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

Jan Bietenbeck[†], Matthew Collins[‡]

October 27, 2021

Abstract

We re-examine the importance of instruction time for student achievement on international assessments. We successfully replicate the positive effect of weekly instruction time in the seminal paper by Lavy (*Economic Journal*, 125, F397-F424) in a narrow sense. Extending the analysis to other international assessments, we find effects that are consistently smaller in magnitude. We provide evidence that this discrepancy might be partly due to a different way of measuring instruction time in the data used in the original paper. Our results suggest that differences in instruction time are less important 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. Corresponding author. Email: jan.bietenbeck@nek.lu.se. Address: Department of Economics, Lund University, P.O. Box 7080, 220 07 Lund, Sweden.

[‡]Lund University.

Online appendix

A Additional results

Online Appendix Table A.1: Effects of imposing categorical measurement of instruction time in PISA waves other than 2006

Wave	Variable Construction	Residual Variance	Residual Variance Factor	Residual Covariance	Covariance Factor	Point Estimate	Point Estimate Factor
2000	Original	0.5165	0.7806	0.0090	0.8750	0.0174	1.1210
	Discretized	0.4032		0.0079		0.0195	
2009	Original	0.3952	0.8219	0.0104	0.9892	0.0264	1.2036
	Discretized	0.3248		0.0103		0.0318	
2012	Original	0.4179	0.7632	0.0133	0.9073	0.0318	1.1889
	Discretized	0.3189		0.0121		0.0378	
2015	Original	0.5541	0.5638	0.0111	0.7517	0.0201	1.3333
	Discretized	0.3124		0.0084		0.0268	
2018	Original	0.4345	0.5818	0.0059	0.7835	0.0135	1.3467
	Discretized	0.2528		0.0046		0.0182	

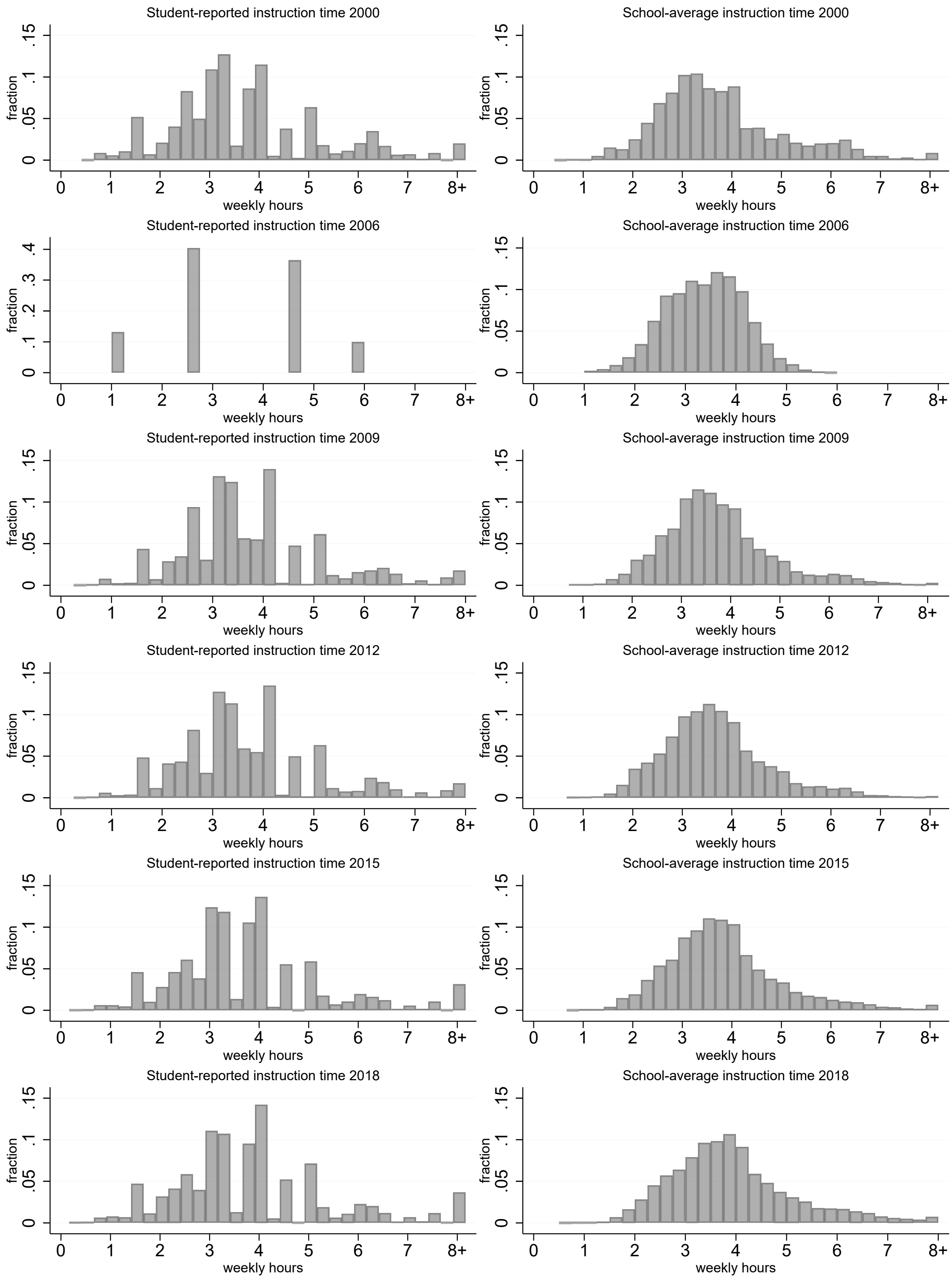
Notes: The table shows how artificially discretizing instruction time in PISA waves other than 2006 affects the variance of the explanatory variable (column 3), the covariance between the dependent and explanatory variables (column 5) and, in turn, the point estimate in our main regression (column 7). Instruction time is discretized by imposing the answer categories used in PISA 2006 and the mid-points used in Lavy (2015). Instruction time and test scores are residualized on student and subject fixed effects. Factors reflect the magnitude of the residual variance (column 4), residual covariance (column 6), and point estimate (column 8) when using discretized instruction time, relative to undiscretized instruction time. For further details, see Figure 2 in the main text.

Online Appendix Table A.2: Estimates for different groups of countries

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: PISA data						
	Orig. data: PISA 2006	PISA 2000	PISA 2009	PISA 2012	PISA 2015	PISA 2018
A.1: Lavy (2015) 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 students	59,005	6,416	61,147	39,062	62,932	64,984
# of countries	14	7	14	14	12	14
A.2: Lavy (2015) 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 students	79,646	5,501	100,371	53,458	60,069	57,170
# of countries	13	6	13	11	8	10
A.3: 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 students	227,445	28,767	273,032	169,342	256,392	241,117
# of countries	40	31	47	43	42	46
A.4: 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 students	91,457	9,997	144,201	76,466	79,796	138,710
# of countries	16	11	26	20	12	26
Panel B: TIMSS data						
	TIMSS 1995	TIMSS 1999	TIMSS 2003	TIMSS 2007	TIMSS 2011	TIMSS 2015
B.1: World Bank high-income economies						
Weekly hours	0.033	0.020	0.011	0.009	0.015	0.007
	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.003)
# of students	79,714	81,722	97,871	95,659	83,833	127,555
# of countries	34	22	26	25	18	27
B.2: World Bank non-high-income economies						
Weekly hours	–0.006	0.008	–0.001	–0.031	0.012	0.021
	(0.011)	(0.004)	(0.004)	(0.003)	(0.009)	(0.005)
# of students	10,024	69,039	81,327	101,358	51,527	80,610
# of countries	4	15	21	25	9	13

Notes: The table shows estimates of the effect of weekly hours of instruction on student achievement separately for different groups of countries. Following Lavy (2015), Panel A.1 restricts the sample to 14 Eastern European countries and Panel A.2 restricts the sample to 13 developing countries. The number of countries is lower in some columns because not all countries participated in all rounds of PISA. Panels A.3 and A.4 (based on PISA data) and Panels B.1 and B.2 (based on TIMSS data) show results for high-income and non-high-income economies as classified by the World Bank as of June 2020. The number of countries included in these regressions varies across samples because not all countries participated in all assessments. All specifications in all panels control for individual and subject fixed effects. Standard errors in parentheses are robust to clustering at the school level.

Online Appendix Figure A.1: Distribution of instruction time by PISA wave



Notes: The figure shows the distribution of student-reported instruction time (left panel) and school-average instruction time (right panel) separately for each PISA wave.

A Additional results

Online Appendix Table A.1: Effects of imposing categorical measurement of instruction time in PISA waves other than 2006

Wave	Variable Construction	Residual Variance	Residual Variance Factor	Residual Covariance	Covariance Factor	Point Estimate	Point Estimate Factor
2000	Original	0.5165	0.7806	0.0090	0.8750	0.0174	1.1210
	Discretized	0.4032		0.0079		0.0195	
2009	Original	0.3952	0.8219	0.0104	0.9892	0.0264	1.2036
	Discretized	0.3248		0.0103		0.0318	
2012	Original	0.4179	0.7632	0.0133	0.9073	0.0318	1.1889
	Discretized	0.3189		0.0121		0.0378	
2015	Original	0.5541	0.5638	0.0111	0.7517	0.0201	1.3333
	Discretized	0.3124		0.0084		0.0268	
2018	Original	0.4345	0.5818	0.0059	0.7835	0.0135	1.3467
	Discretized	0.2528		0.0046		0.0182	

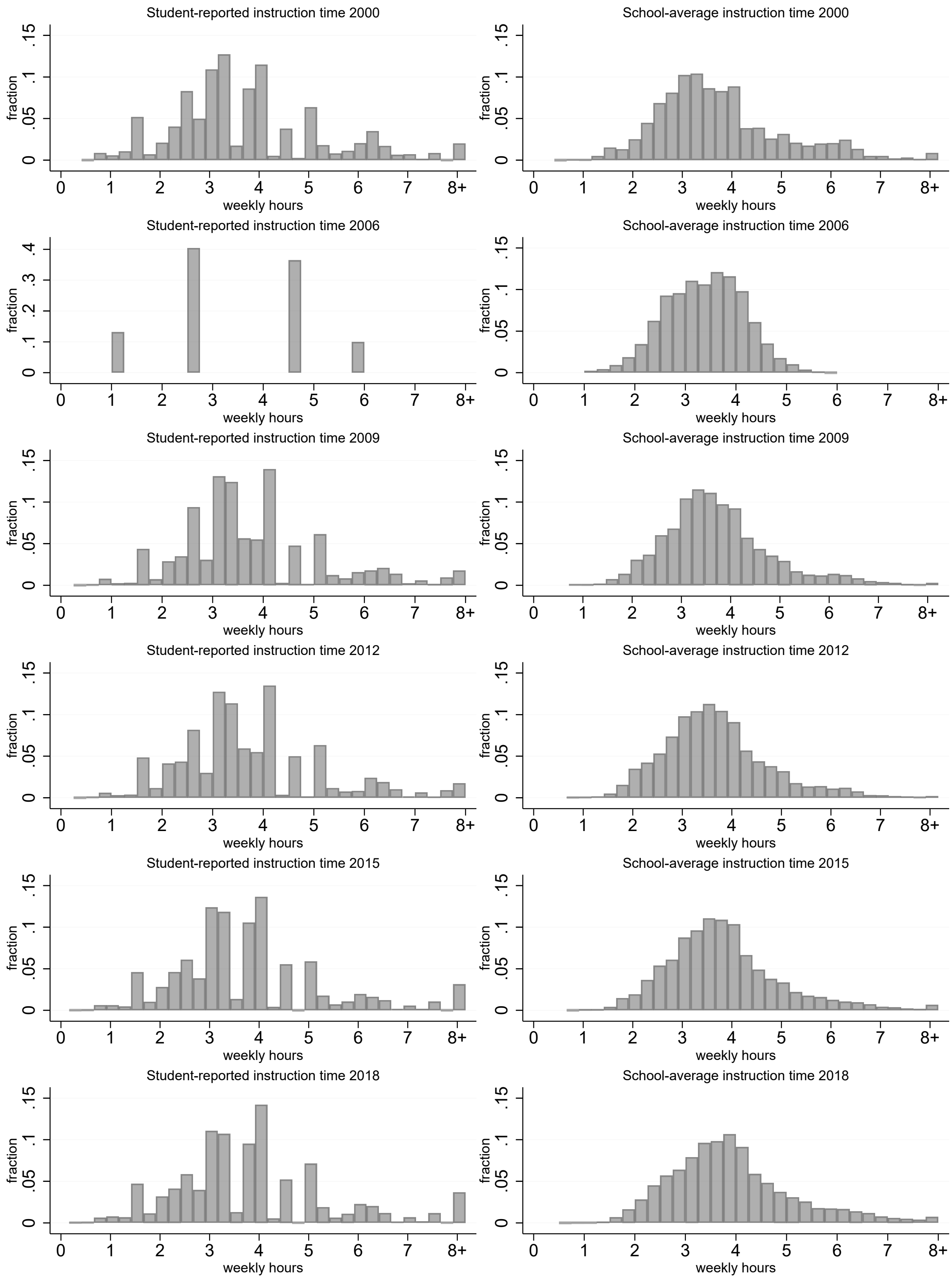
Notes: The table shows how artificially discretizing instruction time in PISA waves other than 2006 affects the variance of the explanatory variable (column 3), the covariance between the dependent and explanatory variables (column 5) and, in turn, the point estimate in our main regression (column 7). Instruction time is discretized by imposing the answer categories used in PISA 2006 and the mid-points used in Lavy (2015). Instruction time and test scores are residualized on student and subject fixed effects. Factors reflect the magnitude of the residual variance (column 4), residual covariance (column 6), and point estimate (column 8) when using discretized instruction time, relative to undiscretized instruction time. For further details, see Figure 2 in the main text.

Online Appendix Table A.2: Estimates for different groups of countries

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: PISA data						
	Orig. data: PISA 2006	PISA 2000	PISA 2009	PISA 2012	PISA 2015	PISA 2018
A.1: Lavy (2015) 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 students	59,005	6,416	61,147	39,062	62,932	64,984
# of countries	14	7	14	14	12	14
A.2: Lavy (2015) 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 students	79,646	5,501	100,371	53,458	60,069	57,170
# of countries	13	6	13	11	8	10
A.3: 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 students	227,445	28,767	273,032	169,342	256,392	241,117
# of countries	40	31	47	43	42	46
A.4: 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 students	91,457	9,997	144,201	76,466	79,796	138,710
# of countries	16	11	26	20	12	26
Panel B: TIMSS data						
	TIMSS 1995	TIMSS 1999	TIMSS 2003	TIMSS 2007	TIMSS 2011	TIMSS 2015
B.1: World Bank high-income economies						
Weekly hours	0.033	0.020	0.011	0.009	0.015	0.007
	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.003)
# of students	79,714	81,722	97,871	95,659	83,833	127,555
# of countries	34	22	26	25	18	27
B.2: World Bank non-high-income economies						
Weekly hours	–0.006	0.008	–0.001	–0.031	0.012	0.021
	(0.011)	(0.004)	(0.004)	(0.003)	(0.009)	(0.005)
# of students	10,024	69,039	81,327	101,358	51,527	80,610
# of countries	4	15	21	25	9	13

Notes: The table shows estimates of the effect of weekly hours of instruction on student achievement separately for different groups of countries. Following Lavy (2015), Panel A.1 restricts the sample to 14 Eastern European countries and Panel A.2 restricts the sample to 13 developing countries. The number of countries is lower in some columns because not all countries participated in all rounds of PISA. Panels A.3 and A.4 (based on PISA data) and Panels B.1 and B.2 (based on TIMSS data) show results for high-income and non-high-income economies as classified by the World Bank as of June 2020. The number of countries included in these regressions varies across samples because not all countries participated in all assessments. All specifications in all panels control for individual and subject fixed effects. Standard errors in parentheses are robust to clustering at the school level.

Online Appendix Figure A.1: Distribution of instruction time by PISA wave



Notes: The figure shows the distribution of student-reported instruction time (left panel) and school-average instruction time (right panel) separately for each PISA wave.

B Differences in estimates due to heterogeneous effects?

In Section 4.2, we mention the possibility that estimates differ across waves due to heterogeneous treatment effects. In this Appendix, we discuss this possibility in more detail. Due to the inherent differences in design between PISA and TIMSS, we concentrate on differences in estimates between PISA waves, with a special focus on the PISA 2006 estimate.

One important dimension of heterogeneity in the effect of instruction time is student background. For example, Lavy (2015) shows that the impact of hours is larger for students with an immigrant background and for students with less educated parents. Similarly, Bingley et al. (2018) find that the effect varies by students' gender and socioeconomic status. If student background differed between samples, this heterogeneity could explain the differences in estimated effects. To explore this possibility, panel A of Online Appendix Table B.1 shows means of students' socio-demographic characteristics separately for each PISA wave. All samples are balanced on gender, but immigration status and parental education trend upwards over time. However, these smooth trends cannot account for the much larger estimate in PISA 2006 compared to all other waves.

The effect of instruction time likely also differs by other, unobserved dimensions of student background. Moreover, it varies with school and class characteristics: for example, Rivkin and Schiman (2015) show that the effect differs by classroom quality. While we cannot determine whether the PISA samples are comparable on all possible dimensions of effect heterogeneity, any changes in such characteristics likely follow similarly smooth time trends as the characteristics observed in panel A of Online Appendix Table B.1 and as such cannot account for the much larger estimate in PISA 2006 compared to the other waves.

A related alternative explanation for the differences in estimates is that the distribution of achievement changes across waves: even in the absence of heterogeneous treatment effects, if the standard deviation of achievement was much larger in PISA 2006, this could explain the higher estimate for this wave. However, panel B of Online Appendix Table B.1 shows that means and standard deviations of test scores are broadly similar across waves, making this explanation unlikely.

Finally, non-linearities in the effect of instruction time could be at play if the distribution of hours changed between waves. In Online Appendix Figure A.1, we show that hours distributions in PISA waves other than 2006 are comparable, and we argue in the main text that the true distribution of instruction time in PISA 2006 likely looks similar. However, due to the very different measurement of instruction time, the observed distribution of hours in that year differs from those in the other years. This means that we unfortunately cannot establish to what extent non-linearities in the effect of instruction time can account for the larger estimate in PISA 2006.

Online Appendix Table B.1: Summary statistics of students' socio-demographic characteristics and achievement by PISA wave

	Orig. data					
	PISA 2006	PISA 2000	PISA 2009	PISA 2012	PISA 2015	PISA 2018
	(1)	(2)	(3)	(4)	(5)	(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.

C Details on sensitivity checks

Section 4.3 summarizes results from sensitivity checks that gauge the extent of bias in our estimates due to subject-specific confounders. In this Appendix, we present additional details on these checks.

C.1 Checks in the PISA data

Our analysis for the PISA data closely follows Lavy (2015) and comprises five sensitivity checks, the results of which are shown in Online Appendix Table C.1. First, we restrict the sample to schools that do not consider students' academic record in the admission process.[6] Intuitively, such schools should be less likely to select students based on subject-specific academic ability, reducing the potential for bias. Panel B presents the corresponding estimates (Panel A reproduces the main estimates from Panel A of Table 2 to facilitate comparison). Second, based on similar reasoning, we restrict the sample to schools that do not consider students' needs or desire for a particular program as a criterion for admission. The results for this check are presented in Panel C.

Third, we restrict the sample to schools that do not practice tracking (in any subject) between or within classes. The intuition is that schools that practice tracking will be more likely to admit students based on subject-specific academic ability. These schools could also place higher-ability students in classes with more instruction time. Panel D shows the results for the sample excluding these schools. Fourth, Panel E presents estimates for the subsample of public schools, for which subject-specific sorting is less of a concern according to Lavy (2015). Finally, we use information on teacher shortages in each subject. For example, schools that are mathematics-oriented might attract more effective math teachers, which could confound the estimates. Panel F presents estimates from regressions which control for an indicator for a lack of qualified teachers in a subject.

Overall, Online Appendix Table C.1 shows that the estimated effect of instruction time is quite similar across the different specifications within a given PISA wave, which suggests that subject-specific confounders do not bias our results. However, one caveat of these checks is that they mostly rely on information that is not subject-specific, and that they therefore might not fully capture the influence of potential subject-specific confounders. As we describe below, the TIMSS data allows us to partially address this concern.

[6] Information on factors considered in the admission process was collected in all PISA waves, but the format of the question posed to principals changed somewhat over time. Question formats also changed for some of the other variables used in our analysis, and in a few cases the question was not asked at all. Whenever information is available, we define our variables such that they most closely resemble the original variables used by Lavy (2015).

C.2 Checks in the TIMSS data

In the TIMSS data, we use detailed background information from surveys to identify potential subject-specific confounders related to schools, teachers, and students. Online Appendix Table C.2 presents the results of our sensitivity checks based on these variables. Note that not all variables are available in all waves. Moreover, the format of the underlying survey questions sometimes changes between waves; in these cases, we define variables as consistently as possible across waves.

Starting with school-related confounders, Panel B shows estimates from regressions in which the sample is restricted to schools that do not use students' academic record in the admission process (Panel A reproduces the main estimates from Panel B of Table 2 for convenience). This sensitivity check is equivalent to one of the checks conducted by Lavy (2015). Panel C shows estimates from specifications that add a control for subject-specific tracking by ability. Intuitively, such tracking could influence school choice and could also be related to instruction time, which in turn could lead to bias in our results. Panel D adds a control for whether the school offers subject-specific enrichment activities and Panel E adds a control for subject-specific remedial teaching. Such special teaching activities are likely to attract students with particularly high or low subject-specific ability, and they might also be related to instruction time. The results in Panels B to E show that our estimates from these checks are virtually identical to our main estimates.

Moving on to teacher-related confounders, Panel F adds a control for whether there is a shortage of teachers in a subject at the school, and Panel G adds a control for whether the school has had difficulties filling open teaching positions in a subject. These controls capture a lack of qualified teachers, which could be related to instruction time and also affect student achievement. Building on this same intuition, Panel H shows results from regressions that control for two observable dimensions of teacher quality: education, measured as an indicator for whether the teacher holds an advanced degree, and experience, measured as an indicator for whether the teacher has been teaching for at least five years. Our estimates in panels F to H are robust to these checks.

Finally, we estimate two specifications that add controls for subject-specific confounders related to students and parents. Panel I uses the fact that in two waves of TIMSS, students were asked to what extent their mother thinks that it is important for them to do well in each subject. This variable proxies for parents' valuation or preferences over subjects and intuitively relates to subject-specific sorting to schools. The specifications in Panel I add this variable as a control to our regressions. In Panel F, we control instead for an indicator for whether a student receives out-of-school extra lessons in a subject. This variable similarly proxies for parents' or students' subject-specific abilities and preferences. The results show that our estimates are robust to both of these sensitivity checks.

Taken together, the estimates in Online Appendix Table C.2 show no indication of bias due to subject-specific confounders related to schools, teachers, or students. However, as we emphasize in the main text, we cannot control for all potential confounders and the key identification assumption in our empirical model is ultimately untestable.

Online Appendix Table C.1: Sensitivity checks in PISA

	Orig. data					
	PISA 2006	PISA 2000	PISA 2009	PISA 2012	PISA 2015	PISA 2018
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: main estimates (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
Panel B: academic record not considered for school admission						
Weekly hours	0.060	0.017	0.025	0.034	0.022	0.013
	(0.005)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)
# of observations	266,769	29,799	265,005	146,370	162,897	171,039
Panel C: students' needs or desire not considered for school admission						
Weekly hours	0.066	0.027	0.035	0.037	0.023	0.015
	(0.006)	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)
# of observations	171,687	22,122	182,931	117,144	124,785	138,258
Panel D: no tracking by ability between or within classes						
Weekly hours	0.052		0.018		0.018	0.009
	(0.007)		(0.004)		(0.003)	(0.003)
# of observations	160,188		173,958		170,250	123,432
Panel E: public schools						
Weekly hours	0.061	0.020	0.031	0.035	0.019	0.011
	(0.004)	(0.004)	(0.003)	(0.003)	(0.002)	(0.002)
# of observations	330,492	37,899	387,117	253,281	271,068	218,034
Panel F: control for lack of qualified teachers in subject						
Weekly hours	0.058		0.027	0.031		
	(0.004)		(0.002)	(0.002)		
# of observations	460,734		493,800	327,891		

Notes: The table shows estimates of the effect of weekly hours of instruction on student achievement from various sensitivity checks. See text for details on these checks. No estimates are available for some specifications in some waves because the necessary information is not available in those waves. All regressions in all panels control for individual and subject fixed effects. Standard errors in parentheses are robust to clustering at the school level.

Online Appendix Table C.2: Sensitivity checks in TIMSS

	TIMSS 1995 (1)	TIMSS 1999 (2)	TIMSS 2003 (3)	TIMSS 2007 (4)	TIMSS 2011 (5)	TIMSS 2015 (6)
Panel A: main estimates (for comparison)						
Weekly hours	0.037	0.037	0.024	0.015	0.019	0.017
	(0.006)	(0.009)	(0.007)	(0.004)	(0.007)	(0.006)
# of observations	83,200	43,036	46,840	41,134	48,322	81,092
Panel B: academic record not considered for school admission						
Weekly hours	0.036	0.037				
	(0.007)	(0.009)				
# of observations	76,704	37,808				
Panel C: control for subject-specific tracking by ability						
Weekly hours	0.036	0.038		0.015		0.018
	(0.006)	(0.009)		(0.004)		(0.006)
# of observations	83,200	43,036		41,134		81,092
Panel D: control for subject-specific enrichment activities						
Weekly hours	0.036	0.037	0.023	0.015		
	(0.007)	(0.009)	(0.007)	(0.004)		
# of observations	83,200	43,036	46,840	41,134		
Panel E: control for subject-specific remedial teaching						
Weekly hours	0.037	0.038	0.023	0.015		
	(0.006)	(0.009)	(0.007)	(0.004)		
# of observations	83,200	43,036	46,840	41,134		
Panel F: control for shortage of teachers in subject						
Weekly hours					0.019	0.017
					(0.007)	(0.006)
# of observations					48,322	81,092
Panel G: control for difficulty of hiring teachers in subject						
Weekly hours			0.024	0.015	0.019	0.018
			(0.007)	(0.004)	(0.007)	(0.006)
# of observations			46,840	41,134	48,322	81,092
Panel H: control for experience and education of subject teacher						
Weekly hours	0.037	0.036	0.024	0.015	0.020	0.017
	(0.006)	(0.009)	(0.007)	(0.004)	(0.007)	(0.006)
# of observations	83,200	43,036	46,840	41,134	48,322	81,092
Panel I: control for mother's stated importance of doing well in subject						
Weekly hours	0.036	0.038				
	(0.006)	(0.009)				
# of observations	83,200	43,036				
Panel J: control for extra lessons in subject						
Weekly hours	0.037	0.036	0.024			
	(0.006)	(0.009)	(0.007)			
# of observations	83,200	43,036	45,433			

Notes: The table shows estimates of the effect of weekly hours of instruction on student achievement from various sensitivity checks. See text for details on these checks. No estimates are available for some specifications because the necessary information is not available in those waves. When information is available in a wave but the value on a control is missing for an observation, we impute this information at the sample mean and include a dummy indicating missing values in the regression. All regressions in all panels control for individual and subject fixed effects. Standard errors in parentheses are robust to clustering at the school level.