LUND UNIVERSITY
School of Economics and Management

Homework assignment and student achievement

Evidence from TIMSS 2019

Jens Mattsson 930522-7036[1]

Second year Master thesis

Supervisor: Petra Thiemann
Department of Economics
Lund University

2021-08-18

[1]E-mail address: jens1mattsson@gmail.com

Abstract

Homework is widely used in most schools across the world to complement classroom teaching, increase knowledge retention, and foster good study habits. The research on the effect of homework assignment on student achievement is, however, inconclusive with earlier research indicating a positive relationship and newer studies giving a more nuanced picture. This thesis uses data from the 2019 wave of the Trends in International Mathematics and Science Study (TIMSS) to examine how teacher-reported homework frequency and homework amount relate to student test scores in 58 countries for grade 4 and 26 countries for grade 8. The TIMSS data is treated as a panel where the effect of homework assignment is identified using the within-student between-subject variation in homework assignment between mathematics and science. Fixed student, class, school, and country characteristics can thereby be controlled for, gaining an estimator that comes significantly closer to a causal estimator than much of the previous research. Results indicate that homework assignment has a null or negative effect on 4^{th} grade students' test scores and a null effect on 8^{th} grade students' test scores. The nonlinear relationship hypothesis, with an optimal level of homework assignment above no homework, is not supported by the results, with relationships either being linear or homogeneous compared to receiving no homework. No heterogeneity is found when splitting the sample into OECD and non-OECD countries for grade 8. In grade 4, the estimated negative effect comes from the non-OECD countries, while a null effect is estimated in the OECD countries. Based on the results, education policy makers are urged to not treat increased homework assignment, without considering the quality of the assigned homework, as a way to increase student achievement.

Keywords: Homework, Student achievement, TIMSS 2019, First difference model

Table of Contents

1. Introduction

Human capital in the form of education level often enter endogenous growth theory models as key variables that countries can affect in order to increase economic growth and the overall welfare of the country (Hanushek and Wößmann 2012; Lucas 1988; Romer 1990). Among ways to increase education levels, homework is one of the most widely debated topics in both academia and in broader society, with no clear consensus on how much, if any, homework should be given to students. Therefore it becomes crucial to determine how homework assignment practices can best be altered by education policy makers to increase education outcomes.

The definition of homework most commonly employed in the literature, and which will be used in this thesis, is "tasks assigned to students by school teachers that are meant to be carried out during nonschool hours" (Cooper 1989, p. 7). This definition thus excludes students' self study time outside of tasks given by teachers.[2] Homework is used extensively by teachers to ensure that students get repetition of what is taught in school and as a way to cover topics outside of what is taught in class. There is, however, large variation across time and between teachers, both within and between countries, in how much homework is given to students (Falch and Rønning 2012; Jerrim, Lopez-Agudo, and Marcenaro-Gutierrez 2020; Zhu 2015). For instance, in recent years there are indications of a decline in the amount of homework given to students in the countries that have participated in Trends in International Mathematics and Science Study (TIMSS) (Zhu 2015, pp. 216–218).

Proponents of homework argue that it helps foster better study habits, independent problem-solving, and discipline, as well as increases the amount of knowledge that students acquire and retention thereof (Bas et al. 2017; Bempechat 2004; Cooper et al. 2006; Zhu and Leung 2012). If this is correct, these features should lead to measurable positive effects on student achievement. Furthermore, if homework has positive effects on students' learning, it is a cost-effective way, compared to alternatives like decreasing class sizes, hiring more teachers, or buying computers to all students, to increase the education level and human capital of a country (Eren and Henderson 2011).

Critics of homework, or its excessive usage, emphasize the possibility of academic satiation, physical and emotional fatigue, and stress that too much homework can result in for the students (Cooper et al. 2006; Jerrim, Lopez-Agudo, and Marcenaro-Gutierrez 2020; Zhu and Leung 2012). They also stress loss of time for leisure and community activities, and unequal circumstances in home environments between students from academic and

[2] See Arregui Alegria and Gatykaev (2019) for a a study on homework defined as all study time outside of school.

non academic backgrounds. Moreover, homework assignment can principally be divided into three parts: assignment frequency, length of the assignment, and quality of the assignment (Trautwein 2007). If the quality of the assigned homework is poor, so that it does not help students increase or retain knowledge and develop important skills, it risks becoming solely a source of stress for students (Dettmers, Trautwein, Lüdtke, et al. 2010; Rosário et al. 2018; Trautwein and Lüdtke 2009). It is also possible that teachers face incentives to outsource the education of their students to the parents via homework assignments. One example of such an incentive is that teachers can then focus on what they think is most important and spend more time on these topics in class. If this is the case, the risk of poor homework quality is likely higher.

Previous research relating homework to student achievement is largely inconclusive. Much of the earlier research, that mostly used data from the US, found large positive effects for secondary school students but null effects for primary school students (Cooper 1989; Cooper et al. 2006). Much of this research was, however, purely correlational and did not control for potentially endogenous variables such as school level variables and student ability and motivation. In recent years the available evidence is more unclear, with some studies finding much smaller positive effects, null effects, or even negative effects (Chin et al. 2020; Dettmers, Trautwein, and Lüdtke 2009; Jerrim, Lopez-Agudo, and Marcenaro-Gutierrez 2020; Scheerens and Hendriks 2014; Zhu and Leung 2012). The true effect of homework on student achievement is also obscured by the fact that researchers use different operationalizations of homework. The most common ones are homework frequency and length reported by students or by teachers and student-reported time or effort spent on homework. For education policy makers, it is easier and more effective to influence the homework assignment practices of teachers, which is why this thesis will focus on teacher-reported homework assignment practices. Furthermore, teacher-reported measures are more exogenous and less prone to measurement errors (Desimone et al. 2010).

In this thesis, data from TIMSS 2019 is used to estimate the effect of homework frequency and homework amount (a combination of frequency and length) on student test scores in 58 countries in 4th grade and 26 countries in 8th grade. To overcome endogeneity issues affecting much of the previous literature, the quasi-panel structure of the TIMSS data, where each student is tested in both mathematics and science, is exploited. The teacher in each subject answered questions on how often homework was given and how extensive it was. Thus, the student fixed effects estimation strategy, which has previously been utilized in similar research (Eren and Henderson 2011; Falch and Rønning 2012; Jerrim, Lopez-Agudo, and Marcenaro-Gutierrez 2020; Rønning 2010), can be used. The estimation strategy identifies the effect of homework on test scores using the within-student between-subject variation in homework assignment. This approach eliminates most of the potential confounding variables, such as students with higher or lower innate ability or motivation self-selecting into schools with more homework, that would bias results of a naive Ordinary Least Squares (OLS) estimation.

The empirical results indicate that homework (both measured as frequency and amount) has a negative or null effect on student test scores in grade 4, and a null effect in grade 8. For grade 4, the results differed for OECD countries (null effect) and non-OECD countries (negative effect). Education policy makers are therefore cautioned to see increased homework as a solution to decreasing student achievement or a way to increase overall stagnant student test scores.

The contribution to the existing literature of this thesis is twofold. First, this is the first time that the within-student between-subjects identification strategy has been used to estimate the effect of teacher-reported homework frequency on student test scores in both 4^{th} and 8^{th} grade, for such a large sample of countries. Second, to the best of the author's knowledge, this is also the first time that the identification strategy has been used to estimate the effect of categorical homework amount, as defined in Martin, Mullis, Gonzalez, et al. (2004) and described in Section 3.2, on student test scores.

The remainder of this thesis is organized as follows. Section 2 explores the previous literature relating homework to student test scores. Section 3 introduces the TIMSS data and describes the operationalization of homework and student achievement. Section 4 describes the empirical strategy. Section 5 presents the regression results and several robustness checks. Section 6 discusses the results and their implications.

2. Review of the literature relating homework to student achievement

There is an extensive literature relating homework to student achievement. Most of this research (especially before 2000) has used data solely from the US. While these early studies mostly find positive effects, results from the last 20 years give a more nuanced picture, to some degree because of the availability of better data, but also because of the use of new microeconometric methods that are able to come closer to a causal estimate.

The majority of the earlier research was summarized in two large meta-studies covering nearly 120 papers written 1960–1986 (Cooper 1989) and a follow up synthesis covering the period 1987–2003 (Cooper et al. 2006). These meta-analyses concluded that there is consistent evidence for a positive relationship between homework and student achievement in secondary school, both on the extensive and the intensive margin, and also both within and across research design types. For the extensive margin papers comparing no homework to any level of homework, the average effect size was $d = 0.21$ in Cooper (1989) and $d = 0.60$ in Cooper et al. (2006). Cohen's "d" is a measure of the standardised difference between two means, often used to summarize results from many studies in meta-analyses. The average correlation coefficient for the intensive margin papers were $r = 0.19$ in Cooper (1989) and $r = 0.24$ in Cooper et al. (2006). However, for elementary school students the relationship was zero or close to zero.

The conclusions drawn in these two meta-analyses have, however, been criticised extensively (among others, Trautwein and Köller 2003; Trautwein 2007; Dettmers, Trautwein, and Lüdtke 2009). Trautwein and Köller (2003), for instance, pointed out three main problems that might affect internal validity for the majority of the studies included in Cooper's meta analyses.[3] First, concerns were raised regarding low sample sizes, the employed randomization procedures, the data handling, and the lack of control for pre-treatment differences in most of the experimental design studies. These factors, they argued, could have lead to systematic bias in the results.

Second, a large amount of the papers included in the meta-analyses were purely correlational, failing to control for potential confounding variables such as innate student ability or teacher quality, and could therefore not make any credible claims to estimate a causal relationship (see also Gustafsson (2013) on this topic). Trautwein (2007), for example, showed that when students' previous test scores (as a proxy for overall ability) were included in the regression, the estimated effect of homework decreased, indicating that there was an upwards bias when not controlling for this variable. Moreover, many studies also failed to control for any hierarchical structure of the data; for example controlling for class and school variables that have been shown to explain a significant percentage of the variation in homework assignment (Rønning 2010; Trautwein and Köller 2003).

Third, most studies used time spent on homework reported by students as the measure of homework. As discussed above, homework is a multifaceted teaching method consisting of at least six main parts: frequency, length, and quality of the assigned homework, as well as student time spent on homework, motivation, and effort to complete the assigned tasks (Trautwein 2007, p. 386). To only consider time spent on homework as reported by students leaves out a large part of what defines homework. It also risks measuring something different than intended; if a student reports spending more time on a given homework task, this might indicate that the student has concentration or motivation problems rather than indicating conscientiousness and hard work that would result in more acquired knowledge (Trautwein 2007, p. 373). Trautwein and Köller (2003) concluded that there likely is a positive relationship between actively working on homework and student achievement. However, the relationship for *time needed* for homework is null or even negative (Trautwein and Köller 2003, p. 132).

Variables that education policy can realistically influence are the frequency, length,

[3] Cooper (1989) and Cooper et al. (2006) also raised concerns regarding some of these design flaws. Still, they maintained that the included studies had different design flaws, making the positive relationship across all design types more credible, warranting their conclusions.

and quality of the assigned homework; not so easily directly affect the time students spend on the given homework. Quality of the assigned homework is a crucial factor since it can be viewed as the mediating component, deciding whether assignments help students develop and/or retain knowledge and important skills, or if it simply acts as a time drain, only leading to increased stress and fatigue for the students (Dettmers, Trautwein, and Lüdtke 2009; Rosário et al. 2018; Trautwein and Lüdtke 2009). Both measuring and affecting homework quality is, nevertheless, extremely difficult and a homework quality measure is seldom available for researchers, especially in large student assessments like TIMSS.

The frequency and length of assigned homework are more easily adjusted through policy, affecting how much time students spend on homework (Trautwein and Köller 2003). Despite this, simply considering only the frequency or length might also be too simplistic. Homework *amount*, i.e. frequency and length combined, could be the more relevant measure to consider (Gustafsson 2013). Two ways in which homework amount can be defined are the number of minutes per week of assigned homework (Eren and Henderson 2011; Gustafsson 2013) and as a categorical variant that groups low length with low frequency and high length with high frequency (Martin, Mullis, Gonzalez, et al. 2004). Both definitions recognize that receiving homework every day with short length is not the same as if the expected time of completion is much longer. Likewise, receiving short homework is not expected to have the same effect if teachers give out assignments once a week compared to every day. What much of previous research does is to only consider either frequency or length separately, which might conceal a more relevant relationship if homework amount is used instead (Jong et al. 2000; Martin, Mullis, Gonzalez, et al. 2004). The second definition, which is preferred by the author, also accounts for the likely difference in effect if homework expected to take 20 minutes to complete is given every day or if homework is only given once per week with a length of 100 minutes. Even though the weekly amount in minutes is the same, in the first case the focus is likely on repetition of topics covered in class while in the second case the assignment is more likely a larger project, possibly on a new topic not covered in class. This distinction is important since the effects of homework have been found to be highest when homework involves practice or rehearsal of what is taught in class as well as rote learning (Hattie 2012).

Another problem of using student-reported homework measures is self-reporting bias. Previous research has shown that the correlations between student- and teacher-reported measures such as goal structures in classrooms, teaching practices (that homework is an example of), teacher-student relationship, motivation, and social engagement are typically quite low, between 0.3 and 0.5 (ACT 2013; Buckley and Krachman 2016; Desimone et al. 2010; Jong et al. 2000).

The meta-analysis on the subject of teacher-reported vs student-reported measures performed by Desimone et al. (2010) concluded that teachers' self-reported teaching practices corresponded with both classroom observations and teacher logs, and that they are "quite valid and reliable in measuring their instruction" (Desimone et al. 2010, p. 270). They further noted that because of the unreliability in student reports, one should exercise caution in using student-reported measures. In summary, teacher reports of homework assignment practices are more reliable and if the goal is to increase student outcomes by altering education policy, which are implemented by teachers, the more relevant measures to consider should be teacher reported homework assignment practices and, if possible, the quality of the assigned homework.

Based on the problems with much of the previous research outlined in Trautwein and Köller (2003) and Trautwein (2007), a new era of research into the relationship between homework and student achievement began in the early 2000s. Advances in the field of microeconometrics, including methods for panel data and causal inference, spilled over to the field of education research. The use of microeconometric strategies and quasi-experimental design methods that allowed for the inclusion of student and teacher fixed effects became more common. In this way, researchers could more confidently claim to control for a larger amount of the possible confounding variables that had plagued previous research, likely coming closer to a causal estimate. During the last 20 years, the available evidence still has a tilt towards homework having a positive effect on student achievement, but the picture is more nuanced, with more papers finding a null or small negative effect. The available evidence is thus still inconclusive.

For the broader perspective, a recent meta-analysis that looked at research papers from 2000–2015 covering the US or Middle eastern countries was conducted by Bas et al. (2017). They found an average effect size of $d = 0.229$, which is smaller than the one found in Cooper et al. (2006) but more comparable to the one found in Cooper (1989). Contrary to the findings of a null effect for primary school students in Cooper (1989) and Cooper et al. (2006), Murillo and Martínez-Garrido (2014) looked at studies in Latin American countries and concluded that time spent on homework had a positive effect on primary-school students' academic achievement, but only when teachers provided feedback in subsequent classes. In contrast, the meta analyses by Scheerens, Luyten, et al. (2007) and Scheerens and Hendriks (2014) found much smaller effect sizes than those in previous meta-analyses. Scheerens, Luyten, et al. (2007) looked at 21 studies and found an average effect size of $d = 0.073$, while Scheerens and Hendriks (2014) considered 27 studies where the average effect size was $d = 0.05$.

When it comes to important individual papers during the last 20 years, Trautwein (2007) used the German sample in TIMSS 2003 to relate student-reported homework time and frequency, as well as student effort on homework with the TIMSS test scores, in grade 8. The main point of the methodology was that previous mathematics achievement could be included in the model, proxying overall student level ability. The results indicated that homework frequency and student effort had a positive effect ($r = 0.11$ standard deviations) on student achievement, but that time spent on homework had a null or negative effect.

Dettmers, Trautwein, and Lüdtke (2009) looked at PISA 2003 data for 40 countries, controlling for the hierarchical structure where students are nested within classes and schools, as well as controlling for socioeconomic status, cognitive abilities, and school track. They found a small but positive average effect of homework time per week reported by the 15 year old students, though the effect varied between countries, with positive effects in some countries and negative in others.

Gustafsson (2013) used TIMSS 2003 and 2007 data for 22 countries in grade 8 and employed two-level regression (class and student level), IV-regression (instrumenting student-reported homework with teacher-reported homework), and a difference-in-differences approach at the country level. In all methods, homework amount per week (in minutes) reported by students had a positive effect on student test scores, with the largest effect size from the IV-estimate indicating an increase of 10 minutes per week is associated with an increase of 4 points on the test.

The National Educational Longitudinal Study of 1988 in the US was used by Eren and Henderson (2011) to relate hours of weekly homework to 8th grade students' test scores. Their results indicated that controlling for unobserved student (teacher) characteristics, using the first difference model, is crucial since the effect of homework otherwise was biased upwards (downwards). In their most complete model, the homework coefficient implied that a one-standard deviation increase in the weekly assigned homework amount was associated with 1.7 % higher test scores relative to the sample mean test score.

Homework frequency reported by teachers was related to student achievement for the Norwegian sample in TIMSS 2007 in Rønning (2010). First, OLS-estimations with a large battery of control variables indicated a positive effect for the 4th grade students with the same teacher in mathematics and science.[4] Next, the within-student between-subject estimation strategy, that allowed for the inclusion of both student and teacher fixed effects, yielded much lower point estimates (only 5 % of a standard deviation), indicating that unobserved omitted variables in the OLS-estimation biased the estimate upwards. Using the same data set (but in this case looking at 16 OECD countries), measure of homework, and empirical approach, Falch and Rønning (2012) came to similar conclusions. The results differed between countries[5], with the pooled sample estimates indicating a linear positive effect of teacher-reported homework frequency. In a simplification of the frequency variable, the estimated effect of assigning homework in all lessons compared to never assigning homework was an increase of student test scores by 4 % of a standard

[4] Rønning (2010) also included OLS-estimations for the 8th grade students, with point estimates similar to those in 4th grade. However, there were not enough variation in homework assignment for the students with the same teacher in both subjects for the first difference approach.

[5] The largest effect was estimated in the US, Austria and Australia (14-21 % of a standard deviation), while most other countries had point estimates close to the pooled sample.

deviation.

Jerrim, Lopez-Agudo, and Marcenaro-Gutierrez (2020) used data for 4th grade students from 24 countries in the 2011 round of Progress in International Reading Literacy Study (PIRLS) and TIMSS to investigate the relationship between teacher reported homework length and student test scores. Using the within-student between-subject approach, controlling for fixed student and teacher characteristics, they found a null effect.

Finally, in connection to the possible positive and negative effects of homework described in the introduction, the positive effects might outweigh the negative ones in only parts of the distribution. It might be the case that if teachers assign too much homework, the negative aspects start outweighing the positive effects. This might result in a reversed relationship with lower positive effects, or even overall negative effects, after the possible optimal homework assignment amount has been exceeded. Previous research has tested for such a relationship by including categorical dummy variables for different levels of homework assignment (e.g. 0 is no homework, 1 is homework assigned once per week, 2 is 2-3 times a week, and 3 is 4-5 times a week) (Cooper et al. 2006; Blazer 2009; Chin et al. 2020). If the 2-3 times a week dummy variable has the highest coefficient in a regression, this would indicate a nonlinear relationship. In the literature, the are some evidence for such a relationship (Cooper et al. 2006; Blazer 2009; Chin et al. 2020), but others have found that the relationship is indeed linear (Falch and Rønning 2012, p. 20).

To summarize, the available research is largely inconclusive. There are, however, indications that if there is a positive relationship, the effect is larger in higher grades than in lower grades. There is also some evidence that there might be a nonlinear relationship, with an optimal level of homework assignment. Further, despite teacher-reported homework assignment practices being the most exogenous and relevant measure of homework for education policy makers, much of the previous research has focused solely on student reported homework measures. In view of these points, in Section 5 teacher-reported homework frequency and homework amount will be related to student achievement separately for grade 4 and grade 8 students. Furthermore, possible nonlinear relationships will be investigated by entering the homework variables as categorical dummy variables.

3. Data

Data from the 2019 wave of the Trends in International Mathematics and Science Study (TIMSS) was used in the empirical analysis. TIMSS is an international mathematics and science test for 4th and 8th grade students, conducted every 4 years since 1995 by the

International Association for the Evaluation of Educational Achievement (IEA).[6] This thesis used the most recent assessment wave from 2019 that included 308,000 students in 58 countries in 4th grade, and 227,000 students in 39 countries in 8th grade (Martin, Mullis, Foy, et al. 2020, p. 2).[7]

TIMSS uses a stratified two-stage cluster sample design to ensure that a representative sample of students are selected in each country (Martin, Davier, et al. 2020, p. 3.11). In the first stage, at least 150 schools in each country are chosen randomly proportional to the number of students in each school. In the second stage, one or more intact classes within each grade and each school are randomly selected to participate. This means that students were clustered at the country, school, and class levels. In the empirical analysis, this complex sampling design was accounted for by using the total student level sampling weights provided by TIMSS.[8][9]

Students were administered standardized tests in mathematics and science[10] and questionnaires were used to gather background information on the participating schools as well as on students and teachers in both subjects. The questionnaires covered a rich set of topics regarding teachers' and students' backgrounds as well as the school environment. The student and teacher questionnaires also had questions on experiences and practices in teaching and learning mathematics and science. Relevant to the analysis in this thesis, the teacher questionnaire administered to teachers in the participating classes included questions on the homework assigning practices used by teachers (see Section 3.2). Students' achievement in mathematics and science could thereby be connected to the homework assigning practices by their teachers in that subject, forming a panel where each student was observed twice.

[6] IEA also administers the international reading test PIRLS, which uses a separate sample of students from that in TIMSS.

[7] Seven so called benchmarking participants, regional entities such as Ontario in Canada or Madrid i Spain, also participated but were not included in this thesis.

[8] TIMSS provides three different student level weights: total student weight that sums to the student population size in each country, senate weights that sums to a weighted sample size of 500 in each country, and house weights that makes the weighted sample size correspond to the actual sample size in each country (Fishbein et al. 2021, p. 83). For the purpose in this thesis, where the research question concerns the population of students in the countries included in the sample, the total student weights should be used (Jerrim, Lopez-Agudo, Marcenaro-Gutierrez, and Shure 2017).

[9] See Martin, Davier, et al. (2020) for further details on the sampling method and weighting, and Jerrim, Lopez-Agudo, Marcenaro-Gutierrez, and Shure (2017) on the importance of using the correct weights when analysing data sets such as TIMSS.

[10] TIMSS 2019 was the first time that a computer based assessment method was used in approximately 50 % of the participating countries (Martin, Mullis, Foy, et al. 2020). Every effort was made by IEA to ensure that the same mathematics and science constructs were measured in both the paper and the computer assessment methods, making comparisons between countries and across time within each country valid.

3.1. Sample selection

While the full samples included 308,000 students in 58 countries in grade 4 and 227,000 students in 39 countries in grade 8, not all observations could be used in the empirical analysis (compare Bietenbeck 2014; Caro et al. 2016; Falch and Rønning 2012). First, those students with more than one teacher in either mathematics or science were removed to ensure that homework assignment could be clearly identified and connected to the students. Second, students with a teacher in one or both subjects that did not answer the homework assignment questions were removed. In these two steps all students from eight countries in grade 8 were removed.[11] Moreover, after the previous steps almost all students were removed in some countries in grade 8. In some countries very few students remained, which could introduce bias in the estimations if there is something systematic for the remaining students or their teachers. Therefore, in a third step all students in countries with less than 500 students left were also removed.[12]

The number of students remaining after each step are presented separately for the grade 4 and grade 8 samples in Appendix I. After these data cleaning procedures were completed, the final estimation samples used in the empirical analysis were the following. In grade 8, 40.89 % of the students were excluded, leaving 134,383 students, in 4,451 schools, from 26 countries. For grade 4, the final sample consisted of 263,118 students, in 10,065 schools, from 58 countries, which means that 14.47 % of the students from the full sample were excluded.

In Table A7 and Table A12, the full and final samples are compared on nine student and teacher characteristics. The final samples seem to represent the full samples, with some small differences in a few variables. In grade 4 these differences occurred in the number of books at home, and in grade 8 in teacher age. Overall, though, the comparison indicates that the external validity of the estimation sample is strong. Table A6 and Table A11 presents the number of students, schools, and the average test scores in math and science in the final samples, divided by country. From the tables, most countries have 3,000–7,000 students in the final samples, though some countries like South Africa and the United Arab Emirates had more than 10,000 students participating in each grade. It is also apparent that the test score in mathematics can differ quite significantly from the test score in science within countries, though most countries only have a differential of 10–20 points.

The regressions presented in Section 5 were run on the whole final samples, pooling

[11] These countries were France, Georgia, Hungary, Kazakhstan, Lebanon, Lithuania, Romania, and Russian Federation.

[12] In the third step, all students from Cyprus, Finland, Morocco, Portugal, and Sweden were removed in the grade 8 sample.

all countries, which will net the average estimate across all countries (compare Falch and Rønning 2012; Gustafsson 2013; Dettmers, Trautwein, and Lüdtke 2009). To give some insight into the heterogeneity of the results, the sample was divided into OECD and non-OECD countries and the regressions were run separately, see Section 5.3. Doing this also makes it possible to compare the grade 4 results to those in Falch and Rønning (2012) who used the same methodology as this thesis, but only studied 16 OECD countries in TIMSS 2007.

3.2. Operationalization of homework

When relating homework to student achievement, teacher-reported homework assignment practices are likely more accurately reported than those reported by students while also being more relevant to education policy makers (see discussion in Section 2). Therefore, homework as reported by teachers were used as the explanatory variables (compare Jerrim, Lopez-Agudo, and Marcenaro-Gutierrez 2020; Rønning 2010). Teachers in both grades were asked questions regarding their homework assignment practices in the teacher questionnaire.[13] The homework frequency and expected time of completion (length) of any given homework were measured via the following two questions:

– A. *How often do you usually assign (subject) homework to the students in this class?*

- I do not assign (subject) homework[14]
- Less than once a week
- 1 or 2 times a week
- 3 or 4 times a week
- Every day

– B. *When you assign (subject) homework to the students in this class, about how many minutes do you usually assign? (Consider the time it would take an average student in your class.)*

- 15 minutes or less
- 16–30 minutes
- 31–60 minutes
- 61–90 minutes
- More than 90 minutes[15]

In the empirical analysis, homework frequency and homework amount were used as explanatory variables. Question A was used for the homework frequency variable while

[13]Interestingly, only teachers were asked about homework assignment frequency and length in 4^{th} grade, while both teachers and students were asked in 8^{th} grade. Thus, the organizers of TIMSS implicitly find teacher-reported homework frequency and length more exogenous, which makes the choices in this thesis more credible.

[14]Teachers who picked this answer did not answer the next question, and were, therefore, coded as giving 0 minutes of homework in that question.

[15]In grade 4, the “61–90 minutes” and “More than 90 minutes” alternatives did not exist. Instead, there is an alternative for “More than 60 minutes”.

homework amount was constructed as a combination of homework frequency (question A) and homework length (question B) (compare Bas et al. 2017; Martin, Mullis, Gonzalez, et al. 2004).

Homework amount was constructed as a categorical variable inspired by Martin, Mullis, Gonzalez, et al. (2004) with four categories: "No homework", "Low amount", "Medium amount", and "High amount" (see Table 3.1). "No homework' was assigned to students with a teacher who answered "I do not assign (subject) homework" in question A. "Low amount" was assigned to those students with a teacher who answered "Less than once a week" or "1 or 2 times a week" on question A, and "15 minutes or less" or "16–30 minutes" on question B. "High amount" was assigned to students with teachers who assigned homework "3 or 4 times a week" or "Every day" and answered "31–60 minutes", "61–90 minutes", or "More than 90 minutes" on question B.[16] All other combinations of question A and B were coded as "Medium amount".

Table 3.1. Combination of homework frequency and homework length into homework amount.

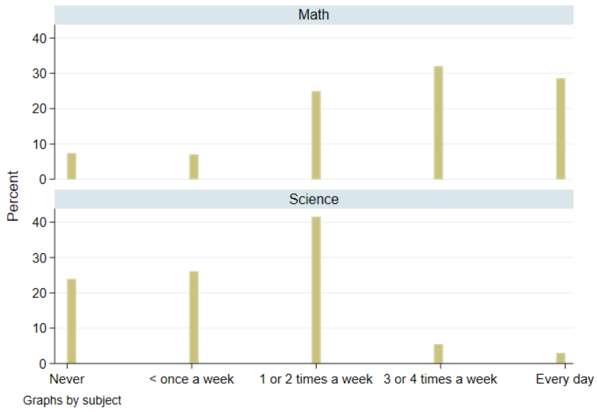
		Homework frequency				
		No homework	Less than once a week	1 or 2 times a week	3 or 4 times a week	Every day
Homework length	0 minutes	No homework				
	15 minutes or less		Low amount	Low amount	Medium amount	Medium amount
	16–30 minutes		Low amount	Low amount	Medium amount	Medium amount
	31–60 minutes		Medium amount	Medium amount	High amount	High amount
	61–90 minutes		Medium amount	Medium amount	High amount	High amount
	More than 90 minutes		Medium amount	Medium amount	High amount	High amount

Notes: This applies to the grade 8 sample. Since the "61–90 minutes" and "More than 90 minutes" alternatives did not exist for grade 4, the "High amount" category was assigned students with teachers who assigned homework "3 or 4 times a week" or "Every day" and answered "31–60 minutes" or "More than 60 minutes" on question B..

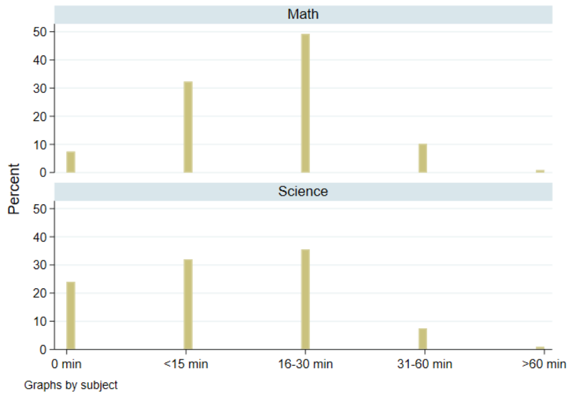
The distribution of students assigned the different levels of homework frequency, homework length, and homework amount are presented in Figure 3.1 for grade 4 and in

[16]Since the "61–90 minutes" and "More than 90 minutes" alternatives did not exist for grade 4, the "High amount" category was assigned students with teachers who assigned homework "3 or 4 times a week" or "Every day" and answered "31–60 minutes" or "More than 60 minutes" on question B.

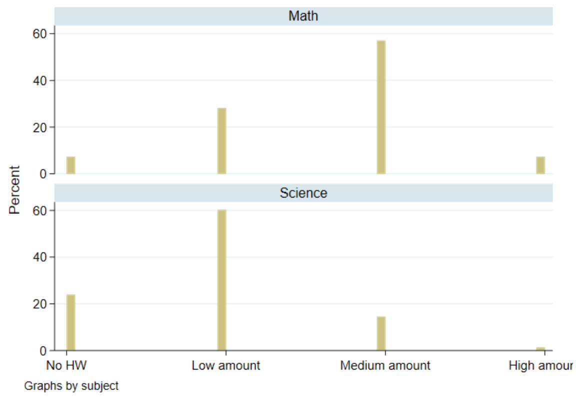
Figure 3.2 for grade 8. The most striking observations that can be made from the figures is that long homework length is rarely assigned, in both grades, and that homework is given more frequently in mathematics than in science, particularly in grade 4.[17] These characteristics translate into a somewhat skewed distribution in homework amount. In the regressions, homework frequency and homework amount were included in separate regressions as the explanatory variable.



(a) Homework frequency



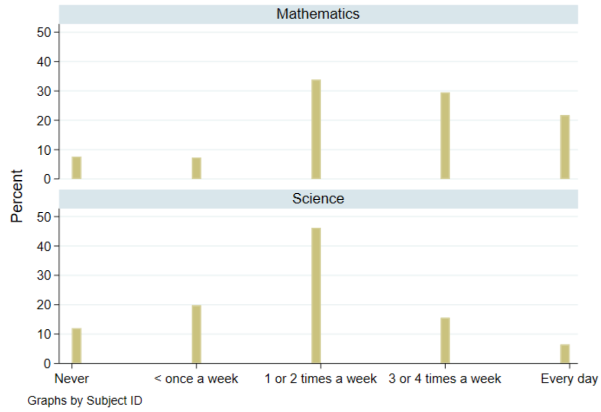
(b) Homework length



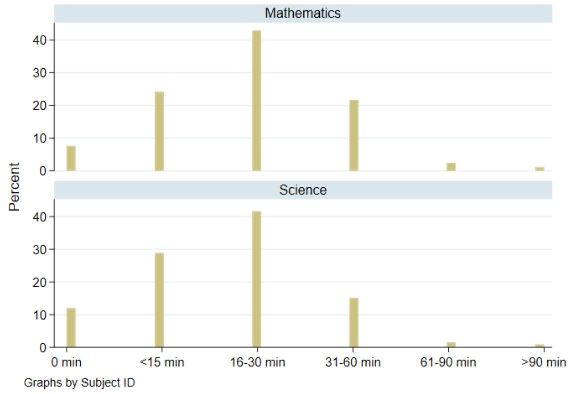
(c) Homework amount

Figure 3.1. Distribution of the homework assignment variables, by subject, in grade 4.

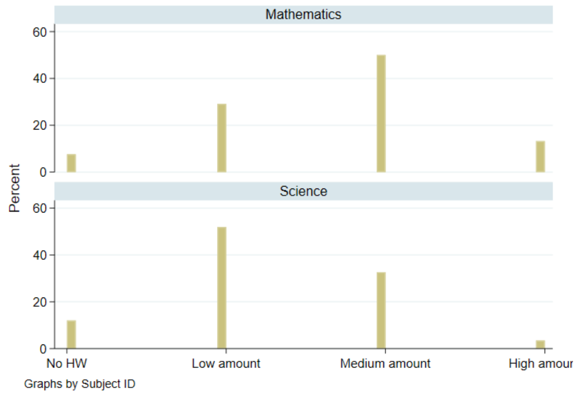
[17]A subject dummy is therefore included in all regressions presented in Section 5 that control for any part of this tendency that affects test scores, as well as other possible subject effects.



(a) Homework frequency



(b) Homework length



(c) Homework amount

Figure 3.2. Distribution of the homework assignment variables, by subject, in grade 8.

3.3. Measuring student achievement

The outcome variable in the empirical analysis is student test score on the standardized tests in mathematics and science that was administered to the participating students. Student achievement in TIMSS, as in other large-scale international student assessment tests such as PISA, is summarized using so called "plausible values" (Martin, Davier, et al. 2020). TIMSS uses an incomplete-booklet design, where each individual student does not complete the same test and is instead given a subset of problems taken from a large question pool. A number of these booklets with a set of questions in each subject are then given randomly to students. IEA then uses Item Response Theory models to construct five plausible values for each student, in each subject. These plausible values can be thought to represent "the range of abilities that a student might reasonably have, given the student's item responses" (Wu 2005, p. 115).[18] All regressions presented in

[18] For further details, see Chapter 11 in Martin, Davier, et al. (2020) and Wu (2005).

Section 5 used these plausible values as the outcome variable.[19]

Finally, in the empirical analysis, the test scores were standardized to have a mean of 0 and a standard deviation of 1 in the full sample so that any effect of homework could be interpreted in terms of standard deviations in test score (compare Falch and Rønning 2012; Jerrim, Lopez-Agudo, and Marcenaro-Gutierrez 2020; Rønning 2010).

4. Empirical strategy

When trying to estimate the effect of homework on student achievement, researchers have to deal with the fact that students (or their parents) and teachers are likely to self-select into cities, schools, and classrooms based on unobserved variables. The likely risk is that unobservable student and teacher characteristics are endogenous, meaning that they correlate with teachers' homework assignment practices. If, for example, students with generally higher motivation or general ability choose schools with higher homework amount, the estimated effect of homework on student test scores using simple cross-sectional data would be biased upwards. If a positive effect is estimated, the researcher cannot know how much of this is caused by homework assignment and how much is caused by the self-selection of these inherently gifted students. Similarly, if better or more motivated teachers choose schools with less focus on homework assignment, and this affects how well students perform in the TIMSS tests through other channels than homework assignment, the effect would be biased downwards. It is also possible that teachers adjust their homework assignment practices based on the general ability or composition of the students in their classes. Teachers with higher achieving, well adjusted students might not feel the need to assign that much homework, which would also lead to attenuation bias in the estimated effect of homework on test scores.

The ideal way to deal with these problems would be to randomly assign students to different homework assignment practices in an experiment. Such random assignment would ensure that a simple OLS approach would produce an unbiased estimate that could gauge the causal effect of homework on student test scores, unaffected by self-selecting students and teachers. However, in most real life situations this is not possible, especially not in such large scale events as TIMSS, with thousands of participating students from many countries.

The next best approach is to try and use a quasi-experiment or to exploit particularities in the data to deal with as much of the endogenous omitted variables problem

[19] In the regressions, the final beta coefficient is obtained by taking the average of the five coefficients acquired from using each individual plausible value (Macdonald 2008)

as possible. One such potential solution is to exploit the quasi-panel structure of the TIMSS data where each student is observed twice, once in mathematics and once in science. This approach has previously been used extensively when relating the effect of homework, as well as teaching practices and instruction time, to student test scores (Bietenbeck 2014; Eren and Henderson 2011; Falch and Rønning 2012; Jerrim, Lopez-Agudo, and Marcenaro-Gutierrez 2020; Lavy 2010; Rønning 2010; Schwerdt and Wuppermann 2011). The so called within-student between-subject estimation strategy developed in Dee (2005) and Dee (2007) (also termed the first difference estimator) treats data sets such as TIMSS as a panel, enabling the inclusion of student fixed effects via a within-student transformation. This estimation strategy thus accounts for the sorting of students to homework assignment practices across schools and classrooms based on fixed student characteristics. Further, under the assumption that teachers are as-good-as randomly assigned to classrooms conditional on fixed student characteristics, and that homework assignment is not correlated with any other unobserved teacher characteristic that drives test scores, a causal effect is obtained.

To obtain the first difference estimator for the final grade 4 and grade 8 estimation samples, the starting point is to assume that the following linear education production function explains student test scores (compare Eren and Henderson 2011; Falch and Rønning 2012; Jerrim, Lopez-Agudo, and Marcenaro-Gutierrez 2020):

$$TS_{iknjsc} = \alpha + \beta HW_{nj} + \gamma T_{nj} + \theta_n + C_c + S_s + \kappa_k + \lambda_i + \varepsilon_{iknjsc} \tag{4.1}$$

where student test score, TS_{iknjsc}, is the achievement of student i, in class k, in subject n ($n = 1$ for mathematics and n = 2 for science) with teacher j, in school s, in country c. HW_{nj} is the homework frequency or homework amount in subject n, assigned by teacher j. T_{nj} is a vector of observable teacher characteristics, θ_n is a subject dummy variable, C_{c} and S_{s} are country and school fixed effects. κ_k is the class fixed effects that includes peer effects and other fixed class characteristics. λ_{i} denotes the student fixed effects, which includes all subject-invariant, student level characteristics (e.g. overall academic ability and motivation, and all fixed background traits such as age, sex, and socio-economic status). Lastly, $\varepsilon_{\text{iknjsc}}$ denotes the student level error term, which includes all unobservable teacher characteristics as well as subject-variable student and class traits that are not captured by λ_{i} and κ_k. These subject variable student traits include, for instance, subject-specific academic ability and motivation which could represent threats to a causal interpretation (see discussion in the next section).

Next, the subject-specific regression equations are defined as follows:

$$TS_{ik1jsc} = \alpha + \beta HW_{1j} + \gamma T_{1j} + \theta_1 + C_c + S_s + \kappa_k + \lambda_i + \varepsilon_{ik1jsc} \tag{4.2}$$

$$TS_{ik2jsc} = \alpha + \beta HW_{2j} + \gamma T_{2j} + \theta_2 + C_c + S_s + \kappa_k + \lambda_i + \varepsilon_{ik2jsc} \tag{4.3}$$

If Equation (4.3) (science test score) is subtracted from Equation (4.2) (mathematics test score), the first-difference estimator is obtained:

$$\begin{aligned} TS_{ik1jsc} - TS_{ik2jsc} = {} & (\alpha - \alpha) + \beta(HW_{1j} - HW_{2j}) + \gamma(T_{1j} - T_{2j}) + (\theta_1 - \theta_2) \\ & + (C_c - C_c) + (S_s - S_s) + (\kappa_k - \kappa_k) + (\lambda_i - \lambda_i) \\ & + (\varepsilon_{ik1jsc} - \varepsilon_{ik2jsc}) \end{aligned} \tag{4.4}$$

As is apparent from Equation (4.4), the student fixed effects drop out since λ_i is the same in both subjects, $\Delta\lambda_\mathrm{i} = 0$. Note that the country fixed effects, school fixed effects, and class fixed effects also drop out ($\Delta C_\mathrm{c} = 0$, $\Delta S_\mathrm{s} = 0$, and $\Delta\kappa_\mathrm{k} = 0$). Therefore, the estimation method controls for student, class, school, and country fixed effects. As a result, the estimated effect of homework on test scores should be significantly closer to a causal estimate compared to a naive OLS regression that is not able to control for these unobserved fixed characteristics. From Equation (4.4), it follows that the first difference model identifies the effect of homework on student test scores by treating the difference in student test scores between mathematics and science as the dependent variable, and the difference in teacher homework assignment as the main explanatory variable:

$$\Delta TS = \beta \Delta HW + \gamma \Delta T + \Delta\theta + \Delta\varepsilon \tag{4.5}$$

Note, however, that teacher fixed effects are not controlled for here, which is why ΔT is included in the estimations. The subject effect, $\Delta\theta$ also remains and must therefore be included when running the regressions. HW will first be entered as homework frequency and then as homework amount in the regressions presented in Section 5. Both were entered as dummy variables for the different categories, with "No homework" as the reference group to try and detect whether there exists a nonlinear relationship or not. Thus, β represents a vector containing the beta-parameters for the individual homework dummy variables. Since HW will be included as dummy variables and test scores are normalized to have a mean of 0 and a standard deviation of 1, the beta-coefficients estimate how much test scores are affected in terms of standard deviations by the different levels of homework frequency or homework amount, compared to getting no homework.

Instead of performing the subtraction described above, creating the delta-variables, and running the model in Equation (4.5), the equivalent approach of treating the sample as a panel with two observations per student and running a student fixed effects model was used. These fixed effects estimations were also contrasted with baseline pooled OLS estimations that include a rich set of student, teacher, class, and school control variables commonly used in the literature (see Table A10 and Table A15 for the full list as well as descriptive statistics). In this way, it is possible to observe the direction and size of any bias induced by unobserved student characteristics. All estimations were performed in both the grade 4 sample and the grade 8 sample.

Finally, the sampling method employed by TIMSS (described in Section 3) where intact classes are sampled implies that independence at the student level could not be assumed. Instead, in the regression results presented in the next section, Jackknife standard errors[20] were clustered at the class level (compare Bietenbeck 2014; Falch and Rønning 2012). Further, the increased uncertainty regarding students' true test scores induced by the use of plausible values, as well as the complex design with student weights, were accounted for by using the Stata package "PV" (Macdonald 2008) for all regressions. This package was specifically designed to analyze student achievement data-sets, such as TIMSS, that uses plausible values and more complex sampling designs.

4.1. Assumptions and threats to the identification strategy

The identification strategy described above has many merits and arguably comes significantly closer to a causal estimate than a naive OLS estimation. There are, however, also some drawbacks with the method and there are still concerns regarding the identifying assumption.

Regarding potential drawbacks with the first difference estimator, Angrist and Pischke (2009) raise two main concerns. First, the risk of bias in the estimate induced by measurement errors is larger since the variation in homework is restricted to within-students. In contrast, for an OLS regression without any control variables, the variation in test scores is related to the variation in homework across the whole sample of students. According to Angrist and Pischke (2009), this typically leads to attenuation bias in first difference estimates. This fact could, however, be interpreted as not entirely detrimental since any estimated positive effect will be conservative and should therefore have a low risk of exaggerating positive effects. On the other hand, an estimated negative or zero effect might mask a positive effect.

[20] Jackknife standard errors is a linear approximation of bootstrapped standard errors preferred by TIMSS (Martin, Davier, et al. 2020). They are calculated by systematically leaving out each observation from the dataset and recalculating the standard error n - 1 times and then averaging these for the final Jackknife standard error.

Second, since the estimated coefficients come only from students with teachers that assign different amounts of homework in mathematics and science[21], the internal validity is restricted to this sub sample. If, for example, only 10 % of the students have teachers that assign different amounts of homework in the two subjects, the estimated effect comes only from these students. Consequently, if these students do not represent the full estimation sample, the estimated effect can only be argued to apply to these 10 %.

Thus, for the estimates to be meaningful, there needs to be a sufficient proportion of students with teachers that assign different amounts of homework in the two subjects, and the $\Delta HW = 0$ students need to be representative of the full estimation sample. It is difficult to claim that there exists a threshold for sufficient variation, though previous research has deemed a proportion of around 50 % non-zero variation in homework assignment between subjects enough to make the identification strategy credible (Jerrim, Lopez-Agudo, and Marcenaro-Gutierrez 2020). The proportion of non-zero difference students for homework frequency (homework amount) were 67 % (60 %) in the grade 8 sample and 77 % (66 %) in the grade 4 sample (see Figure A1 and Figure A2 in Appendix IV). Further, Table A9 and Table A14 present means and standard deviations for nine demographic characteristics in the two grades, contrasting the students with ΔHW amount $= 0$ to those students contributing to the estimations. Overall, the students contributing to the estimations are very similar to those that do not, giving credibility to the identification strategy; although there are some differences in grade 4, there are almost no differences between the two groups in grade 8. The slightly larger, and more prevalent, differences in grade 4, however, mean that these results should be interpreted with more caution.

The identifying assumption of the first difference model is that homework assignment practices are uncorrelated with the error term, conditional on the other regressors (homework must be as good as randomly assigned to students). The first way in which this assumption could be violated is the inability to include teacher fixed effects in the main estimations based on Equation (4.5). Unobserved teacher variables could potentially confound the results if they correlate with the teachers' homework assignment practices. One such example mentioned above is that teachers that are more motivated and/or are better at teaching might give less homework to their students. To alleviate this concern, a rich set of teacher controls (T_{nj}) were included in the estimations. The included teacher variables were sex, age, years of teaching, education level, having a teacher certificate, having majored in the subject taught, teacher feelings regarding teaching, and total instruction time in minutes per week.

Table A8 and Table A13 presents means and standard deviations for these variables by homework amount, as well as information on the percentage of missing values for

[21] If the student receives the same amount of homework in mathematics and science class, ΔHW will be 0 and this student will not contribute to the estimate.

each variable, in both grades. Although small, there are some differences across the different levels of homework amount, once again to a larger extent in grade 4, indicating that also unobserved teacher variables might be associated with homework assignment. As is apparent from the tables, there was an evident problem with missing values for the control variables. Most variables had a small proportion of missing values, but the variable "Teacher certificate" had 10.53 % missing values in grade 4 and 14.49 % missing values in grade 8. Since a large amount of observations had missing values on at least one of the control variables, removing all observations with a missing value on at least one variable would significantly reduce the sample size. Instead, an approach that makes use of all available information, previously used by Bietenbeck (2014), Fuchs and Wößmann (2008), Lavy (2016), and Mattsson (2020), was employed.[22] This method sets all missing values to a constant, here they were set to the median, and includes dummy variables for missing values for each control variable in the regressions. Including these dummy variables ensure that the results are relatively robust to any possible bias caused by imputing the missing values to the median (Fuchs and Wößmann 2008). If observations are not missing values at random, running regressions without the missing value dummy variables should produce different results. In Section 5.4, this notion was tested in a sensitivity test.

Another potential violation of the identifying assumption comes from the fact that the identification strategy assumes β to be the same in both subjects, i.e. that the effect of homework assignment on test scores does not differ between subjects. It is possible that students could have different levels of inherent ability or motivation in the two subjects. This concern is difficult to address, but should be small since mathematics and science arguably require very similar abilities and skills. Moreover, previous research have found high correlations between the two subjects (Clotfelter et al. 2010) and substantial evidence that β is the same across mathematics and science (Jerrim, Lopez-Agudo, and Marcenaro-Gutierrez 2020), which further reduces this concern.

Although the first difference estimator arguably gets significantly closer to a causal estimate than an OLS regression with controls, the remaining concerns should instill caution in those interpreting results from it (Rønning 2010). Nonetheless, gaining such a powerful estimator for such a large data set makes the resulting estimates compelling and should give important insights when compared to previous and future research.

[22]This approach was also used for the control variables used in the OLS regressions. See Table A10 and Table A15 for the percentage of missing values on each control variable.

4.2. Dividing the grade 4 sample as a robustness test

In grade 8, almost all students had different teachers in mathematics and science, while in 4^{th} grade 60 % had the same teacher and 40 % had different teachers in the two subjects. This structure thus allows for the division of the 4^{th} grade sample into two sub samples. Importantly, for the sub sample where students had the same teacher in both subjects it is possible to relax the assumption that homework does not correlate with unobserved fixed teacher characteristics, as this sub sample also enables the inclusion of teacher fixed effects in the estimations (compare Falch and Rønning 2012; Jerrim, Lopez-Agudo, and Marcenaro-Gutierrez 2020). To see this, consider the altered subject-specific regression equations for this sub sample:

$$TS_{ik1jsc} = \alpha + \beta HW_{1j} + \theta_1 + C_c + S_s + \kappa_k + \mu_j + \lambda_i + \varepsilon_{ik1jsc} \tag{4.6}$$

$$TS_{ik2jsc} = \alpha + \beta HW_{2j} + \theta_2 + C_c + S_s + \kappa_k + \mu_j + \lambda_i + \varepsilon_{ik2jsc} \tag{4.7}$$

Everything here is the same as in Equation (4.2) and Equation (4.3) except that the observable teacher characteristics have been replaced with the teacher fixed effects, μ_{j}. Since all variables in T_{nj} are the same for both subjects in the sub sample where students have the same teacher in mathematics and science, these variables are now included in the teacher fixed effects. Moreover, μ_{j} also includes all unobserved subject-invariant teacher characteristics that were previously, in the main specification, included in the error term.

Consequently, for the sub sample where students have the same teacher in both subjects, the within-student transformation, in addition to netting out student fixed effects, will also net out teacher fixed effects ($\Delta\mu_{\mathrm{j}} = 0$). It follows that the resulting first-difference model in this sub sample can, with more credibility, claim to estimate a causal effect since more possible endogenous variables can be controlled for. The error term here only includes student, class, and teacher traits that vary between subjects.

However, even if the identifying assumption is less strict in this case, there are still threats that could violate it. For example, there could be unobserved teacher characteristics that vary between subjects and correlate with homework assignment practices (Rønning 2010). Even if teacher fixed effects are controlled for, a teacher might be inherently more skilled in teaching mathematics than science. In this case, the teacher might assign more homework in science to compensate, which would introduce bias. This concern is, however, likely small since there would have to be a systematic difference in qualification for the two subjects across all teachers.

As a robustness test for the grade 4 sample, the main specifications for the sample containing all grade 4 students was contrasted against estimates in the two sub samples (see Section 5.4). If the estimated effect does not differ significantly between the two sub samples, this would indicate that unobserved fixed teacher characteristics do not drive the results. Such a result would also give some indication that the same might be the case in grade 8, meaning that the assumption that homework does not correlate with unobserved fixed teacher characteristics is viable.

5. Results

The regression results for grade 4 are presented in Table 5.1 and for grade 8 in Table 5.2. The results in each grade are divided into two panels based on whether homework frequency (Panel A) or homework amount (Panel B) was the main explanatory variable for student test scores. In the first four columns the baseline pooled OLS estimates, with more control variables added in each subsequent column, are reported. Column five and six display the student fixed effects estimates based on the model in Equation (4.5), which control for country, school, class, and student fixed effects. All regressions include subject dummy variables, and all OLS estimations include country dummy variables. The reference group for both homework frequency and homework amount is “No homework”. In the regressions, the final beta coefficients are obtained from the average of the five coefficients acquired from using each individual plausible value (Macdonald 2008).

5.1. 4^{th} grade

The first four columns in Table 5.1, that report the pooled OLS regressions for grade 4, indicate a null effect of both homework frequency and homework amount on student test scores, with small statistically insignificant coefficients. These results imply that there is no bias induced by the observable school, class, student, and teacher variables included. Worth noting is that all coefficients (except the “Every day” category for homework frequency) are negative and that the standard errors typically decrease when more control variables are added.

Next, when the student fixed effects estimations are run (column five and six) all coefficients decrease sharply and become statistically significant at the 0.1 % and 5 % levels, except for the “Every day” category dummy. Further, when teacher controls are added in column six, the statistically significant coefficients all become smaller in absolute value, indicating downwards bias induced by these teacher characteristics (similar to the results in Eren and Henderson 2011). In contrast to the OLS estimations, these results imply that there is a significant upwards bias in the effect of homework frequency and homework amount based on unobserved fixed student, school, and class characteristics. The change in the coefficients when teacher controls are added indicates that there is likely to exist additional, unobserved, teacher characteristics that correlate with homework assignment practices, which would make even these student fixed effects estimations biased. Although it is impossible to say what the size of, or in which direction, this bias would go, the included observable teacher variables indicate that there might be further attenuation bias.

Table 5.1. Pooled OLS and fixed effects (FE) estimations of the relationship between homework assignment and student achievement for 4^{th} grade students.

	OLS				FE	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Homework frequency Reference group is "No homework"						
Less than once a week	-0.017	-0.030	-0.020	-0.014	-0.039***	-0.034***
	(0.032)	(0.031)	(0.026)	(0.025)	(0.011)	(0.010)
1 or 2 times a week	-0.016	-0.019	-0.018	-0.021	-0.078***	-0.067***
	(0.030)	(0.027)	(0.023)	(0.023)	(0.012)	(0.011)
3 or 4 times a week	-0.031	-0.022	-0.021	-0.021	-0.127***	-0.107***
	(0.050)	(0.048)	(0.043)	(0.042)	(0.023)	(0.018)
Every day	-0.019	0.005	0.011	0.001	-0.012	-0.015
	(0.065)	(0.060)	(0.048)	(0.042)	(0.023)	(0.019)
Panel B: Homework amount Reference group is "No homework"						
Low amount	-0.015	-0.023	-0.017	-0.016	-0.063***	-0.054***
	(0.028)	(0.025)	(0.022)	(0.022)	(0.012)	(0.009)
Medium amount	-0.026	-0.016	-0.017	-0.017	-0.081***	-0.069***
	(0.041)	(0.040)	(0.034)	(0.033)	(0.018)	(0.014)
High amount	-0.082	-0.098	-0.033	-0.027	-0.073*	-0.063*
	(0.087)	(0.064)	(0.071)	(0.066)	(0.033)	(0.028)
Subject dummies	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	No	No
School and class controls	No	Yes	Yes	Yes	No	No
Student controls	No	No	Yes	Yes	No	No
Teacher controls	No	No	No	Yes	No	Yes
Student fixed effects	No	No	No	No	Yes	Yes
Observations	526,236	526,236	526,236	526,236	526,236	526,236

Notes: The dependent variable in all student fixed effects regressions is the within-student between–subjects difference in standardized test scores. In the pooled OLS regressions the dependent variable is the test score in either mathematics or science. Control variables are listed in Table A10 and Table A15 in Appendix IV. All regressions include dummy variables for missing values on each of the included control variables. All student fixed effects regressions also control for country, school, and class fixed effects. Each coefficient is the average from estimates using each of the five plausible values. Jackknife standard errors that are robust to clustering at the class level are in parentheses. Significance is denoted by * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Interestingly, there seems to be a negative linear effect of homework frequency (up until "Every day"), with coefficients going from -0.034 for "Less than once a week" to -0.107 for "3 or 4 times a week", while the homework amount coefficients indicate a homogeneous negative effect (approximately -0.06) of any homework amount compared to the "No homework" reference. Regardless, the nonlinear relationship hypothesis (Cooper et al. 2006; Blazer 2009; Chin et al. 2020), with an optimal homework assignment level above "No homework", is not supported by these results.

Recall that the student test scores were normalized to have a mean of zero and a standard deviation of one. Hence, in the complete specification in column six the homework frequency coefficients indicate that, compared to getting no homework, getting homework "Less than once a week" is associated with a decrease in test score of 3.4 % of a standard deviation, "1 or 2 times a week" to a decrease of 6.7 % of a standard deviation, and "3 or 4 times a week" to a decrease of 10.7 % of a standard deviation. For homework amount, the results indicate that any amount above "No homework" is related to a decrease in test score of circa 6 % of a standard deviation.

As discussed earlier, an estimated effect of homework frequency is less consequential than if the length of the assigned homework is also considered, as in the homework amount variable.[23] Receiving homework every day that takes 10 minutes to complete should not be expected to have the same effect on students' learning as if it takes an hour to complete. For this reason, the homogeneous negative effect of any homework amount compared to the "No homework" reference is more meaningful and gives more insight into how homework assignment affects students' test scores. This is also likely what causes the contrast between the zero coefficient for the "Every day" category in homework frequency and the "High amount" coefficient (which includes the "Every day" category) that is statistically significant and quantitatively similar to the "Low amount" and "Medium amount" ones.

5.2. 8^{th} grade

In contrast to the results for the grade 4 students, the pooled OLS regressions for grade 8 indicate that there is a linear positive relationship with student test scores for both homework frequency and homework amount. In the first column of Table 5.2, with only country and subject dummies, all homework coefficients are statistically significant at conventional levels. The coefficients are also large in comparison to the negative ones in the grade 4 fixed effects estimations, with the coefficient for "High amount" indicating

[23] When only homework frequency enters the regression, homework length is allowed to vary.

Table 5.2. Pooled OLS and fixed effects (FE) estimations of the relationship between homework assignment and student achievement for 8^{th} grade students.

	OLS				FE	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Homework frequency *Reference group is "No homework"*						
Less than once a week	0.147**	0.104*	0.052	0.049	0.007	0.007
	(0.052)	(0.047)	(0.030)	(0.028)	(0.016)	(0.016)
1 or 2 times a week	0.148**	0.088	0.038	0.027	-0.011	-0.008
	(0.048)	(0.045)	(0.030)	(0.028)	(0.015)	(0.015)
3 or 4 times a week	0.333***	0.220***	0.124***	0.112***	0.003	0.008
	(0.051)	(0.052)	(0.034)	(0.032)	(0.017)	(0.017)
Every day	0.324***	0.211***	0.110**	0.094*	0.011	0.016
	(0.064)	(0.058)	(0.040)	(0.037)	(0.020)	(0.021)
Panel B: Homework amount *Reference group is "No homework"*						
Low amount	0.140**	0.087	0.040	0.033	-0.004	-0.002
	(0.048)	(0.045)	(0.029)	(0.027)	(0.015)	(0.015)
Medium amount	0.239***	0.158***	0.081**	0.070*	-0.003	-0.000
	(0.050)	(0.048)	(0.031)	(0.029)	(0.016)	(0.016)
High amount	0.377***	0.252***	0.143***	0.127***	0.033	0.036
	(0.059)	(0.056)	(0.038)	(0.035)	(0.020)	(0.021)
Subject dummies	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	No	No
School and class controls	No	Yes	Yes	Yes	No	No
Student controls	No	No	Yes	Yes	No	No
Teacher controls	No	No	No	Yes	No	Yes
Student fixed effects	No	No	No	No	Yes	Yes
Observations	268,766	268,766	268,766	268,766	268,766	268,766

Notes: The dependent variable in all student fixed effects regressions is the within-student between–subjects difference in standardized test scores. In the pooled OLS regressions the dependent variable is the test score in either mathematics or science. Control variables are listed in Table A10 and Table A15 in Appendix IV. All regressions include dummy variables for missing values on each of the included control variables. All student fixed effects regressions also control for country, school, and class fixed effects. Each coefficient is the average from estimates using each of the five plausible values. Jackknife standard errors that are robust to clustering at the class level are in parentheses. Significance is denoted by * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

that receiving this amount compared to receiving no homework is associated with 37 % of a standard deviation higher test score. Nevertheless, when more controls are added in columns 2–4 the size of the coefficients decrease substantially. Moreover, only the "3 or 4 times a week" and "Every day" categories for homework frequency, and only the "Medium amount" and "High amount" categories for homework amount, remain statistically significant when all controls have been added in the forth column. These results indicate there exists upwards bias induced by observable school, class, student, and teacher variables, showcasing the importance of controlling for these variables (Trautwein and Köller 2003).

The student fixed effects estimations in column five and six indicate that also unobservable fixed student, class, and school characteristics biased the pooled OLS estimates of the effect of homework assignment upwards. When all of these fixed characteristics are controlled for, the coefficients for all categories of homework frequency and homework amount become statistically insignificant and close to zero. Interestingly, in contrast to the grade 4 student fixed effects regressions, there is virtually no change in the coefficients when teacher controls are added in column six. However, as in the grade 4 regressions there is no indication in the student fixed effects results that there is a nonlinear relationship between homework assignment and student test scores. In fact, the results from the student fixed effects estimations imply that all levels of both homework frequency and homework amount give no further effect on test scores compared to receiving no homework.

5.3. OECD vs non-OECD countries

Heterogeneity between country groups was examined by dividing the samples into OECD countries and non-OECD countries and running the student fixed effects regressions. The OECD and non-OECD countries for both grades are listed in Appendix II.

Table A2 presents the regression results. There are virtually no differences between OECD and non-OECD countries in 8^{th} grade; all significances are the same and all coefficient sizes only differ to a very small extent. In grade 4, however, there are larger and more significant differences between the country groups. The negative effects of both homework frequency and homework amount in the main estimations appears to come from the non-OECD countries (see column 2 in Table A2). In the OECD country regressions, the results indicate a null effect of both homework frequency and homework amount on test scores. Conversely, in the non-OECD country regressions, the results are qualitatively the same as those in Table 5.1, but with lower significances possibly caused by the lower sample size.

5.4. Robustness checks

To deal with the problem of missing values in the control variables, all missing values for each individual control were set to the median of that variable and a missing control dummy variable was included in the regressions. If values were not missing at random, this approach would threaten the internal validity by introducing bias in the estimates. This concern was tested by re-estimating the regressions in Table 5.1 and Table 5.2, this time without the missing value control dummy variables (compare Fuchs and Wößmann 2008). These specifications should produce the same results as the main regressions if observations are missing conditionally at random. The results are presented in Table A3 and Table A4. In both grades, there are almost no differences compared to the main specifications in Table 5.1 and Table 5.2; there are some small changes in coefficient size and some beta-coefficients in the grade 8 OLS regressions have different significance stars, but overall the results are the same both qualitatively and quantitatively, which implies that values are missing at random and that the results are robust in this respect.

There are other ways to define a homework amount variable than the one used in the main estimations. As a robustness test, the homework amount variable was redefined similar to how Gustafsson (2013) did: as the number of minutes of assigned homework per week. The frequency and length variables were first re-coded and then combined into the new variable (Gustafsson 2013, p. 284). The frequency categories from question A were coded as follows: "Every day" = 5, "3 or 4 times a week" = 3.5, "1 or 2 times a week" = 1.5, "Less than once a week" = 0.5, and "No homework" = 0. The categories in the length of the assigned homework, question B, was coded as follows: "Fewer than 15 minutes" = 7.5, "15–30 minutes" = 22.5, "31–60 minutes" = 45, "61–90 minutes" = 75, "More than 90 minutes" = 120, and "No homework" = 0.[24] The continuous homework amount variable was then obtained by multiplying the frequency and length variables.

Note that this definition of homework amount treats all combinations of homework length and homework frequency equally, as long as the multiplication is equal to the same number of minutes. This is different from the homework amount variable used in the main specifications. The categorical amount variable recognizes that homework every day for 25 minutes is different, and is expected to have a different effect, than if homework is only given once per week but the length is 125 minutes. In the first case, students might be given small assignments with focus on repetition of what was covered in class while in the second case the assigned homework might be a larger project or some new topic not covered in class.[25] For this reason, the categorical amount variable should be preferred, though it is still relevant and interesting to consider this continuous definition which is why this robustness test was conducted.

Results of the student fixed effects estimations using continuous homework amount as the main explanatory variable in grade 4 and grade 8 are presented in Table A5. In both grades, the beta coefficients are never statistically significantly different from zero and the coefficients are extremely small. However, the sign of the coefficients correspond

[24] Since the "61–90 minutes" and "More than 90 minutes" alternatives did not exist for grade 4, the "More than 60 minutes" category was coded as 75.

[25] The effects of homework have been found to be highest when homework involves rote learning, practice, or rehearsal (Hattie 2012).

to those in the main specifications, with a negative coefficient in grade 4 and a positive one in grade 8. The interpretation of the coefficient for grade 4 in column 2 (grade 8 in column 4) is that increasing homework amount by 10 minutes per week is associated with a 0.0001 % decrease (0.0016 % increase) of a standard deviation in test score, though neither are statistically significant. Overall, these results corroborate the main results in that no statistically significant positive effect was estimated and that the signs of the coefficients corresponded to those in the main specifications.

As discussed in Section 4.1, a robustness test was run for the grade 4 students where the full estimation sample was divided into two: one where students have different teachers in mathematics and science (105,066 students) and one where they have the same teacher in both subjects (158,052 students). This means that when the student fixed effects regression are run for the sample with students that have the same teacher in both subjects, teacher fixed effects are also controlled for, yielding an estimator that can more credibly be claimed to be causal.

The regression results in Table 5.3 (different teachers in mathematics and science) and Table 5.4 (same teacher in mathematics and science) indicate that the negative effect of homework found in the results based on the whole grade 4 sample come from the students with the same teacher in both subjects. The coefficients and levels of statistical significance in Table 5.4 are very close to those in Table 5.1, while almost all coefficients are statistically insignificant in Table 5.3. The exception is the statistically significant negative coefficient for "1 or 2 times a week" in column five and six in Table 5.3. It is, nonetheless, worth noting that most of the coefficients in column five and six in Table 5.3 are close to those obtained from the students with the same teacher in both subjects. It is also interesting that column five in Table 5.4 suggests that there exists a negative linear relationship between homework amount and student test scores, in contrast to the homogeneous negative effect of any amount of homework implied in Table 5.1.

Regardless, the results in the sample with all grade 4 students seem to be driven by the students with the same teacher in both subjects. The implication of the disparities in the results between Table 5.3 and Table 5.4 is that there seems to exist significant upwards bias induced by unobserved fixed teacher characteristics that correlate with homework frequency and homework amount. Further, while it is not possible to claim with certainty, these results suggest that there are likely unobserved endogenous teacher variables that biased the grade 8 student fixed effects estimates in column five and six of Table 5.2. If the grade 4 results are any indication, such bias implies that the estimated null effect of homework assignment might be too optimistic and that the real relationship is a negative one. However, the fact that the observable teacher controls had virtually no effect on the homework assignment coefficients in grade 8, while they had a sizeable effect in grade 4, could indicate that such bias is likelier to be small in grade 8.

Table 5.3. Pooled OLS and fixed effects (FE) estimations of the relationship between homework assignment and student achievement for 4^{th} grade students with *different teachers* in mathematics and science.

	OLS				FE	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Homework frequency Reference group is "No homework"						
Less than once a week	-0.013	-0.021	0.006	-0.033	0.005	0.024
	(0.067)	(0.063)	(0.053)	(0.055)	(0.027)	(0.026)
1 or 2 times a week	-0.054	-0.062	-0.049	-0.093	-0.099***	-0.070**
	(0.068)	(0.059)	(0.050)	(0.051)	(0.027)	(0.026)
3 or 4 times a week	-0.067	-0.051	-0.023	-0.046	-0.099*	-0.055
	(0.102)	(0.089)	(0.077)	(0.078)	(0.042)	(0.035)
Every day	-0.059	-0.022	-0.005	-0.040	0.066	0.064
	(0.129)	(0.110)	(0.090)	(0.082)	(0.041)	(0.039)
Panel B: Homework amount Reference group is "No homework"						
Low amount	-0.035	-0.040	-0.023	-0.068	-0.046	-0.021
	(0.065)	(0.058)	(0.048)	(0.049)	(0.028)	(0.025)
Medium amount	-0.060	-0.036	-0.023	-0.048	-0.046	-0.017
	(0.090)	(0.082)	(0.068)	(0.066)	(0.036)	(0.031)
High amount	-0.054	-0.103	0.038	0.006	0.027	0.051
	(0.179)	(0.122)	(0.148)	(0.132)	(0.059)	(0.056)
Subject dummies	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	No	No
School and class controls	No	Yes	Yes	Yes	No	No
Student controls	No	No	Yes	Yes	No	No
Teacher controls	No	No	No	Yes	No	Yes
Student fixed effects	No	No	No	No	Yes	Yes
Observations	210,132	210,132	210,132	210,132	210,132	210,132

Notes: The dependent variable in all student fixed effects regressions is the within-student between–subjects difference in standardized test scores. In the pooled OLS regressions the dependent variable is the test score in either mathematics or science. Control variables are listed in Table A10 and Table A15 in Appendix IV. All regressions include dummy variables for missing values on each of the included control variables. All student fixed effects regressions also control for country, school, and class fixed effects. Each coefficient is the average from estimates using each of the five plausible values. Jackknife standard errors that are robust to clustering at the class level are in parentheses. Significance is denoted by * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 5.4. Pooled OLS and fixed effects (FE) estimations of the relationship between homework assignment and student achievement for 4^{th} grade students with the *same teacher* in mathematics and science.

	OLS				FE
	(1)	(2)	(3)	(4)	(5)
Panel A: Homework frequency *Reference group is "No homework"*					
Less than once a week	-0.015	-0.030	-0.029	-0.018	-0.044***
	(0.034)	(0.033)	(0.028)	(0.025)	(0.009)
1 or 2 times a week	0.027	0.022	0.019	0.027	-0.040***
	(0.030)	(0.027)	(0.024)	(0.023)	(0.010)
3 or 4 times a week	0.031	0.017	0.011	0.011	-0.098***
	(0.045)	(0.044)	(0.039)	(0.037)	(0.023)
Every day	0.045	0.047	0.045	0.046	-0.028
	(0.047)	(0.045)	(0.038)	(0.036)	(0.018)
Panel B: Homework amount *Reference group is "No homework"*					
Low amount	0.007	-0.004	-0.004	0.007	-0.045***
	(0.028)	(0.026)	(0.023)	(0.022)	(0.008)
Medium amount	0.019	0.012	0.005	0.007	-0.067***
	(0.038)	(0.037)	(0.032)	(0.031)	(0.017)
High amount	-0.069	-0.085	-0.062	-0.030	-0.086**
	(0.059)	(0.058)	(0.051)	(0.049)	(0.030)
Subject dummies	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	No
School and class controls	No	Yes	Yes	Yes	No
Student controls	No	No	Yes	Yes	No
Teacher controls	No	No	No	Yes	No
Student fixed effects	No	No	No	No	Yes
Teacher fixed effects	No	No	No	No	Yes
Observations	316,104	316,104	316,104	316,104	316,104

Notes: The dependent variable in all student fixed effects regressions is the within-student between–subjects difference in standardized test scores. In the pooled OLS regressions the dependent variable is the test score in either mathematics or science. Control variables are listed in Table A10 and Table A15 in Appendix IV. All regressions include dummy variables for missing values on each of the included control variables. All student fixed effects regressions also control for country, school, and class fixed effects. Each coefficient is the average from estimates using each of the five plausible values. Jackknife standard errors that are robust to clustering at the class level are in parentheses. Significance is denoted by * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

6. Discussion and conclusions

The most significant findings in the previous section are the student fixed effects estimations indicating the existence of a negative effect of homework assignment in grade 4 and a null effect in grade 8. These results differ from those in most of the previous literature. An exception would be that homework in grade 4 is implied to have worse effects on student achievement than in grade 8. This finding corroborates earlier studies that have also typically found a worse effect in lower grades; the effect of homework has often been found to be zero or slightly positive in lower grades and more positive in higher grades (see Section 2). This could imply that the relation between the supposed negative effects of homework (e.g. stress, academic satiation, and loss of time for leisure activities) and the supposed positive effects (e.g. fostering of discipline and study habits, and repetition) could be different for younger students (Jerrim, Lopez-Agudo, and Marcenaro-Gutierrez 2020); it could be the case that the negative effects weigh heavier in relation to the positive effects compared to older students. Another possibility is that the quality of homework is lower in grade 4 (Jerrim, Lopez-Agudo, and Marcenaro-Gutierrez 2020).

Starting with the negative relationship in the main estimations for grade 4, it can be contrasted with the null effect reported in the meta analyses of the early research, summarized in Cooper (1989) and Cooper et al. (2006), and the small positive effect found in Murillo and Martínez-Garrido (2014). Since this thesis could control for more unobserved, potentially endogenous, variables by using the within-student between-subjects estimation strategy, the implication is that the results in these studies are biased upwards.[26] However, the problem with this comparison is that these studies typically used the time spent on homework as reported by students as the operationalization of homework. Regardless, the assigned homework will have some effect on how much time students spend on homework (Trautwein and Köller 2003) which makes the comparison informative and relevant to make.

The two studies resembling this thesis the most, both in terms of methodology and data set, are Falch and Rønning (2012) and Jerrim, Lopez-Agudo, and Marcenaro-Gutierrez (2020) which both used the first difference estimation strategy and data from TIMSS in grade 4. While Falch and Rønning (2012) found a linear positive effect of homework frequency, the results in this thesis imply a linear negative effect. This disparity can not be explained by the fact that Falch and Rønning (2012) also included teacher fixed effects since the results in Table 5.4 indicate that the negative effect comes from the part of the sample that could include teacher fixed effects. The null effect that was estimated

[26]Though it should also be noted that a null effect was estimated when regressions were run for only the OECD countries and when the continuous homework amount variable was used.

when only the OECD countries were included is, however, closer to the results in Falch and Rønning (2012), who only looked at OECD countries. One explanation for the differences might be that they used TIMSS 2007 and this thesis used TIMSS 2019, which would imply that the relationship between homework frequency and student achievement has changed during these 12 years. The null effect found in Jerrim, Lopez-Agudo, and Marcenaro-Gutierrez (2020) would contradict this interpretation since they used TIMSS data from the subsequent assessment wave in 2011, though their operationalization of homework was teacher-reported homework length, and not frequency.

Another possible explanation is that the effects found in Falch and Rønning (2012) and Jerrim, Lopez-Agudo, and Marcenaro-Gutierrez (2020) obscured the more relevant effect of homework amount. It is possible that an estimated positive effect of frequency and a null effect of length could hide a negative effect if these variables were combined into the homework amount variable used in this thesis. The results reported here do, however, not directly support this notion since the only significant difference between the estimated effect of homework frequency and homework amount is the transformation of a linear negative effect to a homogeneous one. Future studies are, nonetheless, recommended to employ a more holistic approach by also considering homework amount whenever possible.

Comparing the estimated null effect for the 8$^{\text{th}}$ grade students to much of the previous literature faces the same problems described above for the grade 4 students. Nevertheless, most previous studies looking at students in secondary school found a positive effect of homework on student test scores (Cooper 1989; Cooper et al. 2006; Dettmers, Trautwein, and Lüdtke 2009; Eren and Henderson 2011; Gustafsson 2013; Scheerens and Hendriks 2014; Trautwein 2007). Even though much of the earlier research seems to have overstated the positive effects of homework, later studies that controlled for more potential bias inducing unobserved variables also came to the conclusion that there exists a positive effect, albeit smaller than previously indicated (Dettmers, Trautwein, and Lüdtke 2009; Eren and Henderson 2011; Gustafsson 2013; Trautwein 2007). Like in grade 4, a possible explanation for the differences could be that the relationship has changed over the years, though this notion is difficult to test. It is also plausible that the parameter estimates in this thesis are biased downwards. One reason to suspect that this might be the case is that Eren and Henderson (2011) found downwards bias induced by unobserved fixed teacher characteristics in 8$^{\text{th}}$ grade. Although the identification strategy used in this thesis could not control for teacher fixed effects in grade 8, the fact that the parameter estimates in the student fixed effects estimations did not change when teacher controls were added makes it less likely that unobserved teacher characteristics should have a large effect. When the clear upwards bias from fixed unobserved teacher characteristics observed for the 4$^{\text{th}}$ grade students (Table 5.3 and Table 5.4) is taken into account, the existence of a strong downwards bias becomes even more implausible. It is also worth noting how remarkably robust the results in grade 8 were to the sensitivity tests performed in Section 5.4.

A remaining concern that might affect the results in both grades is the drawbacks with the first difference model described in Angrist and Pischke (2009): possible attenuation bias caused by measurement errors and threats to the external validity due to differences between the students contributing to the estimates and those with the same frequency or amount of homework assigned in both subjects, $\Delta HW = 0$. The first concern could explain the disparity between the results in this thesis and most of the previous literature if measurement errors were significant enough. It is, unfortunately, not possible to investigate how large this problem might be. Regardless, because of this possible bias, the results should be interpreted as conservative estimates. Regarding the second possible problem, since there are almost no differences between the two groups for the grade 8 students (see Table A14), the estimated null effect should apply to the whole final estimation sample. Furthermore, since this sample has very small differences compared to the full sample, and the full sample represents the whole populations in the participating countries, the results in grade 8 are deemed to have a high external validity. As discussed in Section 4.1, although the two groups in grade 4 are very similar, the differences between the $\Delta HW = 0$ students and those contributing to the estimates are more prevalent than in the grade 8 sample. The results in grade 4 should, therefore, be interpreted with more caution since even though the internal validity might be high, the external validity could be lower than in grade 8.

Overall, despite the possible problems discussed above, the most reasonable interpretation of the results is that homework assignment measured as frequency and amount either has a negative or zero effect in grade 4 and a zero effect in grade 8. At least, it is highly implausible that assigned homework has a significant positive effect on student achievement. If this is the case, it calls into question why homework is used so extensively by teachers. Furthermore, it raises the question of why the supposed positive effects of homework do not overcome the supposed negative effects, not even for small homework frequencies or homework amounts.

Teachers might use homework so extensively since it is a way to outsource some of the teaching to the parents and to the students themselves. In this way, teachers might be able to focus more on topics they find more interesting or important in class. If these incentives are driving homework assignment practices, it could be problematic. As discussed earlier, another reason for the extensive use of homework as a teaching method could be that, even though there is no clear positive effect on student achievement, other positive effects (e.g. development of discipline and independent problem solving) might make it worthwhile. On account of this possibility one should caution against any drastic policy change recommendations because of the results in this thesis. More research is needed to ascertain whether there truly is no positive effect on student achievement and more focus should be directed at examining the possible effect of homework on other factors such as student self-efficacy and self-discipline.

One possible explanation for the absence of an estimated positive effect in this thesis could be that homework does not help students develop the skills measured in TIMSS, and instead targets national curricula and national tests. What speaks against this explanation is that TIMSS worked with the participating countries to ensure that the test items in mathematics and science reflect their national curricular goals (Martin, Mullis, Foy, et al. 2020, pp. 7, 79, 150, 216).

Another, potentially more plausible explanation, is the inability to control for homework quality in the regressions. Previous research has found that the quality of the assigned homework mediates the effect of homework assignment on student achievement (Dettmers, Trautwein, Lüdtke, et al. 2010; Rosário et al. 2018; Trautwein and Lüdtke 2009).[27] If homework quality is poor, the supposed positive effects of homework might not manifest, at least not fully, while the supposed negative effects do. Thus, it is possible that poor homework quality dampens the estimated effect of homework since all regressions in this thesis let this factor vary, unless the quality of homework is the same within schools, across subjects. If there exists a negative correlation between homework assignment and the quality of the homework, there is a downwards bias in the estimated effect of homework on student achievement in this thesis, although a more likely scenario is that the variation in homework quality for any given level of homework assignment results in an overall null effect. Although the effects of homework quality have been examined for both elementary and secondary school students, few countries have been studied and teacher-reported homework amount as defined in this thesis has not been included in previous studies. Future research is therefore urged to further examine the relationship between homework assignment and student achievement, especially in larger data sets such as TIMSS and PISA, including measures of homework quality so as to come closer to the true relationship.

Despite acknowledging the limitations of the empirical method employed and the other potential threats to a causal estimate discussed above, the conclusion of the analysis is that homework assignment, measured as frequency and amount, does not have a positive effect on student achievement. In grade 4, there is even strong evidence for a negative effect, at least in non-OECD countries. If the goal of education policy makers is to increase the education level of a country to affect economic growth in a positive way, they are cautioned that the present study implies that previously reported positive effects of homework may have been overstated, and that homework as a means to increase student achievement may be unhelpful or even counterproductive. Homework often requires significant amounts of invested time, while also inducing stress for teachers, students,

[27] All of these papers used student-reported perception of homework quality as the operationalization. As discussed in Section 2, student-reported measures are more prone to measurement errors and self-report bias, which means that these results should be interpreted with some caution.

and their parents. Assigning large amounts of homework, especially to younger students, should therefore not be done haphazardly without due consideration of the quality of the homework. On the contrary, the results in this thesis indicate that, if anything, the assigned homework amount should be reduced in most schools, particularly for elementary school students.

References

ACT, Inc (2013). *Student- and Teacher-Reported Behavioral Measures: Do They Agree?* Issue Brief. URL: https://eric.ed.gov/?id=ED546856.

Angrist, Joshua and Jorn-Steffen Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. 1st ed. Princeton University Press. URL: https://EconPapers.repec.org/RePEc:pup:pbooks:8769.

Arregui Alegria, Iker and Ruslan Gatykaev (2019). "The Impact of Homework on Student Achievement: Evidence from 57 countries". Master's thesis. Lund University, Department of Economics.

Bas, Gokhan, Cihad Senturk, and Fatih Mehmet Cigerci (2017). "Homework and academic achievement: A meta-analytic review of research". In: *Issues in Educational Research* 27.1, pp. 31–50. URL: http://www.iier.org.au/iier27/bas.pdf.

Bempechat, Janine (2004). "The Motivational Benefits of Homework: A Social-Cognitive Perspective". In: *Theory Into Practice* 43.3, pp. 189–196. URL: https://doi.org/10.1207/s15430421tip4303_4.

Bietenbeck, Jan (2014). "Teaching practices and cognitive skills". In: *Labour Economics* 30, pp. 143–153. URL: http://www.sciencedirect.com/science/article/pii/S0927537114000219.

Blazer, Christie (2009). "Literature review: Homework". In: *Research Services, Miami-Dade County Public Schools*. URL: https://files.eric.ed.gov/fulltext/ED536245.pdf.

Buckley, Katie and Sara Bartolino Krachman (2016). *Patterns in Student Self-Report and Teacher Report Measures of Social-Emotional Mindsets, Skills, and Habits, Initial findings from the Boston Charter Research Collaborative*. Working Paper. Transforming Education. URL: https://www.transformingeducation.org/wp-content/uploads/2017/04/TE-BCRCWorkingPaperFINAL.pdf.

Caro, Daniel H., Jenny Lenkeit, and Leonidas Kyriakides (2016). "Teaching strategies and differential effectiveness across learning contexts: Evidence from PISA 2012". In: *Studies in Educational Evaluation* 49, pp. 30–41. URL: http://www.sciencedirect.com/science/article/pii/S0191491X15300286.

Chin, Joseph Meng-Chun, Hsin-Chih Lin, and Chun-Wei Chen (2020). "Homework and learning achievements: how much homework is enough?" In: *Educational Studies* 0.0, pp. 1–16. URL: https://doi.org/10.1080/03055698.2020.1766423.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor (2010). "Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects". In: *The Journal of Human Resources* 45.3, pp. 655–681. URL: http://www.jstor.org/stable/25703472.

Cooper, Harris, Jorgianne Civey Robinson, and Erika A Patall (2006). "Does Homework Improve Academic Achievement? A Synthesis of Research, 1987–2003". In: *Review of Educational Research* 76.1, pp. 1–62. URL: https://doi.org/10.3102/00346543076001001.

Cooper, Harris M. (1989). *Homework*. White Plains, New York: Longman.

Dee, Thomas S. (May 2005). "A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?" In: *American Economic Review* 95.2, pp. 158–165. URL: https://www.aeaweb.org/articles?id=10.1257/000282805774670446.

— (2007). "Teachers and the Gender Gaps in Student Achievement." In: *Journal of Human Resources* 42.3, pp. 528–554. URL: http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=25852724&site=eds-live&scope=site.

Desimone, Laura, Thomas Smith, and David Frisvold (Mar. 2010). "Survey Measures of Classroom Instruction, Comparing Student and Teacher Reports". In: *Educational Policy - EDUC POLICY* 24, pp. 267–329. URL: https://journals.sagepub.com/doi/10.1177/0895904808330173.

Dettmers, Swantje, Ulrich Trautwein, and Oliver Lüdtke (2009). "The relationship between homework time and achievement is not universal: evidence from multilevel analyses in 40 countries". In: *School Effectiveness and School Improvement* 20.4, pp. 375–405. URL: https://doi.org/10.1080/09243450902904601.

Dettmers, Swantje, Ulrich Trautwein, Oliver Lüdtke, et al. (May 2010). "Homework Works if Homework Quality Is High: Using Multilevel Modeling to Predict the Development of Achievement in Mathematics". In: *Journal of Educational Psychology* 102, pp. 467–482.

Eren, Ozkan and Daniel Henderson (2011). "Are we wasting our children's time by giving them more homework?" In: *Economics of Education Review* 30.5, pp. 950–961. URL: https://www.sciencedirect.com/science/article/pii/S0272775711000549.

Falch, Torberg and Marte Rønning (2012). *Homework assignment and student achievement in OECD countries*. eng. Discussion Papers 711. Oslo. URL: http://hdl.handle.net/10419/192693.

Fishbein, Bethany, Pierre Foy, and Liqun Yin (2021). *TIMSS 2019 User Guide for the International Database*. Lynch School of Education, Boston College: TIMSS & PIRLS International Study Center.

Fuchs, Thomas and Ludger Wößmann (2008). "What accounts for international differences in student performance? A re–examination using PISA data". In: ed. by Machin S. Dustmann C. Fitzenberger B. The Economics of Education and Training. Studies in Empirical Economics. Physica-Verlag HD, pp. 209–240.

Gustafsson, Jan-Eric (2013). "Causal inference in educational effectiveness research: a comparison of three methods to investigate effects of homework on student achievement". In: *School Effectiveness and School Improvement* 24.3, pp. 275–295. URL: https://doi.org/10.1080/09243453.2013.806334.

Hanushek, Eric and Ludger Wößmann (2012). *Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation*. Munich Reprints in Economics. University of Munich, Department of Economics. URL: https://EconPapers.repec.org/RePEc:lmu:muenar:20400.

Hattie, John (2012). *Visible learning for teachers: Maximizing impact on learning*. 1st ed. Routledge.

Jerrim, John, Luis Alejandro Lopez-Agudo, and Oscar D. Marcenaro-Gutierrez (2020). "The association between homework and primary school children's academic achievement. International evidence from PIRLS and TIMSS". In: *European Journal of Education* 55.2, pp. 248–260. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/ejed.12374.

Jerrim, John, Luis Alejandro Lopez-Agudo, Oscar D. Marcenaro-Gutierrez, and Nikki Shure (2017). "What happens when econometrics and psychometrics collide? An example using the PISA data". In: *Economics of Education Review* 61, pp. 51–58. URL: http://www.sciencedirect.com/science/article/pii/S0272775717300869.

Jong, R. de, K.J. Westerhof, and B.P.M. Creemers (2000). "Homework and Student Math Achievement in Junior High Schools". In: *Educational Research and Evaluation* 6.2, pp. 130–157. URL: https://doi.org/10.1076/1380-3611(200006)6:2;1-E;F130.

Lavy, Victor (July 2010). *Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries*. NBER Working Papers 16227. National Bureau of Economic Research, Inc. URL: http://www.nber.org/papers/w16227.

— (Mar. 2016). "What Makes an Effective Teacher? Quasi-Experimental Evidence". In: *CESifo Economic Studies* 62.1, pp. 88–125. URL: https://doi.org/10.1093/cesifo/ifv001.

Lucas, Robert Jr. (July 1988). "On the mechanics of economic development". In: *Journal of Monetary Economics* 22.1, pp. 3–42. URL: https://ideas.repec.org/a/eee/moneco/v22y1988i1p3-42.html.

Macdonald, Kevin (2008). *PV: Stata module to perform estimation with plausible values*. Statistical Software Components, Boston College Department of Economics. URL: https://ideas.repec.org/c/boc/bocode/s456951.html.

Martin, Michael O, Matthias von Davier, and Ina V.S. Mullis (2020). *Methods and Procedures: TIMSS 2019 Technical Report*. Lynch School of Education, Boston College: TIMSS & PIRLS International Study Center.

Martin, Michael O, Ina V.S. Mullis, Pierre Foy, et al. (2020). *TIMSS 2019 International Results in Mathematics and Science*. Lynch School of Education, Boston College: TIMSS & PIRLS International Study Center.

Martin, Michael O, Ina VS Mullis, Eugenio J Gonzalez, et al. (2004). *Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades. TIMSS 2003 International Science Report*. Lynch School of Education, Boston College: TIMSS & PIRLS International Study Center. URL: https://eric.ed.gov/?id=ED494651.

Mattsson, Jens (2020). "Teaching Practices and Collaborative Problem Solving - Evidence from PISA 2015". 1st year Master's thesis. Lund University, Department of Economics.

Murillo, F. Javier and Cynthia Martínez-Garrido (2014). "Homework and primary-school students' academic achievement in Latin America". In: *International Review of Education* 60, pp. 1–23.

Romer, Paul (1990). “Endogenous Technological Change”. In: *Journal of Political Economy* 98.5, pp. 71–102. URL: https://EconPapers.repec.org/RePEc:ucp:jpolec:v:98:y:1990:i:5:p:s71-102.

Rønning, Marte (2010). *Homework and pupil achievement in Norway: evidence from TIMSS*. Statistics Norway, Oslo. URL: https://scholar.google.com/scholar?hl=sv%5C&as_sdt=0%5C%2C5&q=Homework+and+pupil+achievement+in+Norway+Evidence+from+TIMSS+%5C&btnG=.

Rosário, Pedro et al. (2018). “Homework purposes, homework behaviors, and academic achievement. Examining the mediating role of students’ perceived homework quality”. In: *Contemporary Educational Psychology* 53, pp. 168–180. URL: https://www.sciencedirect.com/science/article/pii/S0361476X1730111X.

Scheerens, Jaap and Maria Hendriks (2014). “State of the Art of Time Effectiveness”. In: *Effectiveness of Time Investments in Education: Insights from a review and meta-analysis*. Ed. by Jaap Scheerens. Cham: Springer International Publishing, pp. 7–29. URL: https://doi.org/10.1007/978-3-319-00924-7_2.

Scheerens, Jaap, Johannes W. Luyten, et al. (2007). *Review and meta-analyses of school and teaching effectiveness*. Universiteit Twente, Afdeling Onderwijsorganisatie en -management.

Schwerdt, Guido and Amelie C. Wuppermann (2011). “Is traditional teaching really all that bad? A within-student between–subject approach”. In: *Economics of Education Review* 30.2, pp. 365–379. URL: http://www.sciencedirect.com/science/article/pii/S0272775710001640.

Trautwein, Ulrich (2007). “The homework–achievement relation reconsidered: Differentiating homework time, homework frequency, and homework effort”. In: *Learning and Instruction* 17.3, pp. 372–388. URL: https://www.sciencedirect.com/science/article/pii/S0959475207000369.

Trautwein, Ulrich and Olaf Köller (2003). “The Relationship Between Homework and Achievement—Still Much of a Mystery.” In: *Educational Psychology Review* 15.2, pp. 115–145. URL: https://link.springer.com/article/10.1023/A:1023460414243.

Trautwein, Ulrich and Oliver Lüdtke (2009). “Predicting homework motivation and homework effort in six school subjects: The role of person and family characteristics, classroom factors, and school track". In: *Learning and Instruction* 19.3, pp. 243–258. URL: https://www.sciencedirect.com/science/article/pii/S0959475208000546.

Wu, Margaret (2005). "The role of plausible values in large-scale surveys". In: *Studies in Educational Evaluation* 31.2. Measurement, Evaluation, and Statistical Analysis, pp. 114–128. URL: http://www.sciencedirect.com/science/article/pii/S0191491X05000209.

Zhu, Yan (2015). "Homework and Mathematics Learning: What Can We Learn from the TIMSS Series Studies in the Last Two Decades?" In: *Large-Scale Studies in Mathematics Education*. Ed. by James A. Middleton, Jinfa Cai, and Stephen Hwang. Cham: Springer International Publishing, pp. 209–234. URL: https://doi.org/10.1007/978-3-319-07716-1_10.

Zhu, Yan and Frederick Koon Shing Leung (2012). "Homework and Mathematics Achievement in Hong Kong: Evidence from the TIMSS 2003." In: *International Journal of Science and Mathematics Education* 10.4, pp. 907–925. URL: https://link-springer-com.ludwig.lub.lu.se/article/10.1007/s10763-011-9302-3.

Appendix

I. Data cleaning procedures

Back to Data.

Table A1. Full samples and the number of students left after each data cleaning step.

	Grade 4	Grade 8
Number of students in full sample	308,609	227,345
Step 1. Remove students with more than one teacher in mathematics or science	301,504	158,245
Step 2. Remove students with teachers who did not answer homework questions	263,118	134,981
Step 3. Remove students from countries with too few observations left after step 1 and 2	263,118	134,383
Percent of sample left	85.26 %	59.11 %

II. Heterogeneity regressions

Back to OECD vs non-OECD countries.

- Grade 4 OECD countries:
  - Australia, Austria, Canada, Chile, Czech Republic, Denmark, Finland, France, Germany, Hungary, Ireland, Italy, Japan, South Korea, Latvia, Lithuania, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Spain, Sweden, Turkey, United states, England, Northern Ireland, Belgium (Flanders).
- Grade 4 non-OECD countries:
  - Albania, Azerbaijan, Bahrain, Armenia, Bosnia and Herzegovina, Bulgaria, Chinese Taipei, Croatia, Cyprus, Georgia, Hong Kong, SAR, Iran, Islamic Republic of, Kazakhstan, Kosovo, Kuwait, Malta, Montenegro, Morocco, Oman, Pakistan, Philippines, Qatar, Russian Federation, Saudi Arabia, Serbia, Singapore, South Africa, United Arab Emirates, North Macedonia.
- Grade 8 OECD countries:
  - Australia, Austria, Canada, Chile, Czech Republic, Denmark, Finland, France, Germany, Hungary, Ireland, Italy, Japan, Korea, Latvia, Lithuania, Netherlands, New Zealand, Norway, Portugal, Slovakia, Spain, Sweden, United states, England, Northern Ireland, Belgium (Flanders).

- Grade 8 non-OECD countries:
  - Albania, Azerbaijan, Armenia, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Georgia, Iran, Kazakhstan, Kosovo, Malta, Montenegro, Morocco, Oman, Pakistan, Philippines, Qatar, Russian Federation, Serbia, Singapore, United Arab Emirates, North Macedonia.

Table A2. Fixed effects (FE) estimations of the relationship between homework assignment and student achievement in OECD and non-OECD countries, in both 4^{th} and 8^{th} grade.

	4^{th} grade				8^{th} grade			
	OECD		Non-OECD		OECD		Non-OECD	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Homework frequency *Reference group is "No homework"*								
Less than once a week	0.009	0.008	0.014	0.037	-0.009	-0.005	0.027	0.019
	(0.009)	(0.009)	(0.050)	(0.038)	(0.019)	(0.019)	(0.025)	(0.025)
1 or 2 times a week	0.015	0.015	-0.038**	-0.036**	0.021	0.023	-0.035	-0.030
	(0.011)	(0.009)	(0.014)	(0.013)	(0.017)	(0.017)	(0.025)	(0.026)
3 or 4 times a week	-0.003	-0.005	-0.114*	-0.087*	-0.003	0.005	0.010	0.011
	(0.012)	(0.010)	(0.052)	(0.038)	(0.019)	(0.019)	(0.029)	(0.029)
Every day	0.004	0.001	-0.010	-0007	0.032	0.036	0.002	0.009
	(0.018)	(0.013)	(0.052)	(0.040)	(0.025)	(0.024)	(0.032)	(0.032)
Panel B: Homework amount *Reference group is "No homework"*								
Low amount	0.023	0.022	-0.060*	-0.053*	0.010	0.012	-0.026	-0.025
	(0.015)	(0.008)	(0.028)	(0.025)	(0.016)	(0.016)	(0.026)	(0.026)
Medium amount	0.001	0.001	-0.056**	-0.036*	-0.001	0.005	-0.015	-0.013
	(0.010)	(0.010)	(0.018)	(0.016)	(0.018)	(0.017)	(0.026)	(0.027)
High amount	-0.004	-0.003	-0.080**	-0.072**	0.026	0.035	0.025	0.028
	(0.023)	(0.022)	(0.029)	(0.025)	(0.027)	(0.028)	(0.030)	(0.030)
Subject dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Teacher controls	No	Yes	No	Yes	No	Yes	No	Yes
Student fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	235,720	235,720	290,516	290,516	73,986	73,986	194,780	194,780

Notes: The dependent variable in all student fixed effects regressions is the within-student between–subjects difference in standardized test scores. All regressions include dummy variables for missing values on each of the included control variables. All student fixed effects regressions also control for country, school, and class fixed effects. Each coefficient is the average from estimates using each of the five plausible values. Jackknife standard errors that are robust to clustering at the class level are in parentheses. Significance is denoted by * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

III. Robustness regressions

III.1 Without the missing control dummy variables

Back to Robustness checks.

Table A3. Pooled OLS and fixed effects (FE) estimations of the relationship between homework assignment and student achievement, without the missing control dummy variables, for 4^{th} grade students.

	OLS				FE	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Homework frequency *Reference group is "No homework"*						
Less than once a week	-0.017	-0.032	-0.023	-0.015	-0.039***	-0.036***
	(0.032)	(0.030)	(0.026)	(0.025)	(0.011)	(0.010)
1 or 2 times a week	-0.016	-0.019	-0.018	-0.018	-0.078***	-0.072***
	(0.030)	(0.027)	(0.023)	(0.024)	(0.012)	(0.011)
3 or 4 times a week	-0.031	-0.024	-0.025	-0.023	-0.127***	-0.122***
	(0.050)	(0.048)	(0.043)	(0.042)	(0.023)	(0.022)
Every day	-0.019	0.001	0.011	-0.006	-0.012	-0.010
	(0.065)	(0.061)	(0.050)	(0.044)	(0.023)	(0.020)
Panel B: Homework amount *Reference group is "No homework"*						
Low amount	-0.015	-0.024	-0.017	-0.014	-0.063***	-0.058***
	(0.028)	(0.025)	(0.022)	(0.022)	(0.012)	(0.010)
Medium amount	-0.026	-0.018	-0.020	-0.021	-0.081***	-0.078***
	(0.041)	(0.040)	(0.034)	(0.034)	(0.018)	(0.016)
High amount	-0.082	-0.110	-0.049	-0.051	-0.073*	-0.073*
	(0.087)	(0.066)	(0.076)	(0.066)	(0.033)	(0.030)
Subject dummies	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	No	No
School and class controls	No	Yes	Yes	Yes	No	No
Student controls	No	No	Yes	Yes	No	No
Teacher controls	No	No	No	Yes	No	Yes
Student fixed effects	No	No	No	No	Yes	Yes
Observations	526,236	526,236	526,236	526,236	526,236	526,236

Notes: The dependent variable in all student fixed effects regressions is the within-student between–subjects difference in standardized test scores. In the pooled OLS regressions the dependent variable is the test score in either mathematics or science. Control variables are listed in Table A10 and Table A15 in Appendix IV. All student fixed effects regressions also control for country, school, and class fixed effects. Each coefficient is the average from estimates using each of the five plausible values. Jackknife standard errors that are robust to clustering at the class level are in parentheses. Significance is denoted by * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A4. Pooled OLS and fixed effects (FE) estimations of the relationship between homework assignment and student achievement, without the missing control dummy variables, for 8^{th} grade students.

	OLS				FE	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Homework frequency *Reference group is "No homework"*						
Less than once a week	0.147**	0.108*	0.070*	0.065*	0.007	0.007
	(0.052)	(0.048)	(0.036)	(0.032)	(0.016)	(0.016)
1 or 2 times a week	0.147**	0.094*	0.055	0.042	-0.011	-0.010
	(0.049)	(0.046)	(0.035)	(0.032)	(0.015)	(0.015)
3 or 4 times a week	0.335***	0.223***	0.153***	0.139***	0.003	0.006
	(0.052)	(0.052)	(0.039)	(0.036)	(0.017)	(0.017)
Every day	0.327***	0.216***	0.138**	0.121**	0.013	0.015
	(0.065)	(0.060)	(0.046)	(0.042)	(0.020)	(0.021)
Panel B: Homework amount *Reference group is "No homework"*						
Low amount	0.139**	0.092*	0.055	0.046	-0.004	-0.004
	(0.049)	(0.046)	(0.034)	(0.031)	(0.015)	(0.015)
Medium amount	0.240***	0.164***	0.108**	0.096**	-0.003	-0.001
	(0.051)	(0.049)	(0.037)	(0.033)	(0.016)	(0.016)
High amount	0.379***	0.252***	0.168***	0.150***	0.034	0.035
	(0.061)	(0.058)	(0.045)	(0.040)	(0.020)	(0.021)
Subject dummies	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	No	No
School and class controls	No	Yes	Yes	Yes	No	No
Student controls	No	No	Yes	Yes	No	No
Teacher controls	No	No	No	Yes	No	Yes
Student fixed effects	No	No	No	No	Yes	Yes
Observations	526,236	526,236	526,236	526,236	526,236	526,236

Notes: The dependent variable in all student fixed effects regressions is the within-student between–subjects difference in standardized test scores. In the pooled OLS regressions the dependent variable is the test score in either mathematics or science. Control variables are listed in Table A10 and Table A15 in Appendix IV. All student fixed effects regressions also control for country, school, and class fixed effects. Each coefficient is the average from estimates using each of the five plausible values. Jackknife standard errors that are robust to clustering at the class level are in parentheses. Significance is denoted by * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

III.2 Continuous homework amount

Back to Robustness checks.

Table A5. Fixed effects (FE) estimations of the relationship between continuous homework amount and student achievement, for both grades.

	4^{th} grade		8^{th} grade	
	(1)	(2)	(3)	(4)
Continuous homework amount in minutes per week	-0.00009 (0.00017)	-0.00001 (0.00013)	0.00015 (0.00008)	0.00016 (0.00008)
Subject dummies	Yes	Yes	Yes	Yes
Teacher controls	No	Yes	No	Yes
Student fixed effects	Yes	Yes	Yes	Yes
Observations	523,443	523,443	268,766	268,766

Notes: The dependent variable in all student fixed effects regressions is the within-student between–subjects difference in standardized test scores. Control variables are listed in Table A8 and Table A13. All student fixed effects regressions also control for country, school, and class fixed effects. Each coefficient is the average from estimates using each of the five plausible values. Jackknife standard errors that are robust to clustering at the class level are in parentheses. Significance is denoted by * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

IV. Additional descriptive statistics

Back to Data. Back to Empirical strategy.

IV.1 4th grade

Table A6. Descriptive statistics for the countries in the final grade 4 sample.

Country	Schools	Teachers	Students	Mathematics test score	Science test score
Albania	152	184	3,952	494	489
Armenia	136	234	4,476	498	466
Australia	216	300	4,309	516	533
Austria	174	263	3,906	539	522
Azerbaijan, Republic of	181	278	4,799	515	427
Bahrain	172	398	5,310	480	493
Belgium (Flanders)	119	193	3,492	532	501
Bosnia and Herzegovina	176	315	5,274	452	459
Bulgaria	129	178	3,601	515	521
Canada	586	872	10,737	512	523
Chile	126	182	3,133	441	469
Chinese Taipei	156	357	3,602	599	558
Croatia	152	257	3,710	509	524
Cyprus	149	331	3,923	532	511
Czech Republic	152	322	4,630	533	534
Denmark	125	214	2,364	525	522
England	46	48	1,043	556	537

Finland	147	270	4,065	532	555
France	151	331	3,778	485	488
Georgia	146	286	3,580	482	454
Germany	171	253	2,942	521	518
Hong Kong, SAR	106	227	2,236	602	531
Hungary	148	302	4,036	523	529
Iran, Islamic Republic of	219	219	5,869	443	441
Ireland	148	224	4,458	548	528
Italy	159	223	3,644	515	510
Japan	83	122	2,216	593	562
Kazakhstan	147	186	4,011	512	494
Korea, Republic of	139	204	3,535	600	588
Kosovo	137	204	4,203	444	413
Kuwait	156	320	4,229	383	392
Latvia	149	262	4,247	546	542
Lithuania	195	235	3,517	542	538
Malta	97	227	3,543	509	496
Montenegro	139	340	4,801	453	453
Morocco	244	443	6,954	383	374
Netherlands	92	119	2,222	538	518
New Zealand	135	307	3,983	487	503
North Macedonia	141	217	3,012	472	426
Northern Ireland	121	134	3,031	566	518
Norway	84	134	1,848	543	539
Oman	214	377	6,416	431	435
Pakistan	82	150	2,807	328	290
Philippines	170	281	5,188	297	249
Poland	143	394	4,583	520	531
Portugal	178	301	4,164	525	504
Qatar	235	409	4,799	449	449
Russian Federation	188	188	3,791	567	567
Saudi Arabia	143	287	3,491	398	402
Serbia	158	205	4,215	508	517
Singapore	187	626	5,920	625	595
Slovak Republic	152	333	4,087	510	521
South Africa	276	539	10,936	374	324
Spain	457	665	8,704	502	511
Sweden	129	199	3,112	521	537
Turkey	177	354	3,953	523	526
United Arab Emirates	501	1222	16,610	481	473
United States	274	480	8,121	535	539
Total	10,065	17,725	263,118		

Table A7. Means and standard deviations in full and final sample for the grade 4 students.

	Full sample		Final sample	
	Mean	SD	Mean	SD
Student sex				
Male	0.51	0.50	0.51	0.50
Female	0.49	0.50	0.49	0.50
Student Age	10.36	0.77	10.36	0.75
Books at home				
0—10 books	0.29	0.45	0.35	0.48
11—25 books	0.26	0.44	0.22	0.42
26—100 books	0.25	0.43	0.24	0.43
101—200 books	0.10	0.31	0.08	0.28
more than 200	0.09	0.29	0.10	0.31
Fathers' education				
Did not go to school	0.07	0.25	0.06	0.23
Less than lower secondary	0.12	0.33	0.12	0.32
Lower secondary	0.13	0.34	0.13	0.34
Upper secondary education	0.25	0.44	0.26	0.44
Post-secondary, non-tertiary	0.09	0.29	0.09	0.29
Short-cycle tertiary	0.08	0.28	0.09	0.28
Bachelor's or equivalent	0.13	0.34	0.14	0.34
Postgraduate degree: Master's or Doctor	0.10	0.30	0.10	0.30
Mothers' education				
Did not go to school	0.09	0.28	0.08	0.27
Less than lower secondary	0.12	0.32	0.11	0.32
Lower secondary	0.13	0.33	0.13	0.33
Upper secondary education	0.23	0.42	0.24	0.42
Post-secondary, non-tertiary	0.09	0.28	0.09	0.29
Short-cycle tertiary	0.09	0.29	0.10	0.29
Bachelor's or equivalent	0.13	0.34	0.14	0.34
Postgraduate degree: Master's or Doctor	0.10	0.29	0.10	0.30
Student born in country?				
Yes	0.94	0.23	0.94	0.23
No	0.06	0.23	0.06	0.23
Teacher characteristics				
Teacher sex				
Male	0.23	0.42	0.22	0.41
Female	0.77	0.42	0.78	0.41
Teacher age				
Under 25	0.07	0.26	0.07	0.25
25–29	0.14	0.35	0.13	0.34
30–39	0.25	0.43	0.25	0.43
40–49	0.29	0.45	0.29	0.45
50–59	0.20	0.40	0.20	0.40

60 or more	0.05	0.22	0.05	0.22
Teaching experience				
1 year or less	0.05	0.21	0.04	0.20
2 years	0.05	0.22	0.05	0.22
3-5 years	0.14	0.35	0.14	0.35
>5 years	0.76	0.43	0.76	0.42
Nr of students	301,504		263,118	

Table A8. Descriptive statistics for the teacher controls in the final estimation sample for the grade 4 students.

	No HW		Low amount		Medium amount		High amount		%
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	missing
Teacher sex									
Male	0.19	0.39	0.21	0.41	0.21	0.41	0.33	0.47	0.36
Female	0.81	0.39	0.79	0.41	0.79	0.41	0.67	0.47	0.36
Teacher age									
Under 25	0.04	0.20	0.05	0.22	0.09	0.29	0.12	0.33	0.35
25–29	0.17	0.38	0.13	0.34	0.13	0.33	0.10	0.30	0.35
30–39	0.26	0.44	0.26	0.44	0.24	0.43	0.24	0.43	0.35
40–49	0.30	0.46	0.30	0.46	0.29	0.45	0.26	0.44	0.35
50–59	0.19	0.39	0.21	0.40	0.20	0.40	0.24	0.43	0.35
60 or more	0.05	0.21	0.06	0.24	0.05	0.22	0.04	0.20	0.35
Teacher certificate									
No	0.28	0.45	0.32	0.47	0.31	0.46	0.41	0.49	10.53
Yes	0.72	0.45	0.68	0.47	0.69	0.46	0.59	0.49	10.53
Education level									
Less than Bachelor	0.06	0.24	0.16	0.37	0.24	0.43	0.29	0.45	2.93
Bachelor	0.55	0.50	0.52	0.50	0.44	0.50	0.43	0.49	2.93
Master or doctor	0.39	0.49	0.31	0.46	0.32	0.47	0.29	0.45	2.93
Major match with subject taught?									
No	0.88	0.33	0.76	0.43	0.72	0.45	0.74	0.44	2.35
Yes	0.12	0.33	0.24	0.43	0.28	0.45	0.26	0.44	2.35
Teaching experience									
1 year or less	0.05	0.22	0.04	0.20	0.04	0.19	0.03	0.18	2.30
2 years	0.06	0.24	0.05	0.22	0.04	0.19	0.08	0.28	2.30
3-5 years	0.13	0.34	0.13	0.34	0.15	0.36	0.15	0.36	2.30
>5 years	0.76	0.43	0.77	0.42	0.77	0.42	0.70	0.46	2.30
How often do you feel the following way about being a teacher?									
– Content									
Very often	0.50	0.50	0.58	0.49	0.61	0.49	0.70	0.46	0.36
Often	0.38	0.48	0.34	0.47	0.32	0.47	0.26	0.44	0.36
Sometimes	0.12	0.32	0.07	0.26	0.07	0.25	0.04	0.21	0.36
Never or almost never	0.01	0.10	0.01	0.08	0.01	0.10	0.00	0.02	0.36
– Meaning and purpose									
Very often	0.59	0.49	0.67	0.47	0.71	0.45	0.63	0.48	0.35

Often	0.33	0.47	0.27	0.44	0.24	0.43	0.34	0.47	0.35
Sometimes	0.08	0.27	0.06	0.23	0.04	0.19	0.02	0.15	0.35
Never or almost never	0.00	0.07	0.00	0.05	0.00	0.06	0.00	0.02	0.35
– Enthusiastic									
Very often	0.51	0.50	0.62	0.49	0.66	0.47	0.67	0.47	0.39
Often	0.37	0.48	0.30	0.46	0.28	0.45	0.30	0.46	0.39
Sometimes	0.11	0.31	0.07	0.25	0.05	0.22	0.03	0.17	0.39
Never or almost never	0.01	0.10	0.01	0.08	0.01	0.09	0.00	0.04	0.39
– Work inspires me									
Very often	0.52	0.50	0.60	0.49	0.64	0.48	0.60	0.49	0.48
Often	0.34	0.47	0.30	0.46	0.28	0.45	0.33	0.47	0.48
Sometimes	0.13	0.33	0.09	0.29	0.07	0.25	0.08	0.26	0.48
Never or almost never	0.01	0.10	0.01	0.09	0.01	0.07	0.00	0.04	0.48
– Proud									
Very often	0.60	0.49	0.66	0.47	0.72	0.45	0.67	0.47	0.36
Often	0.32	0.47	0.27	0.44	0.23	0.42	0.27	0.44	0.36
Sometimes	0.08	0.27	0.06	0.24	0.05	0.22	0.05	0.22	0.36
Never or almost never	0.00	0.06	0.01	0.08	0.00	0.07	0.01	0.08	0.36
Too many students in the classes?									
Agree a lot	0.35	0.48	0.35	0.48	0.31	0.46	0.36	0.48	2.78
Agree a little	0.35	0.48	0.33	0.47	0.37	0.48	0.33	0.47	2.78
Disagree a little	0.19	0.39	0.20	0.40	0.18	0.38	0.19	0.39	2.78
Disagree a lot	0.10	0.31	0.12	0.33	0.13	0.34	0.12	0.33	2.78
Hours in professional development past two years									
None	0.82	0.38	0.69	0.46	0.64	0.48	0.68	0.47	0.55
Less than 6 hours	0.05	0.22	0.08	0.27	0.10	0.30	0.05	0.22	0.55
6–15 hours	0.05	0.22	0.09	0.28	0.08	0.27	0.08	0.27	0.55
16–35 hours	0.03	0.18	0.06	0.23	0.07	0.25	0.06	0.25	0.55
More than 35 hours	0.05	0.21	0.09	0.29	0.11	0.31	0.13	0.33	0.55
Total instruction time (min/week)	330.27	63.22	306.53	73.59	306.30	67.21	294.79	68.40	6.46

Table A9. Means and standard deviations for the zero-difference students vs the non zero-difference students in the final sample for grade 4.

	No zero-difference		Zero-difference	
	Mean	SD	Mean	SD
Student characteristics				
Student sex				
Male	0.50	0.50	0.52	0.50
Female	0.50	0.50	0.48	0.50
Student Age	10.36	0.68	10.36	0.85
Books at home				
0—10 books	0.30	0.46	0.42	0.49
11—25 books	0.22	0.42	0.22	0.42

26—100 books	0.27	0.44	0.20	0.40
101—200 books	0.09	0.29	0.07	0.25
more than 200	0.12	0.32	0.08	0.28
Fathers' education				
Did not go to school	0.05	0.21	0.07	0.26
Less than lower secondary	0.09	0.29	0.16	0.36
Lower secondary	0.12	0.33	0.15	0.36
Upper secondary education	0.26	0.44	0.25	0.44
Post-secondary, non-tertiary	0.10	0.30	0.07	0.26
Short-cycle tertiary	0.09	0.29	0.08	0.27
Bachelor's or equivalent	0.15	0.36	0.12	0.32
Postgraduate degree: Master's or Doctor	0.12	0.32	0.08	0.26
Mothers' education				
Did not go to school	0.06	0.24	0.11	0.31
Less than lower secondary	0.09	0.28	0.15	0.36
Lower secondary	0.12	0.32	0.14	0.35
Upper secondary education	0.24	0.43	0.23	0.42
Post-secondary, non-tertiary	0.10	0.30	0.07	0.25
Short-cycle tertiary	0.10	0.31	0.08	0.27
Bachelor's or equivalent	0.15	0.35	0.12	0.33
Postgraduate degree: Master's or Doctor	0.12	0.32	0.07	0.26
Student born in country?				
Yes	0.94	0.23	0.94	0.23
No	0.06	0.23	0.06	0.23
Teacher characteristics				
Teacher sex				
Male	0.20	0.40	0.25	0.43
Female	0.80	0.40	0.75	0.43
Teacher age				
Under 25	0.05	0.21	0.11	0.31
25–29	0.13	0.34	0.14	0.35
30–39	0.23	0.42	0.28	0.45
40–49	0.31	0.46	0.27	0.44
50–59	0.23	0.42	0.17	0.37
60 or more	0.06	0.24	0.04	0.20
Teaching experience				
1 year or less	0.04	0.19	0.05	0.21
2 years	0.05	0.21	0.06	0.23
3-5 years	0.13	0.33	0.17	0.38
>5 years	0.79	0.41	0.72	0.45
Percent of estimation sample	66.42 %		33.58 %	

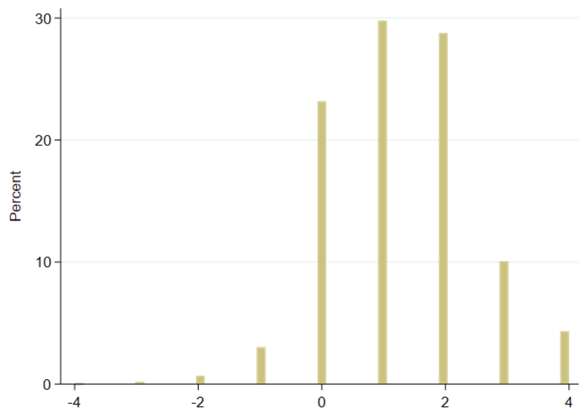
Table A10. Descriptive statistics for the OLS controls in the final estimation sample for the grade 4 students.

	Mean	SD	% missing
Student level controls			
Student sex			
Male	0.51	0.50	1.24
Female	0.49	0.50	1.24
Student Age	10.36	0.75	0.14
Student language at home same as test language?			
Always	0.60	0.49	18.40
Almost always	0.09	0.28	18.40
Sometimes	0.26	0.44	18.40
Never	0.05	0.23	18.40
Books at home			
0—10 books	0.35	0.48	18.10
11—25 books	0.22	0.42	18.10
26—100 books	0.24	0.43	18.10
101—200 books	0.08	0.28	18.10
more than 200	0.10	0.31	18.10
Computer at home?			
No	0.24	0.42	2.32
Yes	0.76	0.42	2.32
Own room?			
No	0.40	0.49	2.47
Yes	0.60	0.49	2.47
Fathers' education			
Did not go to school	0.06	0.23	23.13
Less than lower secondary	0.12	0.32	23.13
Lower secondary	0.13	0.34	23.13
Upper secondary education	0.26	0.44	23.13
Post-secondary, non-tertiary	0.09	0.29	23.13
Short-cycle tertiary	0.09	0.28	23.13
Bachelor's or equivalent	0.14	0.34	23.13
Postgraduate degree: Master's or Doctor	0.10	0.30	23.13
Mothers' education			
Did not go to school	0.08	0.27	25.96
Less than lower secondary	0.11	0.32	25.96
Lower secondary	0.13	0.33	25.96
Upper secondary education	0.24	0.42	25.96
Post-secondary, non-tertiary	0.09	0.29	25.96
Short-cycle tertiary	0.10	0.29	25.96
Bachelor's or equivalent	0.14	0.34	25.96
Postgraduate degree: Master's or Doctor	0.10	0.30	25.96
Student born in country?			
Yes	0.94	0.23	2.64

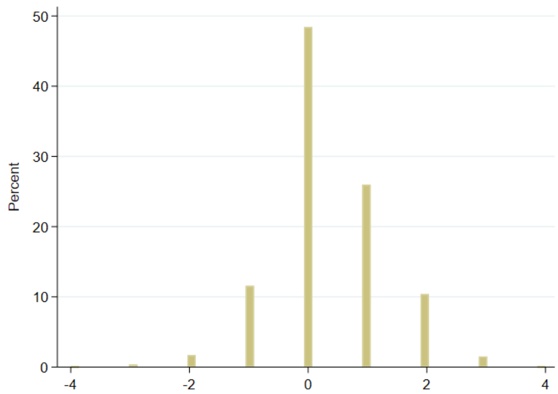
No	0.06	0.23	2.64
Father born in country?			
Yes	0.80	0.40	3.20
No	0.13	0.33	3.20
I don't know	0.06	0.24	3.20
Mother born in country?			
Yes	0.78	0.42	4.08
No	0.14	0.35	4.08
I don't know	0.07	0.26	4.08
Extra lessons in mathematics?			
No	0.72	0.45	23.67
Yes, to excel	0.13	0.34	23.67
Yes, to keep up	0.15	0.36	23.67
Extra lessons in science?			
No	0.81	0.39	24.75
Yes, to excel	0.09	0.28	24.75
Yes, to keep up	0.10	0.30	24.75
Teacher level controls			
Teacher sex			
Male	0.22	0.41	0.36
Female	0.78	0.41	0.36
Teacher age			
Under 25	0.07	0.25	0.35
25–29	0.13	0.34	0.35
30–39	0.25	0.43	0.35
40–49	0.29	0.45	0.35
50–59	0.20	0.40	0.35
60 or more	0.05	0.22	0.35
Teacher certificate			
No	0.31	0.46	10.53
Yes	0.69	0.46	10.53
Education level			
Less than Bachelor	0.18	0.39	2.93
Bachelor	0.49	0.50	2.93
Master or doctor	0.33	0.47	2.93
Major match with subject taught?			
No	0.76	0.43	2.35
Yes	0.24	0.43	2.35
Teaching experience			
1 year or less	0.04	0.20	2.30
2 years	0.05	0.22	2.30
3-5 years	0.14	0.35	2.30
>5 years	0.76	0.42	2.30
How often do you feel the following way about being a teacher?			
– Content			
Very often	0.58	0.49	0.36

Often	0.33	0.47	0.36
Sometimes	0.08	0.27	0.36
Never or almost never	0.01	0.09	0.36
– Meaning and purpose			
Very often	0.67	0.47	0.35
Often	0.27	0.45	0.35
Sometimes	0.05	0.22	0.35
Never or almost never	0.00	0.06	0.35
– Enthusiastic			
Very often	0.62	0.49	0.39
Often	0.31	0.46	0.39
Sometimes	0.07	0.25	0.39
Never or almost never	0.01	0.08	0.39
– Work inspires me			
Very often	0.60	0.49	0.48
Often	0.30	0.46	0.48
Sometimes	0.09	0.28	0.48
Never or almost never	0.01	0.08	0.48
– Proud			
Very often	0.67	0.47	0.36
Often	0.26	0.44	0.36
Sometimes	0.06	0.24	0.36
Never or almost never	0.01	0.07	0.36
Too many students in the classes?			
Agree a lot	0.34	0.47	2.78
Agree a little	0.35	0.48	2.78
Disagree a little	0.19	0.39	2.78
Disagree a lot	0.12	0.33	2.78
Hours in professional development past two years			
None	0.69	0.46	0.55
Less than 6 hours	0.08	0.27	0.55
6–15 hours	0.08	0.27	0.55
16–35 hours	0.06	0.23	0.55
More than 35 hours	0.09	0.29	0.55
Total instruction time (min/week)	309.70	69.82	6.46
School level controls			
Percentage of students from economically disadvantaged homes			
0 to 10%	0.26	0.44	6.24
11 to 25%	0.22	0.41	6.24
26 to 50%	0.17	0.38	6.24
More than 50%	0.35	0.48	6.24
Percentage of students with test language as native language			
More than 90%	0.49	0.50	4.21
76 to 90%	0.11	0.31	4.21
51 to 75%	0.07	0.25	4.21
26 to 50%	0.04	0.21	4.21

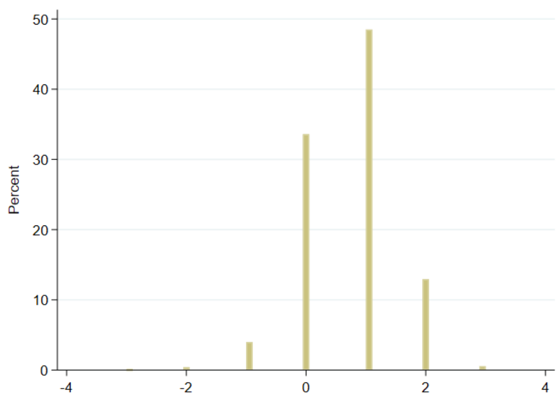
25% or less	0.29	0.45	4.21
Population of area that school is located			
More than 500,000 people	0.20	0.40	5.28
100,001 to 500,000 people	0.16	0.37	5.28
50,001 to 100,000 people	0.11	0.31	5.28
30,001 to 50,000 people	0.07	0.26	5.28
15,001 to 30,000 people	0.10	0.30	5.28
3,001 to 15,000 people	0.20	0.40	5.28
3,000 people or fewer	0.16	0.36	5.28
Area type that school is located in			
Urban	0.25	0.43	3.31
Suburban	0.17	0.38	3.31
Medium size city or large town	0.22	0.41	3.31
Small town or village	0.25	0.43	3.31
Remote rural	0.11	0.32	3.31
Class level controls			
Number of students in the class	28.04	11.81	5.77



(a) Homework frequency



(b) Homework time



(c) Homework amount

Figure A1. Distribution of the within-student difference in the homework assignment variables in grade 4.

IV.2 8th grade

Back to Data. Back to Empirical strategy.

Table A11. Descriptive statistics for the countries in the final grade 8 sample.

Country	Schools	Teachers	Students	Mathematics test score	Science test score
Australia	251	856	6,837	517	528
Bahrain	100	375	4,873	481	486
Chile	119	233	2,904	441	462
Chinese Taipei	188	478	4,427	612	574
Egypt	149	298	6,275	413	389
England	59	114	1,600	515	517
Hong Kong, SAR	117	263	2,737	578	504
Iran, Islamic Republic of	214	426	5,827	446	449
Ireland	145	800	3,410	524	523
Israel	120	511	2,795	519	513
Italy	156	412	3,571	497	500
Japan	88	189	2,702	594	570
Jordan	222	444	6,756	420	452
Korea, Republic of	98	217	2,239	607	561
Kuwait	154	310	4,097	403	444
Malaysia	176	511	6,718	461	460
New Zealand	124	504	4,693	482	499
Norway	89	153	2,320	503	495
Oman	214	450	6,320	411	457
Qatar	125	300	3,128	443	475
Saudi Arabia	159	331	4,323	394	431
Singapore	153	578	4,789	616	608
South Africa	437	874	17,729	389	370
Turkey	177	354	4,004	496	515
United Arab Emirates	397	1081	12,592	473	473
United States	220	720	6,717	515	522
Total	4,451	11,782	134,383		

Table A12. Means and standard deviations in full and final sample for the grade 8 students.

	Full sample		Final sample	
	Mean	SD	Mean	SD
Student characteristics				
Student sex				
Male	0.50	0.50	0.49	0.50
Female	0.50	0.50	0.51	0.50
Student Age	14.34	0.68	14.21	0.67

Books at home				
0—10 books	0.23	0.42	0.26	0.44
11—25 books	0.29	0.45	0.28	0.45
26—100 books	0.25	0.43	0.24	0.42
101—200 books	0.12	0.32	0.11	0.32
more than 200	0.11	0.31	0.11	0.32
Fathers' education				
Less than lower secondary	0.07	0.26	0.10	0.30
Lower secondary	0.10	0.30	0.12	0.32
Upper secondary education	0.18	0.38	0.19	0.39
Post-secondary, non-tertiary	0.06	0.25	0.05	0.21
Short-cycle tertiary	0.08	0.27	0.08	0.27
Bachelor's or equivalent	0.16	0.37	0.14	0.35
Postgraduate degree: Master's or Doctor	0.09	0.29	0.09	0.29
I don't know	0.24	0.42	0.21	0.41
Mothers' education				
Less than lower secondary	0.06	0.23	0.07	0.26
Lower secondary	0.11	0.31	0.12	0.32
Upper secondary education	0.17	0.38	0.19	0.39
Post-secondary, non-tertiary	0.07	0.25	0.05	0.22
Short-cycle tertiary	0.07	0.25	0.07	0.26
Bachelor's or equivalent	0.15	0.36	0.15	0.35
Postgraduate degree: Master's or Doctor	0.07	0.26	0.07	0.26
I don't know	0.28	0.45	0.25	0.43
Student born in country?				
Yes	0.95	0.21	0.95	0.21
No	0.05	0.21	0.05	0.21
Teacher characteristics				
Teacher sex				
Male	0.37	0.48	0.43	0.49
Female	0.63	0.48	0.57	0.49
Teacher age				
Under 25	0.03	0.17	0.03	0.16
25–29	0.12	0.32	0.13	0.33
30–39	0.28	0.45	0.34	0.47
40–49	0.29	0.45	0.29	0.45
50–59	0.22	0.41	0.18	0.38
60 or more	0.07	0.25	0.04	0.20
Teaching experience				
1 year or less	0.04	0.19	0.04	0.18
2 years	0.04	0.19	0.04	0.19
3-5 years	0.11	0.31	0.12	0.32
>5 years	0.81	0.39	0.81	0.40
Number of students	227,345		134,383	

Table A13. Descriptive statistics for the teacher controls in the final estimation sample for the grade 8 students, by homework amount.

	No HW		Low amount		Medium amount		High amount		%
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	missing
Teacher sex									
Male	0.47	0.50	0.44	0.50	0.41	0.49	0.42	0.49	0.61
Female	0.53	0.50	0.56	0.50	0.59	0.49	0.58	0.49	0.61
Teacher age									
Under 25	0.04	0.19	0.02	0.14	0.03	0.17	0.01	0.12	0.43
25–29	0.17	0.38	0.12	0.33	0.12	0.33	0.09	0.29	0.43
30–39	0.35	0.48	0.35	0.48	0.34	0.47	0.26	0.44	0.43
40–49	0.24	0.43	0.30	0.46	0.28	0.45	0.33	0.47	0.43
50–59	0.16	0.37	0.17	0.37	0.18	0.38	0.24	0.43	0.43
60 or more	0.03	0.18	0.03	0.18	0.04	0.21	0.07	0.25	0.43
Teacher certificate									
No	0.29	0.45	0.27	0.44	0.30	0.46	0.31	0.46	14.49
Yes	0.71	0.45	0.73	0.44	0.70	0.46	0.69	0.46	14.49
Education level									
Less than Bachelor	0.02	0.14	0.03	0.18	0.05	0.21	0.06	0.24	3.34
Bachelor	0.67	0.47	0.67	0.47	0.62	0.49	0.53	0.50	3.34
Master or doctor	0.31	0.46	0.29	0.46	0.33	0.47	0.41	0.49	3.34
Major match with subject taught?									
No	0.42	0.49	0.32	0.47	0.30	0.46	0.24	0.43	1.05
Yes	0.58	0.49	0.68	0.47	0.70	0.46	0.76	0.43	1.05
Teaching experience									
1 year or less	0.06	0.24	0.03	0.18	0.03	0.18	0.01	0.11	0.85
2 years	0.06	0.24	0.04	0.19	0.04	0.20	0.02	0.13	0.85
3-5 years	0.16	0.36	0.12	0.32	0.12	0.32	0.09	0.29	0.85
>5 years	0.72	0.45	0.81	0.39	0.81	0.39	0.88	0.33	0.85
How often do you feel the following way about being a teacher?									
– Content									
Very often	0.48	0.50	0.52	0.50	0.55	0.50	0.51	0.50	0.68
Often	0.36	0.48	0.34	0.47	0.34	0.47	0.40	0.49	0.68
Sometimes	0.14	0.35	0.13	0.34	0.11	0.31	0.07	0.26	0.68
Never or almost never	0.02	0.12	0.01	0.09	0.01	0.09	0.02	0.13	0.68
– Meaning and purpose									
Very often	0.52	0.50	0.58	0.49	0.63	0.48	0.66	0.47	0.73
Often	0.35	0.48	0.32	0.47	0.29	0.45	0.27	0.44	0.73
Sometimes	0.11	0.32	0.09	0.28	0.08	0.26	0.07	0.26	0.73
Never or almost never	0.01	0.10	0.00	0.06	0.00	0.06	0.01	0.07	0.73
– Enthusiastic									
Very often	0.49	0.50	0.54	0.50	0.59	0.49	0.55	0.50	0.9
Often	0.35	0.48	0.36	0.48	0.31	0.46	0.37	0.48	0.9
Sometimes	0.14	0.34	0.10	0.30	0.10	0.30	0.07	0.25	0.9
Never or almost never	0.02	0.14	0.00	0.07	0.01	0.08	0.01	0.07	0.9

– Work inspires me									
Very often	0.47	0.50	0.47	0.50	0.54	0.50	0.53	0.50	0.81
Often	0.35	0.48	0.39	0.49	0.33	0.47	0.37	0.48	0.81
Sometimes	0.16	0.36	0.13	0.34	0.13	0.33	0.10	0.29	0.81
Never or almost never	0.02	0.13	0.00	0.07	0.01	0.09	0.01	0.08	0.81
– Proud									
Very often	0.57	0.50	0.62	0.49	0.65	0.48	0.63	0.48	0.68
Often	0.31	0.46	0.31	0.46	0.28	0.45	0.33	0.47	0.68
Sometimes	0.11	0.31	0.06	0.24	0.07	0.25	0.04	0.19	0.68
Never or almost never	0.01	0.09	0.01	0.09	0.01	0.09	0.00	0.04	0.68
Too many students in the classes?									
Agree a lot	0.34	0.47	0.35	0.48	0.35	0.48	0.35	0.48	4.1
Agree a little	0.34	0.48	0.35	0.48	0.36	0.48	0.33	0.47	4.1
Disagree a little	0.19	0.39	0.20	0.40	0.18	0.38	0.19	0.39	4.1
Disagree a lot	0.13	0.34	0.11	0.31	0.11	0.31	0.13	0.34	4.1
Hours in professional development past two years									
None	0.15	0.36	0.13	0.34	0.13	0.33	0.10	0.30	0.77
Less than 6 hours	0.19	0.39	0.16	0.37	0.16	0.37	0.14	0.35	0.77
6–15 hours	0.19	0.40	0.30	0.46	0.28	0.45	0.27	0.45	0.77
16–35 hours	0.20	0.40	0.21	0.40	0.23	0.42	0.24	0.43	0.77
More than 35 hours	0.26	0.44	0.20	0.40	0.19	0.40	0.25	0.43	0.77
Total instruction time (min/week)	339	70	340	61	335	59	337	54	8.4

Table A14. Means and standard deviations for the zero-difference in homework amount students vs the non zero-difference students in the final sample for grade 8.

	No zero-difference		Zero-difference	
	Mean	SD	Mean	SD
Student characteristics				
Student sex				
Male	0.50	0.50	0.48	0.50
Female	0.50	0.50	0.52	0.50
Student Age	14.21	0.67	14.21	0.68
Books at home				
0—10 books	0.24	0.43	0.29	0.45
11—25 books	0.28	0.45	0.28	0.45
26—100 books	0.24	0.43	0.22	0.41
101—200 books	0.12	0.32	0.11	0.31
more than 200	0.12	0.32	0.10	0.30
Fathers' education				
Less than lower secondary	0.10	0.29	0.10	0.30
Lower secondary	0.11	0.32	0.12	0.32
Upper secondary education	0.19	0.40	0.19	0.39
Post-secondary, non-tertiary	0.04	0.21	0.05	0.22

Short-cycle tertiary	0.08	0.27	0.08	0.27
Bachelor's or equivalent	0.15	0.36	0.14	0.34
Postgraduate degree: Master's or Doctor	0.10	0.30	0.09	0.28
I don't know	0.21	0.41	0.22	0.41
Mothers' education				
Less than lower secondary	0.07	0.25	0.07	0.26
Lower secondary	0.12	0.32	0.12	0.33
Upper secondary education	0.19	0.40	0.19	0.39
Post-secondary, non-tertiary	0.05	0.22	0.05	0.22
Short-cycle tertiary	0.08	0.26	0.07	0.26
Bachelor's or equivalent	0.15	0.36	0.14	0.34
Postgraduate degree: Master's or Doctor	0.08	0.26	0.07	0.26
I don't know	0.25	0.43	0.26	0.44
Student born in country?				
Yes	0.95	0.21	0.95	0.21
No	0.05	0.21	0.05	0.21
Teacher characteristics				
Teacher sex				
Male	0.42	0.49	0.44	0.50
Female	0.58	0.49	0.56	0.50
Teacher age				
Under 25	0.03	0.16	0.03	0.16
25–29	0.12	0.33	0.13	0.34
30–39	0.33	0.47	0.36	0.48
40–49	0.29	0.45	0.28	0.45
50–59	0.18	0.39	0.17	0.37
60 or more	0.04	0.20	0.04	0.19
Teaching experience				
1 year or less	0.03	0.18	0.04	0.19
2 years	0.04	0.19	0.04	0.20
3-5 years	0.12	0.33	0.12	0.32
>5 years	0.81	0.39	0.80	0.40
Percent of estimation sample	59.18 %		40.82 %	

Table A15. Descriptive statistics for the OLS controls in the final estimation sample for the grade 8 students.

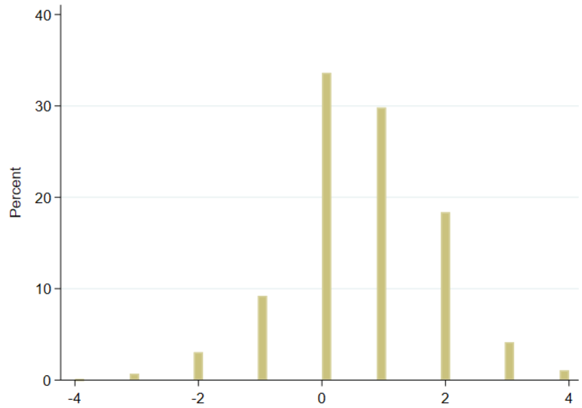
	Mean	SD	% missing
Student level controls			
Student sex			
Male	0.49	0.50	0.6
Female	0.51	0.50	0.6
Student Age	14.21	0.67	0.21
Student language at home same as test language?			

Always	0.68	0.47	1.2
Almost always	0.14	0.35	1.2
Sometimes	0.15	0.36	1.2
Never	0.04	0.19	1.2
Books at home			
0—10 books	0.26	0.44	1.55
11—25 books	0.28	0.45	1.55
26—100 books	0.24	0.42	1.55
101—200 books	0.11	0.32	1.55
more than 200	0.11	0.32	1.55
Computer at home?			
No	0.17	0.37	1.33
Yes	0.83	0.37	1.33
Own room?			
No	0.27	0.44	3.47
Yes	0.73	0.44	3.47
Fathers' education			
Less than lower secondary	0.10	0.30	3.95
Lower secondary	0.12	0.32	3.95
Upper secondary education	0.19	0.39	3.95
Post-secondary, non-tertiary	0.05	0.21	3.95
Short-cycle tertiary	0.08	0.27	3.95
Bachelor's or equivalent	0.14	0.35	3.95
Postgraduate degree: Master's or Doctor	0.09	0.29	3.95
I don't know	0.21	0.41	3.95
Mothers' education			
Less than lower secondary	0.07	0.26	4.28
Lower secondary	0.12	0.32	4.28
Upper secondary education	0.19	0.39	4.28
Post-secondary, non-tertiary	0.05	0.22	4.28
Short-cycle tertiary	0.07	0.26	4.28
Bachelor's or equivalent	0.15	0.35	4.28
Postgraduate degree: Master's or Doctor	0.07	0.26	4.28
I don't know	0.25	0.43	4.28
Student born in country?			
Yes	0.95	0.21	1.18
No	0.05	0.21	1.18
Father born in country?			
Yes	0.86	0.34	2.11
No	0.11	0.31	2.11
I don't know	0.02	0.14	2.11
Mother born in country?			
Yes	0.84	0.37	2.83
No	0.12	0.33	2.83
I don't know	0.03	0.17	2.83
Extra lessons in mathematics?			
No	0.55	0.50	22.04

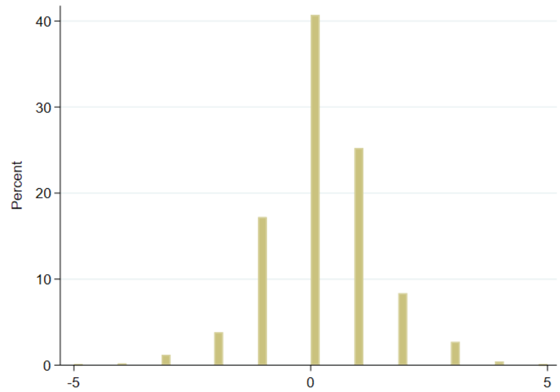
Yes, to excel	0.29	0.45	22.04
Yes, to keep up	0.17	0.37	22.04
Extra lessons in science?			
No	0.69	0.46	21.93
Yes, to excel	0.19	0.39	21.93
Yes, to keep up	0.12	0.32	21.93
Teacher level controls			
Teacher sex			
Male	0.43	0.49	0.61
Female	0.57	0.49	0.61
Teacher age			
Under 25	0.03	0.16	0.43
25–29	0.13	0.33	0.43
30–39	0.34	0.47	0.43
40–49	0.29	0.45	0.43
50–59	0.18	0.38	0.43
60 or more	0.04	0.20	0.43
Teacher certificate			
No	0.29	0.45	14.49
Yes	0.71	0.45	14.49
Education level			
Less than Bachelor	0.04	0.19	3.34
Bachelor	0.64	0.48	3.34
Master or doctor	0.32	0.47	3.34
Major match with subject taught?			
No	0.32	0.46	1.05
Yes	0.68	0.46	1.05
Teaching experience			
1 year or less	0.04	0.18	0.85
2 years	0.04	0.19	0.85
3-5 years	0.12	0.32	0.85
>5 years	0.81	0.40	0.85
How often do you feel the following way about being a teacher?			
– Content			
Very often	0.53	0.50	0.68
Often	0.35	0.48	0.68
Sometimes	0.12	0.32	0.68
Never or almost never	0.01	0.09	0.68
– Meaning and purpose			
Very often	0.60	0.49	0.73
Often	0.31	0.46	0.73
Sometimes	0.09	0.28	0.73
Never or almost never	0.00	0.07	0.73
– Enthusiastic			
Very often	0.55	0.50	0.9
Often	0.34	0.47	0.9

Sometimes	0.10	0.30	0.9
Never or almost never	0.01	0.08	0.9
– Work inspires me			
Very often	0.50	0.50	0.81
Often	0.36	0.48	0.81
Sometimes	0.13	0.34	0.81
Never or almost never	0.01	0.09	0.81
– Proud			
Very often	0.63	0.48	0.68
Often	0.30	0.46	0.68
Sometimes	0.07	0.25	0.68
Never or almost never	0.01	0.09	0.68
Too many students in the classes?			
Agree a lot	0.35	0.48	4.1
Agree a little	0.35	0.48	4.1
Disagree a little	0.19	0.39	4.1
Disagree a lot	0.11	0.32	4.1
Hours in professional development past two years			
None	0.13	0.34	0.77
Less than 6 hours	0.16	0.37	0.77
6–15 hours	0.28	0.45	0.77
16–35 hours	0.22	0.41	0.77
More than 35 hours	0.21	0.41	0.77
Total instruction time (min/week)	337.73	60.94	8.4
School level controls			
Percentage of students from economically disadvantaged homes			
0 to 10%	0.20	0.40	6.01
11 to 25%	0.20	0.40	6.01
26 to 50%	0.25	0.43	6.01
More than 50%	0.36	0.48	6.01
Percentage of students with test language as native language			
More than 90%	0.65	0.48	3.89
76 to 90%	0.13	0.33	3.89
51 to 75%	0.06	0.24	3.89
26 to 50%	0.03	0.16	3.89
25% or less	0.13	0.34	3.89
Population of area that school is located			
More than 500,000 people	0.23	0.42	4.37
100,001 to 500,000 people	0.20	0.40	4.37
50,001 to 100,000 people	0.13	0.33	4.37
30,001 to 50,000 people	0.09	0.29	4.37
15,001 to 30,000 people	0.11	0.31	4.37
3,001 to 15,000 people	0.17	0.37	4.37
3,000 people or fewer	0.08	0.28	4.37
Area type that school is located in			
Urban	0.27	0.44	3.57

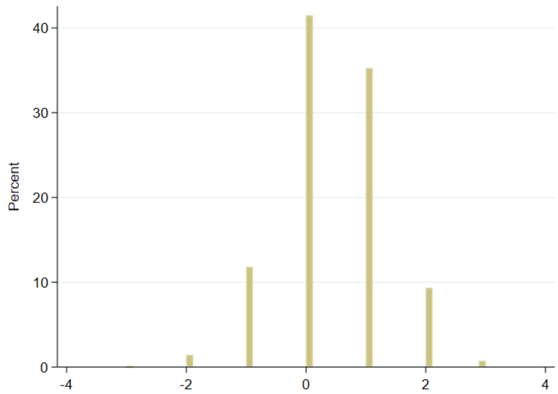
Suburban	0.20	0.40	3.57
Medium size city or large town	0.23	0.42	3.57
Small town or village	0.22	0.41	3.57
Remote rural	0.08	0.27	3.57
Class level controls			
Number of students in the class	31.82	13.48	2.84



(a) Homework frequency



(b) Homework time



(c) Homework amount

Figure A2. Distribution of the within-student difference in the homework assignment variables in grade 8.