Working Paper 2020:18

Department of Economics
School of Economics and Management

New Evidence on the Importance of Instruction Time for Student Achievement on International Assessments

Jan Bietenbeck
Matthew Collins

September 2020

Lund University

New Evidence on the Importance of Instruction Time for Student Achievement on International Assessments*

Jan Bietenbeck[†], Matthew Collins[‡]

September 3, 2020

Abstract

We revisit and substantially extend the evidence on the importance of instruction time for student achievement on international assessments. We first successfully replicate the estimate of a positive effect of weekly instruction time in the seminal paper by Lavy (*Economic Journal*, 125, F397-F424) in a narrow sense. We then extend the analysis to data from other international student assessments and find effects that are consistently smaller in magnitude. We provide suggestive evidence that this divergence is partly due to different measurement of instruction time in the data used in the original paper. Our results suggest that differences in instruction time play a less important role than previously thought for explaining international gaps in student achievement.

Keywords: instruction time; student achievement; PISA; TIMSS
JEL code: I21

*We thank Kaveh Majlesi, Derek Neal, Therese Nilsson, Luca Repetto, and audiences at Lund University, the autumn 2019 Copenhagen Education Network Workshop, the 2019 PhD Workshop in Education Economics at the University of Stavanger, and the 2020 RGS Doctoral Conference in Economics for helpful comments.

[†]Lund University and CESifo, DIW Berlin, IZA.

[‡]Lund University.

1 Introduction

Student achievement on international assessments differs widely across countries, and research shows that these achievement gaps are important drivers of cross-country differences in economic growth (Hanushek and Woessmann, 2012). This has spurred interest in the question of what explains international variation in student achievement, with one line of research focusing on the importance of instruction time. In the seminal study in this literature, Lavy (2015) uses student-level data from the Programme for International Student Achievement (PISA) to show that weekly instruction time positively affects achievement. Given large variation in weekly instruction time across countries, this suggests that international achievement gaps are partly due to differences in the amount of hours students spend learning in the classroom.

In this paper, we revisit and substantially extend the international evidence on the importance of instruction time. We first successfully replicate the results of Lavy (2015) in a narrow sense, using the same student fixed effects specification and data from the 2006 wave of PISA for a sample of 22 OECD developed countries. We then document that the effect of instruction time is qualitatively similar but smaller in data from five further waves of PISA conducted between 2000 and 2018: whereas Lavy (2015) estimates that a one-hour increase in weekly instruction time raises achievement by 0.058 standard deviations (SD), the average estimate across the other waves is only 0.022 SD. We provide suggestive evidence that this divergence is partly due to different measurement of instruction time in PISA 2006. In additional analyses, we confirm our smaller estimate using data from another international student assessment, the Trends in International Mathematics and Science Study (TIMSS), and we show that the effect is of similar magnitude in a wider sample of high-income countries, but even smaller in a sample of non-high-income countries.

Our paper adds to the existing literature on the effect of instruction time on student achievement on international assessments. Besides the seminal study by Lavy (2015), this research includes work by Rivkin and Schiman (2015), who use data from PISA 2009 and find an effect of between 0.02 SD and 0.03 SD per weekly hour. Importantly, because the authors use a different methodology and a sample that includes both OECD and non-OECD countries, those estimates are not directly comparable to those in Lavy (2015).[1] Cattaneo, Oggenfuss, and Wolter (2017) also estimate the

[1] As we describe in Section 2 below, Lavy (2015) identifies the effect of instruction time from differences across

impact of instruction time using data from PISA 2009 but focus only on Switzerland. Further studies such as Bingley et al. (2018) and Lavy (2020) similarly estimate effects on student achievement on national tests from individual countries. Our contribution to this literature lies in providing comparable evidence on the impact of instruction time on achievement on international student assessments using many different datasets and considering various groups of countries.

2 Empirical strategy

PISA is an international repeated cross-sectional study that assesses the competencies of 15-year-old students in math, reading, and science. To estimate the causal effect of instruction time in the resulting individual-level data, Lavy (2015) exploits the fact that each student is observed in three subjects in the following student fixed effects specification:

$$A_{iks} = \beta WeeklyHours_{ks} + \mu_i + \eta_k + \varepsilon_{iks} \quad (1)$$

Here, i denotes students, k denotes subjects (math, reading, science), and s denotes schools. A_{iks} is the achievement of student i in subject k. $WeeklyHours_{ks}$ are the weekly hours of instruction received in subject k, measured at the school level. μ_i is a student fixed effect, which controls for all student-level determinants of achievement that do not vary across subjects. η_k is a subject fixed effect, which controls for any level differences in achievement across subjects. ε_{iks} is the error term. Lavy (2015) estimates this specification by ordinary least squares and computes standard errors that are robust to clustering at the school level.

The regression in Equation 1 identifies the effect of instruction time from differences between subjects. A causal interpretation of the coefficient of interest β relies on the key assumption that students do not sort into schools based on subject-specific determinants of achievement, which are not captured by the student fixed effects. A further assumption in Equation 1 is that the effect of instruction time is the same for all subjects. In the original paper, Lavy (2015) provides evidence in favor of both of these assumptions, thus supporting a causal interpretation of the results.

subjects at the *school* level. Rivkin and Schiman (2015) use two different identification strategies, the most similar of which exploits variation across subjects at the *school-by-grade* level. A further difference in that paper is that the authors focus on achievement in math and reading only, whereas Lavy (2015) additionally considers science.

3 Narrow replication and extension using PISA data

3.1 Background on PISA

PISA was first conducted by the OECD in 2000 and has since been repeated every three years. The number of countries participating in the study differs somewhat between waves, but it usually includes more than 50 developed and developing countries. In each wave, PISA draws nationally representative samples of 15-year-old students and assesses them on their math, reading, and science skills using standardized tests. It also asks students to complete a questionnaire which, among other things, collects information about the weekly amount of instruction time received in each subject.

The tests in math, reading, and science measure students' ability to use their knowledge of the subject to solve real-life problems. In practice, PISA administers several different versions of the tests, which are referred to as booklets and which are randomly assigned to students. While some booklets include questions on all three subjects, others only contain questions on one or two of the subjects. Using Item Response Theory, PISA converts the answers from all booklets to subject-specific scores on a common scale. As part of this conversion, scores are imputed for students whose booklets did not contain any questions on a particular subject. The outcome of this process is a set of five (ten since PISA 2015) multiple imputations of achievement scores, called plausible values, for each student in each subject. These scores are scaled to have mean 500 and SD 100 across OECD countries in PISA 2000. Scores from other countries and from later waves are then put onto the same scale, such that student achievement is comparable across countries and over time.

The PISA student questionnaire collects information on the amount of school-based instruction time received in each subject. In the 2006 wave of the study used by Lavy (2015), this information was gathered by asking students how much time they typically spend per week attending school lessons in each subject, with possible answers being "no time," "less than 2 hours," "2 or more but less than 4 hours," "4 or more but less than 6 hours," and "6 or more hours." In the other waves, students are instead asked open-ended questions on the number of class periods per week they have in a given subject and how long a typical class period lasts. Appendix Table 1 gives an overview of the exact questions used to measure instruction time in the different PISA waves.

3.2 Sample selection and construction of variables

Our analysis draws on publicly available data from the OECD. For the narrow replication, we use data from PISA 2006 as in Lavy (2015). For our extension, we use data from five further waves of PISA conducted between 2000 and 2018. We do not analyze data from PISA 2003 because instruction time was measured only in math in that wave, which means that we cannot identify its impact using the student fixed effects model in Equation 1. Following the original paper, we restrict our samples to students who are observed with achievement scores and who answered the questions on instruction time in all subjects, resulting in a balanced panel with three observations per student. The main analysis further restricts attention to a group of 22 OECD developed countries.[2]

The key independent variable measures the weekly hours of school-based instruction received in a given subject. In the PISA 2006 data used by Lavy (2015), we follow the author and transform the categorical answers into continuous hours by recoding "no time" to missing, "less than 2 hours a week" to 1 hour, "2 or more but less than 4 hours" to 2.5 hours, "4 or more but less than 6 hours" to 4.5 hours, and "6 or more hours" to 6 hours.[3] In the data from the other PISA waves used in our extension, we multiply the number of class periods by the number of minutes per class period and divide by 60. For all waves, we then average instruction time at the school-by-subject level.

Following Lavy (2015), the outcome variable in our regressions is student achievement as measured by the first plausible value for each student in each subject in the data. As in the original paper, we transform these values into z-scores by subtracting 500 and dividing by 100. In this way, we can interpret the estimated effects in terms of standard deviations of the test score distribution among OECD countries participating in the first PISA assessment in 2000.

[2] The 22 OECD developed countries included in the main sample in Lavy (2015) and in our replication are: Australia, Austria, Belgium, Canada, Germany, Denmark, Spain, Finland, France, Greece, Ireland, Iceland, Italy, Japan, Luxembourg, Netherlands, Norway, New Zealand, Portugal, Sweden, Switzerland, United Kingdom. No data on instruction time are available for Norway in PISA 2000 and no data on reading achievement are available for Spain in PISA 2018, such that our samples for these two waves include only 21 countries.

[3] In his paper, Lavy (2015) writes that he merges the "no time" and "less than 2 hours a week" categories, but the publicly available code on the journal website actually changes "no time" to missing. We choose to follow the code in order to replicate exactly the original estimates, but in practice this makes very little difference.

3.3 Main results for OECD countries

We begin our analysis by replicating the headline result by Lavy (2015) in a narrow sense. To this end, column 1 of Table 1 shows the estimated effect of instruction time on student achievement using data from PISA 2006 for the main sample of 22 OECD developed countries. The results indicate that a one-hour increase in weekly instruction time raises student achievement by 0.058 SD. This estimate is exactly identical to the one reported in the original paper and thus constitutes a successful replication in a narrow sense.

In the remaining columns of Table 1, we extend the existing evidence on the importance of instruction time for student achievement on international assessments by reporting estimates from five further waves of PISA conducted between 2000 and 2018. The impact of instruction time in these samples is also positive, but smaller in magnitude: across columns 2 to 6, the estimates range from 0.014 SD to 0.031 SD, and their average is less than half the original estimate at 0.022 SD. In what follows, we consider three potential explanations for this divergence.

First, one possible reason why the estimate is larger for PISA 2006 is that the effect of instruction time is heterogeneous, and that the composition of students included in the estimation sample differs between waves.[4] To explore this possibility, we computed means of socio-demographic characteristics of students included in the estimation sample separately by wave. The results in Appendix Table 2 show that while there are some differences between waves, students in PISA 2006 do not stand out in any particular way, suggesting that variation in sample composition is not behind the larger impact of instruction time in that wave.

Second, a related potential explanation for the discrepancy in effect sizes is that the distribution of student achievement differs across waves. For example, if the standard deviation of achievement was much larger in PISA 2006, this could account for the higher impact of instruction time. However, Appendix Table 2 shows that the means and standard deviations of student achievement are roughly constant across waves, ruling out this explanation.

Third, another possible reason is different measurement of instruction time. Recall that only in PISA 2006, instruction time was recorded using categorical answers and then transformed into

[4] In line with the idea of a heterogeneous treatment effect, Lavy (2015) shows that the impact of instruction time is larger for students with an immigrant background and for students with less educated parents.

continuous hours. Table 1 shows that the mean instruction time in that wave is lower at 3.4 hours than the means in all other waves, which range from 3.7 to 3.9 hours. To test whether this difference can account for the discrepancy in effect sizes, we artificially discretized the responses to the instruction time questions in the other waves using the PISA 2006 answer categories and then re-constructed a continuous measure of hours from this variable in the same way as in Lavy (2015).[5] According to this measure, the mean of instruction time is very similar across all waves at 3.3 to 3.4 hours, see Appendix Table 3. Moreover, when we used this measure as our independent variable, the estimated effect of instruction time in the other PISA waves became larger, with a range from 0.018 SD to 0.037 SD and an average of 0.027 SD. While not conclusive, this evidence suggests that different measurement of instruction time can account for some, though not all, of the difference in effect sizes between PISA 2006 and the other waves.

3.4 Robustness checks

In recent work, Rutkowski et al. (2010) and Jerrim et al. (2017) highlight some conceptual and econometric problems that commonly arise in secondary analyses of data from international student assessments. In brief, these papers suggest testing the robustness of regression results in three ways: (1) estimating effects on all plausible values, rather than just a single plausible value; (2) applying sampling weights; and (3) using test scores in a given subject only for students whose booklets actually included questions on that subject.[6] We implemented these suggestions in Appendix Table 4. The main takeaway from these results is that the estimate by Lavy (2015) and our estimates for the other PISA waves are broadly robust to these modifications. In particular, in most specifications the effect of a one-hour increase in instruction time hovers around 0.025 SD, with the exception of regressions based on PISA 2006, in which the impact is typically at least twice as large.

[5] For example, if a student reported to have four class periods of 45 minutes per week, we categorized the answer as "2 or more but less than 4 hours." As in the original paper, we then coded this answer as 2.5 hours when re-constructing the continuous measure.

[6] For further details on the underlying problems, we refer to the original papers. We also note that Jerrim et al. (2017) replicate part of the analysis by Lavy (2015) in a narrow sense, but focus mostly on a specification without student fixed effects. In contrast, we replicate and test the robustness of the main specification in Lavy (2015) and provide new evidence on the effect of instruction time using other datasets.

3.5 Additional results for other groups of countries

An interesting question is whether the effect of instruction time differs between developed and developing countries. In the original paper, Lavy (2015) shows that the impact is similar to the one found for OECD developed countries in a sample of 14 Eastern European countries at 0.061 SD, but that it is substantially smaller in a sample of 13 developing countries at 0.030 SD. In Appendix Table 5, we successfully replicate these results in a narrow sense. Mirroring the results for OECD developed countries in Table 1, we find that the estimated effects in samples from PISA waves other than 2006 are almost always smaller in magnitude.

To provide further insights into how the effect of instruction time differs between countries, Table 2 shows estimates separately for high-income and non-high-income economies as classified by the World Bank.[7] This more comprehensive classification, together with the different coverage across the different waves of PISA, means that our results include a total of 49 high-income and 35 non-high-income countries. Ignoring the estimates for PISA 2006, which are again much larger for both groups of countries, we find that a one-hour increase in instruction time raises student achievement by between 0.009 SD and 0.024 SD (with an average estimate of 0.017 SD) in high-income economies, a result which is similar to the one for OECD developed countries. In contrast, the impact of instruction time is smaller among non-high-income economies at -0.003 SD to 0.013 SD (with an average estimate of 0.006 SD). Our results thus support the conclusion by Lavy (2015) that instruction time is more productive in developed countries.

4 Extension using TIMSS data

4.1 Background on TIMSS

TIMSS is an international assessment of the math and science knowledge of fourth- and eighth-grade students. The study has been conducted by the International Association for the Evaluation of Educational Achievement (IEA) every four years since 1995 and usually covers more than 40 developed and developing countries. In each participating country, nationally representative samples

[7] We use the classification as of June 2020, which is available at https://datahelpdesk.worldbank.org/knowledgebase/articles/906519.

of students are assessed using standardized tests. Using a questionnaire, the study also collects information on weekly instruction time from teachers.

The standardized tests in TIMSS are designed to measure students' knowledge of the common part of the math and science curricula of participating countries. TIMSS administers different versions of these tests, but includes math and science questions in all test booklets. Similarly to PISA, students' answers from all booklets are scored on a common scale using Item Response Theory, resulting in five plausible values for each student and subject. These scores are scaled to have mean 500 and SD 100 across countries participating in TIMSS 1995, with scores from later waves put onto the same scale.

TIMSS asks all math and science teachers of students participating in the study to complete a questionnaire. Among other things, teachers are asked how many minutes per week they typically teach their subject to the sampled students' class. Appendix Table 6 gives a detailed overview of the questions measuring instruction time in the different TIMSS waves.

4.2 Sample selection and construction of variables

Our regressions use publicly available data from the six waves of TIMSS conducted between 1995 and 2015. When selecting our samples and constructing our variables, we follow the original analysis by Lavy (2015) as closely as possible. Thus, we focus on eighth-grade students, who are similar in age to students participating in PISA at about 13.5 years on average. We further restrict our data to a balanced panel of students observed with achievement scores and instruction time in both subjects. We provide estimates both for the countries included among the 22 OECD developed countries studied by Lavy (2015) and for World Bank high-income and non-high-income economies.

To measure instruction time, we first assign each student the total hours received in each subject as reported by her teachers and then compute the average at the school-by-subject level. We construct our achievement measure using the first plausible value for each student in each subject in the data. We transform these values into z-scores by subtracting 500 and dividing by 100. This implies that we can interpret the estimated effects in terms of standard deviations of the test score distribution among all countries participating in TIMSS 1995.

4.3 Results

Table 3 reports estimates of the effect of instruction time on student achievement in TIMSS. Panel A focuses on those countries among the 22 OECD developed countries studied by Lavy (2015) that also participated in TIMSS. The effect of a one-hour increase in weekly instruction time in these samples ranges from 0.015 SD to 0.037 SD, with an average of 0.025 SD. Panel B instead shows estimates for World Bank high-income economies that range from 0.007 SD to 0.033 SD (with an average estimate of 0.019 SD), and Panel C shows estimates for World Bank non-high-income economies that range from -0.031 SD to 0.021 SD (with an average estimate of 0.000 SD). The main takeaway from this analysis is thus that the estimates based on TIMSS are very similar to the corresponding estimates for PISA other than the 2006 wave.

5 Conclusion

We revisit the question of the importance of instruction time for student achievement on international assessments. We successfully replicate the estimate of a positive effect of weekly instruction time in the seminal study by Lavy (2015) in a narrow sense. However, when we extend the analysis to data from other international student assessments, we find effects that are consistently smaller in magnitude than the ones reported in the original paper. We provide suggestive evidence that this divergence is partly due to different measurement of instruction time in the PISA 2006 data used by Lavy (2015). Taken together, our results suggest that differences in instruction time play a less important role than previously thought for explaining international gaps in student achievement.

References

Bingley, P., E. Heinesen, K.F. Krassel, and N. Kristensen. 2018. "The Timing of Instruction Time: Accumulated Hours, Timing and Pupil Achievement." IZA Discussion Paper No. 11807.

Cattaneo, M.A., C. Oggenfuss, and S.C. Wolter. 2017. "The More, the Better? The Impact of Instructional Time on Student Performance." *Education Economics* 25:433–445.

Hanushek, E.A., and L. Woessmann. 2012. “Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation.” *Journal of Economic Growth* 17:267–321.

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

Lavy, V. 2015. “Do Differences in Schools’ Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries.” *The Economic Journal* 125:F397–F424.

—. 2020. “Expanding School Resources and Increasing Time on Task: Effects on Students’ Academic and Noncognitive Outcomes.” *Journal of the European Economic Association* 18:232–265.

Rivkin, S.G., and J.C. Schiman. 2015. “Instruction Time, Classroom Quality, and Academic Achievement.” *The Economic Journal* 125:F425–F448.

Rutkowski, L., E. Gonzalez, M. Joncas, and M. von Davier. 2010. “International large-scale assessment data: Issues in secondary analysis and reporting.” *Educational Researcher* 39:142–151.

Table 1: Effect of weekly hours of instruction time on student achievement in 22 OECD developed countries: replication and extension using PISA data

	Replication	Extension: data from further PISA waves				
	PISA 2006 (1)	PISA 2000 (2)	PISA 2009 (3)	PISA 2012 (4)	PISA 2015 (5)	PISA 2018 (6)
Weekly hours	0.058	0.019	0.027	0.031	0.020	0.014
	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
# of observations	460,734	65,577	493,800	327,891	420,186	342,288
# of students	153,578	21,859	164,600	109,297	140,062	114,096
# of schools	6,577	4,352	7,176	7,774	6,204	6,070
# of countries	22	21	22	22	22	21
Mean weekly hours	3.4	3.8	3.7	3.7	3.8	3.9

Notes: The table shows estimates of the effect of weekly hours of instruction time on student achievement in a sample of 22 OECD countries. Estimates in column 1 use data from PISA 2006 like in Lavy (2015). Estimates in columns 2-6 use data from five further waves of PISA as indicated in the column headers. The sample in column 2 includes only 21 countries because data on instruction time are not available for Norway in PISA 2000. The sample in column 6 includes only 21 countries because data for reading achievement are not available for Spain in PISA 2018 and in line with Lavy (2015), the sample is restricted to students observed in all three subjects. All specifications control for individual and subject fixed effects. Standard errors in parentheses are robust to clustering at the school level.

Table 2: Effect of weekly hours of instruction time on student achievement in 49 high-income and 35 non-high-income economies: extension using PISA data

	Orig. data	Data from further PISA waves				
	PISA 2006	PISA 2000	PISA 2009	PISA 2012	PISA 2015	PISA 2018
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: World Bank high-income economies						
Weekly hours	0.054	0.024	0.017	0.023	0.014	0.009
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
# of observations	682,335	86,301	819,096	508,026	769,176	723,351
# of students	227,445	28,767	273,032	169,342	256,392	241,117
# of schools	9,707	5,672	11,557	11,742	10,624	11,869
# of countries	40	31	47	43	42	46
Mean weekly hours	3.4	3.7	3.8	3.7	3.9	4.0
Panel B: World Bank non-high-income economies						
Weekly hours	0.063	–0.003	0.003	0.013	0.005	0.010
	(0.007)	(0.005)	(0.002)	(0.003)	(0.003)	(0.002)
# of observations	274,371	29,991	432,603	229,398	239,388	416,130
# of students	91,457	9,997	144,201	76,466	79,796	138,710
# of schools	4,326	2,013	6,715	5,778	3,206	6,811
# of countries	16	11	26	20	12	26
Mean weekly hours	3.2	3.3	3.7	3.5	3.6	3.8

Notes: The table shows estimates of the effect of weekly hours of instruction time on student achievement in samples of 49 high-income economies (Panel A) and 35 non-high-income economies (Panel B) as classified by the World Bank. Estimates use data from six waves of PISA as indicated in the column headers. The individual samples do not include all economies due to countries not participating in the corresponding PISA wave or due to missing data on achievement or instruction time. All specifications control for individual and subject fixed effects. Standard errors in parentheses are robust to clustering at the school level.

Table 3: Effect of weekly hours of instruction time on student achievement in different groups of countries: extension using TIMMS data

	TIMSS 1995 (1)	TIMSS 1999 (2)	TIMSS 2003 (3)	TIMSS 2007 (4)	TIMSS 2011 (5)	TIMSS 2015 (6)
Panel A: OECD developed countries as in Lavy (2015)						
Weekly hours	0.037	0.037	0.024	0.015	0.018	0.017
	(0.006)	(0.009)	(0.007)	(0.004)	(0.007)	(0.006)
# of observations	83,200	43,036	46,840	41,134	47,060	81,092
# of students	41,600	21,518	23,420	20,567	23,530	40,546
# of schools	1,770	949	915	804	895	1,324
# of countries	16	6	6	5	6	8
Mean weekly hours	3.3	3.2	3.1	3.0	3.3	3.3
Panel B: World Bank high-income economies						
Weekly hours	0.033	0.020	0.011	0.009	0.014	0.007
	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.003)
# of observations	159,428	163,444	195,742	191,276	154,168	255,110
# of students	79,714	81,722	97,871	95,638	77,084	127,555
# of schools	3,535	3,052	3,453	3,187	2,734	4,162
# of countries	34	22	26	25	18	27
Mean weekly hours	3.3	3.5	3.6	3.5	3.7	3.8
Panel C: World Bank non-high-income economies						
Weekly hours	–0.006	0.007	–0.002	–0.031	0.012	0.021
	(0.011)	(0.004)	(0.004)	(0.003)	(0.009)	(0.005)
# of observations	20,048	137,950	162,396	202,698	99,382	161,212
# of students	10,024	68,975	81,198	101,349	49,691	80,606
# of schools	440	2,156	2,593	3,609	1,533	2,481
# of countries	4	15	21	25	9	13
Mean weekly hours	3.7	3.9	4.0	3.6	3.4	3.8

Notes: The table shows estimates of the effect of weekly hours of instruction time on student achievement based on data from six waves of TIMMS 8th grade as indicated in the column headers. Estimates in Panel A focus on countries included among the 22 OECD developed countries studied by Lavy (2015). Estimates in Panels B and C focus on high-income and non-high-income economies as classified by the World Bank, respectively. The individual samples do not include all countries or economies due to countries not participating in the corresponding TIMSS 8th grade assessment. All specifications control for individual and subject fixed effects. Standard errors in parentheses are robust to clustering at the school level.

Appendix for online publication

Appendix Table 1: Questions about instruction time in PISA

Study	Respondent	Question	Type
PISA 2000	Student	In the last full week you were in school, how many <class periods> did you spend in *subject*?	open-ended
	Principal	How many instructional minutes are there in the average single <class period>?[a]	open-ended
PISA 2006	Student	How much time do you typically spend per week studying the following subjects? Regular lessons in *subject* at my school?	categorical[b]
PISA 2009	Student	How many <class periods> per week do you typically have for the following subjects?	open-ended
	Student	How many minutes, on average, are there in a <class period> for the following subjects?	open-ended
PISA 2012	Student	How many <class periods> per week do you typically have for the following subjects?	open-ended
	Student	How many minutes, on average, are there in a <class period> for the following subjects?	open-ended
PISA 2015	Student	How many <class periods> per week are you typically required to attend for the following subjects?	open-ended
	Student	How many minutes, on average, are there in a <class period>?[a]	open-ended
PISA 2018	Student	How many <class periods> per week are you typically required to attend for the following subjects?	open-ended
	Student	How many minutes, on average, are there in a <class period>?[a]	open-ended

Notes: The table gives an overview of the questions used to measure instruction time in each wave of PISA. The term "<class period>" is translated to the locally used term in each country. The term "*subject*" is replaced by "mathematics", "reading", or "science" for the corresponding subject-specific question. [a]This question is not asked separately for each subject. [b]Answer categories for this question: no time, less than 2 hours a week, 2 or more but less than 4 hours a week, 4 or more but less than 6 hours a week, 6 or more hours a week.

Appendix Table 2: Summary statistics of students' socio-demographic characteristics and achievement in the PISA estimation samples of 22 OECD developed countries

	Orig. data	Extension: data from further PISA waves				
	PISA 2006 (1)	PISA 2000 (2)	PISA 2009 (3)	PISA 2012 (4)	PISA 2015 (5)	PISA 2018 (6)
Panel A: means of socio-demographic characteristics						
Female	0.51	0.51	0.50	0.50	0.50	0.51
First-generation immigrant	0.05	0.05	0.06	0.07	0.07	0.08
Second-generation immigrant	0.05	0.05	0.06	0.07	0.08	0.10
Father has college education	0.24	–[a]	0.26	0.28	0.33	0.36
Mother has college education	0.22	–[a]	0.25	0.28	0.34	0.39
Panel B: mean and standard deviation of achievement						
Mean	513.42	521.62	509.76	513.17	509.38	510.34
Standard deviation	93.28	96.12	92.69	90.92	92.56	93.20

Notes: The table shows means of students' socio-demographic characteristics (Panel A) and the mean and standard deviation of student achievement (Panel B) separately by PISA wave as indicated in the column headers. Statistics for each wave are computed for the students included in the estimation sample of 22 OECD countries. [a]Data on parental education in PISA 2000 are not directly comparable to data in the other waves because of a change in the PISA student questionnaire after this wave.

Appendix Table 3: Effect of weekly hours of instruction time on student achievement in 22 OECD developed countries: alternative measurement of hours using PISA data

	Orig. data	Extension: data from further PISA waves				
	PISA 2006 (1)	PISA 2000 (2)	PISA 2009 (3)	PISA 2012 (4)	PISA 2015 (5)	PISA 2018 (6)
Weekly hours	0.058	0.021	0.033	0.037	0.027	0.018
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
# of observations	460,734	65,577	493,800	327,891	420,186	342,288
# of students	153,578	21,859	164,600	109,297	140,062	114,096
# of schools	6,577	4,352	7,176	7,774	6,204	6,070
# of countries	22	21	22	22	22	21
Mean weekly hours	3.4	3.3	3.3	3.3	3.3	3.4

Notes: The table shows estimates of the effect of weekly hours of instruction time on student achievement in a sample of 22 OECD countries. For this table, weekly hours are constructed by first discretizing instruction time as in PISA 2006 and then re-constructing a continuous measure in the same way as in Lavy (2015), see text for details. The samples and specifications are otherwise identical to those in Table 1. All specifications control for individual and subject fixed effects. Standard errors in parentheses are robust to clustering at the school level.

Appendix Table 4: Effect of weekly hours of instruction time on student achievement in 22 OECD developed countries: robustness using PISA data

	Orig. data	Extension: data from further PISA waves				
	PISA 2006	PISA 2000	PISA 2009	PISA 2012	PISA 2015	PISA 2018
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: main results (for comparison)						
Weekly hours	0.058	0.019	0.027	0.031	0.020	0.014
	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
# of observations	460,734	65,577	493,800	327,891	420,186	342,288
# of students	153,578	21,859	164,600	109,297	140,062	114,096
Panel B: using the average plausible value as an outcome						
Weekly hours	0.058	0.017	0.026	0.031	0.019	0.015
	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
# of observations	460,734	65,577	493,800	327,891	420,186	342,288
# of students	153,578	21,859	164,600	109,297	140,062	114,096
Panel C: applying student weights						
Weekly hours	0.035	0.020	0.022	0.025	0.024	0.015
	(0.006)	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)
# of observations	460,734	65,577	493,800	327,891	420,186	342,288
# of students	153,578	21,859	164,600	109,297	140,062	114,096
Panel D: applying senate weights						
Weekly hours	0.069	0.022	0.048	0.032	0.022	0.017
	(0.006)	(0.008)	(0.007)	(0.003)	(0.002)	(0.002)
# of observations	460,734	65,577	493,800	327,891	420,186	342,288
# of students	153,578	21,859	164,600	109,297	140,062	114,096
Panel E: using only booklets containing questions on all subjects						
Weekly hours	0.060	0.019	0.028	0.029	0.023	0.014
	(0.004)	(0.003)	(0.002)	(0.003)	(0.004)	(0.002)
# of observations	212,253	65,433	228,171	151,389	17,916	342,288
# of students	70,751	21,811	76,057	50,463	5,972	114,096

Notes: The table shows estimates of the effect of weekly hours of instruction time on student achievement in a sample of 22 OECD countries. Panel A reproduces the results from Table 1. Each of the subsequent panels implements a robustness check suggested by Rutkowski et al. (2010) or Jerrim et al. (2017). Regressions in Panel B use the average of all available plausible values, rather than only the first plausible value used by Lavy (2015), as an outcome. Regressions in Panel C apply student sampling weights, which scale the sample up to the size of the population within each country. Regressions in Panel D apply senate weights, which sum up to the same constant value within each country (note that because our sample restrictions affect countries to different extents, this sum actually varies somewhat across countries in the final estimation sample). Panel E drops students whose assigned test booklets did not contain questions on all three subjects from the sample. All specifications control for individual and subject fixed effects. Standard errors in parentheses are robust to clustering at the school level.

Appendix Table 5: Effect of weekly hours of instruction time on student achievement in 14 Eastern European and 13 developing countries: replication and extension using PISA data

	Replication	Extension				
	PISA 2006 (1)	PISA 2000 (2)	PISA 2009 (3)	PISA 2012 (4)	PISA 2015 (5)	PISA 2018 (6)
Panel A: 14 Eastern European countries						
Weekly hours	0.061	0.023	0.004	0.018	0.005	0.009
	(0.006)	(0.007)	(0.004)	(0.004)	(0.004)	(0.003)
# of observations	177,015	19,248	183,441	117,186	188,796	194,952
# of students	59,005	6,416	61,147	39,062	62,932	64,984
# of schools	2,606	1,195	2,586	2,741	2,700	3,373
# of countries	14	7	14	14	12	14
Mean weekly hours	3.0	3.1	3.2	3.1	3.3	3.3
Panel B: 13 developing countries						
Weekly hours	0.030	0.004	0.003	–0.005	0.005	0.033
	(0.008)	(0.006)	(0.003)	(0.003)	(0.004)	(0.003)
# of observations	238,938	16,503	301,113	160,374	180,207	171,510
# of students	79,646	5,501	100,371	53,458	60,069	57,170
# of schools	3,990	1,206	4,642	4,222	2,523	3,070
# of countries	13	6	13	11	8	10
Mean weekly hours	3.2	3.6	3.9	3.9	3.9	4.1

Notes: The table shows estimates of the effect of weekly hours of instruction time on student achievement in samples of 14 Eastern European countries (Panel A) and 13 developing countries (Panel B). Estimates in column 1 use data from PISA 2006 like in Lavy (2015). Estimates in columns 2-6 use data from five further waves of PISA as indicated in the column headers. The samples in columns 2 to 6 sometimes include fewer countries than in column 1 due to countries not participating in the corresponding PISA wave or due to missing data on achievement or instruction time. All specifications control for individual and subject fixed effects. Standard errors in parentheses are robust to clustering at the school level.

Appendix Table 6: Questions about instruction time in TIMSS

Study	**Respondent**	**Question**	**Type**
TIMSS 1995	Teacher	How many minutes per week do you teach mathematics/science to your mathematics/science class?	open-ended
TIMSS 1999	Teacher	How many minutes per week do you teach mathematics/science to your mathematics/science class?	open-ended
TIMSS 2003	Teacher	How many minutes per week do you teach mathematics/science to the TIMSS class?	open-ended
TIMSS 2007	Teacher	How many minutes per week do you teach mathematics/science to the TIMSS class?	open-ended
TIMSS 2011	Teacher	In a typical week, how much time to you spend teaching mathematics/science to the students in this class?	open-ended
TIMSS 2015	Teacher	In a typical week, how much time to you spend teaching mathematics/science to the students in this class?	open-ended

Notes: The table gives an overview of the questions used to measure instruction time in each wave of TIMSS.