SCHOOL OF
ECONOMICS AND
MANAGEMENT

Does Teacher Training Make Teachers More Effective?

Evidence from TIMSS.

by
Lukas Maschmann
&
Christina Schwarz

May 2021

Master's Programme in Economics

Supervisor: Jan Bietenbeck

Abstract

Several countries have recently increased their efforts to improve teaching quality by focusing their attention on pre- and in-service teacher training. This raises the question whether such training makes teachers actually more effective. Yet, the existing literature on teacher training is limited and inconclusive. Using cross-sectional data from the Trends in International Mathematics and Science Study (TIMSS), we follow a within-student fixed effects approach and show that in-service teacher training does significantly benefit student achievements. More specifically, teachers' participation in professional development increases students' standardized test scores by 0.031-0.040 of a standard deviation. Moreover, we identify a dose response pattern suggesting that increasing participation hours in such programs are associated with rising teacher effectiveness. In contrast, we do not find any evidence that pre-service teacher training has a significant impact on student performance.

Keywords: teacher training, student achievement, professional development, education, TIMSS

Table of Contents

1. Introduction

How can teachers become more effective? This question is widely discussed in the economic literature not least due to its policy relevant implications. In the past few years policymakers in different countries have introduced new policies in order to improve teaching quality and consequently student performance. In that context, some countries have tightened their entry requirements for teacher educational programs or teaching certifications (Mullis, Martin & Loveless, 2016). For example, France has raised the requirement to commence teacher qualification programs from a Bachelor's to a Master's degree in 2010. Focusing on teachers' initial educational training, Italy requires students to obtain 24 university credits in pedagogy since 2017 (Kelly et al., 2020). Generally, not only the degree obtained but also the acquisition of content-specific or pedagogical knowledge receives major attention from policymakers. Moreover, a large proportion of available resources for teachers' development tends to be invested in pre-service teacher training (OECD, 2018).

As teachers usually stay in their profession for the majority of their career, professional development is crucial to keep them up to date to changes in their field of teaching or new insights on pedagogical methods. Therefore, several countries have increased their efforts to support teachers by introducing extensive measures for formal on-the-job training in the past few years (Mullis, Martin & Loveless, 2016; OECD, 2020). For instance, in 2014 the Ministry of Education and Culture in Finland funded a new professional development program with the goal of further educating 50,000 trained teachers over two years (Mullis, Martin & Loveless, 2016). Although participation in pre- and in-service training is considered to increase teachers' confidence in their knowledge and skills (OECD, 2020), there is only little consensus whether those teacher credentials are effective at boosting student performance.

In this essay, we test whether teachers' subject-specific education studies at the university (pre-service teacher training) and participation in professional development programs (in-service teacher training) increase student test scores. For our empirical analysis we draw data from the Trends in International Mathematics and Science Study (TIMSS). This database contains mathematics and science test scores for 4th and 8th grade students conducted in an international assessment survey (Mullis et al., 2016a). Applying a within-student fixed effects approach, we focus on 8th grade students in the five Anglo-Saxon countries Australia, Canada, England, Ireland and New Zealand in 2015. Our main results suggest that pre-service teacher training does not have a significant effect on student achievement. However, we do find evidence that in-service teacher training significantly improves student test scores by 0.031-0.040 of a standard deviation. Our findings further show that only high-intense professional development programs with 6-15 or more hours of participation significantly enhance student performance. Moreover, the estimates follow a dose-response pattern indicating that increasing participation hours of in-service training are associated with rising teacher effectiveness.

One concern about our results are unobserved confounders such as teachers' motivation or self-selection processes. For instance, highly motivated teachers might participate in professional development programs more frequently. Likewise, students with high preferences for one subject might self-select themselves to well-trained teachers. We tackle these potential omitted variable biases in two ways. First, we follow Oster's (2019) approach on selection on observables and unobservables. Second, we include an indicator for students' participation in extra lessons outside the schools' instructional time. Addressing students' self-selection to specific teachers, this control variable models parental and student preferences for a subject. As our results survive both sensitivity checks, we conclude that omitted variable bias does not tend to drive our findings.

This essay is organized as follows. Section 2 reviews the literature on teachers' contribution to student achievement at schools, especially on the effects of pre- and in-service teacher training. Section 3 describes the data used in our empirical analysis as well as details on our sample selection. Section 4 presents our empirical strategy and Section 5 outlines our main results as well as heterogeneity analyses. Section 6 discusses the robustness checks and Section 7 concludes.

2. Literature Review

A well-known literature on teachers' contribution to student learning is conclusive that teachers' effectiveness does explain a non-negligible variation in student performance (Jackson, Rockoff & Staiger, 2014). Overall, this literature can be divided into two strands which use distinct empirical approaches.

The first strand estimates the overall contribution of teachers to students' test scores using longitudinal data. This method is known as the "value-added" approach and suggests that a teacher's value-added can represent a teacher's overall aptitude to increase students' human capital (Chetty, Friedman & Rockoff, 2014a; Jackson, Rockoff & Staiger, 2014). However, this teacher effect is unobserved and not associated with measures of inputs into the educational production function (Jackson, Rockoff & Staiger, 2014). Using least square regressions, the value-added framework estimates a teacher's relative productivity relying on test score variations across students linked to an identical teacher. Consequently, it circumvents the identification of specific traits being important for teachers' quality. (Hanushek & Rivkin, 2010; Jackson, Rockoff & Staiger, 2014). Furthermore, including students' prior test scores supports this method in capturing the causal impact of teachers on student test scores (Chetty, Friedman & Rockoff, 2014a).

The literature on teachers' value-added agrees on a non-negligible variation in teacher effectiveness (Hanushek & Rivkin, 2010; Jackson, Rockoff & Staiger, 2014). In that context, Chetty, Friedman and Rockoff (2014a) suggest that the value-added approach helps to identify which teachers indicate the greatest aptitude to increase students' performance. Moreover, teachers' value-added is not only limited to effects on student achievements at school but also indicates substantial impacts on long-term attainments (Chetty, Friedman & Rockoff, 2014b). Analyzing administrative data on students and teachers in grades 3-8 linked to selected tax record data in the U.S., Chetty, Friedman and Rockoff (2014b) find that students assigned to teachers with high value-added are more likely to attend college, have a steeper earnings trajectory in their 20s and live in a higher quality neighborhood as adults.

While the value-added approach can provide unbiased estimates of teachers' impact on student learning, it does not give insights about the characteristics to make some teachers more effective than others. However, this information is useful for policymakers and school principals when establishing certification requirements or conducting hiring processes. Both parties usually observe teacher characteristics such as education, teaching experience and training rather than teachers' value-added. Introducing professional requirements for teachers based on those observable traits can help to ensure a minimum level of teaching quality (Hanushek & Rivkin, 2006). Moreover, a solid knowledge of the effects of teacher characteristics on student achievement could help policymakers to re-assess the prevailing policies on teacher recruitment and existing incentives for teachers to acquire specific credentials, such as a higher degree (Clotfelter, Ladd & Vigdor, 2010).

Therefore, the second strand of literature focuses on the effects of teachers' observable characteristics on student outcomes, in particular student performance on standardized tests. The most commonly studied factors are education and teaching experience as those usually determine teachers' wage (Clotfelter, Ladd & Vigdor, 2010; Hanushek & Rivkin, 2006). Several studies suggest that experience matters as teachers accumulate their pedagogical skills and subject knowledge predominantly over time by on-the-job learning, especially in the first years of their profession (Clotfelter, Ladd & Vigdor, 2006; Harris & Sass, 2011; Rivkin, Hanushek & Kain, 2005). In terms of education, holding a master's degree does not seem to increase teacher effectiveness (Clotfelter, Ladd & Vigdor, 2007; Hanushek & Rivkin, 2006).

As part of this literature, there are studies that explore the impacts of pre-service teacher training on student performance. However, the literature on this pre-service pedagogical teacher instruction at universities is limited. Most studies tend to focus on the degree level or the type of teacher certification. Furthermore, available research often concentrates on the effects of either subject-specific knowledge (see, e.g., Metzler & Woessmann, 2012; Bietenbeck, Piopiunik & Wiederhold, 2018) or the pedagogical training (see, e.g., Campbell et al., 2014; Goldhaber & Brewer, 2000), but rarely combines both subjects of investigations. Additionally, even when using the same variables of interest, they are often derived differently across the studies. While some studies are focusing on the amount of subject-specific education coursework a teacher underwent at the university (see Boyd et al., 2009; Harris & Sass, 2011; Monk, 1994), others like Campbell et al. (2014) define pedagogical knowledge by incorporating their own criteria in the form of dedicated questionnaires to teachers. Consequently, the work on this matter is lacking a definite consensus both on the approach and on the results. Despite the different approaches, the studies rarely find significant effects. Depending on the econometric specification strategy, the effect of a major in education on student achievements can take every form, from either negative or positive to no significant impact at all (see Betts, Zau & Rice, 2003; Campbell et al., 2014; Goldhaber & Brewer, 2000). Furthermore, studies that conclude any sort of significant effects usually find them for math or reading rather than for science (Goldhaber & Brewer, 2000; Monk, 1994; Boyd et al., 2009). Keeping these outcomes in mind, it is noteworthy that apart from Boyd et al. (2009), Harris and Sass (2011) and Monk (1994) there is no current literature that has investigated the formal subject-specific pedagogical background of teachers. While Boyd et al. (2009) and Monk (1994) do find significant positive effects on additional math pedagogy courses at the university, Harris and Sass (2011) specifically do not find evidence on this matter for math but for reading. Moreover, the majority of studies in the existing literature is mainly focusing on the United States. Hereby, some of this research is working with samples containing only one city (Betts, Zau & Rice, 2003) or one federal state (Harris & Sass, 2011).

Another group of studies, also part of the second strand of literature, investigates the impact of in-service teacher training i.e., the effect of professional development programs on student achievement. Although teachers' participation in professional development is well recognized in many countries (Mullis et al., 2016b), there is no consensus of the actual effects of such formal training. Angrist and Lavy (2001) analyze the impact of in-service teacher training on students' mathematics and reading performance in elementary schools in Jerusalem. Using different econometric models such as differences-in-differences, regression analysis and matching, they find evidence for significant positive effects on the overall student achievement for both subjects. Complementing these findings, Bressoux, Kramarz and Prost (2009) use a quasi-experimental design to investigate primary schools in France. However, their results slightly contrast Angrist and Lavy (2001) as they only find significant effects on teachers' aptitude to enhance student performance for mathematics but not for reading. In contrast, Harris and Sass (2011) focus their attention not only on elementary, but also on middle and high school math as well as reading achievements. Relying on school administrative data from Florida, their results suggest that professional development is either associated with no significant or even negative impacts on teacher effectiveness, with exception for middle school math. Moreover, the quantity of professional development undertaken appears to show heterogeneous effects. In their evidence review on the relationship between in-service teacher training and student achievement, Yoon et al. (2007) identify that only studies considering more than 14 hours of professional development detect a positive and significant impact on the overall student performance.

However, the preceding evidence is confronted by Jacob and Lefgren (2004) as well as Garet et al. (2008) and Garet et al. (2011). Following a regression discontinuity design, Jacob and Lefgren (2004) analyze teacher effectiveness in context of professional development programs in probation elementary schools in Chicago. They report no discernable effects of additional training hours neither on math nor reading achievements. Similarly, in both of their studies, Garet et al. (2008) and Garet et al. (2011) exploit an experimental design and investigate the effect of professional development programs in the U.S. on student reading and math test scores, respectively. Although treated teachers show substantial knowledge gains in both cases, no significant enhancement of the overall student performance is evident.

Our work intends to attach to the rather limited and inconclusive literature on pre- and in-service teacher training by investigating the effects of teachers' credentials acquired before (pre-service university education) and after entering the profession (in-service teacher training). In terms of pre-service training, we explicitly focus on the subject-specific pedagogical knowledge the teacher has gained during her studies at a university, by following Harris and Sass (2011). Doing so, we investigate the combination of two prevailing suggestions in the literature, namely education majors influencing student performance and subject-specific studies posing a significant impact on this matter. We further explore the effects of formal on-the-job professional development as a second source of pedagogical and subject-area knowledge available to teachers. Additionally, and contrary to the vast majority of the prevailing literature, our analysis is explicitly dedicated to a larger sample of five Anglo-Saxon countries containing Australia, Canada, England, Ireland and New Zealand, rather than only to the USA. We exploit the fact that those countries share similar professional requirements for teachers, such as the prescribed educational path. Furthermore, those economies indicate mostly the same teaching language and a similar educational system (Mullis et al., 2016b).

3. Data

The following subsections present TIMSS data used in our empirical analysis as well as details on our sample selection. We further provide detailed summary statistics for teacher, class and subject characteristics employed as controls in our student fixed effect specifications.

3.1 TIMSS Database

We obtain our data on students' performance in mathematics and science from the Trends in International Mathematics and Science Study (TIMSS). This study is conducted every four years by the International Association for the Evaluation of Educational Achievement (IEA), with a total of seven waves and 90 countries since 1995. In our analysis we focus on the 2015 wave for 8th grade students in the five Anglo-Saxon countries Australia, Canada, England, Ireland and New Zealand. These countries predominantly have English as a teaching language, share similar education systems and teach science as an integrated subject that combines biology, chemistry, earth science or physics rather than separately (Mullis et al., 2016b). This facilitates the linking of teachers to their students in the TIMSS data.

TIMSS employs a clustered sample design in two stages. First, a random sample of schools is drawn for each participating country. In the second stage, one or more classes are selected from their respective schools. TIMSS samples classes instead of individual students, as the study focuses specifically on the students' exposure to differences at the class level, such as teacher characteristics or instructional experiences. Background information of students at the individual level, of teachers at the class level and of principals or department heads at the school level are collected using detailed questionnaires. The study observes students twice, once in their mathematics and once in their science class. As not all classes in a school are sampled by TIMSS it can happen that students from the same sampled math class are split across different science classes. Therefore, the science classes can contain students from sampled as well as not sampled math classes. Importantly, only students from the sampled math class participate in mathematics and science tests conducted by TIMSS. This increases the possible variation of teacher characteristics between the subjects which is favorable for our identification strategy. However, this also imposes subsequent sampling constraints that need to be accounted for by using student weights representing the probability by which a student was selected within a class to ensure an adequate analysis (Bietenbeck, 2014; LaRoche, Joncas & Foy, 2016). Therefore, we make use of such student weights provided by TIMSS throughout the whole analysis.

In order to measure student achievement, TIMSS provides a rich set of standardized tests for different domains of the respective subject. Total subject test scores are derived afterwards by combining scores of the individual domains. As not all students are administered with the same set of questions, TIMSS uses a method of plausible values to account for the uncertainty of reported statistics resulting from differences in the distributed tasks (Foy & LaRoche, 2016).

3.2 Sample Selection

We first consider a pooled sample containing 36,755 8th grade student observations attending a total of 905 math and 943 science classes taught by 5,012 teachers in 998 schools. We focus on 8th grade rather than 4th grade students assessed by TIMSS, as students at the primary school level usually have the same teacher for all subjects. This would remove the between-subject and between-teacher variation from the sample which is necessary for our empirical identification strategy (Bietenbeck, 2014; Schwerdt & Wuppermann, 2011).

To draw our final sample, we omit all observations where we could not successfully link students to their math or science teachers. Furthermore, it is important for our identification strategy to relate students to exactly one teacher per subject as violating this condition erases the unique assignment of teacher as well class characteristics. Thus, this step also excludes those few students who attend science classes where the corresponding subjects are taught separately instead of integrated science. Moreover, we drop all student observations which have the same teacher in both subjects as those would, similarly to the 4th graders, introduce a lack of within-student between-subject variation. As a final step, we consider only those teachers who do have a formal education beyond upper secondary studies. Completing these steps, our final pooled sample thus comprises 34,843 students in 888 mathematics classes and 895 science classes as well as 4,540 teachers in 965 schools.

Following the literature on teacher training we define pre-service training as teachers' university degree specialization obtained before entering the profession and in-service teacher training as participation in professional development offers after being an employed teacher (see Harris & Sass, 2011). We use the information gathered in TIMSS questionnaires for math and science teachers to derive those two treatment variables. First, we create a dummy variable as an indicator for pre-service teacher training. It takes a value of 1 if a teacher is holding a subject-specific education major in the subject she teaches. For example, mathematics teachers who have studied mathematics and majored in education during their post-secondary studies are assigned a value of 1. Whereas teachers who do not fulfill this criterion are assigned a value of 0.[1] Additionally, we introduce a basic control in our analyses which makes teachers who hold a subject-specific specialization degree for either math or science comparable to those who have not studied the respective subject.[2]

Second, we introduce a treatment variable for in-service teacher training. As an indicator for in-service teacher training TIMSS measures the hours a teacher spent in formal professional development programs (e.g., workshops or seminars) during the last two years.[1] For our main analysis we define a dummy taking the value 0 if a teacher did not attend any in-service training measures and a value of 1 if the teacher spent any time in such programs. In Section 5.2 we will then use the actual participation hours in professional development to investigate possible heterogeneous treatment effects. We further consider a broad set of control variables consisting of teacher and class characteristics to address possible confounders in our empirical strategy as we will describe in detail in Section 4. In terms of teacher characteristics, we include variables to control for a teacher's gender, age categorized in ranges, whether the teacher holds a post-secondary degree, experience categorized in years of teaching as well as a variable indicating the assignment of homework by the respective teacher. Class and subject characteristics are included as instruction time in hours per week as well as class size. Additionally, we add a dummy for ability tracking at the class level, as many schools in our sample practice this method of forming classes based on students' ability only for one of the two subjects. We further introduce the difficulty to fill open vacancies for math or science as a factored variable.[3] In order to ensure a maximum amount of variation in our main variables of interest and a sizable sample, we impute all missing values. Additionally, we include a dummy for each variable with missing observations taking the value 1 indicating imputation and 0 otherwise. The rationale for this

[1] For detailed information on questions and possible response options in the TIMSS questionnaire see Table A1 in the appendix.
[2] In the teacher questionnaire TIMSS declares majors or main areas of studies apart from mathematics or science as "others".
[3] Categorical variables are categorized as presented in Table 1.

Table 1: Means and Standard Deviations of Pre- and In-Service Teacher Training, Teacher Characteristics as well as Class and Subject Characteristics

	Mean	SD
Treatment variables		
Subject-specific education studies (1 = yes)	0.498	0.500
Professional development (1 = yes)	0.888	0.316
Teacher characteristics		
Teaching experience		
1-2 years	0.084	0.277
3-4 years	0.073	0.260
5-10 years	0.223	0.416
more than 10 years	0.620	0.485
Teacher's gender (1 = female)	0.568	0.495
Teacher's age		
under 25 years	0.034	0.180
25-29 years	0.126	0.332
30-39 years	0.285	0.452
40-49 years	0.277	0.448
50-59 years	0.219	0.414
60 or more years	0.058	0.234
Postgraduate degree (1 = yes)	0.295	0.456
Homework assigned (1 = yes)	0.945	0.228
Class and subject characteristics		
Class size	25.518	5.021
Instruction time (hours per week)	3.394	0.843
Ability tracking (1 = yes)	0.523	0.499
Open vacancies		
No vacancies to fill	0.427	0.495
Easy to fill vacancies	0.302	0.459
Somewhat difficult to fill	0.201	0.401
Very difficult to fill	0.070	0.254

Notes: The table presents means and standard deviations of both treatment variables as well as teacher, class and subject characteristics. Teaching experience starts at one year, as the questionnaire was administered at the end of the school year.

step is that we prevent imputed data from being responsible for our results while still keeping a high variation in our controls (Hanushek, Link & Woessmann, 2013).

Table 1 presents the descriptive statistics for our final pooled sample including means and standard deviations for both treatment variables (pre-service and in-service teacher training) as well as teacher, class and subject characteristics. According to this summary, roughly 50 % of teachers in our final sample have graduated holding a major in education with a specialization in the subject they teach. Moreover, about 89 % have participated in teacher training programs offered on the job regardless of their post-secondary education. It is noteworthy that in Australia and England the share of teachers who finished subject-specific studies at the university with a major in education is above the sample average with 64 % and 52 %, respectively. In terms of professional development those two countries also strongly contribute positively to the average proportion: 90 % of teachers in Australia and 91 % of teachers in England have participated in formal on-the-job teacher training programs in the last two years.[4]

About 30 % of math and science teachers in our data have a postgraduate degree, over 50 % indicate a teaching experience of more than ten years and 95 % stated to assign homework. Approximately 91 % of all teachers are between 25 and 59 years, while 3 % are under 25 years old and 6 % are 60 or above. 57 % of all science and math teachers are females. In terms of class and subject characteristics, we observe an average class size of 25.518 students per class with an average instruction time of 3.394 hours per week in mathematics and science. In almost 50 % of all classes, students are assigned to the respective science and math classes based on their achievements. Furthermore, 7 % of all schools in our sample had great difficulties to fill free vacancies in mathematics and science for the analyzed school year. 30 % of the schools stated that it was easy to hire teachers and nearly 43 % were at capacity in terms of teacher demand for both subjects.

4. Empirical Strategy

Investigating the effect of teacher training on student achievement using a non-experimental approach, one will potentially face a bias of the coefficients of interest. The assignment of teachers and students to classes is usually based on self-selection rather than on random mechanisms (Bietenbeck, 2014; Lavy, 2015). Considering the concerns of recent studies (see, e.g., Bietenbeck, 2014; Lavy, 2015; Schwerdt & Wuppermann, 2011), we identify the following possible confounders. First, students might

[4] For detailed information on the descriptive statistics of the treatment variables for each country in the final sample see Table A2 in the appendix.

follow their own as well as their parents' preferences on specific teacher credentials and sort themselves into schools or classrooms which attach a great importance to specialized education majors or participation in formal on-the-job teacher training. In that context, students with a low unobserved ability might especially choose schools where teachers tend to graduate in subject-specific studies at the university including a major in education or attend professional development programs frequently. This would result in a downward bias of the true estimated effect of professional development and subject-specific education studies on student test scores.[5] Second, the degree of participation in pre- and in-service teacher training is potentially correlated with unobserved teacher characteristics, such as motivation and ability. Teachers with a high passion for their profession and subject are likely to make a conscious decision to embark on a career as a teacher, including a specialized major in education. Furthermore, teachers might adapt the amount of professional development they acquire to the students they face. Assuming this relation between in-service teacher training programs and students' unobserved cognitive as well as academic ability will again lead to a biased estimate of the investigated effect.

The literature suggests that one way to address such unobserved confounders is to use an empirical strategy based on within-student between-subject variation. With this approach, students' observed and unobserved characteristics, such as race, family background, motivation and ability, are held constant by introducing student fixed effects. This empirical strategy follows a panel data approach demanding each student to be observed at least twice. However, instead of focusing on a student's observation at two different points in time, the same student should hereby be detected in two different subjects but at the same point of time (Bietenbeck, 2014; Clotfelter, Ladd & Vigdor, 2010; Dee, 2007; Grönqvist & Vlachos, 2016; Lavy, 2015). As TIMSS incorporates this feature for the subjects mathematics and science, we follow this approach and formulate a within-student fixed effect model which additionally allows the educational outcome of a student to be a function of observable teacher and class characteristics. We consequently introduce an educational production function of the form:

$$S_{ijk} = a + \beta_1 PTT_{ijk} + \beta_2 ITT_{ijk} + \boldsymbol{T}_{jk}\gamma + \delta_i + \varepsilon_{ijk}, \qquad (1)$$

where S_{ijk} is student i 's standardized test score in subject k taught by teacher j. PTT_{ijk} is the treatment

[5] A vice versa effect can be considered for students with a high unobserved ability and strong preferences for excellent grades. If those students would tend to sort themselves into schools and classrooms where teachers acquire both types of teacher training, the actual effect on student achievement will be biased upwards.

dummy for pre-service teacher training (subject-specific education studies), while ITT_{ijk} represents the treatment dummy for in-service teacher training (professional development). Therefore, β_1 and β_2 are our parameters of interest. $\boldsymbol{T}_{jk}$ contains observable teacher and class characteristics. δ_i captures the student fixed effects in form of observable and unobservable subject-invariant student traits, while ε_{ijk} describes the unobserved error term. Additionally, an advantageous feature of this approach is that δ_i does not only contain student fixed effects but also subject-invariant school fixed effects, as we observe the same student in the same school. Therefore, it also controls for any school-specific factors that might determine test scores (Bietenbeck, 2014). Consequently, we are able to remove all observable and unobservable subject-invariant variations within students and schools, requiring the added control variables to capture possible differences between teacher and class attributes for one student in both subjects.

In terms of our key identifying assumption, namely the uncorrelatedness of the error term and our treatment variables, we find the necessity to address different concerns. First, note that we specifically control for ability tracking of students and difficulties to fill vacancies for mathematics and science at the class level, as those potentially pose a threat to the integrity of our estimation due to inherent self-selection.[6] Students with a strong interest and ability for math or science could tend to sort themselves into tracking schools and classes which might also require teachers with a higher qualification with regard to teacher training. Conversely, schools which incorporate ability tracking select and therefore admit students according to their subject-specific strengths. Moreover, schools which do have difficulties to fill vacancies in either mathematics or science may be in particular specialized in this subject. Those math- and science-orientated schools tend to comprise an over-concentration of more effective teachers who specifically self-select themselves in such schools aiming to teach extraordinarily motivated and able students in both subjects. Furthermore, teachers with a well-developed pedagogical and context-specific knowledge might consider such specialized schools to have above-average standards for math and science increasing their professional recognition (Clotfelter, Ladd & Vigdor, 2006; Lavy, 2015). If schools would incentivize those teachers to participate in professional development programs or especially attract teachers with subject-specific education studies, the true effect of pre- and in-service teacher training would be biased by teachers'

[6] Note that both control variables are considered at the class rather than at the school level as some schools in the data set only conduct ability tracking for either science or math and not necessarily for both classes. Contrary to Clotfelter, Ladd and Vigdor (2010) or Bietenbeck (2014) we therefore explicitly consider ability tracking as a control variable in our approach in order to capture possible differences between math and science classes which are not erased by within-student and therefore school fixed effects.

quality and also student ability. As mentioned above, we tackle these two potential threats of distortion to our effect of pre- and in-service teacher training on student achievement by including various controls.

Second, subject-specific unobserved characteristics determining student test scores would jeopardize our identifying assumption if they were also correlated with our treatment variables. One could imagine subject-specific requirements to students' cognitive skills or that the subjects are stimulating students' motivation unequally. Such differences in subject characteristics might further demand different types of teacher education or emphasize an exceptional amount of professional development to meet the requirements faced by students. On the one hand, this would imply that teachers sort into specific subjects as a result of their particular form of pre-service training. On the other hand, it could be necessary to adjust the amount of in-service teacher training according to the subject they are teaching. Such multidimensional alternating effects would pose a threat to our key identifying assumption as they lead to a distortion of our treatment effect due to the inherent selection bias. We argue that even though there might be variation between mathematics and science, the subjects demand very similar cognitive skills and are likely to address students' motivation in a quite similar way (Bietenbeck, 2014). Furthermore, both subjects likely pose similar requirements to teachers' qualifications in terms of pre- and in-service teacher training. Consequently, this concern is mitigated for our further analysis. Moreover, we assume that the effect of our treatment variables is equal for mathematics and science leading to the same parameter β_1 and β_2 of interest for both subjects.

Third, pre-service teacher training in the form of a subject-specific educational major can be considered as a fixed characteristic of a teacher entering the profession. However, participation in professional development might vary depending on students' observed and unobserved skills. Teachers who face students with difficulties to process mathematical or scientific content might increase their efforts by participating in formal on-the-job training in order to improve their pedagogical or context-specific expertise. Using the student fixed effects approach, we account for any subject-invariant student traits which potentially impact student achievement. Therefore, this mitigates a potential bias arising if teachers would adapt their participation in professional development depending on students' ability.

Finally, strongly motivated and passionate teachers might select themselves into math- or science-specific education studies and put a greater emphasis on participation in professional development. These correlations between unobserved teacher traits and pre- as well as in-service teacher training could induce additional impact channels on student achievement beside the actual treatments and lead to an omitted variable problem overestimating the actual effect. In the literature on the impacts of teacher characteristics on student achievement, this problem is discussed broadly without proposing an explicit solution. Similar to Bietenbeck (2014), Clotfelter, Ladd and Vigdor (2010), Lavy (2015) as well as Schwerdt and Wuppermann (2011) we address this confounder by controlling for a rich set of observed teacher characteristics. Taking one step further, we investigate the limit of this omitted variable bias by checking the coefficient stability. More specifically, we apply Oster's (2019) approach on unobservable selection as a robustness check to our analysis.

5. Results

In the following section we will present our results regarding the effects of subject-specific education studies and formal on-the-job professional development on students' standardized test scores. In the first step, these results will be discussed generally. In the second subsection, we will further investigate the existence of heterogeneous effects for varying intensity in terms of participating hours in professional development programs and additionally, among different subgroups of students in our sample.

5.1 Effects of Teacher Training on Overall Student Achievement

In order to identify the effects of pre- and in-service teacher training on overall student achievement, we run several single- and multi-treatment regressions as specified in Equation 1. As TIMSS provides student test scores in the form of five plausible values per subject, we randomly draw one of the resulting 25 possible combinations for mathematics and science to measure students' performance. This method has been established as standard in the literature (see Jerrim et al., 2017; Lavy, 2015) and generates robust estimates as we will show in Section 6.1 by using TIMSS' advertised jackknife repeated replication method.[7] We further standardize the drawn scores with a mean of zero and a standard deviation of one. This is advantageous as it allows us to interpret the estimated coefficients for both treatment variables in respect of the standard deviation of the pooled score distribution. Moreover, all regressions are adjusted using student sampling weights provided by TIMSS to mitigate the selection

[7] Beside the jackknife repeated replication method described in Section 6.1, we conduct an additional robustness check of our existing results and draw a random combination of another two plausible values for mathematics and science. Table A3 shows virtually identical coefficients for both treatment variables in terms of magnitude and significance.

problem mentioned in Section 3.1 and include subject fixed effects. We further introduce a control in all specifications indicating whether a teacher has studied the subject she teaches in order to account for differences in the teachers' fields of study. Finally, we address potential correlations in the error terms within classes by clustering standard errors of all subsequent results at the class level.

Table 2 presents the estimated coefficients of pre- and in-service teacher training from standardized test score regressions using student fixed effects. Both treatments indicate a positive impact on overall student achievement regardless of the econometric specification. However, the effect of subject-specific education studies with 0.008 of a standard deviation is comparatively small in its magnitude and statistically insignificant in the single-treatment model without any controls (column 1). Extending this analysis to a multi-treatment specification in column 3 and including teacher characteristics in column 4 as well as class and subject characteristics in column 5 does not lead to noteworthy changes in the estimated parameter of interest or its significance.

The overall insignificant estimates for subject-specific education studies are in line with the results found in previous literature on teacher training as well as teacher credentials. Similar to Harris and Sass (2011) as well as Monk (1994), we find that specialized course selection in education at the university level does not significantly impact teacher effectiveness and thus overall student performance. Moreover, our findings for formal pre-service teacher training consistently complement the existing research on teacher's college major choice (see Betts, Zau & Rice, 2003; Goldhaber & Brewer, 2000).

In contrast, the estimated coefficient of participating in professional development is significant either at the 5 % or 1 % level while also virtually identical in columns 2-4 but being slightly reduced in column 5. According to Oster (2019) this coefficient stability might indicate a limited omitted variable bias in our specifications. We investigate this robustness in a more technical manner by applying her econometric approach for selection on observables and unobservables in Section 6.2. Furthermore, it is evident that all specifications presented in Table 2 show noticeably larger estimates for in-service teacher training compared to pre-service training. Considering the single-treatment model without any controls, our results suggest an increase of 0.038 of a standard deviation in the overall student test score if teachers acquire in-service training compared to those who did not participate in any professional development in the last two years (column 2). Adding the indicator for pre-service teacher training to this specification almost does not alter the effect in terms of its magnitude keeping the significance still at the 5 % level (column 3). In our preferred specification (column 5) which includes both forms of teacher training and the full set of controls mentioned above,

Table 2: Estimated Effects of Pre-Service and In-Service Teacher Training on Standardized Student Test Scores

	(1)	(2)	(3)	(4)	(5)
Subject-specific education studies (1 = yes)	0.008		0.007	0.006	0.009
	(0.011)		(0.011)	(0.011)	(0.010)
Professional development (1 = yes)		0.038**	0.038**	0.040***	0.031**
		(0.015)	(0.015)	(0.015)	(0.014)
Teacher controls	NO	NO	NO	YES	YES
Class and subject controls	NO	NO	NO	NO	YES
Number of students	34,843	34,843	34,843	34,843	34,843

Notes: Results from weighted student fixed effects regressions with student standardized test scores as dependent variable. Test scores are standardized with mean zero and standard deviation of one. Robust standard errors clustered at the class level are in parentheses. All regressions include subject fixed effects and a control for teachers' subject-specific specialization degree in either math or science. Constants are included. Teacher, class and subject controls are used as specified in Table 1. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

the coefficient implies that teachers' participation in professional development is associated with an 0.031 increase of a standard deviation in the overall student performance. In line with Angrist and Lavy (2001) as well as Harris and Sass (2011)[8], our results suggest that in-service teacher training does influence the overall student achievement. Therefore, our findings are in contrast to several other studies on this subject which find either no significant or only small effects of professional development in that context (see Jacob & Lefgren, 2004; Garet et al., 2008; Garet et al., 2011).

5.2 Heterogeneous Treatment Effects

The literature suggests various possible starting points for heterogeneous effects of teacher credentials and characteristics on student achievements to unfold. On the one hand, Clotfelter, Ladd and Vigdor (2006) as well as Grönqvist and Vlachos (2016) emphasize the importance of such heterogeneity among different types of students. They investigate whether contrasting subgroups stratified by certain student characteristics respond differently to teachers' credentials. On the other hand, several studies focus mainly on the treatment variable itself and find positive and significant effects of in-service teacher training on student performance only for more than 14 hours of participation (Yoon et al., 2007). In order to take up the literature's propositions, we build our investigation of heterogeneous treatment effects on these two separate approaches and pay special attention to identifying potential

[8] Angrist and Lavy (2001) report positive effects of in-service teacher training programs on students' reading and mathematics achievements in elementary schools in Jerusalem. They find that students who attended non-religious schools where teachers participated in in-service teacher training indicated higher test scores by 0.25-0.5 of a standard deviation compared to the control group. Harris and Sass (2011) only find positive effects for middle school math teachers in Florida with an effect size of approximately 0.006 of a standard deviation.

nonlinearities of our estimations.

We first focus on the *amount* of professional development a teacher underwent in the last two years. We investigate this relation by again utilizing TIMSS' measurement of in-service teacher training as in our previous analysis. However, now we are using a categorical treatment variable instead of a dummy. Consequently, we consider the following five categories of participating hours in professional development defined by TIMSS: none, less than 6 hours, 6-15 hours, 16-35 hours and more than 35 hours.[9]

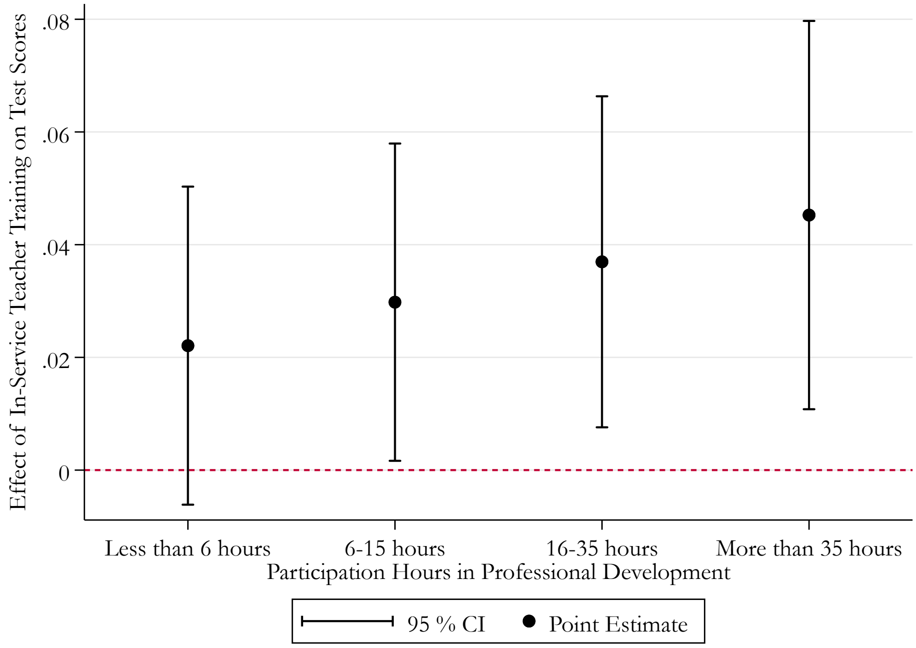
Figure 1 presents the effects of in-service teacher training on student achievement for different hours of professional development undergone by a teacher. Hereby, we focus on our preferred specification with the full set of controls as depicted in column 5 of Table 2. The graph shows a clear dose-response pattern as the effect on the overall student achievement increases with the undergone hours of in-service teacher training. The estimates are all positive and range from 0.022 to 0.045 of a standard deviation. However, our results yield significant effects only for 6-15 hours of professional development or higher.[10] This supports findings in the existing literature that fruitful professional development programs are intense in terms of participation hours (Jackson, Rockoff & Staiger, 2014). Further, our results are partly in line with Yoon et al. (2007) who report significant effects for professional development for studies with more than 14 hours of in-service teacher training acquired.

Next, we explore the effects of pre- and in-service teacher training among different subgroups of students. In particular, we examine whether participation in professional development[11] and subject-specific education studies interact with students' gender and parental educational background as those are important determinants of student performance (Grönqvist & Vlachos, 2016; Björklund & Salvanes, 2011). We therefore consider two econometric specifications using interaction terms of our treatment variables with the mentioned student characteristics while further including the full set of teacher and class controls.[12] All student characteristics are drawn from the TIMSS student questionnaire.

In terms of students' gender, column 1 of Table 3 shows a significant increase in test scores by 0.042 of a standard deviation among male students if teachers have participated in professional

[9] For detailed information on questions and possible response options in the TIMSS questionnaire see Table A1 in the appendix.
[10] It is noteworthy that the estimates for pre-service teacher training still do not indicate any significant effects. For detailed information on coefficients of interest and the respective standard errors see Table A4 in the appendix.
[11] Contrary to our previous analysis presented in Figure 1, we again consider participation in professional development as a dummy rather than a categorical treatment variable.
[12] See again Table 1 for a detailed description of control variables used.

Figure 1: Estimated Coefficients for In-Service Teacher Training by Participation Hours



Notes: This figure is based on the results from a weighted student fixed effects regression with student standardized test scores as dependent variable and a full set of controls as specified in Table 1. Test scores are standardized with mean zero and standard deviation of one. The regression includes subject fixed effects and a control for teachers' subject-specific specialization degree in either math or science. as well as the pre-service teacher training variable. The spikes depict the 95 % confidence intervals on each estimated coefficient, with standard errors clustered at the class level. The regression coefficients as well as corresponding standard errors are presented in Table A4 in the appendix.

development programs during the last two years. Similar to our estimates in previous sections, a teacher's formal pedagogical specialization in the subject she teaches does not have further impact on the overall performance in this subgroup. However, our results suggest a negative but insignificant interaction between pre- and in-service teacher training and 8th grade girls. Consequently, teachers with given subject-specific education studies and participation in professional development seem to be equally effective among female and male students.

We now investigate the role of parental education for the effects of pre- and in-service teacher training on the student performance. The rationale behind this step is two-sided. First, the literature on student education and family background suggests parental education as a proxy for student socio-economic background (Björklund and Salvanes, 2011). Consequently, we can exploit this fact to gain insights on how effective teacher training is among groups with different socio-economic backgrounds. Second, highly educated parents might especially encourage their children with regard to school related tasks e.g., by providing help with homework or school projects. Those students could additionally benefit from parental subject-specific knowledge in science and mathematics besides teachers' pedagogical and content-specific qualifications (Bietenbeck, Piopiunik & Wiederhold, 2018).

Table 3: Heterogeneous Effects of Pre- and In-Service Teacher Training on Student Achievement by Student Characteristics

	(1)	(2)
Subject-specific education studies (1 = yes)	0.012	0.009
	(0.012)	(0.014)
x female	-0.006	
	(0.016)	
x higher educated parents		-0.006
		(0.018)
Professional development (1 = yes)	0.042***	0.018
	(0.016)	(0.017)
x female	-0.022	
	(0.021)	
x higher educated parents		0.014
		(0.023)
Teacher controls	YES	YES
Class and subject controls	YES	YES
Number of students	34,833	20,545

Notes: Results from weighted student fixed effects regressions with student standardized test scores as dependent variable. Test scores are standardized with mean zero and standard deviation of one. Robust standard errors clustered at the class level are in parentheses. Pre- and in-service teacher training is interacted with female students (column 1) and with students having higher educated parents (column 2). We specifically excluded 10 students in column 1 and 1,111 students in column 2 due to missing values. Additionally, we excluded 13,187 students in column 2 from our analysis as they answered the question regarding their parental education with “don’t know”. All regressions include subject fixed effects and a control for teachers’ subject-specific specialization degree in either math or science. Teacher, class and subject controls are used as specified in Table 1. Constants are included. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

In order to address this possible heterogeneity, we use data on parental education collected in the TIMSS student questionnaire. We create a dummy variable which takes a value of 1 if at least one parent has a university degree or higher education, and 0 otherwise. Interacting both, pre- and in-service teacher training with this indicator for parental education yields opposing results (see column 2 of Table 3). On the one hand, students with highly educated parents indicate slightly lower effects of subject-specific education studies on test scores with a reduction by 0.006 of a standard deviation compared to students with less educated parents. On the other hand, the impact of participation in professional development is increased by 0.014 of a standard deviation in this subgroup. However, none of the interactions are statistically significant suggesting that students with highly and less educated parents do not respond differently to the same teacher’s qualifications in terms of pre- and in-service teacher training.

Overall, our analyses show unambiguous results: a teacher's effectiveness increases significantly with the growing amount of in-service training hours she acquires. However, our investigation does not suggest any heterogeneous treatment effects for different subgroups of students stratified by gender and parental education. We consequently conclude that teachers who completed a subject-specific education major are equally effective among different types of students. Moreover, teachers' participation in professional development does not seem to have a significantly different impact on the overall student performance among students with those certain characteristics.

6. Robustness Checks

This section is going to provide our key findings with regard to the robustness of our results. In the first step, we will address possible drawbacks resulting from not using TIMSS' advertised estimation method for the standard errors in our analysis. However, we will show that our results are robust to the use of this resampling technique, in particular to the jackknife repeated replication method. Second, we use Oster's (2019) approach of analyzing coefficient stability in order to obtain additional insights on the robustness of our estimations to omitted variable bias. Finally, in accordance with our considerations in Section 4 regarding parental preferences and possible selection bias we compare our results from our preferred specification of the full controlled student fixed effects regression to including a control for students' extra lessons in this model.

6.1 Jackknife Repeated Replication Method

For our main analysis in Section 5.1, we renounced the use of resampling methods in order to keep calculations convenient. This approach is already used in the literature (see e.g., Lavy, 2015) and considered to be an unbiased alternative to more complex resampling techniques (Jerrim et al., 2017). However, there are also studies which argue strongly in favor of such methods as the estimated standard errors might generally be underestimated (Rutkowski et al., 2010). Therefore, we test the robustness of our regression results and introduce the jackknife repeated replication technique (hence JRR) as advertised by TIMSS.

The necessity to use such a resampling method arises from the nature of TIMSS data collection structure. To keep the burden exposed to a single student to a minimum, TIMSS relies on an incomplete-booklet design. This means that every student is only assigned with a subset of the full pool of items available. Consequently, this resource efficient method comes with the downside of uncertainty regarding estimates developed from this incomplete data. This uncertainty regarding the actual achieved test score is referred to as imputation variance. In order to calculate this variance, we first run regressions on all 25 possible combinations of the five plausible values that TIMSS provides for each subject and take the average over the estimates (Foy & LaRoche, 2016). We then compute the imputation variance as:

$$Var_{\mathrm{IMP}}(\beta_0) = (1 + \frac{1}{M}) \sum_{m=1}^{M} \frac{(\beta_m - \beta_0)^2}{M-1} \tag{2}$$

where β_m is the regression coefficient for pre- and in-service teacher training estimated with the overall student weights for plausible value combination m and M as the total number of plausible value combinations. $\beta_0 = \frac{1}{M} \sum_{m=1}^{M} \beta_m$ is the final average regression coefficient. (Foy & LaRoche, 2016).

Furthermore, as TIMSS is not sampling every school in the respective countries but only a few, there is also uncertainty about the representability of the sample with respect to the target population. This uncertainty is called sampling variance. Obtaining this variance requires the use of a suitable resampling scheme such as JRR. Hereby, the sample is repeatedly cut, and the weights of remaining units are readjusted according to the scheme. These so-called replication weights can then be used to compute the statistics of interest. In order to calculate such weights, the resampling technique pairs the sampled schools from each country into 75 dedicated sampling zones.[13] In the next step, the JRR procedure is applied. This involves the repeated drawing of two subsamples, where one school of the respective subsample will be removed while the other is included with a weight adjustment to cope for the missing school. This process is reiterated for all 75 zones, leading to 150 replicate sampling weights that can further be used to calculate the weighted estimates for pre- and in-service teacher training (Foy & LaRoche, 2016). The sampling variance can be calculated according to:

$$Var_{\mathrm{JRR}}(\beta_m) = \frac{1}{2} \sum_{h=1}^{75} \sum_{j=1}^{2} (\beta_{mhj} - \beta_m)^2 \tag{3}$$

[13] TIMSS samples 150 schools in most of the available countries in the database. In the cases where fewer schools had been sampled, TIMSS fills those missing spots with quasi-schools which represent splits of the schools that have already been sampled. In cases where more than 150 schools have been sampled, TIMSS collapses the emerging additional zones into the first 75 (Foy & LaRoche, 2016).

where β_m is the regression coefficient for pre- and in-service teacher training estimated with the overall student weights for plausible value combination *m* and *M* as the total number of plausible value combinations. In contrast, β_{mhj} presents the final average regression coefficient for plausible value *m* but from regressions conducted with replicate sampling weights derived from sampling zone *h* (*h* = *1,…,75*), where one of the j^{th} schools ($j = 1, 2$) in this zone is included and the other removed, and vice versa (Foy & LaRoche, 2016).

Table 4: Effects of Pre- and In-Service Teacher Training on Standardized Student Test Scores using the Jackknife Repeated Replication Technique as a Robustness Check

	Main results	JRR
	(1)	(2)
Subject-specific education studies (1 = yes)	0.009	0.010
	(0.010)	(0.008)
Professional development (1 = yes)	0.031**	0.030***
	(0.014)	(0.010)
Teacher controls	YES	YES
Class and subject controls	YES	YES
Number of students	34,843	34,843

Notes: Column 1 presents the results from our preferred student fixed effect specification as depicted in column 5 of Table 2. Column 2 shows the average coefficients estimated from regressions for all 25 possible combinations of plausible values for this specification using the JRR technique. Robust standard errors clustered at the class level and adjusted for imputation variance for the model in column 2 are in parentheses. All regressions include subject fixed effects and a control for teachers' subject-specific specialization degree in either math or science. Constants are included. Teacher, class and subject controls are used as specified in Table 1. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

In the last step, an unbiased estimate of the standard errors can be obtained by addressing both prevailing uncertainties. We do so by taking the square root of the combined variances.[14] To obtain the coefficients for pre- and in-service teacher training we first run the regressions for all 25 possible combinations of plausible values for mathematics and science separately using student sampling weights and afterwards take the average over the estimates.

In order to verify the robustness of our main findings from Table 2, we focus our attention on the fully controlled specification as depicted in column 5 of Table 2 and contrast those estimates for pre- and in-service teacher training to the coefficients resulting from the JRR method. Table 4 reports

[14] Hereby we sum up the sampling and the imputation variance whereas the sampling variance is the average sampling variance for all 25 plausible value combinations

the estimates and standard errors for both measurement techniques whereas column 1 contains the coefficients for our treatment variables derived under Equation 1. Column 2 refers to the results of applying the JRR procedure to our sample. Comparing the two columns, it is evident that the coefficients almost do not differ from each other. For subject-specific education studies, using JRR increases the estimate by 0.001 while decreasing the standard error by 0.002. However, this doesn't change the fact that pre-service teacher training shows no significant effect on student performance. Regarding on-the-job professional development, it is evident that although the JRR-coefficient only slightly decreased from 0.031 of a standard deviation to 0.030, the standard errors shift was proportionally larger. This finally leads the coefficient to a higher level of significance at 1 % compared to the former 5 % level. As this is the only difference, we conclude that our findings are robust estimates of the effect of pre- and in-service teacher training on student achievement.

6.2 Oster (2019) Analysis – Investigating Coefficient Stability

In order to address the possible correlation between unobserved teacher traits and pre- as well as in-service teacher training, we added a set of controls to our analysis as presented in column 5 of Table 2. As mentioned in Section 5.1, our estimated coefficients of interest remain unchanged in terms of their significance.[15] Moreover, the estimated effect of a teacher's participation in professional development changed only slightly in its magnitude when including teacher as well as class and subject characteristics to our regression. This coefficient stability may suggest that the added controls capture the most relevant selection e.g., in terms of teachers' motivation or ability, implicating a limitation of an omitted variable bias in our analysis (Oster, 2019). In this section, we follow the approach built by Altonji, Elder and Taber (2005) and further extended by Oster (2019) in order to demonstrate formal evidence for the reasoning above. We investigate the robustness to omitted variable bias under the hypothesis that the interrelation between teacher training and unobservables can be derived from the relationship between teacher training and observables (Oster, 2019).

Oster (2019) emphasizes that coefficient changes alone are not an adequate statistic to identify bias. More specifically, omitted variable bias corresponds to coefficient shifts only if such changes are accompanied by movements in R^2 (Oster, 2019).[16] Therefore, we follow Oster (2019) and compare

[15] In particular, only participation in professional development yielded a significant effect on the overall student achievement.

[16] Oster (2019) demonstrates that the explanatory power of a control variable is crucial. Assume that student achievement can be fully explained by two variables whereas one indicates a higher variance than the other. However, in practice one can only observe the variable with the lower variance and therefore include this as a control to the regression of interest.

coefficient movements to corresponding changes in R^2 after including the full set of controls in our student fixed effects regressions. Hereby, we focus our attention on teachers' participation in professional development programs as only this treatment yields significant effects on students' performance.

Oster (2019) suggests two possibilities of investigating the coefficient stability of an econometric regression. In the first approach, one would estimate the proportional relationship between selection on observables and unobservables (δ) which would explain away the effect of in-service teacher training. Hence, we estimate the extent of omitted variable bias necessary to drive the coefficient of professional development to zero ($\beta_2 = 0$ in Equation 1). Hereby, a $\delta > 1$ means that unobservables would need to be disproportionately important compared to the observables i.e., our included teacher and class characteristics, in order to explain away the effect on the overall student achievement (Oster, 2019). Therefore, both Oster (2019) and Altonji, Elder and Taber (2005) suggest a $\delta > 1$ as an appropriate critical value to identify robustness.

In the second approach, one would estimate the identified set $\Delta = [\tilde{\beta}, \; \beta^*(min\{R_{max}, 1\}, 1)]$ which contains the coefficient for in-service teacher training of the full controlled regression as specified in column 5 of Table 2 as a lower bound ($\tilde{\beta}$) and the unrestricted bias-adjusted estimator for $\delta = 1$ (β^*) as an upper bound.

If in that case the identified set includes zero, the emerging coefficient stability after including observed controls could not be considered as robust to omitted variable bias, as the unrestricted estimator would change its sign remarkably compared to $\tilde{\beta}$. Importantly, the two approaches used to analyze coefficient stability rely on an arbitrary chosen R_{max} which is the R^2 from a theoretical regression where all controls, observed and unobserved, are included (Oster, 2019). According to Oster (2019) one should set R_{max} to $1.3\tilde{R}$ in practice.[17] Besides incorporating this suggestion in our analysis, we additionally set R_{max} to a more conservative value of $1.6\tilde{R}$ as suggested by Bietenbeck (2020).

Table 5 presents the controlled and uncontrolled effect of professional development on student test scores including the corresponding R^2 (column 1 and 2, respectively). Note that we use the within R^2 which indicates how much variation of the test scores within students is explained by

The appearing coefficient stability would now be misleading as it is not driven by smaller bias but by the fact that less of the student outcome is explained by the included control variables (Oster, 2019).

[17] After analyzing the outcomes of several randomized studies from top journals, Oster (2019) concludes that at least 90 % of the results survive the cutoff of $R_{max} = 1.3\tilde{R}$.

Table 5: Selection on Observables: Investigating the Coefficient Stability

	Uncontrolled effect	Controlled effect
	(1)	(2)
Professional development (1 = yes)	0.038**	0.031**
	(0.015)	(0.014)
Teacher controls	NO	YES
Class controls	NO	YES
Number of students	34,843	34,843
R^2 (within)	0.001	0.012
A. $R_{max} = 1.3\tilde{R}$		
δ for $\beta = 0$		12.150
Identified set: $\Delta = [\tilde{\beta},\ \beta^*(min\{R_{max},1\},1)]$		[0.031; 0.029]
Does identified set exclude zero?		YES
B. $R_{max} = 1.6\tilde{R}$		
δ for $\beta = 0$		6.136
Identified set: $\Delta = [\tilde{\beta},\ \beta^*(min\{R_{max},1\},1)]$		[0.031; 0.026]
Does identified set exclude zero?		YES

Notes: Results from weighted student fixed effects regressions with student standardized test scores as dependent variable. Test scores are standardized with mean zero and standard deviation of one. Robust standard errors clustered at the class level are in parentheses. All regressions include subject fixed effects and a control for teachers' subject-specific specialization degree in either math or science. as well as the pre-service teacher training variable. Teacher, class and subject controls are used as specified in Table 1. Panel A considers a R_{max} of $1.3\tilde{R}$ and panel B a R_{max} of $1.6\tilde{R}$. Both panels present the degree of selection on unobservables relative to observables (δ) needed for the treatment coefficients β to be zero as postulated by Oster (2019). Additionally, the identified set is calculated for both specifications with $\delta = 1$. If the identified set excludes zero, the determined effects from the controlled student fixed effects regression are said to be robust to omitted variable bias. For calculations of the identified set and δ we used the Stata command -psacalc-. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

our fixed effect model.[18] Panel A considers an R_{max} of $1.3\tilde{R}$ as suggested by Oster (2019) and the analysis in panel B is based on the more conservative choice of $R_{max} = 1.6\tilde{R}$. In both cases the estimated δ lies with 12.150 and 6.136 above the critical value of one. This implicates that the selection on unobservables would need to be more than 12 or 6 times as large as selection on the observed teacher and class characteristics in order to drive the treatment effect of professional development to zero. According to Oster (2019) this can be considered as implausible. Additionally, the identified set

[18] Dee (2007) and Schwerdt and Wuppermann (2011) who also used the within-fixed-effect approach indicate a within R^2 in a range from 0.002 to 0.040. Our within R^2 in the fully controlled regression is 0.012 which lies within the observed range in the literature.

excludes zero for both values of R_{max} indicating no considerable sign changes for the unrestricted bias-adjusted estimator. Overall, our findings suggest that omitted variable bias is not the driving force of our results (Oster, 2019).

6.3 Controlling for extra Lessons

Resuming the reasoning about parental preferences and their influence on their child's educational decisions in Section 4, we investigate further threats of this matter towards the estimates of pre- and in-service teacher training. Asymmetric preferences expressed by parents or their children in favor of either mathematics or science might incentivize the students to visit supplementary classes besides the regular instruction time at school. Due to this demand for high mathematics or science performance students might further tend to self-select themselves to well trained teachers i.e., teachers who specifically hold an education major in the subject they teach or teachers who participate in professional development programs frequently. Moreover, by taking extra lessons outside of school, students access an additional source of tuition which could confound our estimates.

We address these threats to our estimated effects of pre- and in-service teacher training by exploiting TIMSS' student questionnaire. More specifically, students were asked whether they have attended extra lessons or tutoring in mathematics or science not provided by the school during the last 12 months. Including an indicator for extra lessons participation should enable us to control indirectly for eventual asymmetric preferences and thus for possible selection to teachers with specific qualifications. However, the general, underlying issue with controlling for tutoring lies within the uncertainty regarding the point in time at which the decision to attend those additional classes was made. Extra lessons might not only be an indicator for the student's or her parent's preferences on a specific subject before the assignment of the treatment. They might also hint towards the inability of the teacher due to missing pedagogical studies or the lack of professional development. A student facing a teacher with less pronounced traits might feel the need to take extra tuition regardless of her preferences. Likewise, a teacher who attended professional development programs or holds a subject-specific education major might erase the need of extra hours for a student that usually would take such lessons in all subjects. Therefore, controlling for supplementary classes could potentially invoke a bad control problem which would lead to a bias of our results. To mitigate this risk, we did not include this control in our main analysis.

Exploring the results presented in Table 6 suggest that the bad control problem is seemingly

Table 6: Estimated Effects of Pre- and In-Service Teacher Training on Standardized Test Scores with an included Control for extra Lessons in the Regression

	Main results	Extra lessons as additional control
	(1)	(2)
Subject-specific education studies (1 = yes)	0.0092	0.0087
	(0.0100)	(0.0100)
Professional development (1 = yes)	0.0309**	0.0305**
	(0.0142)	(0.0141)
Teacher controls	YES	YES
Class and subject controls	YES	YES
Number of students	34,843	34,843

Notes: Results from weighted student fixed effects regressions with student standardized test scores as dependent variable. Column 1 presents the coefficients from our preferred student fixed effect specification as depicted in column 5 of Table 2. Column 2 additionally considers a student's participation in extra lessons beside regular school instruction time as a control. Test scores are standardized with mean zero and standard deviation of one. Robust standard errors clustered at the class level are in parentheses. All regressions include subject fixed effects and a control for teachers' subject-specific specialization degree in either math or science. Constants are included. Teacher, class and subject controls are used as specified in Table 1. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

not a threat to our investigation. Column 1 of Table 6 shows the estimates for pre- and in-service teacher training from our preferred specification with full controls (see column 5 of Table 2). In contrast, column 2 uses the same specification but additionally includes the indicator for a student's participation in extra lessons in math or science. Comparing the effects of subject-specific education studies and participation in professional development yields that the added control for supplementary tutoring does not distort the coefficient stability for both treatment variables as the estimates are virtually identical for both specifications. Moreover, our results suggest that students' participation in extra lessons outside of school is not a potential source of (selection) bias regarding our main estimates.

7. Conclusion

Our results contribute to the limited and inconclusive literature on the merits of teacher training for student learning. More specifically, we test whether teachers' subject-specific education studies (pre-service training) and participation in professional development (in-service training) increase students' test scores. Using data from the Trends in International Mathematics and Science Study (TIMSS) we follow a within-student fixed effects approach. Our results suggest that a teacher's major in education with a specialization in the subject she teaches does not show significant effects. In contrast, in-service teacher training contributes significantly positive to student outcomes. We find evidence that teachers' participation in professional development increases student test scores by 0.031-0.040 of a standard deviation. Furthermore, our analysis shows clear evidence for a dose-response pattern for in-service teacher training as the effect increases proportionally with the hours participated in such programs. This supports the overall findings in the literature that high-intense in-service teacher training programs (6-15 hours or more) tend to be more effective.

From a policy perspective, these results underline the importance of teachers' on-the-job learning. Although pre-service teacher training is usually a mandatory requirement to enter a teaching profession, teachers tend to accumulate their pedagogical skills and subject knowledge predominantly while already serving, especially in the first years of their profession (Clotfelter, Ladd & Vigdor, 2006; Harris & Sass, 2011; Rivkin, Hanushek & Kain, 2005). Moreover, due to teachers' usually long service it is crucial for them to update their knowledge on the ongoing innovations in their teaching area as well as new pedagogical methods (Mullis, Martin & Loveless, 2016). One possibility to support teachers to cope with such new developments would be the supply of high-intense professional development programs as those tend to be more fruitful towards teachers' effectiveness. Additionally, the insignificant estimates of pre-service teacher training suggest a rethinking of the current pre-service requirements of teachers. As teachers holding a subject-specific education major do not appear to be more effective than others, it seems to be reasonable to reassess the current structure of formal teaching preparation. It could give rise to alternative certification programs as an additional way into teaching without having a subject-specific education major as a mandatory requirement (Harris & Sass, 2011).

The policy implications above need to be considered with caution. Although the within-student fixed effects approach tackles most of the possible confounders, we cannot perfectly rule out all potential biases. For instance, due to unobserved correlated teacher characteristics that might not be captured by our already included controls. Some teachers could be more motivated than others and therefore participate in professional development programs more frequently. Likewise, teachers with a high passion for their profession and subject are likely to make a conscious decision to embark on a career as a teacher, including a specialized major in education. Another concern of our analysis is that we have incomplete information on the actual structure of subject-specific studies with a major in education. More specifically, there is no data on how many credits were necessary to achieve such a degree. It could be that some teachers were required to attend less pedagogical and subject-specific courses than others at the university to major in education with a specialization in the subject they teach. However, following Oster's (2019) approach on selection of observables and unobservables we have shown that such omitted variable biases does not seem to be the driving force of our results.

Finally, this essay could not address which specific content of professional development programs is crucial for teachers' effectiveness. It would be interesting to gain additional knowledge whether e.g., training that emphasizes subject-specific content or rather pedagogy has a positive effect on student test scores. Hereby, further studies could provide valuable insights supporting policymakers and school principals to rethink and readjust the structure of such programs in order to maximize their benefits.

References

Angrist, J.D., & Lavy, V. (2001). Does Teacher Training Affect Pupil Learning? Evidence from Matched Comparisons in Jerusalem Public Schools, *Journal of Labor Economics*, vol. 19, no. 2, pp. 343-369

Altonji, J.G., Elder, T.E., & Taber, C.R. (2005). Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools, *Journal of Political Economy*, vol. 113, no. 1, pp. 151-184

Bietenbeck, J. (2020). Own Motivation, Peer Motivation, and Educational Success, *IZA Discussion Papers no. 13872*, Institute of Labor Economics (IZA)

Bietenbeck, J., Piopiunik, M., & Wiederhold, S. (2018). Africa's Skill Tragedy: Does Teachers' Lack of Knowledge Lead to Low Student Performance?, *Journal of Human Resources*, vol. 53, no. 3, pp. 553-578

Bietenbeck, J. (2014). Teaching Practices and Cognitive Skills, *Labour Economics*, vol. 30, pp. 143-153

Betts, J.R., Zau, A.C., & Rice, L.A. (2003). Determinants of Student Achievement: New Evidence from San Diego, San Diego: Public Policy Institute of California

Björklund, A., & Salvanes, K.G. (2011). Education and Family Background: Mechanisms and Policies, in E.A. Hanushek, S. Machin & L. Woessmann (eds), *Handbook of the Economics of Education*, vol. 3, pp. 201-247

Boyd, D.J., Grossman, P.L., Lankford, H. Loeb, S., & Wyckoff, J. (2009). Teacher Preparation and Student Achievement, *Educational Evaluation and Policy Analysis*, vol. 31, no. 4, pp. 416-440

Bressoux, P., Kramarz, F., & Prost, C. (2009). Teachers' Training, Class Size and Students' Outcomes: Learning from Administrative Forecasting Mistakes, *The Economic Journal*, vol. 119, no. 536, pp. 540-561

Campbell, P.F., Nishio, M., Smith, T.M., Clark, L.M., Conant, D.L., Rust, A.H., Neumayer DePiper, J., Frank, J.T., Griffin, M.J., & Choi, Y. (2014). The Relationship Between Teachers' Mathematical Content and Pedagogical Knowledge, Teachers' Perceptions, and Student Achievement, *Journal for Research in Mathematics Education*, vol. 45, no. 4, pp. 419-459

Chetty, R., Friedman, J.N., & Rockoff, J.E. (2014a). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates, *American Economic Review*, vol. 104, no. 9, pp. 2593-2632

Chetty, R., Friedman, J.N., & Rockoff, J.E. (2014b). Measuring the Impacts of Teachers II: Evaluating Bias in Teacher Value-Added Estimates, *American Economic Review*, vol. 104, no. 9, pp. 2633-2679

Clotfelter, C.T., Ladd, H.F., & Vigdor, J.L. (2010). Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects, *The Journal of Human Resources*, vol. 45, no. 3, pp. 655-681

Clotfelter, C.T., Ladd, H.F., & Vigdor, J.L. (2006). Teacher-Student Matching and the Assessment of Teacher Effectiveness, *The Journal of Human Resources*, vol. 41, no. 4, pp. 778-820

Clotfelter, C.T., Ladd, H.F., & Vigdor, J.L. (2007). Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects, *Economics of Education Review*, vol. 26, no. 6, pp. 673-682

Dee. T.S. (2007). Teachers and the Gender Gaps in Student Achievement, *Journal of Human Resources*, vol. 42, no. 3, pp. 528-554

Foy, P., & LaRoche, S. (2016). Estimating Standard Errors in the TIMSS 2015 Results, in M.O. Martin, I.V.S. Mullis & M. Hooper (eds), *Methods and Procedures in TIMSS 2015,* TIMSS & PIRLS International Study Center, Boston College, Chestnut Hill, pp. 4.1-4.69

Garet, M.S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., Uekawa, K., Falk, A., Bloom, H.S., Doolittle, F., Zhu, P., & Sztejnberg, L. (2008). The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement [pdf], NCEE 2008-4031, *Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S.*

Department of Education, Available at: https://ies.ed.gov/ncee/pdf/20084030.pdf [Accessed 26 May 2021]

Garet, M., Wayne, A., Stancavage, F., Taylor, J., Eaton, M., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S., and Doolittle, F. (2011). Middle School Mathematics Professional Development Impact Study: Findings after the First Year of Implementation [pdf], NCEE 2011-4024, *Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education*, Available at: https://ies.ed.gov/ncee/pubs/20104009/pdf/20104009.pdf [Accessed 26 May 2021]

Goldhaber, D.D., & Brewer, D.J. (2000). Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement, *Educational Evaluation and Policy Analysis*, vol. 22, no. 2, pp. 129-145

Grönqvist, E., & Vlachos, J. (2016). One Size Fits All? The Effects of Teachers' Cognitive and Social Abilities on Student Achievement, *Labour Economics*, vol. 42, pp. 138-150

Hanushek, E.A., Link, S., & Woessmann, L. (2013). Does School Autonomy Make Sense Everywhere? Panel Estimates from PISA, *Journal of Development Economics*, vol. 104, pp. 212-232

Hanushek, E.A., & Rivkin, S.G. (2010). Generalizations about Using Value-Added Measures of Teacher Quality, *American Economic Review*, vol. 100, no. 2, pp. 267-271

Hanushek, E.A., & Rivkin, S.G. (2006). Teacher Quality, *Handbook of the Economics of Education*, vol. 2, pp. 1051-1078

Harris, D.N., & Sass, T.R. (2011). Teacher Training, Teacher Quality and Student Achievement, *Journal of Public Economics,* vol. 95, no. 7-8, pp. 798-812

Jackson, C.K., Rockoff, J.E., & Staiger, D.O. (2014). Teacher Effects and Teacher-Related Policies, *Annual Review of Economics*, vol. 6, pp. 801-825

Jacob, B.A., & Lefgren, L. (2004). The Impact of Teacher Training on Student Achievement: Quasi-Experimental Evidence from School Reform Efforts in Chicago, *The Journal of Human Resources*, vol. 39, no. 1, pp. 50-79

Jerrim, J., Lopez-Agudo, L.A., Marcenaro-Gutierrez, O.D., & Shure, N. (2017). What Happens When Econometrics and Psychometrics Collide? An Example Using the PISA Data, *Economics of Education Review*, vol. 61, pp. 51–58

Kelly, D.L., Centurino, V., Martin, M.O., & Mullis, I.V.S. (eds). (2020). TIMSS 2019 Encyclopedia: Education Policy and Curriculum in Mathematics and Science. *TIMSS & PIRLS International Study Center, Boston College*, Available online: https://timssandpirls.bc.edu/timss2019/encyclopedia/ [Accessed 26 May 2021]

LaRoche, S., Joncas, M., & Foy, P. (2016). Sample Design in TIMSS 2015, in M.O. Martin, I.V.S. Mullis & M. Hooper (eds), *Methods and Procedures in TIMSS 2015,* TIMSS & PIRLS International Study Center, Boston College, Chestnut Hill, pp. 3.1-3.38

Lavy, V. (2015). Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries, *The Economic Journal*, vol. 125, no. 588, pp. 397-424

Metzler, J., & Woessmann, L. (2012). The Impact of Teacher Subject Knowledge on Student Achievement: Evidence from Within-Teacher Within-Student Variation, *Journal of Development Economics*, vol. 99, no. 2, pp. 486-496

Monk, D.H. (1994). Subject Area Preparation of Secondary Mathematics and Science Teachers and Student Achievement, *Economics of Education Review*, vol. 13, no. 2, pp. 125-125

Mullis, I.V.S., Martin, M.O., & Loveless, T. (2016). 20 Years of TIMSS: International Trends in Mathematics and Science Achievement, Curriculum, and Instruction [pdf], *TIMSS & PIRLS International Study Center, Boston College*, *Chestnut Hill*, Available online: http://timssandpirls.bc.edu/timss2015/international-results/timss2015/wp-content/uploads/2016/T15-20-years-of-TIMSS.pdf

[Accessed 26 May 2021]

Mullis, I.V.S., Cotter, K.E., Fishbein, B.G., & Centurino, V.A.S. (2016a). Developing the TIMSS 2015 Achievement Items, in M.O. Martin, I.V.S. Mullis & M. Hooper (eds), *Methods and Procedures in TIMSS 2015,* TIMSS & PIRLS International Study Center, Boston College, Chestnut Hill, pp. 1.1-1.22

Mullis, I.V.S., Martin, M.O., Goh, S., & Cotter, K. (eds.). (2016b). TIMSS 2015 Encyclopedia: Education Policy and Curriculum in Mathematics and Science. *TIMSS & PIRLS International Study Center, Boston College*, Available online: http://timssandpirls.bc.edu/timss2015/encyclopedia/ [Accessed 26 May 2021]

OECD (2020). TALIS 2018 Results (Volume II): Teachers and School Leaders as Valued Professionals, *TALIS*, OECD Publishing, Paris

OECD (2018). What Makes High-Performing School Systems Different, *World Class: How to Build a 21st-Century School System*, OECD Publishing, Paris

Oster, E. (2019). Unobservable Selection and Coefficient Stability: Theory and Evidence, *Journal of Business & Economic Statistics*, vol. 37, no. 2, pp. 187-204

Rivkin, S.G., Hanushek, E.A., & Kain, J.F. (2005). Teachers, Schools and Academic Achievement, *Econometrica*, vol. 73, no. 2, pp. 417-458

Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International Large-Scale Assessment Data: Issues in Secondary Analysis and Reporting, *Educational Researcher*, vol. 39, no. 2, pp. 142–151

Schwerdt, G., & Wuppermann, A.C. (2011). Is Traditional Teaching Really all that Bad? A Within-Student Between-Subject Approach, *Economics of Education Review*, vol. 30, no. 2, pp. 365-379

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). Reviewing the Evidence on how Teacher Professional Development Affects Student Achievement, Issues & Answers Report

[pdf], REL 2007 – no. 033, *Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest*, Available online: https://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL_2007033.pdf [Accessed 26 May 2021]

Appendix

Table A1: Questions about Pre- and In-Service Teacher Training in TIMSS

Respondent	Question	Response Options
Teacher	During your <post-secondary> education, what was your major or main area(s) of study?	Mathematics; Biology, Physics, Chemistry, <Earth Science>; Education - Mathematics; Education - Science; Education - General; Other
Teacher - Science	In the past two years, how many hours in total have you spent in formal <in-service/professional development> (e.g., workshops, seminars, etc.) for mathematics?	None; Less than 6 hours; 6-15 hours; 16- 35 hours; More than 35 hours
Teacher - Math	In the past two years, how many hours in total have you spent in formal <in-service/professional development> (e.g., workshops, seminars, etc.) for science?	None; Less than 6 hours; 6-15 hours; 16- 35 hours; More than 35 hours

Notes: The table depicts the questions provided with TIMSS wave 2015 to classify teachers field of study and their attendance in professional development.

Table A2: Means and Standard Deviations of Pre- and In-Service Teacher Training by Country

Country	Treatment variables: Subject-specific education studies (1 = yes) (1)	Treatment variables: Professional development (1 = yes) (2)
Australia	0.643 (0.479)	0.900 (0.300)
Canada	0.432 (0.495)	0.874 (0.332)
England	0.517 (0.500)	0.911 (0.285)
Ireland	0.452 (0.498)	0.872 (0.334)
New Zealand	0.419 (0.493)	0.887 (0.316)

Notes: The table presents means and standard deviations of subject-specific education studies (column 1) and participation in professional development (column 2) for each Anglo-Saxon country. Standard deviations are in parentheses.

Table A3: Estimated Effects of Pre- and In-Service Teacher Training on Standardized Test Scores using an another randomly drawn Combination of Plausible Values for Mathematics and Science

	Main results	Another PV combination
Subject-specific education studies (1 = yes)	0.0092	0.0094
	(0.0100)	(0.0101)
Professional development (1 = yes)	0.0309**	0.0306**
	(0.0142)	(0.0142)
Teacher controls	YES	YES
Class and subject controls	YES	YES
Number of students	34,843	34,843

Notes: Results from weighted student fixed effects regressions with student standardized test scores as dependent variable. Column 1 presents the coefficients from our preferred student fixed effect specification as depicted in column 5 of Table 2. Column 2 considers the same specification as column 1 but uses an another randomly drawn combination of plausible values for math and science. Test scores are standardized with mean zero and standard deviation of one. Robust standard errors clustered at the class level are in parentheses. All regressions include subject fixed effects and a control for teachers' subject-specific specialization degree in either math or science. Constants are included. Teacher, class and subject controls are used as specified in table 1. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A4: Estimated Effects of Pre- and In-Service Teacher Training on Standardized Student Test Scores using different Participation Hours for Professional Development

Subject-specific education studies (1 = yes)	0.008
	(0.010)
Professional development (base: 0 hours in the past two years)	
Less than 6 hours	0.022
	(0.014)
6-15 hours	0.030**
	(0.014)
16-35 hours	0.037**
	(0.015)
More than 35 hours	0.045***
	(0.018)
Teacher controls	YES
Class and subject controls	YES
Number of students	34,843

Notes: Results from weighted student fixed effects regression with student standardized test scores as dependent variable. Test scores are standardized with mean zero and standard deviation of one. Robust standard errors clustered at the class level are in parentheses. The regressions include subject fixed effects and a control for teachers' subject-specific specialization degree in either math or science. Constant is included. Teacher, class and subject controls are used as specified in table 1. Significance levels: *** $p<0.01$, ** $p<0.05$, * $p<0.1$.