponds to a student-subject combination. All regressions include an imputation dummy for teacher subject-specific qualifications. In column 1, I report the result using the sampling weights. I use "senate weights", i.e., rescaled weights such that each country carries the same weight, in column 2 and no weights in column 3. Standard errors (in parentheses) have been clustered at the classroom level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A10: Plausible Values and JRR

	Std. Score	PV1	PV1-PV5	PV1-PV5 & JRR
	(1)	(2)	(3)	(4)
Teacher Subject-Specific Qualifications	0.035***	4.370***	4.343***	4.343***
	(0.004)	*(0.532)*	*(0.597)*	*(0.655)*
Subject, Student, Teacher FE	YES	YES	YES	YES
Observations	897,760	897,760	897,760	897,760

Note: The table reports OLS estimation using subject, student, and teacher fixed effects. The outcome of interest is the standardized subject-specific (biology, chemistry, physics, and earth science) test score (column 1), the first subject-specific plausible value (column 2) and all five subject-specific plausible values (column 3 and 4). In column 4, I perform the Jackknife Repeated Replication (JRR) method to account for the sampling variance. The explanatory variable is teacher subject-specific qualifications. An observation corresponds to a student-subject combination. All regressions include weights and an imputation dummy for teacher subject-specific qualifications. Standard errors (in parentheses) have been clustered at the classroom level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A11: Mediation Analysis

	(1)	(2)
Panel A: Effect of Mediator on Student Test Scores		
Teacher Subject-Specific Qualifications	0.035***	0.028***
	(0.004)	*(0.004)*
Share Topics Confident to Teach		0.050***
		(0.006)
Subject, Student, Teacher FE	YES	YES
Observations	897,760	897,760
R-squared	0.94	0.94
Panel B: Effect of Teacher Subject-Specific Qualifications on Mediator		
Teacher Subject-Specific Qualifications	0.142***	
	(0.009)	
Mean (SD) of Dep. Variables	0.54 (0.37)	
Subject, Student, Teacher FE	YES	
Observations	897,760	
R-squared	0.64	

Note: The table reports OLS estimation using subject, student and teacher fixed effects. In Panel A, the outcome of interest is the standardized subject-specific (biology, chemistry, physics, and earth science) test score. Test scores have been standardized within each subject. The explanatory variable is teacher subject-specific qualifications. In column 1, I report the effect of teacher subject-specific qualifications on student test scores. I then include the mediator, the share of topics a teacher feels confident to teach, in column 2. In Panel B, the outcome of interest is the subject-specific (biology, chemistry, physics, and earth science) share of topics that a teacher feels confident to teach. The explanatory variable is teacher subject-specific qualifications in a subject. In all regressions, an observation corresponds to a student-subject combination. All regressions include weights and an imputation dummy for the explanatory variables. Standard errors (in parentheses) have been clustered at the classroom level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$