Routledge
Taylor & Francis Group

Education Economics

ISSN: 0964-5292 (Print) 1469-5782 (Online) Journal homepage: http://www.tandfonline.com/loi/cede20

The more, the better? The impact of instructional time on student performance

Maria A. Cattaneo, Chantal Oggenfuss & Stefan C. Wolter

EDUCATION ECONOMICS, 2017
http://dx.doi.org/10.1080/09645292.2017.1315055

Routledge
Taylor & Francis Group

Check for updates

The more, the better? The impact of instructional time on student performance

Maria A. Cattaneo[a], Chantal Oggenfuss[a] and Stefan C. Wolter[a,b,c,d]

[a]Swiss Coordination Center for Research in Education, Aarau, Switzerland; [b]Department of Economics, University of Bern, Berne, Switzerland; [c]CESifo, Ifo Institute, München, Germany; [d]IZA, Bonn, Germany

ABSTRACT

Building on earlier work that explored within-student variation in hours of instruction across school subjects, we investigate the impact of instruction time on student test scores in Switzerland, as measured by the PISA 2009 test. Our results confirm the results of previous studies of a positive effect of instruction time on student performance. Moreover, we find considerable heterogeneity in the effectiveness of instructional time across ability-related tracks, with the more able students benefitting more. Additional instruction time increases the within-school variance of subject-specific test scores, indicating an increase in educational inequality.

ARTICLE HISTORY
Received 20 June 2016
Accepted 30 March 2017

KEYWORDS
Instruction time; PISA; fixed-effect models; educational inequality

JEL-CODES
C21; I21; I25

1. Introduction

> Some scientists and educational scholars find the concept of instructional time to be intellectually unexciting, so commonsensical, and of such obvious importance that it only leads to trivial understandings and to findings that have the status of truisms (e.g. students who spend more time studying learn more) … (Berliner 1990, 3)

Although the assessment of David Berliner of the state of research dates back to the beginning of the 1990s, the empirical contributions to the questions of the use and the effectiveness of instructional time in educational production have – with a few recent exceptions – remained relatively limited (Hanushek 2015). This is astonishing given that instructional time is not only an important but most of all a scarce resource in the educational production function. The time that students can spend on education is limited by the hours in a day, the days in a week and the weeks in a school year. Every hour of instruction comes at a high cost and as educational budgets are limited, the money spent on instruction lacks for other potential inputs in the education production.

Besides the importance of instructional time, both in terms of real and monetary input in the educational process, the lack of research is surprising for at least one further reason. Total instruction time and the allocation of time to specific subjects vary greatly between countries,[1] but even a very superficial glance at the small correlations between average instruction times and student test scores at the cross-national level (e.g. Scheerens 2014) is sufficient to cast a doubt that there is a simple relationship between the two.

In this paper we analyze the effectiveness of the use of instructional time in terms of student performance. The intention is to contribute to the existing literature in five different albeit related ways. First, we provide evidence on the average effectiveness of instructional time, replicating for Switzerland the methodology applied by Lavy (2015). Contrary to international comparisons and their limitations (see e.g. Hanushek and Woessmann 2011 or Goldstein 2004) the use of only one country has

CONTACT Stefan C. Wolter stefan.wolter@vwi.unibe.ch
Supplemental data for this article can be accessed here http://dx.doi.org/10.1080/09645292.2017.1315055

the advantage that we can compare the impact of different uses of instruction time in a setting where educational goals and curricula are very similar and therefore not a source of potential bias. Furthermore and contrary to schools in many of the countries included in the previous analysis of Lavy (2015), Swiss schools do not have enough freedom to autonomously set subject-specific instruction time, which excludes at least this non-exogenous source of variation. Second, we try to overcome potential misreporting by students by using not only student self-reported data but also the official instruction times prescribed by the educational authorities. Third, we refine the analysis by controlling for extra time spent on specific subjects either during school or after school (enrichment, remedial courses or paid private tutoring). Fourth, we conduct separate analyses per school track to investigate track-specific heterogeneity in the effectiveness of instruction time on student achievements. Fifth, we analyze the impact of additional instruction time on the within-school variance of student test scores to assess potential changes in educational inequality that may arise from an ability-dependent effectiveness of learning.

We focus here on variations in instruction time on specific subjects arising mainly from variations in the number of lessons taught weekly. This is, however, not the only source of variation in instruction time. Instruction time can also vary considerably due to different durations in the time of lessons, the number of school weeks or even the number of school years. These latter sources for variations in instruction time may have very different impacts on student learning (for an overview of the literature, see OECD 2016), but due to data limitations these aspects are beyond the scope of this paper.

Our analysis shows that one additional hour of instruction per week increases the PISA score by between 0.05 and 0.06 standard deviations (SDs). However, the returns of one additional hour vary greatly by school track. For students attending schools with advanced requirements, one extra hour of instruction increases the PISA score by between 0.06 and 0.08 SDs, while the increase in PISA score is only between 0.03 and 0.04 SDs for students in tracks with basic requirements. These differences can be the consequence of many factors, such as different school environments, different teachers' attitudes or behavioral aspects like school discipline. However, the differences in the effectiveness of instructional time can also be the consequence of the difference in aptitude (time needed to learn) of pupils. If this is the case, then pupils with different abilities benefit to a different extent from additional instruction time.

The remainder of the paper is organized as follows. Section 2 provides a brief overview of the Swiss education system. Section 3 reviews the existing literature. Section 4 describes the PISA 2009 survey and the data in more detail. Section 5 introduces the empirical model and explains our identification strategy. Section 6 presents our findings of the impact of instruction time on student outcomes. Section 7 concludes the paper.

2. The Swiss education system

Switzerland provides an excellent laboratory for the study of educational policies for at least two reasons. First, the Swiss education system comprises different cantons[2] with independent educational policies, which leads to a considerable variation in many relevant parameters of the education production function. This is also the case for subject-specific instruction times. Second, despite their high degree of freedom in educational policy, including the regulation of instruction times, due to strict admission criteria for post-compulsory education the cantonal systems are under pressure to achieve similar outcomes at the end of compulsory school. In other words, considerable variation in instruction time, which is relevant in our case, paired with high pressure to achieve comparable outcomes make Switzerland an ideal setting to study the impact and the effectiveness of school resources.

The instruction times vary from canton to canton. The reasons for that have evolved historically and are based on financial, political and pedagogical rationales. The cantonal authorities stipulate not only the compulsory number of subject-specific lessons per week, but they also determine the duration of a school lesson and the total weeks per school year. Switzerland is one of the countries where tracking takes place rather early (in sixth grade, ages 11 or 12). Students are assigned according to their ability at the end of primary school to one of the two or three different ability tracks in lower secondary education, among which the subject-specific instruction times differ as well. Depending on the canton, the admission procedure might be based on the performance in the last year of primary school, on the teachers' recommendations, on an admission test, or any combination thereof. Students in all cantons are allocated in accordance with their residential address to the closest school in the school district, except for children attending Baccalaureate schools (upper track). This indicates that all students in the same ability track in one canton should have the same number of instruction hours irrespective of which school they go to.

3. A short literature review

Lavy (2015), the OECD (2016) and Rivkin and Schiman (2015) provide excellent and recent overviews of the related literature. For this reason, we only highlight here selected studies that are directly linked to our paper.

Lavy (2015) examines international gaps in students' achievements, estimating the effects of instructional time using PISA 2006 data. The study exploits within-student and within-school variation by subject (reading, math and science), thereby controlling for student and school fixed effects. The results suggest that instructional time has a positive and significant effect on test scores. By comparing different country groups, Lavy finds evidence for lower effects of instructional time in developing countries and a higher productivity of instructional time in countries with school characteristics like accountability measures and autonomy in budgetary decisions. Lavy also considers different subsamples distinguishing between different school tracking policies. The reported effects are significantly lower for schools without tracking than for schools that have some kind of tracking. The results are overall in line with Lavy (2012), who provides evidence for the impact of increasing the length of school weeks and subject-specific instructional time per week in Israel.

Building on the method used by Lavy (2015), Rivkin and Schiman (2015) run similar regressions using the international PISA 2009 database, adding controls for school quality. They find that school circumstances are important determinants of the benefits of additional instruction time.

Woessmann (2003) presents results based on the international student-level database TIMSS, using a cross-country setting and controlling institutional characteristics of different school systems. He finds that instructional time is positively related to student performance. Mandel and Suessmuth (2011) estimate state-fixed effects within Germany for the cumulated instructional time and find positive effects on student performance.

There are also studies that find no significant relationship between instructional time and school outcomes. Woessmann (2010) uses cross-state variation in Germany to eliminate unobserved state-specific factors and he does not find a statistically significant effect of instructional time on student test scores. In addition, some studies that look at the length of the school year and its impact on later earnings (Grogger 1996 or Pischke 2007) do not find statistically significant results. A study analysing the duration of the academic baccalaureate in Switzerland (Skirbekk 2006) estimates the impact of the canton-based variation of these programs on the TIMSS 2006 scores and does not find an effect of time spent in school on student achievement after controlling for school and student characteristics.

There is a small number of studies that explore exogenous variation in instructional hours. For example, Marcotte (2007) and Goodman (2014) use weather-related school closures or absences to instrument for instruction time and find negative effects of lost instructional time on student achievements. Dahmann (2015) and Huebener, Kuger, and Marcus (2016) use a large German education reform that increased the number of instruction hours per week in academic track schools and find positive effects on student performance. The effects are rather small, which might be explained by the type of intervention and content of the additional instruction hours. Dahmann (2015) finds that the quantity of instructional time matters more than the timing of instruction and Huebener, Kuger, and Marcus (2016) find stronger effects for the higher skilled students and smaller effects for the lowest skilled students, leading to a wider gap between low- and high-performing students.

Several other studies focusing on educational tracking and ability grouping provide evidence that the gains in achievement among students differ from track to track (for an overview on tracking see Betts 2011, and on instruction time and tracking Robertson and Symons 2003; Nomi and Allensworth 2009; Angelone and Moser 2012). The empirical estimates of this relationship do not necessarily constitute causality, but they still suggest that the effect of instructional time may vary between tracks.

4. Data

In order to analyze how hours of instruction affect the pupils' performance, we use data from the fourth PISA test in Switzerland, conducted by the Organization for Economic Cooperation and Development (OECD) in 2009. PISA is a standardized test administered to 15-year-old students in OECD member countries and other participating countries who are enrolled in grades seven and above. Students are assessed in three domains, reading, math and science, by tasks that are in the context of rather general everyday situations, avoiding a strong link to the country-specific curriculum framework. The sample is drawn using a two-stage stratification design. First, schools within the country are randomly selected. Second, a random sample of students is selected from within each of the schools. In addition to the test results, PISA includes a student questionnaire with detailed family, socio-demographic and economic background information, and a school questionnaire with information about school type and school demographics. In Switzerland, an additional representative PISA sample of students in grade 9 (ages 13–18), i.e. in the last year of compulsory education, was collected in 2009. This was done only in 12 out of 26 cantons.[3] This so-called national PISA sample of 12 cantons will be the main data source for our analysis, first because a comparison of students attending the same grade, instead of students of the same age but not necessarily of the same grade, is more adequate for our purposes, and, second, because the over-sampling in the national PISA sample increases the number of observations considerably.

Besides the information on test performance, the PISA 2009 database provides detailed information on the hours of instruction. Specifically, students were asked, 'How many minutes, on average, are there in a <class period> for the following subjects?' and 'How many <class periods> per week do you typically have for the following subjects?'. We combined these two questions from the PISA data with information on the number of school weeks per year per canton to create a variable for the number of hours of instruction per year, and the weekly number of hours of instruction. In order to make our results comparable to Lavy (2015) we do not use the individual reported time for the estimations but we use the school-level averages of the individual answers.

While school-level averages are less prone to misreporting, there might still be measurement error due to self-reporting and recall bias of students. We try to overcome this issue by not only using the student self-reported information on instruction time (PISA) but also explore an additional source of information for mandatory hours of instruction, coming from the official cantonal administration (EDK 2008). This data source contains the number of lessons and minutes per lesson per subject by canton and type of school (ability track).[4] This data should avoid some of the individual misreporting from the PISA data. However, there are still some potential other issues to consider when using this data as we cannot be sure that each teacher actually follows the official time table approved by the cantonal authorities. We believe that the true measure of hours of instruction is probably somewhere in between the official and the self-reported hours. This is the reason why we will use both in the analysis and compare the results.

Our initial PISA sample consists of 13,605 ninth graders (ages 13–18). After deleting observations with missing values for hours of instruction, we have a final sample of 11,433 students for our empirical analysis with an average reading score of 511 points, an average math score of 546 points and an

Table 1. Summary of PISA test scores and weekly hours of instruction by subject.

	Mean	Std.Dev.	Lower track	Middle track	Upper track
Test scores (PISA data)					
All subjects	529.1	85.3	460.1	526.2	573.7
Reading	511.4	81.8	442.1	508.1	558.8
Math	546.2	87.9	477.2	544.9	591.6
Science	530.0	82.6	461.3	525.4	570.2
Number of pupils	11,433		2870	4103	4460
Self-reported hours (PISA data)					
All subjects[a]	3.42	0.91	3.46	3.45	3.36
Reading	3.85	0.57	3.85	3.81	3.90
Math	3.72	0.39	3.74	3.74	3.68
Science	2.56	1.07	2.59	2.70	2.42
Number of pupils	11,433		2870	4103	4460
Official hours (Council of cantonal ministers)					
All subjects[a]	3.41	1.15	3.43	3.44	3.38
Reading	3.91	0.61	3.85	3.87	3.98
Math	3.73	0.43	3.66	3.74	3.76
Science	2.47	1.56	2.55	2.59	2.31
Number of pupils	11,433		2870	4103	4460

Notes: To standardize the number of hours per week in different schools, we multiplied the minutes of instruction time per week with the number of weeks in the canton and divide this by 38 (which is the mean number of school weeks in Switzerland). Calculation of mean numbers of test scores are based on the five plausible values of the test scores per subject using the methods outlined in OECD (2014).

[a]Average instructional time per week per subject for all three subjects.

average science score of 530. The average score in all subjects for the gross sample is slightly lower than for the final analytical sample.

The average hours of instruction are very similar in both data sets. However, the information differs significantly between PISA self-reported instruction time and the official cantonal guidelines for reading and science (see Table 1). A possible reason for the observed difference in science is that pupils take more lessons than the minimum suggested by the cantonal authorities because the number of optional subjects offered by schools tends to be larger in that area. In the case of reading, the pupils' and the authorities' definition of reading lessons, and what they enclose, might vary, leading pupils to think that they have fewer hours of instruction than they actually have.

The Swiss education system with its different tracks at the lower-secondary school level allows us to compare the efficiency of one extra hour of instruction per week for pupils with different academic abilities and within different school environments. In the majority of cantons, pupils are sorted into different school tracks after sixth grade (ages 11 or 12) according to their intellectual abilities (see also section 2). The PISA database provides detailed information on the cantonal track that students are assigned to, which we can use for our estimation. The upper track with advanced requirements teaches the more intellectually demanding courses, middle track offers courses with intermediate requirements, and the lower track offers basic-level courses. All tracks teach general skills but at different levels of academic requirements.

5. Empirical strategy

The starting point for our empirical analysis of the effect of instructional time on PISA scores is the following linear regression model:

$$y_{ijk} = \beta_j I_{jk} + \kappa Z_i + \gamma X_{ij} + \psi S_k + \alpha_i + \lambda_j + \sigma_k + u_{ijk} \quad (1)$$

where y_{ijk} is the PISA score of pupil i in subject j in school k. I_{jk} is the instructional time for subject j in school k. X_{ij} is a vector of individual characteristics that vary by subject, Z_i is a vector of individual characteristics that are the same over all subjects, and X and Z are assumed to not vary across schools. The vector S_k contains school specific characteristics. The parameters α_i, λ_j, and σ_k represent individual, subject and school specific effects, respectively. u_{ijk} is the unobserved error term.

One of the greatest difficulties when estimating the effect of hours of instruction time on test scores, like PISA, is the possibility that unobservable individual and school characteristics might be correlated with the number of school lessons. Following Lavy (2015), we take advantage of the PISA data structure and explore the within-student variation over subjects to estimate the effects of instructional time on test scores. This method accounts for potential confounding factors such as students' ability and school quality that are constant over the three subjects.

The key identifying assumption needed for providing unbiased estimates of the effects of hours of instruction on test scores using the within-student variation of instruction hours is that the average effect of one hour of instruction is the same over all subjects. This would not be the case if for example children had a systematically higher learning productivity in one subject than in another. In order to test if the assumption of homogenous effects is fulfilled, we apply different approaches, including the one suggested by Metzler and Woessmann (2012) with a parametric specification of the unobserved pupil effect and an equation systems approach as in Chamberlain (1982), as well as the tests proposed by Lavy (2012, 2015) on the subsample analyses by pairs of subjects. Overall, we do not find evidence for a violation of the assumption of equal effects, which is consistent with the results of Lavy (2012, 2015). For further details and the results of these tests, see Appendix A3 (online supplemental file).

Another important assumption is that children are not choosing or being sorted into schools based on different subject instruction times. This would be the case if for example children with a special talent for math could select schools that specialized in math, providing extra instruction time on the subject. In this case, the unobservables u_{ijk} would be correlated with instruction hours even after controlling for pupil, subject and school fixed effects. However, we do not believe that this selection could be an issue in our context because there are no subject-specialized schools in Switzerland, except for some Baccalaureate schools, which we are able to identify and exclude from the sample. Appendix A4 (online supplemental file) shows the results with those schools excluded from the sample, which are very similar to those in our main table (compare section 6.1 below).

Under the assumption of equal effects of instructional time on test scores, i.e. $\beta_j = \beta$, we can apply a within-student transformation over the three subjects in Equation (1), which yields the following restricted fixed effects model:

$$\ddot{y}_{ijk} = \beta \ddot{I}_{jk} + \gamma \ddot{X}_{ij} + \lambda_j + \ddot{u}_{ijk} \tag{2}$$

where $\ddot{y}_{ijk} = y_{ijk} - \sum_j y_{ijk}/3$ and similarly for $\ddot{I}_{jk}$, $\ddot{X}_{ij}$ and $\ddot{u}_{ijk}$. The vector X in the fixed effects regression includes control variables for attending enrichment, remedial or extra private lessons, and we add subject fixed effects through λ_j. Estimation of Equation (2) is straightforward using standard software packages with fixed effects capabilities. We account for the five plausible values of the test scores per subject using the methods suggested in OECD (2009a, 2009b). In the within-student FE regression, this leads to $5^3 = 125$ possible combinations of the five plausible values per three subjects. Standard errors are calculated based on the variability that comes from the imputation plus the sampling variability. As an alternative and computationally simpler approach, we also use the means per subject of the plausible values, as for example in Kunz (2016), see Table A1 in the appendix (online supplemental file). It turns out that our results are not sensitive to the choice of the two methods, so we will confine ourselves to the averages only for all other results. For the sampling variability, we cluster the standard errors at the school level for both the self-reported instruction hours and the official instruction hours as reported by the cantonal authorities.[5]

6. Results

In this section, we present our empirical results looking at the average impact of instruction time on student performance (section 6.1), at the impact of instruction time per school track (6.2), and the impact of instruction time on the variance of student performance within schools (6.3). We provide some robustness checks to support our results at the end (6.4).

6.1. Impact of instruction time on student performance

Lavy (2015) reports a student fixed-effects result of 0.058, i.e. one additional hour of instruction time per week increases PISA scores by 0.058 standard deviations (SDs) on average. Lavy uses the international mean of 500 points and a standard deviation of 100 to standardize the scores. The results of our replication of the Lavy (2015) estimates, focusing on the Swiss case, are presented in Table 2. Using the national mean of 518 points and the national standard deviation of 92, we find exactly the same coefficient for Switzerland as in Lavy (2015) when using the self-reported data on instruction hours. If we standardize PISA scores using the international mean and standard deviation, then the estimated effect of one additional hour of instruction time decreases slightly to 0.054 SDs. When using the official cantonal data on instruction time, the estimated coefficient decreases to 0.046 SDs.[6,7]

If we use these results for a small back-of-the envelope calculation comparing the effect of an average hour of instruction time with the effect of an additional hour of instruction, then we find that an additional hour has only between one third and less than half of the effect of an average hour on PISA scores.[8]

6.2. Impact of instruction time on student performance per track

Even if one hour of instruction time has the same productivity for all subjects, it may be that different students benefit differently from extra instructional time, as suggested by the descriptive statistics in Table 1. In order to examine whether students with different academic skills levels have different returns to instructional time, we split the sample, based on information from the PISA data, in students who attended an upper-level school track, students in a middle-level school track and students in a lower-level school track.

The results in Table 2 show that one additional hour of instruction per week has in fact a significantly higher effect for pupils in a more demanding track.[9] This result holds independently of the data source for instruction time. We can think of two possible explanations for this: First, pupils in lower-level tracks might have indeed lower learning capabilities. Second, the inputs in the different tracks might be different. This may be the case if for example teachers are less qualified or the discipline is worse in the lower-level tracks. This would be in line with results from Rivkin and Schiman (2015) who show that returns to additional instruction time depend on the quality of the learning environment, and considerations of the quality of teachers and schools.[10]

The PISA data provides some information on the learning environment in the form of indices for attitudes towards school, teacher-student relations, disciplinary climate, student-related and

Table 2. Effect of hours of instruction on mean PISA test scores.

	All	Lower track	Middle track	Higher track
Self-reported hours (PISA data)				
Hours of instruction	0.059	0.038	0.060	0.080
	(0.007)	(0.012)	(0.010)	(0.014)
Official hours (Council of cantonal ministers)				
Hours of instruction	0.046	0.034	0.043	0.062
	(0.006)	(0.008)	(0.008)	(0.010)
Number of pupils	11,433	2870	4103	4460
Number of observations	32,411	7851	11,637	12,923

Notes: Reported numbers are based on within-student fixed effects regressions over all three subjects (reading, math, science). All regressions control for subject fixed effects and additional hours of instruction taken, which include enrichment, remedial and private tutoring lessons. Test scores are standardized using the mean and standard deviation for Switzerland of 518 and 92, respectively. Estimates and standard errors are obtained using the five plausible values of test scores per subject and the methods outlined in OECD (2014). Standard errors in parentheses are clustered at the school level.

teacher-related factors affecting school climate (for a detailed description of the indices see OECD 2010b). The indices can be used to analyze the potential differences in inputs and in the learning environment among different school tracks. All values of the indices are lower in the lower-level tracks, with the exception of the teacher-student relations. Especially the values of disciplinary climate and student-related factors affecting school climate are significantly lower in the lower-level tracks. Table A5 in the appendix (online supplemental file) presents the results using the different indices by school track interacted with the instruction hours. The results show that the interaction effects overall are rather small, but where significant they indicate that a better learning environment increases the returns to instruction time. This might provide supportive evidence for the second explanation why the efficiency of one hour of instruction is lower in the basic-level tracks.

A related issue is that tracks may differ in terms of teaching styles. Teachers in lower tracks need to prevent that some pupils do not even reach basic competencies by the end of compulsory school. Therefore, teachers have an incentive to focus on the weakest pupils by reducing the intensity and speed of instruction. In the highest track, however, the goal is to make pupils ready for the academic baccalaureate schools and therefore the focus is on excellence and the brightest pupils. If some pupils are not able to follow the increased speed of instruction, they can always be re-directed to the intermediate track, where the teaching style and speed focuses more on the median pupil.

6.3. Impact of instruction time on the variance of student performance within schools

It is easy to understand why school and class climate can influence the effectiveness of instruction time. Another source that could potentially explain the differences in the effectiveness of hours of instruction between tracks, however, could be the differences of the students in terms of time needed to learn. Students have been sorted into the different tracks due to their different potential and abilities, and therefore, it is possible that students in the highest track can make more productive use of an hour of instruction time than less talented students sorted into the lower tracks – all other things being equal.

If time needed to learn is the explanation for the differences in the effectiveness of instructional time between the different ability tracks, then we can assume that the effectiveness of instructional time also differs within tracks and schools. In this case, additional instructional time would not only increase the average performance of students but it would also increase the performance gaps between students in the same class or school. Of course, this assumption depends on how extra time would be used by teachers. If teachers used the extra hours to compensate the weaker students and helping them to catch up with the better students, instead of teaching additional or new content, then the extra time would produce more homogenous class results.

We investigate the last issue by regressing further statistics than the mean values describing the test score distribution within schools (like the variance, skewness and indicators whether the test scores per subject exceed a certain cutoff) on the number of instruction hours. Once again we exploit the within school variation in those statistics and instructional time across the three subjects. This allows us to identify the effect of instruction time by keeping constant unobserved school characteristics common over the subjects, such as general school or teacher quality. In order for this to work, we need to assume that the effect of one additional hour on the score gap between the best and the worst students as measured by the different statistics describing the distribution of test scores, is the same for all subjects.

Table 3 shows that one extra hour of instruction time per week increases the score variance by almost 150 points squared when using PISA information and almost 110 when using the official cantonal hours. Increasing the hours of instruction does, however, not affect the skewness of the results within a school, which could be expected if additional time would have been used specifically for a particular part of the ability distribution of students, either by only compensating the weakest or by using the extra time only for additional content for the gifted students. The considerable impact on

Table 3. Effect of hours of instruction on heterogeneity in PISA test scores.

	Self-reported (PISA)		Official hours	
	Variance	Skewness	Variance	Skewness
Hours of instruction	149.21	0.008	107.05	0.005
	(50.37)	(0.011)	(34.33)	(0.008)
Number of schools	321		321	
Number of observations	963		963	

Notes: Within-school regressions of variance and skewness in PISA scores on hours of instruction, including subject fixed effects. Standard errors in parentheses are clustered at the school level.

the within school variance and the non-effect on the skewness of the results is therefore a hint that teachers gravitate towards the median students in their teaching.

We further exploit the impact of instruction hours on the score gap by plotting the cumulative distribution of test scores by three categories of instruction hours (less than 3.5 h per week, between 3.5 and 4.5 h, and more than 4.5 h; see Table 5 below for further description of the categories). Figure 1 shows that the distribution of test scores net of the controls and fixed effects has a smaller mean/median and a smaller spread for the category less than 3.5 h per week than for the other two categories of instruction time, indicating that the score gap tends to increase with more instruction hours. This result is confirmed in Table 4, which depicts the results of within-student fixed effects regressions of different indicators for the test score exceeding percentiles of the test score distribution per subject. The effects of one additional hour of instruction time are smallest for the lowest percentiles, largest for the 25% to 75% percentiles, and then again get smaller for the highest percentiles (but are still significantly larger than for the lowest percentiles).

A potential issue might be the existence of simultaneity in determining hours of instruction and score heterogeneity, for example if cantons or schools with more heterogeneity increase instructional time to reduce the score gap.[11] However, we do not think that this might be an issue here. The number of hours is regulated by the cantonal authorities and schools therefore do not have much freedom. At the same time authorities cannot easily change the number of hours already set. On the one hand, there are teacher unions who advocate teachers' interests, and on the other hand, there is a budgetary aspect of increasing expenditures. We therefore conclude that increasing the number of hours of instructional time tends to increment performance heterogeneity.

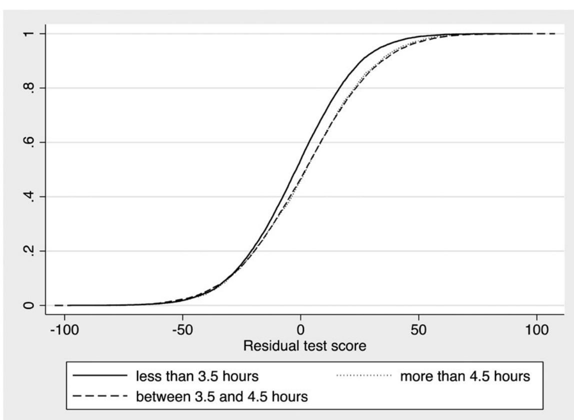

Figure 1. Distribution of PISA test scores by categories of self-reported instruction time.

Notes: Horizontal axis: Residuals of PISA test scores. Vertical axis: Cumulative distribution function by categories of self-reported instruction time. Residuals are obtained from within-student FE regression of test scores on extra hours of instruction time (enrichment, remedial courses, paid private tutoring) and subject fixed effects. All subjects and tracks are used for the graph. For a description of the categories of instruction time, see Table 4.

Table 4. Effect of hours of instruction on thresholds in the PISA test score distribution.

	Percentiles						
	5%	10%	25%	50%	75%	90%	95%
Self-reported hours (PISA data)							
Hours of instruction	0.003	0.007	0.022	0.030	0.024	0.016	0.010
	(0.002)	(0.002)	(0.003)	(0.004)	(0.003)	(0.003)	(0.002)
Official hours (Council of cantonal ministers)							
Hours of instruction	0.004	0.007	0.017	0.020	0.018	0.013	0.009
	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Number of pupils	11,433						
Number of observations	32,411						

Notes: Reported numbers are based on within-student fixed effects regressions over all three subjects (reading, math, science). The dependent variables are binary indicators if the PISA test score exceeds the percentile indicated at the top of the table in the distribution per subject. All regressions control for subject fixed effects and additional hours of instruction taken, which include enrichment, remedial and private tutoring lessons. Estimates and standard errors are based on the mean of the five plausible values of test scores per subject. Standard errors in parentheses are clustered at the school level.

Table 5. Percent of students by subject and number of hours.

	Less than 3.5 h	3.5–4.5 h	More than 4.5 h
Self-reported (PISA)			
All subjects[a]	45.8%	45.9%	8.3%
Reading	30.3%	51.4%	18.3%
Math	28.8%	69.4%	1.8%
Science	83.8%	11.9%	4.3%
Number of pupils	11,433		
Official hours			
All subjects[a]	40.0%	45.8%	14.2%
Reading	23.5%	47.5%	29.0%
Math	23.6%	75.5%	0.9%
Science	78.3%	9.1%	12.6%
Number of pupils	11,433		

[a]Average time per week per subject for all three subjects.

6.4. Robustness checks

It is possible that the relationship between PISA scores and hours of instruction is non-linear. In order to investigate potential non-linearities, we estimate a model including a square term of the hours of instruction, which also allows us to test whether there are diminishing returns to additional instruction time. Second, we estimate a model including the hours of instruction as categorical variables. Since the variation in instructional time in the data is mainly between 3 and 5 h per week, the

Table 6. Non-linearities in effect of hours of instruction on mean PISA test scores.

	Self-reported (PISA)		Official hours	
	(1)	(2)	(1)	(2)
Hours of instruction	0.110		0.142	
	(0.033)		(0.023)	
Hours of instruction squared	−0.008		−0.013	
	(0.005)		(0.003)	
3.5–4.5 h of instruction		0.093		0.106
		(0.015)		(0.017)
>4.5 h of instruction		0.089		0.105
		(0.021)		(0.014)
Number of pupils	11,433			
Number of observations	32,411			

Notes: The reference category for hours of instruction is less than 3.5 h (see also Table 4). Scores are standardized using the mean and standard deviation for Switzerland of 518 and 92 respectively. Standard errors in parentheses are clustered at the school level.

categories were created as follows: time of instruction up to 3.5 h per week, between 3.5 and 4.5 h per week, and more than 4.5 h per week. The distribution of students by categories of the number of hours of instruction and by subjects is presented in Table 5.

The results of the regressions including a second order polynomial in the hours of instruction as well as the instruction hours in categorical form are presented in Table 6. The coefficient of the squared hours in specifications 1 and 2 is negative, with a maximum at about 7–8 h of instruction per week, which indicates diminishing returns to instruction time. However, the significance depends on the data used. Specifications 3 and 4 include the indicators for instruction time. The results using official data and PISA data do not differ significantly and both suggest that the effect of more than 4.5 h of instruction relative to less than 3.5 h is the same as that for between 3.5 and 4.5 h, which confirms the result of diminishing returns.

7. Conclusions

Summarizing the results found in this paper, we would like to highlight three important findings:

First, our replication of Lavy's (2015) study using the within student variation in instruction time within Switzerland to assess the impact of instruction time on student test scores shows remarkably similar results. Controlling for additional information on extra in-school hours for remedial or enrichment purposes or out-of-school private tutoring does not affect our results. However, using official, prescribed school hours instead of self-reported data lowers the impact of instruction time on test scores somewhat. While we find a difference in the two measures of instruction time, we cannot say whether the one is better than the other. Self-reported data is obviously prone to measurement error but prescribed hours might also not adequately depict the reality in schools.

Second, in all our specifications we find a significant impact of additional instructional time on learning outcomes measured with PISA test scores, but the effectiveness of an additional hour of instruction is only between 35% and 50% of the impact we would expect from an average hour of instruction. In other words, variations of instructional time should be considered carefully by educational authorities as the marginal gains from more instruction time might be relatively low compared to alternative uses of time and the financial resources that are needed for additional instruction time.

Third, the heterogeneity of the effectiveness of instructional time between different ability tracks and the impact of additional instruction time on the within school variance of student test scores puts the individual student time needed to learn at the center of attention. John Carroll, when revisiting the literature and findings that had emerged after he had developed, what later had become known as the Carroll Model, had stressed the same idea:

> The models emphasis on aptitude as a determinant of time needed for learning suggests that increased efforts be placed on predicting student potentialities and designing instruction appropriate to those potentialities, if ideals of equal opportunity to learn are to be achieved within a diversity of educational objectives. (Carroll 1989, 26)

Interestingly, in discussions with educational practitioners and policy makers, the need for additional instruction time is often motivated by the argument that more time is needed to close the gap between individual student performances. Our results support this view insofar that we find indications that less able students would indeed need more instruction time to achieve similar results as the more able students. But our additional finding that more instruction time increases the variance of test results rather than reduces it, reveals that apparently the additional instruction time in schools is not used to compensate the weak. Therefore, the use of instruction time in schools has to be reconsidered if more equal results want to be achieved.

Notes

1. According to the latest OECD statistics (2015), the average hours per school year in lower secondary education range from 754 (Sweden) to 1167 h (Mexico) with an OECD average of 916 h.

2. The Swiss cantons are comparable with US states, the German 'Bundesländer' or the Canadian provinces in terms of their degree of autonomy in educational policy.
3. These 12 cantons were Zurich, Berne, Aargau, Fribourg, Appenzell, St. Gallen, Schaffhausen, Ticino, Valais, Vaud, Geneva, Jura.
4. For this reason, the within-canton variation per subject and school type is zero for this variable. This does not hold for the self-reported hours, where the within-canton variation is on average 0.33 h per week for reading, 0.28 h per week for math, and 0.51 h per week for science. See Table 1 below for the overall variation in the self-reported and official hours of instruction time.
5. Our conclusions do not change if we alter the clustering of standard errors to the cantonal level for both self-reported and official hours of instruction time.
6. We also considered learning as a cumulative process (Mandel and Suessmuth 2011) by estimating the model using official data on the cumulated hours of instruction between seventh and ninth grade. The coefficient of hours of instruction is still positive and significant, even though smaller at about 0.039 SDs (se = 0.004). This leads us to conclude that the hours of instruction in the current year (ninth grade) are more strongly related to the PISA test scores than those in the previous years, but there is still an overall positive effect.
7. Appendix Table A1 (online supplemental file) shows the main results without the control variables for the additional hours of instruction time for attending enrichment, remedial or extra private lessons. If we do not drop observations with missing information in these variables, the main coefficients of the model also do not alter and are 0.046 for the estimation using the official data, and 0.059 for the estimation using the PISA self-reported data in the full sample (based on 12,446 pupils and 35,150 subject-pupil observations, and similar for the subsamples).
8. According to the OECD, an average student should progress by 39 PISA points per year of schooling (see OECD 2010a, 27). On average (all subjects), a Swiss student had 3.4 h of subject specific weekly schooling. Therefore, if an average Swiss student would also progress by 39 PISA points per year, an average weekly hour of instruction would generate a gain of 11.47 PISA points. Our FE results indicate, however, only a gain of between 4.2 and 5.3 PISA points for an additional hour of instruction.
9. Lavy (2015) presents the estimated effects of hours of instructional time by tracking status, comparing systems that have no tracking at all with systems that track by class or in class. This is not the same as in our analyses, where we compare the effects between different secondary school track levels.
10. There are several studies that show a direct relation between classroom climate and student achievement, e.g. Lazear (2001), Benner, Graham, and Mistry (2008) and Arens, Morin, and Watermann (2015).
11. As the PISA test is based on a school sample and not a class sample, the sample of bigger schools might comprise students of different tracks. In order to test whether this would bias our results, we also ran our regressions with mono-track schools only. The number of observations is considerably smaller but the coefficients are almost identical (results available upon request).

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Angelone, D., and U. Moser. 2012. "More Hours Do Not Necessarily Pay Off. The Effect of Learning Time on Student Performance at Different School Types in Switzerland." In *Research on PISA. Research Outcomes of the PISA Research Conference 2009*, edited by M. Prenzel, M. Kobarg, K. Schöps, and S. Rönnebeck, 129–144. Dordrecht: Springer.

Arens, A. K., A. J. S. Morin, and R. Watermann. 2015. "Relations between Classroom Disciplinary Problems and Student Motivation: Achievement as a Potential Mediator?" *Learning and Instruction* 39: 184–193. doi:10.1016/j.learninstruc.2015.07.001.

Benner, A. D., S. Graham, and R. S. Mistry. 2008. "Discerning Direct and Mediated Effects of Ecological Structures and Processes on Adolescents' Educational Outcomes." *Developmental Psychology* 44: 840–854. doi:10.1037/0012-1649.44.3.840.

Berliner, D. C. 1990. "What's All the Fuss about Instructional Time?" In *The Nature of Time in School. Theoretical Concepts, Practitioners Perceptions*, edited by M. Ben-Peretz and R. Bromme, 3–35. New York: Columbia University Teachers College Press.

Betts, J. R. 2011. "The Economics of Tracking in Education." In *Handbook of the Economics of Education*, edited by E. A. Hanushek, S. J. Machin, and L. Woessmann, Vol. 3, 341–381. Amsterdam: Elsevier.

Carroll, J. B. 1989. "The Carroll Model: A 25-year Retrospective and Prospective View." *Educational Researcher* 18 (1): 26–31. doi:10.3102/0013189X018001026.

Chamberlain, G. 1982. "Multivariate Regression Models for Panel Data." *Journal of Econometrics* 18 (1): 5–46. doi:10.1016/0304-4076(82)90094-X.

Dahmann, S. 2015. "How Does Education Improve Cognitive Skills? Instructional Time Versus Timing of Instruction." SOEP Papers, No. 769.

EDK [Swiss Conference of Cantonal Ministers of Education]. 2008. *Stundentafeln der Volksschule*. Bern: Schweizerische Konferenz der kantonalen Erziehungsdirektoren.

Goldstein, H. 2004. "International Comparisons of Student Attainment: Some Issues Arising from the PISA Study." *Assessment in Education: Principles, Policy & Practice* 11 (3): 319–330. doi:10.1080/0969594042000304618.

Goodman, J. S. 2014. "Flaking Out: Student Absences and Snow Days as Disruptions of Instruction Time." NBER Working Paper, 20221.

Grogger, J. 1996. "Does School Quality Explain the Recent Black/White Wage Trend?" *Journal of Labor Economics* 14: 231–253. doi:10.1086/209810.

Hanushek, E. A. 2015. "Time in Education: Introduction." *The Economic Journal* 125 (588): F394–F396. doi:10.1111/ecoj.12266.

Hanushek, E. A., and L. Woessmann. 2011. "The Economics of International Differences in Educational Achievement." In *Handbook of the Economics of Education*, edited by E. A. Hanushek, S. J. Machin, and L. Woessmann, Vol. 3, 89–200. Amsterdam: Elsevier.

Huebener, M., S. Kuger, and J. Marcus. 2016. "Increased Instruction Hours and the Widening Gap in Student Performance." DIW Discussion Paper 1561.

Kunz, J. S. 2016. "Analyzing Educational Achievement Differences between Second-generation Immigrants: Comparing Germany and German-Speaking Switzerland." *German Economic Review* 17: 61–91. doi:10.1111/geer.12062.

Lavy, V. 2012. "Expanding School Resources and Increasing Time on Task: Effects of a Policy Experiment in Israel on Student Academic Achievement and Behavior." NBER Working Paper Number w18369.

Lavy, V. 2015. "Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries." *The Economic Journal* 125 (588): F397–F424. doi:10.1111/ecoj.12233.

Lazear, E. P. 2001. "Educational Production." *The Quarterly Journal of Economics* 116 (3): 777–803. doi:10.1162/00335530152466232.

Mandel, P., and B. Suessmuth. 2011. "Total Instructional Time Exposure and Student Achievement: An Extreme Bounds Analysis Based on German State-level Variation." CESifo Working Paper Series. CESifo Group Munich.

Marcotte, D. E. 2007. "Schooling and Test Scores: A Mother-natural Experiment." *Economics of Education Review* 26 (5): 629–640. doi:10.1016/j.econedurev.2006.08.001.

Metzler, J., and L. Woessmann. 2012. "The Impact of Teacher Subject Knowledge on Student Achievement: Evidence from within-teacher within-student Variation." *Journal of Development Economics* 99: 486–496. doi:10.1016/j.jdeveco.2012.06.002.

Nomi, T., and E. Allensworth. 2009. "Double-dose Algebra as an Alternative Strategy to Remediation: Effects on Students' Academic Outcomes." *Journal of Research on Educational Effectiveness* 2 (2): 111–148.

OECD. 2009a. *PISA 2009 Technical Report*. Paris: OECD.

OECD. 2009b. *PISA Data Analysis Manual*. Paris: OECD.

OECD. 2010a: *PISA 2009 Results: Overcoming Social Background Equity in Learning Opportunities and Outcomes* (Vol. II). Paris: OECD.

OECD. 2010b. *PISA 2009 Results: What Makes a Schools Successful? - Resources, Policies and Practices* (Vol. IV). Paris: OECD.

OECD. 2014. *PISA 2012 Technical Report*. Pisa: OECD.

OECD. 2015. *Education at a Glance*. Paris: OECD.

OECD. 2016. *Student Learning Time: A Literature Review*. OECD Education Working Paper No. 127. Paris: OECD.

Pischke, J.-S. 2007. "The Impact of Length of the School Year on Student Performance and Earnings: Evidence from the German Short School Years." *The Economic Journal* 117 (4): 1216–1242. doi:10.1111/j.1468-0297.2007.02080.x.

Rivkin, S. G., and J. C. Schiman. 2015. "Instruction Time, Classroom Quality, and Academic Achievement." *The Economic Journal* 125 (588): 425–448. doi:10.1111/ecoj.12315.

Robertson, D., and J. Symons. 2003. "Do Peer Groups Matter? Peer Group Versus Schooling Effects on Academic Attainment." *Economica* 70: 31–53. doi:10.1111/1468-0335.t01-1-00270.

Scheerens, J. 2014. *Effectiveness of Time Investments in Education. Insights from a Review and Meta-Analysis*. SpringerBriefs in Education. doi:10.1007/978-3-319-00924-7.

Skirbekk, V. 2006. "Does School Duration Affect Student Performance? Findings from Canton-Based Variation in Swiss Educational Length." *Swiss Journal of Economics and Statistics* 142 (1): 115–145.

Woessmann, L. 2003. "Schooling Resources, Educational Institutions and Student Performance: The International Evidence." *Oxford Bulletin of Economics and Statistics* 65: 117–170. doi:10.1111/1468-0084.00045.

Woessmann, L. 2010. "Institutional Determinants of School Efficiency and Equity: German States as a Microcosm for OECD Countries." *Journal of Economics and Statistics (Jahrbücher für Nationalökonomie und Statistik)* 230: 234–270.