

Wedel, Katharina

Working Paper

Instruction time and student achievement: The moderating role of teacher qualifications

ifo Working Paper, No. 344

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Wedel, Katharina (2021) : Instruction time and student achievement: The moderating role of teacher qualifications, ifo Working Paper, No. 344, ifo Institute - Leibniz Institute for Economic Research at the University of Munich, Munich

This Version is available at:

<https://hdl.handle.net/10419/229891>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Instruction Time and Student Achievement: The Moderating Role of Teacher Qualifications

Katharina Wedel

Imprint:

ifo Working Papers

Publisher and distributor: ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49(0)89 9224 0, Telefax +49(0)89 985369, email ifo@ifo.de

www.ifo.de

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

www.ifo.de

Instruction Time and Student Achievement: The Moderating Role of Teacher Qualifications*

Abstract

Recent evidence suggests a positive effect of the quantity of instruction on student achievement. In this paper, I focus on the interaction between the quantity and the quality of instruction. Using international TIMSS data, I exploit within-student between-subject variation. I find that on average, an additional hour of instruction time leads to an increase of 0.03 standard deviations in students' test scores across all countries. Importantly, these effects of instruction time are significantly larger for students with better qualified teachers, resulting in an increase in test scores of 0.04 to 0.05 standard deviations. While on average, instruction time has no significant effect in developing countries, it increases test scores by 0.02 standard deviations when taught by a high-qualified teacher also in developing countries.

JEL Code: I21, I25, C21

Keywords: Instruction time, student achievement, teacher qualifications, education production function, TIMSS

Katharina Wedel
ifo Institute – Leibniz Institute for
Economic Research
at the University of Munich
Poschingerstr. 5
81679 Munich, Germany
wedel@ifo.de

* I gratefully acknowledge comments from seminar participants at the ifo Center for the Economics of Education in Munich, participants at the EEA Virtual Congress 2020 and at the 76th Annual Congress of the IIPF. This work was supported by the Smith Richardson Foundation.

1 Introduction

Quantity and quality of instruction are essential for students' educational achievement. On the one hand, instructional quantity has a positive impact on a student's achievement (e.g. Lavy, 2015; Rivkin & Schiman, 2015). On the other hand, instructional quality has proven to be important for student achievement (Hanushek & Rivkin, 2006). Hence, there may be a trade-off between the quality and quantity of instruction: more instruction time will probably only be of benefit if teachers use the additional time efficiently, e.g. by covering new or revising old content instead of using the time for classroom management or administrative tasks. Thus, more instruction time only leads to better performance if teachers actively use the time for teaching. Furthermore, it is important how well a teacher knows the subject and how well she is able to explain it to her students (Carroll, 1989). Patall, Cooper and Allen (2010) also state that "the effectiveness of instruction" (p. 430) can influence whether additional time at school affects students' outcomes, and if so, whether the effect is positive or negative.

The effect of instruction time might go in different directions. On the one hand, additional instruction time in a subject might give the teacher the opportunity to cover more material, analyze and discuss it in more detail, take the time to answer students' questions and combine concepts that arise in different classes (National Center on Time and Learning, 2017). On the other hand, students might get to a point where more instruction time and thus more input is harmful since they are unable to absorb further information. Some pupils might even get bored, especially the already high-performing students. Andersen, Humlum and Nandrup (2016) argue that students need to be motivated to follow and focus on what they are being taught to achieve long-term success so that they can benefit from the extra instruction time. The authors explain that students need self-control to focus. However, students' ability to concentrate decreases with more instruction time, making it more difficult for them to pay attention and control their emotions and thoughts. Ultimately, students might become less focused and even more aggressive (Andersen, Humlum & Nandrup, 2016).

In this paper, I study to what extent the effect of instruction time on student performance is moderated by the quality of teachers. I add instructional time as an input to the education production function and interact its effect with teacher qualifications. I use a student

fixed effects model which accounts for observable and unobservable individual-specific factors, such as unobserved ability, and exploit within-student between-subject variation to identify the effect of an additional hour of instruction time on test scores.

I use data from the 2015 Trends in International Mathematics and Science Study (TIMSS) which contains two observations for each student, one in math and one in science. My main outcome variable is students' test scores in math and science as a measure of their cognitive skills. The main independent variable is the instruction time in these subjects, measured in hours per week. Instruction time is defined as the "amount of time during which students receive instruction from a classroom teacher in a school context" (UNESCO, 2019). It does not include teacher training days, holidays, breaks at school or learning time outside of school, such as time for homework and tutoring.

My results show that teacher qualifications play a moderating role for the effect of instruction time on student achievement. In the student fixed effects model, I regress student test scores on instruction time as well as on teacher qualifications and their interactions. On average, across all countries, one hour more instruction time leads to 0.03 standard deviations higher test scores, with boys (0.038 standard deviations) benefiting more than girls (0.025 standard deviations). It is important to note that the effect varies according to teachers' formal qualifications: it is larger for students with better qualified teachers. On average, the effect of instruction time on student achievement by a highly qualified teacher is 0.04 to 0.05 standard deviations, depending on which teacher qualification is considered. These teacher qualifications are measured in terms of participation in professional development, a Bachelor's degree (or higher) with the relevant subject as a major subject as well as teacher education with a specialization in the relevant subject.

The moderating role of teacher quality is particularly telling in understanding the effect of instruction time in developing countries. As in a previous analysis of PISA data by Lavy (2015), I find that the effect of instruction time is larger for students in developed countries than in developing countries. In fact, for an extended sample of developing countries covered by TIMSS, including countries in the Middle East, the average effect of instruction time in developing countries is not statistically significant and close to zero. However, also in developing

countries, instruction time by a highly qualified teacher increases test scores by 0.02 standard deviations.

I verify the robustness of my results using a series of varying specifications. First, I apply a within-teacher specification to rule out bias due to unobserved teacher characteristics. The results are robust: the coefficients are slightly larger than those in the main specification. Furthermore, the results are robust to including a squared term of instruction time, as well as to restricting the sample to schools which do not track students into different classes according to their ability.

In addition to test scores, I also examine whether more instruction time in a subject affects students' motivation and attitude towards that subject. On the one hand, additional instruction time in a subject can cause a student to become tired of that subject which does not necessarily lead to lower test scores but instead students may develop an aversion to the subject. On the other hand, more instruction time may encourage students to enjoy the subject even more since they are able to go into more detail about specific content. My results suggest that more instruction time leads to a more positive attitude towards the subject. However, this attitude effect is not influenced by the qualifications of teachers.

Recent literature on the impact of instruction time on students' test scores mostly finds a positive impact. This paper mainly relates to two recent articles which use international data. First, Lavy (2015) examines the effect of instruction time in 50 different countries using the 2006 Programme for International Student Assessment (PISA) study, which evaluates students in math, science and reading. Applying student fixed effects, Lavy (2015) finds that an increase in instruction time of one hour on average leads to 0.06 standard deviations higher test scores.¹ Second, Rivkin and Schiman (2015) use 2009 PISA data and apply a student fixed effects model that exploits variation in instruction time within schools, across subjects or across grades. They find that increasing the weekly instruction time by one hour leads to test scores that are 0.02 to 0.03 standard deviations higher. In addition, they show that better classroom

¹ Cattaneo, Oggenfuss and Wolter (2017) replicate Lavy's (2015) study, applying the same methodology to PISA data for Switzerland. Their focus on only one country has the advantage of allowing them to examine the effects of distinct uses of instruction time in a context with very similar curricula and educational objectives which do not cause biases (Cattaneo, Oggenfuss & Wolter, 2017). They also deviate from Lavy's (2015) approach by using official teaching times recommended by the education authorities instead of data reported by students.

environment in terms of student behavior and teacher-student interaction enhances the positive effect of additional instruction time (Rivkin & Schiman, 2015).

The second strand of literature to which this paper refers uses national data. Bingley et al. (2018) apply the same methodology as Lavy (2015) and Rivkin and Schiman (2015) to Danish administrative data. Bingley et al. (2018), in comparison, use data for three cohorts of students in each grade throughout their entire compulsory education. Hence, they focus on the accumulated time from one grade to the next. These effects are about twice as large as those based on the time of only one grade. Other studies use exogenous variation in instruction time due to school closures or weather-related absences as an instrument for instruction time (Marcotte, 2007; Goodman, 2014), and find adverse effects of a lack of instruction time on student achievement. Further studies exploit school reforms, for example in Germany, which lead to an increase in the number of weekly hours of instruction in academic-track high schools (Dahmann, 2017; Huebener, Kuger & Marcus, 2017). These authors find small but positive effects of instructional hours. Andersen, Humlum and Nandrup (2016) use a randomized controlled trial (RCT) in which some schools were randomly assigned more instruction time. They find that a student's reading performance increases by 0.15 standard deviations due to an increase in instruction time.

To the best of my knowledge, none of these previous studies addresses the trade-off between the quantity and quality of instruction. Hence, I contribute to the literature by exploring the interaction between instruction time and teacher quality, measured by the formal qualifications of teachers. Using a new data source, the 2015 TIMSS data, allows me to show results for developing countries that were not considered in previous studies on instruction time. These are especially countries from the Middle-East, such as Saudi Arabia, the United Arab Emirates and Oman, as well as Singapore and Kazakhstan. Hence, I present new evidence on the smaller effect of instruction time in developing countries, and how teacher qualifications play a moderating role for this effect. Moreover, I focus on fourth-graders, which is especially relevant since young children are particularly sensitive to interventions and the return on investment in human capital is higher (Cunha et al., 2006). In addition, I examine not only the impact on cognitive skills, but also on students' motivation and their attitude towards the subject.

This paper is organized as follows: Section 2 describes the data, followed by a description of the empirical strategy in Section 3. The empirical results are presented in Section 4. Section 5 concludes.

2 Data

2.1 Trends in International Mathematics and Science Study

I use data from the Trends in International Mathematics and Science Study (TIMSS) which are a repeated cross-section. The TIMSS & PIRLS International Study Center at the International Association for the Evaluation of Educational Achievement (IEA) conducts assessments of students' achievements in math and science and reading, which are internationally comparable. In general, this study is conducted in more than 60 countries (TIMSS, 2018). In addition to information about a student's achievement, the data also include information about students' attitudes, teachers, school resources and instructional practices (TIMSS, 2018). This information is collected in separate questionnaires: there is a student questionnaire, a home questionnaire which parents are asked to fill out, a teacher questionnaire, and a school and curriculum questionnaire that is filled out by the school principal.

TIMSS uses a two-stage random sample design: in the first stage, a sample of schools is determined, and in the second stage, one or more classes within a school are selected (Martin, Mullis & Hooper, 2016). Since TIMSS monitors students' instructional and curricular experiences at the classroom level, TIMSS samples classes rather than individual students (Martin, Mullis & Hooper, 2016). TIMSS monitors student achievement along with the other previously mentioned information for two grades: fourth and eighth grade. The national samples of students in TIMSS are constructed to describe the target population. Theoretically, the two-stage random sample design generates samples of students with the same probability of selection (Martin, Mullis & Hooper, 2016). In practice, however, a varying number of selected classes and differential non-response can result in different probabilities of selection, requiring individual sampling weights for the students. I use the senate weights which can be used when differences across countries are examined. In this way each country receives the same weight.

Since the pool of questions and items in TIMSS is too large to be answered by one student, students receive only a subset of questions (a so-called booklet) to answer. Each booklet contains both math and science questions. This is different from the PISA data where some students only answer questions in some subject areas, but not in all three. In PISA data, students were assigned scores according to their performance in other domains (Jerrim et al., 2017). According to Jerrim et al. (2017), the results could then be driven by a random imputation error. However, my results are not affected by this since all students answer both math and science questions. Aggregating the results of all booklets yields results for the overall assessment. Then, plausible values are imputed to obtain five estimates of a student's achievement (Martin, Mullis & Hooper, 2016).

2.2 Analysis Sample and Variables

I use data on fourth grade students from the 2015 survey wave since interventions matter more for younger students.² The final sample includes 115,071 students in 1,623 classes and 4,529 schools in 42 countries. Since every student is evaluated twice - once in math, once in science - the number of observations amounts to 230,142. I present descriptive statistics in Table 1.

The dependent variable is the *test score* of a student in math or science. The test score variable in the data set is the plausible value for math and science.³ I standardize the test scores so that the mean is zero and the standard deviation is one. These test scores measure a student's cognitive attainment in math and science (Woessmann, 2003).

To evaluate whether additional instruction time affects a student's motivation and attitude, I generate an index of four variables for each subject to obtain a measure of a student's motivation and attitude towards a subject. For this, I calculate an index *like subject* as an al-

² In the analysis I use only those countries for which information on test scores and instruction time in both math and science is available and where science is taught as a separate subject. Science comprises life science, physical science and earth science. In general, math test scores are available for all countries. There are no science test scores for Indonesia, Iran, Kuwait, Morocco, Belgium. A list of participating countries with available data on instruction time in both subjects can be found in Appendix Table A.3. Besides, I drop observations with more than one teacher in either science or math. Otherwise there would be four or more observations per student. Since there is no information about why there is a change of teacher, I decide not to use these observations.

³ In particular, I choose the 1st plausible value in both subjects since this is often used in the literature, for example in Rivkin and Schiman (2015). The results are robust to using the other four plausible values. This is in line with Jerrim et al. (2017) stating that using one plausible value or all five does not alter the results.

ternative dependent variable. Questions are for example "I enjoy learning mathematics." and "I learn many interesting things in mathematics." as well as the corresponding questions for science. The students can choose from the following answers: "agree a lot", "agree a little", "disagree a little" and "disagree a lot".⁴

The main independent variable of interest is *instruction time* either in math or in science, depending on whether the observation is for student i in math or science. The underlying question for teachers in order to determine this variable is the following: "In a typical week, how much time do you spend teaching mathematics to the students in this class? (minutes)" (TIMSS, 2015). The same question is asked for science. To make them comparable to other studies, I convert these variables into hours.⁵ Following Lavy (2015), I aggregate instruction time on the school by subject level. This is done to overcome potential problems due to sorting and tracking since schools might sort or track students into classes based on subject-specific instruction time. This can potentially lead to bias: An upward bias, for example, might occur when students who excel in math sort themselves into schools with more instruction time in math (Bingley et al., 2018).

Differences in instruction time occur both across and within countries. According to the curriculum data provided by TIMSS for each participating country, in some countries the curricula prescribe the percentage of math and science lessons as a proportion of the total instruction time, while in other countries there is no such official document. Overall, 70 (60) percent of the countries indicate that the curriculum prescribes a certain percentage of instruction time in math (science). In some countries it varies by state or school. Other countries define a range of percentages that should be devoted to instruction time in a given subject.

The information about a student's background, obtained from the student questionnaire, also includes the gender of the student (the variable *female*).

The data also contain information about the teacher: a variable *teacher female* (1 if the

⁴ I reverse the four labels of the original variable in TIMSS such that 1 equals "disagree a lot". The original variables used in TIMSS are shown in Appendix Table A.4. I generate the index according to Kling, Liebman & Katz (2006) by first standardizing each variable, then adding the four measures and dividing by four, the number of questions.

⁵ On average across all countries, the mean of instruction time is 4.10 hours in math and 2.04 hours in science.

Table 1: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
instruction time	2.955	1.513	0.017	10	230,142
female	0.489	0.5	0	1	230,142
teacherfemale	0.832	0.374	0	1	230,142
teacheraage	42.502	9.67	25	60	230,142
PD	0.493	0.5	0	1	230,142
education specialization	0.273	0.446	0	1	230,142
major degree	0.236	0.425	0	1	230,142
experience	0.519	0.5	0	1	230,142
tracking	0.141	0.348	0	1	219,271
remote	0.328	0.47	0	1	223,282
developed	0.635	0.481	0	1	230,142
developing	0.365	0.481	0	1	230,142

test scores and like subject are standardized (mean 0, std. dev. 1)

Notes: 4th grade sample in TIMSS 2015. Senate weights are used. PD stands for professional development.

teacher is female), the teacher's *age* as well as several indicators of the *teacher's qualifications*.⁶ The formal qualifications of teachers can be seen as one of two parts of teacher quality (Nilsen, Scherer & Blömeke, 2018).⁷ The formal qualifications of teachers consist of their educational background, measured by the years of experience, the highest level of education, i.e. highest degree, specialization in a subject and participation in professional development (Nilsen, Scherer & Blömeke, 2018).

Of the large amount of questions in the TIMSS teacher questionnaire, I use the following variables for teachers' formal qualifications in my analysis. First, the teachers are asked how many years they have been teaching (*experience*). On average, teachers have been teach-

⁶ One approach in the literature used to determine the quality of a teacher is the teacher value added, an approach introduced among others by Hanushek (1971). However, the teacher value added cannot be measured with the TIMSS data since students and teachers are only observed during one year. Besides, TIMSS only measures the cognitive abilities of students and not those of teachers. Thus, I cannot use teachers' skills or subject knowledge as a measure for teacher quality (e.g. Metzler & Woessmann, 2012).

⁷ The competencies of teachers are the second part of teacher quality. TIMSS also assesses teachers' competencies in teaching by asking them about the collaboration with other teachers, their motivation, their satisfaction with their job, their level of preparation, and their confidence (Nilsen, Scherer & Blömeke, 2018). For example, there is a number of item questions asking for the *preparedness* of the teacher in their subject. This measure covers a broad spectrum and teachers might be less likely to misreport the level of preparation if they can differentiate between several subtopics. However, this measure, like the other self-reported teacher competencies, might suffer from an endogeneity problem: A teacher might be better prepared if she has to teach several hours than if she only has to teach a few hours. This is why, I only focus on the formal qualifications of the teachers.

ing for 17.4 years across all countries, with a maximum of 60 years.⁸ I divide this variable by the median, which yields a variable indicating whether a teacher has more than 15 years of experience. Teachers are also asked about the highest level of formal education they have completed (according to the ISCED classification⁹), whether they majored in math or science and whether they specialized in math or science when their major or main area of study was teacher education. I generate the following variables: *Major degree* indicates whether the teacher has a Bachelor's degree (or higher) and whether she majored in math or science. Hence, this variable is an indicator of a teacher's knowledge about the subject: It can be assumed that a teacher knows the content of a subject when she majored in that subject. Studying mathematics as a major, for example, provides knowledge of the content required for teaching mathematics to students (Blömeke, Olsen & Suhl, 2016). *Education specialization* indicates whether a teacher has a specialization in the relevant subject if she has an educational background, i.e. a major in teacher education. This variable accounts for the fact that teacher education and pedagogy are also relevant. When teaching the fourth grade, it is especially important that teachers know how to teach and are good pedagogues (and not only know about the content of the subjects). The fourth aspect is whether a teacher has participated in *professional development* (PD) in the respective subject in the last two years. Categories of professional development are subject content, subject pedagogy/instruction or subject curriculum.

Overall, 24 percent of the teachers have a Bachelor's degree (or higher) with the relevant subject as their main subject and 27 percent have an educational background with a specialization in the subject. About half of all teachers participated in professional development in the last two years. However, these numbers conceal differences between developed and developing countries. In developing countries, more teachers have a Bachelor's degree (or higher) with the relevant subject as their main subject (33 percent) and more teachers have an educational background with a specialization in the relevant subject (44 percent). On the other hand, fewer teachers participated in professional development (45 percent). On average, 41 percent of the teachers in developing countries have more than 15 years of experience while

⁸ Less than 10 percent have zero to three years of experience which is often considered to matter the most.

⁹ The ISCED classification is suitable for indicating the level of education or the highest level of qualification. This facilitates international comparison.

about 53 percent in developed countries have more than 15 years of experience.

The data also contain information about the geographical area of the schools, i.e. whether it is a *remote* or urban region, and whether *tracking* is used. Headmasters are asked the following question (separately for math and science): "As a general school policy, is student achievement used to assign fourth grade students to classes?"

3 Empirical Strategy

Identifying the causal effect of instructional time on student attainment requires exogenous variation which is unrelated to any unobservable differences in schools and students (Rivkin & Schiman, 2015). Therefore, I use a student fixed effects model and exploit variation within-students between subjects. Due to their panel-like nature, the TIMSS data are particularly suitable since they provide two observations per student: A student's attainment, as well as the instruction time are reported for math and science. Using a student fixed effects model controls for unobservable student characteristics, such as unobserved ability and school differences in both subjects (Rivkin & Schiman, 2015). The attractiveness of this model lies in the fact that the students taking two subjects have the same general skill level and that the school environment is the same for both subjects (Rivkin & Schiman, 2015). Hence, no heterogeneity in terms of ability, habits or school quality will introduce biases to the estimates. Only subject-specific factors might have confounding effects. Therefore, I include subject-specific effects in the specification.

Based on this method, my regression equation is presented in equation 1.

$$testscore_{ijk} = \beta_1 H_{kj} + \beta_2 X_{ij} + \beta_3 Q_{lj} + \mu_i + \varepsilon_j + \eta_k + u_{ijk} \quad (1)$$

where $test\ score_{ijk}$ is the test score for student i in school j in subject k ($k \in \text{math, science}$). H_{kj} is the instruction time (in hours) in school j in subject k . X_{ij} are student char-

acteristics of student i in school j and Q_{lj} are teacher characteristics of teacher l in school j .¹⁰ μ_i are student fixed effects, capturing for example unobserved student and family background. η_k are unobserved subject-specific characteristics and ε_j are unobserved school characteristics. Controlling for student fixed effects already controls for school fixed effects.

The effect of instruction time might differ according to the quality of the teacher. An additional hour of instruction by an unqualified teacher or a teacher who does not know the subject matter well might not result in higher test scores since students might not learn more during this time. It might be more important how time at school is spent, how good teachers are at teaching and how motivated students are to learn and not so much the absolute amount of time (OECD, 2014). To assess this, I interact the instruction time variable H_{kj} with a quality indicator of teacher Q_{lj} , measured by her formal qualifications (equation 2).

$$test\ score_{ijk} = \beta_1 H_{kj} + \beta_2 X_{ij} + \beta_3 Q_{lj} + \beta_4 H_{kj} Q_{lj} + \mu_i + \varepsilon_j + \eta_k + u_{ijk} \quad (2)$$

The formal teacher qualifications consist of participation in professional development in the relevant subject, a Bachelor's degree (or higher) with the relevant subject as a major subject, teacher education with a specialization in the subject, as well as teaching experience (in years) as described in section 2.2. Regressions are run separately for each qualification.

Although I can difference out many unobservable factors with this approach, some unobserved factors that may confound the estimates could remain: students might choose themselves which school they want to attend or are assigned to schools based on subject-specific time of instruction (Lavy, 2015), which would invalidate the identification strategy. An upward bias can occur if students with high level of interest and ability in math sort themselves into schools that offer more instruction hours in math. A downward bias, on the other hand, can occur if students with low ability in math (or their parents) choose schools offering more math instruction hours because they need more instruction time to understand the content (Bingley et al., 2018). However, this is not an issue here, since instruction time is measured at the school

¹⁰ However, most student characteristics, such as gender or the number of books in the household, cannot be used as X_{ij} since these characteristics are the same across both subjects and hence do not change within students. Teacher characteristics, however, can change since a student might have different teachers in the two subjects.

by subject level. Thus, this approach helps to overcome the bias that might arise from the non-random allocation of instruction time. Tracking also appears to be more common within and not across schools, and often in higher grades when students choose a specialization (Lavy, 2015). Reassuringly, less than one fifth of the schools in the sample use tracking as a school policy.

Nevertheless, it must not go unnoticed that the estimates of the causal effect of instruction time on educational attainment might not be unbiased (Rivkin & Schiman, 2015) and that the strength of the identification of causal relationships might be lower than in randomized controlled trials, regression discontinuity designs, or for instrumental variables. Besides, measurement errors might be present in the self-reported measure of instruction time. A fixed effects model might reinforce these measurement errors (Angrist & Pischke, 2008).

A potential limitation of this strategy is that the effect of an additional hour of instruction is assumed to be the same in both subjects (math and science): β_1 does not vary by subject (Bingley et al., 2018). Another assumption is that the impact of instruction time does not take into account spillovers from instruction time in other subjects (Lavy, 2015): Instruction time in science does not influence a student's test score in math, and vice versa. Positive spillovers, for example, would lead to an underestimation of the effect. However, such spillovers might not be a cause for concern since the content of math and science lessons differs more in lower grades. At this stage, not many calculations and other mathematical concepts are used in science lessons.

4 Results

4.1 The Impact of Instruction Time

To determine the effect of instruction time on student achievement, one could regress test scores on instruction time using OLS. The results of this regression are presented in Table A.1: An increase in instruction time by one hour *ceteris paribus* leads to a decrease in test scores of about 0.02 to 0.04 standard deviations. However, many unobserved confounding factors will most likely bias the results of such a regression, even after controlling for students' demographic characteristics (column 3). Hence, the negative effect of instruction time is probably due to the

bias resulting from omitted variables.¹¹ If students sort (or are sorted) into schools or classes by ability, estimates might be biased upward (Bingley et al., 2018). On the other hand, compensatory teaching might lead to downward biased estimates (Bingley et al., 2018). Besides, instruction time might be correlated with unobserved factors influencing a student's achievement: Parents can decide where the family lives and which schools their children should attend. This decision is likely to be based on the perceived quality of the school, which also includes instruction time (Bingley et al., 2018).

To overcome this problem, I perform a regression as presented in equation 1 in the previous section. The dependent variable is the test score of a student, either in math or science. Table 2, columns 1 and 2 present the estimates of the fixed effects model. Standard errors are clustered at the school level and both specifications include subject fixed effects. All regressions are weighted using senate weights.

The results show that an additional hour of instruction time increases students' test scores by 0.03 standard deviations. Hence, students benefit in terms of test scores from additional instruction time. This finding is in line with the results of previous studies. In column 2, teacher characteristics are added.¹² The coefficient on instruction time remains almost exactly the same and is thus robust to including these additional variables.

By interacting instruction time with a student's gender, I examine whether there are heterogeneous effects with respect to a student's gender. The coefficient on instruction time for boys is slightly higher than in the baseline specification (Table 2, column 3), indicating that an additional hour of instruction time leads to an increase in test scores of male students by 0.038 standard deviations. The effect for girls is 0.025. Hence, boys seem to benefit more from additional instruction time. This finding is surprising since returns to schooling are often lower for boys than for girls, especially in low-income countries (Montenegro & Patrinos, 2014). A tentative explanation for this result could be that girls generally study more for school outside of school hours while boys study less in their time off from school. Consequently, boys might need to spend more time studying with a teacher than girls to improve their test scores. For

¹¹ Lavy (2012) finds the same negative effect for a naive OLS approach when analyzing instruction time in Israel.

¹² Note: student characteristics are not included in the FE specification since these cannot be estimated since they are the same for each student in both subjects.

girls, the time they spend on homework (and not instruction time) might play a greater role than for boys.

As explained above, additional instruction time can affect a student's attitude towards a subject. A student might become tired of a subject, leading to an aversion to the subject. Alternatively, a student might enjoy a subject even more when additional instruction time is used to deal with more specific content. Using the same specification as before, I therefore choose the variable *like subject*, an index comprising four questions, as an alternative dependent variable. The results are shown in Table 2, column 4: They suggest that additional instruction time leads to a more positive attitude towards the subject.

Table 2: Baseline results

VARIABLES	(1) test score	(2) test score	(3) test score	(4) like subject	(5) like subject
Instruction time	0.032*** (0.003)	0.032*** (0.003)	0.038*** (0.003)	0.051*** (0.006)	0.064*** (0.006)
female x instruction time			-0.013*** (0.002)		-0.026*** (0.004)
Observations	230,142	230,142	230,142	214,102	214,102
R-squared	0.923	0.923	0.923	0.608	0.608
Student FE	Yes	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes	Yes
Teacher Controls	No	Yes	Yes	Yes	Yes

Clustered standard errors at school level in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: Dependent variable: TIMSS student test score in maths and science in 4th grade in 2015. Instruction time is aggregated on school-by-subject level. Teacher controls are *teacher female* and *teacher age*. Regressions run as in equation 1 and senate weights are used.

4.2 Interaction between Instruction Time and Teacher Qualifications

The effect of instruction time might differ according to the quality of the teacher: If teachers do not actively use their time for teaching and if their teaching is not of high quality, the additional instruction time might not result in achievement gains for students. To assess this, I interact the

instruction time variable H_{kj} with a quality indicator of the teacher Q_{lj} , measured by her formal qualifications, according to equation 2, specified in section 3.

The results in Table 3 suggest that having a teacher who participates in professional development (column 1), having a teacher with a teacher training background and a specialization in the subject (column 2) and having a teacher who completed the relevant subject as the main subject with a Bachelor's degree (or higher) (column 3) enhance the positive effect of instruction time for students.

Table 3: Results for teachers' formal qualifications

VARIABLES	(1) test score	(2) test score	(3) test score	(4) test score
Instruction time	0.017*** (0.004)	0.027*** (0.003)	0.027*** (0.003)	0.036*** (0.003)
PD x instruction time	0.025*** (0.004)			
education specialization x instruction time		0.022*** (0.005)		
major degree x instruction time			0.027*** (0.005)	
experience x instruction time				-0.012*** (0.004)
Observations	230,142	230,142	230,142	230,142
R-squared	0.923	0.923	0.923	0.923
Student FE	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes
Teacher Controls	Yes	Yes	Yes	Yes
effect for high qualification	0.042*** (0.00351)	0.050*** (0.00501)	0.054*** (0.00475)	0.024*** (0.00397)

Clustered standard errors at school level in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Notes: Dependent variable: TIMSS student test score in maths and science in 4th grade in 2015. Instruction time is aggregated on school-by-subject level. Teacher controls are *teacher female* and *teacher age*. Regressions run as in equation 2 and senate weights are used. PD stands for professional development. Effect for high qualification shows the coefficient on instruction time when the respective teacher qualification (PD, education specialization, major degree, experience) equals 1.

The coefficients on the interaction terms are statistically significant and range from 0.022 to 0.027 standard deviations. The impact of one hour more instruction time is 0.027

standard deviations if the teacher does not have an educational background with a specialization, while it is 0.050 when having a teacher with exactly such a background (column 2). Similarly, the effect is 0.054 standard deviations when having a teacher who has a Bachelor's degree (or higher) with the relevant subject as their major (Table 3, column 3).

Figure 1 suggests that a teacher with a pedagogical background and specialization in the subject who teaches three hours has the same impact on student achievement as a teacher who teaches four hours but does not meet these criteria.¹³ When a teacher teaches the same students for many hours, i.e. more than three hours, it is especially important for the effect of instruction time on test scores whether the teacher is highly qualified.

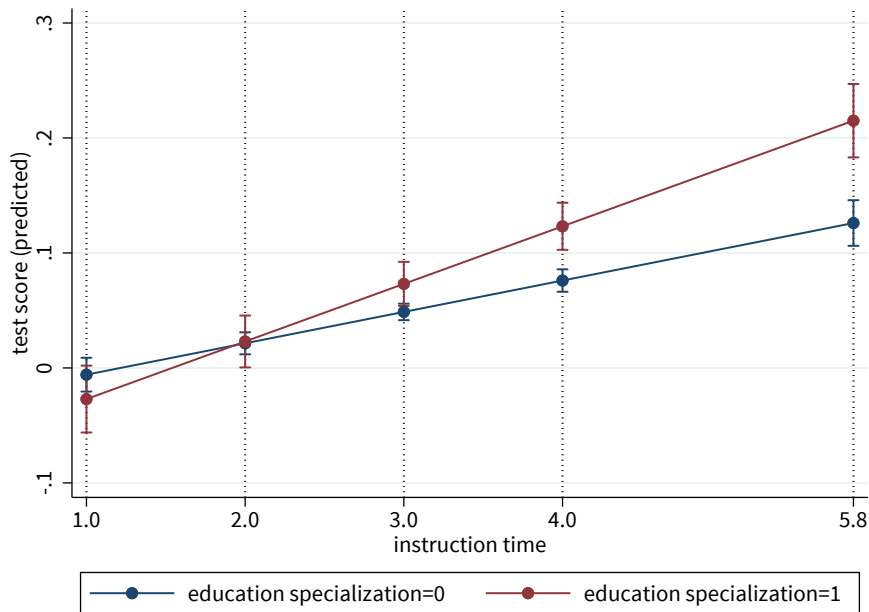


Figure 1: Marginal effects when education specialization is used as the teacher qualification measure in a regression as in equation 2 with senate weights used.

The results suggest that across all countries the impact of instruction time is enhanced by the fact that a teacher majored in the relevant subject, i.e. knowledge of the content, and that

¹³ The difference in the coefficients at three hours is statistically significant at the 10% level. The graph looks almost identical for having a teacher who has a Bachelor's degree (or higher) with the relevant subject as their main subject.

she has an educational background, i.e. pedagogy. However, in the case of having a teacher who has been teaching for many years, i.e. a more experienced teacher, the effect seems to be reduced (Table 3, column 4). This is surprising in that a more experienced teacher is expected to know how to use the time in a way that benefits the students. Clotfelter et al. (2007), for example, find that more experienced teachers improve student achievement more effectively than less experienced teachers. However, the results do not change when the number of years of experience is restricted to 40 years to exclude outliers.¹⁴

These results can complement the study by Rivkin and Schiman (2015): Instead of teacher qualifications, they examine the quality and environment of the classroom and find that this increases the effect of additional instruction time. Hence, it seems that both teacher quality and student behavior in the classroom play an important role.

4.3 Country Analysis

In addition, I examine whether the effect varies across country groups. Various countries from different continents participate in TIMSS. Hence, these countries differ in their culture, particularly in their educational culture and educational system. One main difference is that some participating countries are developing countries (such as Chile, Oman, and Saudi Arabia), while others are developed countries (such as France, the United States of America and Japan). The results for developing and developed countries are presented in Table A.2¹⁵: The effect of instruction time on students' test scores is statistically significant and higher in developed countries (0.06 standard deviations, column 1) than in developing countries (not statistically significant, column 3). The magnitude of the coefficient on instruction time in developed countries is similar to the coefficient determined by Lavy (2015) for OECD countries.

In both country groups, the effect is smaller for girls, and even negative for girls in developing countries (Table A.2, columns 2 and 4). A potential reason for this is the fact that education for girls is still not compulsory in developing countries. Hence, girls might react

¹⁴ When splitting the number of years of experience between very inexperienced teachers (less than three years) and experienced teachers, the interaction term is not statistically significant.

¹⁵ I group countries according to the WESP classification (United Nations (UN), 2014) which uses an exchange rate based method for aggregation. I combine countries in transition and developing countries. In the following, the term "developing countries" also includes countries in transition.

negatively to spending more time in school, knowing that they are needed at home for work or that they have to earn income that their families depend on (Glewwe & Kremer, 2006). Another reason might be that teachers spend the additional instruction time on boys and less on girls, leading to higher gains for boys than for girls.

More importantly, I also analyze how teacher qualifications interact with instruction time in these two groups of countries since the influence of teacher quality might vary from one educational system to another (Blömeke, Olsen & Suhl, 2016). Table 4 shows that in developing countries, having a teacher who participated in professional development (column 1) or having a teacher who completed the relevant subject as a main subject with a Bachelor's degree (or higher) (column 3) enhances the impact of instruction time.¹⁶ The impact of instruction time is negative when having a teacher who does not have a degree or who did not participate in professional development, but it becomes positive when having a better qualified teacher: Instruction time by a highly qualified teacher increases test scores by 0.016 (participation in professional development) or 0.024 (a Bachelor's degree (or higher) with the relevant subject as a major subject) standard deviations, while it seems to decrease test scores when having an unqualified teacher.

One potential reason for this might be that in developing countries students are needed at home to work. If these students have to stay at school longer with a teacher without good qualifications, they become distracted and unfocused, which leads to worse outcomes. An educational background with a specialization, however, does not have a moderating function for the effect of instruction time in developing countries (column 2). Hence, the results suggest that having studied the relevant subject as major subject is more important than having an educational background. In view of the observation that teachers lack adequate knowledge and that the quality of school education in developing countries is often low (Glewwe & Kremer, 2006), having studied the subject in question as a major could thus be an indicator of more substantial content knowledge about the subject and hence plays a moderating role for the impact of instruction time.

¹⁶ The coefficients on the interaction terms are larger in magnitude than in the specification including all countries.

Table 4: Results for developing countries: teacher qualifications

VARIABLES	(1) test score	(2) test score	(3) test score	(4) test score
Instruction time	-0.028*** (0.009)	-0.004 (0.008)	-0.014** (0.007)	0.008 (0.007)
PD x instruction time	0.043*** (0.008)			
education specialization x instruction time		0.010 (0.008)		
majordegree x instruction time			0.038*** (0.008)	
experience x instruction time				-0.018** (0.008)
Observations	86,902	86,902	86,902	86,902
R-squared	0.940	0.940	0.940	0.940
Student FE	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes
Teacher Controls	Yes	Yes	Yes	Yes
effect for high qualification	0.016** (0.00632)	0.007 (0.00691)	0.024*** (0.00711)	-0.010 (0.00776)
Clustered standard errors at school level in parentheses *** p<0.01, ** p<0.05, * p<0.1				

Notes: Sample restricted to developing countries: Armenia, Chile, Chinese Taipei, Georgia, Hong Kong SAR, Kazakhstan, Rep. of Korea, Oman, Qatar, Russian Federation, Saudi Arabia, Serbia, Singapore, United Arab Emirates and Turkey. Dependent variable: TIMSS student test score in maths and science in 4th grade in 2015. Instruction time is aggregated on school-by-subject level. Regressions run as in equation 2 and senate weights are used. PD stands for professional development. Effect for high qualification shows the coefficient on instruction time when the respective teacher qualification (PD, education specialization, major degree, experience) equals 1.

In developed countries, by comparison, having a teacher with an educational background (Table 5, column 2) seems to enhance the impact, as does having a teacher who completed the relevant subject as a major with a Bachelor's degree (or higher) (column 3). The coefficients are of the same magnitude and hence suggest that both subject knowledge and knowledge about pedagogical elements can enhance the impact of instruction time in developed countries. The coefficient on the variable participation in professional development is positive but not statistically significant.¹⁷

¹⁷ The equality of the interaction coefficients between developed and developing countries can be rejected for professional development, major degree (both at 1% level) and teacher education (at 10% level).

Table 5: Results for developed countries: teacher qualifications

VARIABLES	(1) test score	(2) test score	(3) test score	(4) test score
Instruction time	0.059*** (0.004)	0.060*** (0.003)	0.060*** (0.004)	0.063*** (0.004)
PD x instruction time	0.003 (0.005)			
education specialization x instruction time		0.017*** (0.006)		
major degree x instruction time			0.017*** (0.006)	
experience x instruction time				-0.005 (0.004)
Observations	143,240	143,240	143,240	143,240
R-squared	0.907	0.907	0.907	0.907
Student FE	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes
Teacher Controls	Yes	Yes	Yes	Yes
effect for high qualification	0.063*** (0.00410)	0.077*** (0.00692)	0.077*** (0.00606)	0.059*** (0.00430)
Clustered standard errors at school level in parentheses *** p<0.01, ** p<0.05, * p<0.1				

Notes: Sample restricted to developed countries: Australia, Bulgaria, Canada, Croatia, Cyprus, Czech Republic, Denmark, England, Finland, France, Germany, Hungary, Ireland, Italy, Japan, Lithuania, Netherlands, New Zealand, Northern Ireland, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden and the United States. Dependent variable: TIMSS student test score in maths and science in 4th grade in 2015. Instruction time is aggregated on school-by-subject level. Regressions run as in equation 2 and senate weights are used. PD stands for professional development. Effect for high qualification shows the coefficient on instruction time when the respective teacher qualification (PD, education specialization, major degree, experience) equals 1.

4.4 Robustness Checks

As a first robustness check, to rule out bias from unobserved teacher characteristics, I apply a within-teacher specification: I exploit the fact that students in primary schools are often taught by the same teacher in both subjects which is true for about 70 percent of the sample. In this specification, the estimates are no longer biased by any unobserved teacher traits that are subject invariant. Both, the coefficient on instruction time as well as the coefficients on the interaction terms including the measure of teacher qualifications, are robust in this variant of

the specification (Table 6). The coefficients on instruction time are marginally larger than the ones in the main specification.

Table 6: Results - within teacher specification

VARIABLES	(1) test score	(2) test score	(3) test score	(4) test score	(5) test score
Instruction time	0.035*** (0.003)	0.030*** (0.004)	0.033*** (0.003)	0.031*** (0.003)	0.039*** (0.004)
PD x instruction time		0.010** (0.004)			
education specialization x instruction time			0.016*** (0.005)		
major degree x instruction time				0.028*** (0.005)	
experience x instruction time					-0.010*** (0.004)
Observations	163,994	163,994	163,994	163,994	163,994
R-squared	0.922	0.922	0.922	0.922	0.922
Student FE	Yes	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes	Yes
Teacher Controls	No	No	No	No	No

Clustered standard errors at school level in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: Sample restricted to students with only one teacher in both subjects. Dependent variable: TIMSS student test score in maths and science in 4th grade in 2015. Instruction time is aggregated on school-by-subject level. Regressions run as in equation 2 and senate weights are used. PD stands for professional development. Effect for high qualification shows the coefficient on instruction time when the respective teacher qualification (PD, education specialization, major degree, experience) equals 1.

Further, I assess the robustness of my results by piecewise excluding each country individually from the main analysis to see if outliers drive the results, both for instruction time and teacher qualifications. Both the coefficients on instruction time and teacher qualifications remain stable across all 42 regressions. The impact of instruction time on student achievement ranges from 0.03 (excluding Northern Ireland) to 0.04 (excluding Chinese Taipei) (detailed results are available on request).

I also investigate whether the relationship between instruction time and test scores is non-linear. Similar to Rivkin and Schiman (2015), I examine whether the returns to additional instruction time are diminishing. For this purpose, I add a squared term of instruction time (in

hours) to the baseline specification (Table 7, column 1). The coefficient on the squared term is slightly positive but not statistically significant. Therefore, this result does not support the hypothesis of diminishing returns. The coefficient on the level of instruction time is slightly smaller but is still robust to including the squared term.

A threat to identification and a source of selection bias might arise from tracking students into different classes within school (Lavy, 2015). To assess whether the results are robust to this threat, I present results for two sub-samples, that are characterized by whether schools pursue this school policy of tracking. Overall, the incidence of using tracking as a school policy is low with 16.88 percent across all schools in the sample. Columns 2 and 3 of Table 7, show that additional instruction time has a positive, statistically significant effect in both sub-samples: Reassuringly, the coefficient in the non-tracking sample is similar to that in the main specification (column 3).

Table 7: Robustness Checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	test score	tracking: yes test score	tracking: no test score	remote: yes test score	remote: no test score	test score	test score
Instruction time	0.019** (0.008)	0.040*** (0.008)	0.029*** (0.004)	0.038*** (0.006)	0.027*** (0.004)	0.029*** (0.003)	0.040*** (0.006)
Squared instruction time	0.002 (0.001)						
Observations	230,142	24,374	173,202	58,448	164,834	230,142	66,148
R-squared	0.923	0.944	0.918	0.915	0.925	0.923	0.928
Student FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Clustered standard errors at school in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: Dependent variable: TIMSS student test score in maths and science in 4th grade in 2015. Instruction time is aggregated on school-by-subject level in columns 1 to 5 and 7, not in column 6. Teacher controls are *teacher female* and *teacher age*. Regressions run as in equation 1 and senate weights are used.

Furthermore, I check whether there are differences with respect to the type of geographic area in which the school is located. In remote or rural areas, students or rather their parents often have no alternative in the choice of school as there is often only one school in places with 3,000 or fewer inhabitants. This mitigates the problem of non-random allocation to

schools. The results in column 4 of Table 7 demonstrate that the effect in rural areas is about 0.038 standard deviations, which is slightly higher than the effect in the main specification.

The use of an alternative specification for instruction time, i.e. the exact report of the teacher (no aggregation on school by subject level) yields a coefficient that is almost identical to that in the main specification (Table 7, column 6). Overall, the findings in columns 2 to 6 indicate that the results are robust to sorting and tracking.

A possible concern might be that some teachers do not teach the subject they are supposed to teach. For example, a teacher might be teaching science when she should be teaching math, possibly because she prefers science over math or because she has majored in science but is required to also teach math. To alleviate this concern, I only focus on the subsample of students who have two separate teachers for math and science. It can be assumed that in this case a teacher only teaches the subject that she is supposed to teach. Conversely, a teacher could spend more time teaching math if she teaches both subjects to the same students. The coefficient on instruction time in the subsample which includes only the students taught by two separate teachers increases slightly (Table 7, column 7). This suggests that the main effect that I estimate in Table 2 might be slightly downward biased.

5 Conclusion

Using a fixed effects model and within-student between-subject variation, I show that instruction time positively affects students' test scores. On average across all countries, I find that an additional hour of instruction time leads to 0.03 standard deviations higher test scores. More importantly, I find that teacher quality, measured by teachers' formal qualifications, such as teacher training with a specialization in the relevant subject, a Bachelor's degree (or higher) with the relevant subject as their main subject, and participation in professional development, plays a moderating role for the effect of instruction time on student achievement: The effect is larger for students with better qualified teachers. This is especially relevant in developing countries, where the effect of instruction time on student achievement is on average not statistically significant and close to zero. However, instruction time with a highly qualified teacher also increases test scores by 0.02 standard deviations in developing countries.

The estimates on instruction time are about the same magnitude as those of Rivkin and Schiman (2015), but slightly smaller than those of Lavy (2015) who finds effect sizes of 0.06 to 0.08 standard deviations. In line with Lavy (2015), I also find that the impact of instruction time is lower for developing countries, even close to zero and statistically insignificant.

Some further points are important to consider when interpreting my results. The first thing to consider is whether extending the instruction time in a subject increases the overall time that students spend in school and whether this is at the expense of reducing breaks and vacation time (Farbman (2015), Jarett et al. (1998)). For example, more instruction time in math at the expense of instruction time in another subject, e.g. arts and music, might improve test scores in math, especially if the lessons are given by a highly qualified teacher. On the other hand, this could affect students' development in terms of creativity, physical activity and health, especially in primary school and especially students from lower socioeconomic backgrounds since they often do not have access to voluntary education outside of school.

In addition to the education and training of a teacher, other school inputs and the behavior of teachers and students might affect the productivity of instruction time and make it rather endogenous (Lavy, 2015). Another factor might be how autonomously a school principal can make decisions. According to Lavy (2015), greater autonomy in decisions about hiring or firing teachers can lead to a better fit between schools and teachers, and thus to teachers making more effort. As already mentioned, Rivkin and Schiman (2015) look at classroom quality and environment and find that this is important for student achievement. This finding is in line with my finding that higher qualified teachers have a larger impact on student achievement. Thus, the classroom environment in terms of student behavior as well as the quality of the teacher matter for the effect of instruction time.

When assessing the effectiveness of schools, more instruction time is not the only relevant component. However, my results and those of previous research suggest that instruction time is one of the key factors in promoting student achievement and that the quality of teachers, in particular the qualifications of teachers, can enhance the influence of instruction time on student achievement. The positive effect of instruction time on students' test scores and its interaction with teachers' qualifications is of particular importance for policy decisions.

A slight increase in instruction time is most likely simple to implement. Hence, increasing instruction time would be an easy way to improve student achievement. However, since additional instruction time has to be financed, policymakers need to know whether this money is being invested effectively. My results suggest that it is the combination between instruction time and the quality of a teacher that is relevant to student achievement.

References

- Andersen, S. C., Humlum, M. K., & Nandrup, A. B. (2016). Increasing Instruction Time in School Does Increase Learning. *Proceedings of the National Academy of Sciences*, vol. 113, no. 27, pp. 7481-7484
- Angrist, J. D., & Pischke, J. S. (2008). Mostly Harmless Econometrics: An empiricist's companion. Princeton University Press
- Blömeke S., Olsen R.V., & Suhl U. (2016). Relation of Student Achievement to the Quality of Their Teachers and Instructional Quality. In: Nilsen T., Gustafsson JE. (eds) Teacher Quality, Instructional Quality and Student Outcomes. IEA Research for Education (A Series of In-depth Analyses Based on Data of the International Association for the Evaluation of Educational Achievement (IEA)), vol. 2, Springer, Cham.
- Bingley, P., Heinesen, E., Krassel, K. F., & Kristensen, N. (2018). The Timing of Instruction Time: Accumulated Hours, Timing and Pupil Achievement, *IZA Discussion Papers*, No. 11807
- Carroll, J. B. (1989). The Carroll Model: A 25-year retrospective and prospective view. *Educational Researcher*, vol. 18, no. 1, pp. 26-31
- Cattaneo, M. A., Oggenfuss, C., & Wolter, S. C. (2017). The More, the Better? The impact of instructional time on student performance. *Education Economics*, vol. 25, no. 5, pp. 433-445
- 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
- Cunha, F., Heckman, J. J., Lochner, L., & Masterov, D. V. (2006). Interpreting the Evidence on Life Cycle Skill Formation. *Handbook of the Economics of Education*, 1, pp. 697-812
- Dahmann, S. C. (2017). How Does Education Improve Cognitive Skills? Instructional time versus timing of instruction. *Labour Economics*, vol. 47, pp. 35-47
- Farbman, D. (2015). The Case for Improving and Expanding Time in School: A Review of Key Research and Practice. *National Center on Time & Learning*
- Glewwe, P., & Kremer, M. (2006). Schools, teachers, and education outcomes in developing countries. *Handbook of the Economics of Education*, vol. 2, pp. 945-1017
- Goodman, J. (2014). Flaking Out: Student absences and snow days as disruptions of instructional time. *NBER working paper series*, No. w20221. National Bureau of Economic Research
- Hanushek, E. A. (1971). Teacher Characteristics and Gains in Student Achievement: Estimation using micro data. *American Economic Review*, vol. 61, no. 2, pp. 280-288
- Hanushek, E. A., & Rivkin, S. G. (2006). Teacher quality. *Handbook of the Economics of Education*, vol. 2, pp. 1051-1078.

- Huebener, M., Kuger, S., & Marcus, J. (2017). Increased Instruction Hours and the Widening Gap in Student Performance. *Labour Economics*, vol. 47, pp. 15-34
- Jarrett, O. S., Maxwell, D. M., Dickerson, C., Hoge, P., Davies, G., & Yetley, A. (1998). Impact of Recess on Classroom Behavior: Group effects and individual differences. *The Journal of Educational Research*, vol. 92, no. 2, pp. 121-126
- 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
- Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, vol. 75, no. 1, pp. 83-119
- Lavy, V. (2012). Expanding School Resources and Increasing Time on Task: Effects of a policy experiment in Israel on student academic achievement and behavior. *NBER working paper series*, No. w18369. National Bureau of Economic Research
- 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. F397-F424
- Marcotte, D. E. (2007). Schooling and Test Scores: A mother-natural experiment. *Economics of Education Review*, vol. 26, no. 5, pp. 629-640
- Martin, M. O., Mullis, I. V. S., & Hooper, M. (Eds.). (2016). *Methods and Procedures in TIMSS 2015*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <http://timssandpirls.bc.edu/publications/timss/2015-methods.html> [Accessed: 28 May 2019]
- 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.
- Montenegro, C. E., & Patrinos, H. A. (2014). Comparable Estimates of Returns to Schooling Around the World. The World Bank
- National Center on Time and Learning. (2017). Why Time Matters. Available online: <http://www.timeandlearning.org/?q=why-time-matters-0> [Accessed: 22 April 2019]
- Nilsen, T., Scherer, R., & Blömeke, S. (2018). 3. The relation of science teachers' quality and instruction to student motivation and achievement in the 4th and 8th grade: A Nordic Perspective. *Northern Lights on TIMSS and PISA 2018*, pp. 61-93
- Organisation for Economic Co-operation and Development (OECD). (2014). Education Indicators in Focus: How much time do primary and lower secondary students spend in the classroom? Paris: OECD

- Patall, E. A., Cooper, H., & Allen, A. B. (2010). Extending the School Day or School Year: A systematic review of research (1985–2009). *Review of Educational Research*, vol. 80, no. 3, pp. 401-436
- Rivkin, S. G., & Schiman, J. C. (2015). Instruction Time, Classroom Quality, and Academic Achievement. *The Economic Journal*, vol. 125, no. 588, pp. F425-F448
- TIMSS. (2015). Supplement 1: International Version of the TIMSS 2015 Context Questionnaires. Available online: https://timss.bc.edu/timss2015/international-database/downloads/T15_UG_Supplement1.pdf [Accessed: 28 May 2019]
- TIMSS. (2018). About TIMSS & PIRLS International Study Center. Available online: <https://timss.bc.edu/about.html> [Accessed: 27 March 2019]
- United Nations (UN). (2014). World Economic Situation and Prospects 2014 - Country Classification. Available online: https://www.un.org/en/development/desa/policy/wesp/wesp_current/2014wesp_country_classification.pdf [Accessed: 28 May 2019]
- UNESCO. (2019). Instructional Time Definition. Available online: <http://uis.unesco.org/en/glossary-term/instructional-time> [Accessed: 22 April 2019]
- Woessmann, L. (2003). Schooling Resources, Educational Institutions and Student Performance: The international evidence. *Oxford Bulletin of Economics and Statistics*, vol. 65, no. 2, pp. 117–170

A Appendix

A.1 Tables

Table A.1: OLS Regression

VARIABLES	(1) test score	(2) test score	(3) test score
Instruction time	-0.031*** (0.008)	-0.027*** (0.008)	-0.017** (0.007)
Observations	230,142	230,142	226,250
R-squared	0.001	0.004	0.114
Student FE	No	No	No
Subject FE	Yes	Yes	Yes
Teacher Controls	No	Yes	Yes
Student Controls	No	No	Yes

Clustered standard errors at school level in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Notes: Dependent variable: TIMSS student test score in maths and science in 4th grade in 2015. Instruction time is aggregated on school-by-subject level. Simple OLS regressions are run and senate weights are used.

Table A.2: Country Analysis

VARIABLES	(1) developed countries test score	(2) developed countries test score	(3) developing countries test score	(4) developing countries test score
Instruction time	0.062*** (0.003)	0.068*** (0.004)	0.001 (0.006)	0.008 (0.006)
female x instruction time		-0.013*** (0.002)		-0.013*** (0.004)
Observations	143,240	143,240	86,902	86,902
R-squared	0.907	0.907	0.940	0.940
Student FE	Yes	Yes	Yes	Yes
Subject FE	Yes	Yes	Yes	Yes
Teacher Controls	Yes	Yes	Yes	Yes

Clustered standard errors at school level in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Notes: Dependent variable: TIMSS student test score in maths and science in 4th grade in 2015. Instruction time is aggregated on school-by-subject level. Regressions run as in equation 1 and senate weights are used. Countries are grouped into developed and developing countries according to the WESP classification. Developed countries: Australia, Bulgaria, Canada, Croatia, Cyprus, Czech Republic, Denmark, England, Finland, France, Germany, Hungary, Ireland, Italy, Japan, Lithuania, Netherlands, New Zealand, Northern Ireland, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden and the United States. Developing countries: Armenia, Chile, Chinese Taipei, Georgia, Hong Kong SAR, Kazakhstan, Rep. of Korea, Oman, Qatar, Russian Federation, Saudi Arabia, Serbia, Singapore, United Arab Emirates and Turkey.

A.2 List of Participating Countries & Variables used from TIMSS

Table A.3: List of Participating Countries

Country No.	Country Name	WESP class.	OECD country
51	Armenia	in transition	non-OECD
36	Australia	developed	OECD
100	Bulgaria	developed	non-OECD
124	Canada	developed	OECD
152	Chile	developing	OECD
158	Chinese Taipei	developing	non-OECD
191	Croatia	developed	non-OECD
196	Cyprus	developed	non-OECD
203	Czech Republic	developed	OECD
208	Denmark	developed	OECD
926	England	developed	OECD
246	Finland	developed	OECD
250	France	developed	OECD
268	Georgia	in transition	non-OECD
276	Germany	developed	OECD
344	Hong Kong SAR	developing	non-OECD
348	Hungary	developed	non-OECD
372	Ireland	developed	OECD
380	Italy	developed	OECD
392	Japan	developed	OECD
398	Kazakhstan	in transition	non-OECD
410	Korea, Rep. of	developing	OECD
440	Lithuania	developed	OECD
528	Netherlands	developed	OECD
554	New Zealand	developed	OECD
928	Northern Ireland	developed	OECD
578	Norway	developed	OECD
512	Oman	developing	non-OECD
616	Poland	developed	OECD
620	Portugal	developed	OECD
634	Qatar	developing	non-OECD
643	Russian Federation	in transition	non-OECD
682	Saudi Arabia	developing	non-OECD
688	Serbia	in transition	non-OECD
702	Singapore	developing	non-OECD
703	Slovak Republic	developed	OECD
705	Slovenia	developed	OECD
724	Spain	developed	OECD
752	Sweden	developed	OECD
784	United Arab Emirates	developing	non-OECD
792	Turkey	developing	non-OECD
840	United States	developed	OECD

Notes: Countries are grouped into developed and developing countries according to the WESP classification.

Table A.4: List of variables for a student's attitude and teacher qualifications

TIMSS Variable Name	Question Text	Answer Choices (adjusted labels)
<i>like subject</i>		
ASBM01A	I enjoy learning mathematics	<i>Answer choices for each question:</i> - Agree a lot (4) - Agree a little (3) - Disagree a little (2) - Disagree a lot (1)
ASBM01D	I learn many interesting things in mathematics	
ASBM01E	I like mathematics	
ASBM01I	Mathematics is one of my favorite subjects	
ASBS04A	I enjoy learning science	
ASBS04D	I learn many interesting things in science	
ASBS04E	I like science	
ASBS04I	Science is one of my favorite subjects	
<hr/>		
<i>Professional development:</i> In the past two years, have you participated in professional development in any of the following?		
ATBM09A/ATBS08A	Mathematics/science content	<i>Answer choices for each question:</i> - Yes - No
ATBM09B/ATBS08B	Mathematics/science pedagogy / instruction	
ATBM09C/ATBS08C	Mathematics/science curriculum	
<hr/>		
<i>Specialization:</i> During your <post-secondary> education, what was your major or main area(s) of study?		
ATBG05AC	Mathematics	<i>Answer choices for each question:</i> Yes or No
ATBG05AD	Science	
<hr/>		
<i>Specialization:</i> If your major or main area of study was education, did you have a <specialization> in any of the following?		
ATBG05BA	Mathematics	<i>Answer choices for each question:</i> Yes or No
ATBG05BB	Science	
<hr/>		
<i>Experience</i>		
ATBG01	By the end of this school year, how many years will you have been teaching altogether?	< years >
<hr/>		
<i>Degree</i>		
ATBG04	What is the highest level of formal education you have completed?	<ul style="list-style-type: none">- did not complete upper secondary education (ISCED level 3)- upper secondary education (ISCED level 3)- post-secondary, non-tertiary education (ISCED level 4)- short-cycle tertiary education (ISCED level 5)- Bachelor's or equivalent level (ISCED level 6)- Master's or equivalent level (ISCED level 7)- Doctor or equivalent level (ISCED level 8)

Notes: Questions on students' motivation and attitude (student questionnaire) as well as on teacher qualifications (teacher questionnaire) from 4th grade sample in TIMSS 2015.