NBER WORKING PAPER SERIES

DO DIFFERENCES IN SCHOOL'S INSTRUCTION TIME EXPLAIN INTERNATIONAL ACHIEVEMENT GAPS IN MATH, SCIENCE, AND READING? EVIDENCE FROM DEVELOPED AND DEVELOPING COUNTRIES

Victor Lavy

Working Paper 16227
http://www.nber.org/papers/w16227

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2010

Special thanks go to Katherine Eyal and Alexander Zablotsky for their outstanding research assistance. I benefited from comments at seminars at the LSE, Rome Tor Vergata, LACEA 2009 conference in Buenos Aires, Itau Bank conference in Rio, Hebrew University and Paris School of Economics. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Do Differences in School's Instruction Time Explain International Achievement Gaps in Math, Science, and Reading? Evidence from Developed and Developing Countries
Victor Lavy
NBER Working Paper No. 16227
July 2010
JEL No. I21,I28

ABSTRACT

There are large differences across countries in instructional time in schooling institutions. Can these differences explain some of the differences across countries in pupils' achievements in different subjects? While research in recent years provides convincing evidence about the effect of several inputs in the education production function, there is limited evidence on the effect of classroom instructional time. Such evidence is of policy relevance in many countries, and it became very concrete recently as President Barrack Obama announced the goal of extending the school week and year as a central objective in his proposed education reform for the US. In this paper, I estimate the effects of instructional time on students' academic achievement in math, science and reading. I estimate linear and non-linear instructional time effects controlling for unobserved heterogeneity of both pupils and schools. The evidence from a sample of 15 year olds from over fifty countries that participated in PISA 2006 consistently shows that instructional time has a positive and significant effect on test scores. The effect is large relative to the standard deviation of the within pupil test score distribution. I obtain similar evidence from a sample of 10 and 13 year olds in Israel. The OLS results are highly biased upward but the within student estimates are very similar across groups of developed and middle-income countries and age groups. Evidence from primary and middle schools in Israel is similar to the evidence from OECD countries. However, the estimated effect of instructional time in the sample of developing countries is much lower than the effect size in the developed countries. I also show that the productivity of instructional time is higher in countries that implemented school accountability measures, and in countries that give schools autonomy in hiring and firing teachers.

Victor Lavy
Department of Economics
Hebrew University
Mount Scopus
91905 Israel
and Royal Holloway
University of London
and NBER
msvictor@mscc.huji.ac.il

I. Introduction

There are large differences across countries in instructional time in public schooling institutions. For example, among European countries such as Belgium, France and Greece, pupils aged 15 have an average of over a thousand hours per year of total compulsory classroom instruction while in England, Luxembourg and Sweden the average is only 750 hours per year.[1] For children aged 7-8 in England, Greece, France and Portugal average instructional time is over 800 hours per year while in Finland and Norway it is less than 600 hours. Similar differences among countries exist in the number of classroom lessons per week in different subjects as evident from the PISA 2006 data. For example, pupils aged 15 in Denmark are exposed to 4 hours of instruction per week in math and 4.7 in reading, while pupils of the same age in Austria have only 2.7 hours of weekly classroom lessons in math and 2.4 in reading. Overall, total weekly hours of instruction in math, reading and science is 55 percent higher in Denmark (11.5 hours) than in Austria (7.4 hours). Similar magnitudes of disparities in instructional time are observed among the Eastern European and developing countries that are included in PISA 2006. Can these large differences explain some of the differences across countries in pupils' achievements in different subjects? This question is of policy relevance in many countries, for example, it became very concrete recently in the US as President Barrack Obama argued that American children should go to school longer, either to stay later in the day or into the summer. He announced the objective of extending the school week and year as a central element in his proposed education reform for the US.[2]

---

[1] Source: OECD Education at Glance, 2006: http://www.oecd.org/dataoecd/6/47/37344903.xls.

[2] President Barrack Obama said recently (March 10, 2009, at a speech to the U.S. Hispanic Chamber of Commerce) that American children should go to school longer — either stay later in the day or into the summer — if they're going to have any chance of competing for jobs and paychecks against foreign kids. He said "We can no longer afford an academic calendar designed when America was a nation of farmers who needed their children at home plowing the land at the end of each day. That calendar may have once made sense, but today, it puts us at a competitive disadvantage. Our children spend over a month less in school than children in South Korea. That is no way to prepare them for a 21st Century economy. The challenges of a new century demand more time in the classroom. If they can do that in South Korea, we can do it right here in the United States of America." He urged administrators to "rethink the school day" to add more class time. He proposed longer class hours as part of a broader effort to improve U.S. schools that he said are falling behind foreign competitors. "Despite resources that are unmatched anywhere in the world, we have let our grades slip, our schools crumble, our teacher quality fall short, and other nations outpace us," Obama said. "In 8th grade math, we've fallen to 9th place. Singapore's middle-schoolers outperform ours three to one. Just a third of our 13- and 14-year olds can read as well as they should."

The simple correlations and the simple regression relationship between classroom instructional time per week and test scores of pupils aged 15 in the 2006 PISA exams in math, science and reading are positive and highly significant. Of course, these correlations do not represent causal relationships because of potential selection and endogeneity. In this paper, I investigate the causal relationship between instructional time and pupils' knowledge in these subjects. While research in recent years provides convincing evidence about the effect of several inputs in the education production function[3], there is limited evidence on the effect of classroom instructional time. This evidence can be very important for policy because it is relatively simple to increase instructional time, provided resources are available, and there is much scope for such an increase in many countries. In the last section of the paper I also examine what factors can explain part of the variation across countries in the marginal productivity of classroom instructional time. I focus in this analysis on structural features such as accountability, autonomy and governance of schools.

I use two sources of data in this study. The first are the results of the PISA 2006. PISA is a triennial survey of the knowledge and skills of 15-year-olds. It is the product of collaboration between participating countries through the Organization for Economic Co-operation and Development (OECD), and draws on leading international expertise to develop valid comparisons across countries and cultures. More than 400,000 students from 57 countries constituting close to 90% of the world economy took part in PISA 2006. The study focused on science but the assessment also included reading and mathematics and collected data on student, family and institutional factors, including information about the amount of instructional time per week in each of the subjects tested. The second source of data is a testing and survey program of 5th and 8th grade pupils in Israel in 2002-2005 in math, science, English and Hebrew. These two sources of administrative data have two key features that I exploit to identify the effect of instructional time on academic achievement. Firstly, both data sets include test scores in multiple subjects

[3] For example, studies on class size (e.g. Angrist and Lavy, 1999, Kruger 1999, Hoxby 2000), teachers' training and certification (Angrist and Lavy, 2001, Kane Rockoff and Staiger, 2007), remedial education (Jacob and Lefgren, 2004, Lavy and Schlosser, 2005), teacher effect (Rockoff, 2004, Rivkin, Hanushek, and Kain, 2005), computer aided instruction (Angrist and Lavy. 2002, Barrow, Markman and Rouse, 2009), students' incentives (Angrist.J, P. Lang, and P. Oreopoulos, 2009, Angrist and Lavy, 2009), teachers' incentives (Lavy, 2009).

for each student, and there is relatively large variation in instructional time across subjects. This allows me to use within student estimation of the effect of instructional time while controlling for individual time invariant characteristics that equally affect performance across subjects such as the individual's underlying ability, parental and family background, lagged achievements and lagged and current school resources and characteristics. Secondly, there is considerable within student variation in instructional time. For example, among the OECD developed countries, the minimum mean classroom instructional time in math is 2.4 hours per week in the Netherlands and the maximum is 4.2 in Iceland. The respective figures in science are 1.8 in the Netherlands and 3.6 in New Zealand, while in reading they are 2.4 in Austria and 4.7 in Germany. I use this significant variation to test whether the effect of instructional time is non-linear and whether it differs among developed and developing countries. The disadvantage of this identification approach is that I assume that the effect of instructional time is the same for all three subjects. This assumption is common in many studies that pool cross sectional data across subjects. However, in this study I can assess how restrictive it is by comparing estimates obtained based on pooling only sub-groups or all three or four subjects together.

The use of the data from Israel provides evidence for different age groups than the PISA data, and it has the additional advantage of offering longitudinal data based on following pupils from fifth to 8th grade. Although this is possible only for a sub-sample of students (for whom I can link their records in 2002 and 2005), these data permit identification based on a student fixed effect due to a change in instructional time over time. Another advantage to using the Israeli data is that it permits estimation of the effect for each subject separately, based on within-pupil variation, while with the OECD data it is only possible to use the within-pupil variation by pooling together some or all subjects.

There are numerous studies about the effect of time spent in school on student achievement and earnings. For example, Grogger (1996), and Eide and Showalter (1998), estimated the effect of the length of the school year in the US and found insignificant effects, perhaps due to limited variation in this variable across schools or also due to correlated omitted variables. Rizzuto and Wachtel (1980), Card and Krueger (1992), and Betts and Johnson (1998) used State level data in the US to examine the same effect and found a positive significant effect on earnings, perhaps because they study earlier periods where there was more variation in length of the school year and because the effect of unobserved heterogeneity may also be less of an issue with state level data. Card and Krueger also present results controlling for state effects. The positive effect of year length vanishes within states and conditional on other school quality variables. Lee and Barro (2001) examine the effect of the amount of time spent in school during the year on student performance across countries while controlling for a variety of measures for school resources. They find no effects of the length of the school year on internationally comparable test scores. A more recent study by Wößmann (2003), which also analyzes cross-country test score data, corroborates this finding. He finds a significant effect of instructional time, but the size of the effect is negligible. However, these two studies attempt to identify the effect of instructional time on test scores by controlling for many characteristics and resources in each school and country. This method cannot rule out biases (due to school and country unobserved heterogeneity) that are correlated with instructional time and test scores. A more recent study, Pishke (2007), overcomes potential selection and endogeneity problems by using the variation introduced by the West-German short school years in 1966-67 as a natural experiment, which exposed some students to a total of about two thirds of a year less of schooling while enrolled. The study reports that the short school years increased grade repetition in primary school, and led to fewer students attending higher secondary school tracks. On the other hand, the short school years had no adverse effect on earnings and employment later in life.

The results I present in the paper show that instructional time has a positive and significant effect on the academic achievements of pupils. These results derived both from the 2006 PISA data of pupils aged 15 and the Israeli data of pupils in 5th and 8th grade. The size of the estimated effects is modest to large. On average an increase of one hour of instruction per week in math, science or reading raises the test score in these subjects by 0.15 of a standard deviation of the within student distribution of test scores. The size of the effect is larger for girls, for pupils from low socio economic status families and for immigrants. Estimates based on the sample of the former Soviet block eastern European countries are very similar to the average effect obtained from the sample of OECD developed countries. The evidence based on a sample of developing countries suggests a much lower effect of instructional time on test scores, on average one additional instructional hour improves test scores by 0.075 standard deviation of the within pupil test score distribution. In similarity to the OECD results, the effect is much larger for girls, for pupils from disadvantaged backgrounds and for immigrants. Overall, the main results presented in the paper are very robust to a variety of robustness checks with respect to the identification assumptions and to threats to their validity. The evidence from Israel add to the credibility of the results based on the PISA data as they yield very similar estimates of the effect of instructional time on pupil achievements. The estimates are consistent across primary and middle schools and across the various methods of identification and estimation.

In the second part of the paper, I investigate whether the estimated effect of instructional time varies by certain characteristics of the labor market for teachers and of the school environment. I use information from PISA 2006 about school accountability measures and the degree of school autonomy such as the role of schools in hiring and firing teachers and in determining wages of teachers. The main effects of these characteristics, which vary by school, are absorbed in the estimation by the school fixed effect but I am able to estimate the effect of their interactions with instructional time in each subject. The evidence suggest that the productivity of instructional time is higher in schools that operate under well defined accountability measures, and in schools that enjoy extensive autonomy in budgetary decisions and in hiring and firing teachers.

The rest of the paper is organized as follows: Section II describes the identification strategy. Section III discusses the data, the construction of the analysis samples, and presents various pieces of evidence that assess the validity of the identification strategy. Section IV reports the pupil cross section fixed effects estimates of the effect of instructional time in each subject using the three international samples of countries, while section V presents evidence based on Israeli data. Section VI shows results about the correlations of the average productivity of instructional time with schools and teachers' labor market characteristic. Section VII concludes.

II. Empirical Strategy

The effects of unobserved correlated factors usually confound the effect of instructional time on students' outcomes. Such correlations could result if self-selection and sorting of students across schools are affected by school resources or if there is a correlation between school instructional time and other characteristics of the school that may affect students' outcomes. One possible method to account for both sources of confounding factors in the estimation of instructional time is to rely on within-student variations in instructional time across various subjects of study. Based on this approach, I examine whether differences in a student between subjects are systematically associated with differences between subjects in instructional time. The basic idea for identification is that the student's characteristics, ability, and the school environment are the same for all three subjects except for the fact that some subjects have more instructional time than the other subjects do. Of course, it could be that at the school level, such variation is not purely random but the cause of such selection across schools is constant for each student in school and therefore does not vary within each student. Based on this approach I present within student estimates of the effect of instructional time on individual test scores using the following panel data specification,

$$A_{ijk} = \mu_i + \gamma H_{kj} + \beta X_{ij} + \delta S_j + (\varepsilon_j + \eta_k) + u_{ijk} \qquad (1)$$

Where A_{ijk} is the achievement of the i^{th} student, in the j^{th} school, in the k^{th} subject, H_{kj} is instructional time in the k^{th} subject in the j^{th} school, X is a vector of characteristics of the i^{th} student in the j^{th} school and S_j is a vector of characteristics of the j^{th} school. ε_j and η_k represent the unobserved characteristics of the school and the subject, respectively, and u_{ijk} is the remaining unobserved error term. The student fixed effect μ_i captures the individual's family background, underlying ability, motivation, and other constant non-cognitive skills. Note that by controlling for this individual fixed effect, using within-student across subjects' variation in test scores, I also control for the school fixed effect ε_j. Therefore, exploiting within-student variation allows for the controlling of a number of sources of potential biases related to unobserved characteristics of the school, the student or their interaction. Firstly, students might be placed or be sorted according to their ability across schools that provide more (less) instructional time in some subjects. If, for example, more able students attend better schools who provide more instructional hours overall in each subject, it would cause γ to be downward biased unless the effect of student and school fixed effects are accounted for. The bias will have an opposite sign if the less able students are exposed to more instructional time. Identification of the effect of instructional time based on a comparison of the performance of the same student in different subjects is therefore immune to biases due to omitted school level characteristics, such as resources, peer composition and so on, or to omitted individual background characteristics, such as parental schooling and income.[4]

I should make here three important remarks about this identification strategy. First, the necessary assumption for this identification strategy is that the effect of instructional time is the same for all subjects, implying that γ cannot vary by subject. Although this restriction is plausible, in the analysis that follows I will provide some evidence to support this conjecture. Second, the effect of instructional time is 'net' of instructional time spillovers across subjects, (e.g. instruction time in English might influence pupils' test scores in Mathematics). Third, the pupil fixed effect framework does not preclude the possibility that pupils select or are sorted across schools partly based on subject-specific instructional time. Stated differently, pupils who are high ability, for example, in math may select or be placed in a school that specialize in math and have more instructional time in math. This concern may be less relevant in the sample that I use for two reasons. First, such tacking is mostly within schools and I measure instructional time in each subject by the school level means and not by the class means or even the within school program level means. Secondly, the pupils in the sample are age 15 and therefore most are still in 9th grade. In most countries, 9th grade is part of middle school or lower secondary school while schools that specialize in a given subject are mostly upper secondary schools, from 10th grade on. Moreover, I am able to stratify the sample according to good proxies of weather the school sorts and

[4] Since the treatment variable, instructional time, is measured at the school level, the error term, u_{ijk}, is clustered by school to capture common unobservable shocks to students at the same school.

selects students based on subject specific considerations. For example, I observe in the PISA data information of whether the school considers for admission the student's academic record, weather it uses tracking in forming classes and weather it is a public or a private school. I assume that a school that do not use academic ability as criterion for admission, nor does it use any form of tracking by ability, and that it is public, will most likely not select students on subject specific considerations. Indeed, the results that I present below are very similar across the various stratified samples based on these school characteristics.

I also address the issue of subject specific selection based on the Israeli data. First, by first using data at the primary school where there is not at all any kind of sorting by subject specific pupil's ability or subject specific specialization. Second, by using panel data that allows to account for such sorting by controlling for lagged test scores in primary school (5th grade) of all subjects in the within-pupil estimation. The identifying assumption here is that the lagged test scores in each subject effectively capture any unobserved heterogeneity that lead to sorting into school according to subject specific considerations such as expected school hours of instruction in a given subject. I can control for lagged test scores in a very flexible way by including in the specification at the same time *same*-subject lagged test scores (e.g. looking at 8th grade English test score for pupil *i* controlling for his/her 5th grade English achievement), as well as *cross*-subject test scores (e.g. looking at pupil *i*'s 8th grade English test score controlling for his/her 5th grade test score in Mathematics). Additionally, I can interact lagged test scores with subject-specific dummies, so that 5th grade achievements can exhibit different effects on 8th grade outcomes in different subjects. The specific specification that I use in this context is presented in section IV.

III. Data

PISA is an acronym for the "Program for International Student Assessment". It provides regular data on the knowledge and skills of OECD country students and education systems. The first survey was in 2000, the second in 2003 and the third in 2006. More than 60 countries have taken part in PISA so far and it is the only international education survey to measure the knowledge and skills of 15-year-olds, an age at which students in most countries are nearing the end of their compulsory time in school. Rather than examine mastery of specific school curricula, PISA looks at students' ability to apply knowledge and skills in key subject areas and to analyze, reason and communicate effectively as they examine, interpret and solve problems. PISA measures student performance in reading, mathematics and science literacy and asks students about their motivations, beliefs about themselves and learning strategies. All OECD member countries participated in the first three PISA surveys, along with certain partner countries. In total, 43 countries took part in PISA 2000, 41 in PISA 2003 and 58 in PISA 2006. Countries who are interested in participating in PISA contact the OECD Secretariat. The PISA Governing Board then approves membership according to certain criteria. Participating countries must have the technical expertise necessary to administer an international assessment and must be able to meet the full costs of participation. To take part in a cycle of PISA, countries must join two years before the survey takes place.

Each OECD country participating in PISA has a representative on the PISA Governing Board, appointed by the country's education ministry. Guided by the OECD's education objectives, the Board determines the policy priorities for PISA and makes sure that these are respected during the implementation of each PISA survey. For each survey, an international contractor (usually made up of testing and assessment agencies) has been responsible for the survey design and implementation. Working with the OECD Secretariat, the PISA Governing Board and the international contractor, the PISA National Project Managers oversee the implementation of PISA in each participating country. PISA has Subject Matter Expert Groups for its three key areas of testing – reading, mathematics and science literacy – as well as for other subjects when appropriate (problem solving in PISA 2003, for example). These groups include world experts in each area. They design the theoretical framework for each PISA survey.

The international contractor randomly selects schools in each country. The tests are administered to students who are between 15 years 3 months and 16 years 2 months of age at the time of the test, rather than to students in a specific year of school. This average age of 15 was chosen because at this age young people in most OECD countries are nearing the end of compulsory education. The selection of schools aims to be representative of the respective population of schools and students. To date, PISA has used pencil-and-paper tests. The tests are made up of both multiple-choice questions and questions requiring students to construct their own responses. The material is organized around texts and sometimes includes pictures, graphs or tables setting out real-life situations. Each PISA survey includes about seven hours of test material. From this, each student takes a two-hour test, with the actual combination of test materials different for every student. All PISA countries are invited to submit questions to the international contractor; in addition, the international contractor also writes some questions. The questions are reviewed by the international contractor and by participating countries and are carefully checked for cultural bias. Only those questions that are unanimously approved are used in PISA.

Students answer a background questionnaire, providing information about themselves, their attitudes to learning and their homes. It takes 20-30 minutes to complete. In addition, school principals are given a 20-minute questionnaire about their schools.

Each country has its own group of test markers, overseen by the country's National Project Manager. They mark the PISA tests using a guide developed by the international contractor and the PISA Subject Experts (with input from all participating countries). Other experts crosscheck the corrections. The results are then sent to the international contractor, who in turn transmits the final data to the OECD Secretariat. The average score among OECD countries is 500 points and the standard deviation is 100 points. The results from PISA can be compared across the surveys, as can some of the background questionnaire items.

Table 1 reports the distribution of instructional time in each of the three international samples of countries in the 2006 PISA based on the pupil level data. Each pupil replied to a question about hours of instruction in each subject: less than two hours, between two and three hours, between four and five hours and over six hours. I computed the school average in each subject using the mid values of each range. The means of instructional time in the developed OECD countries in math, science and reading are 3.53, 3.06 and 3.54 hours, respectively. In the Eastern European's sample mean instructional time in math are 3.30, in science 2.77 and in reading 3.08, lower than in the OECD countries in all subjects. Mean instructional time in the sample of developing countries is similar to the means in the Eastern Europe sample, math - 3.48, science - 2.97 and reading - 3.24. Surprisingly, the respective means of instructional time in four new-industrialized East Asian countries, Macao-China, Korea, Hong Kong and Chinese Taipei, are much higher, and they are similar to the means of the developing countries sample.

Appendix Tables A1-A3 present the mean instructional time in each of the subjects for each of the countries included in the three samples of the 2006 PISA. These tables present as well the country means of the test scores in each of the subjects for each of the countries. In Table A4, I present the means of the test score and the instruction time variables for all three samples of countries. The average test score in the developed OECD countries is 513.4, the standard deviation in test scores between pupils is 84.4, and most relevant for our analysis, the within student standard deviation in test scores is almost half as large, 38.8. In short, there is considerable variation in test scores within the same pupil to explain. The average instructional time per subject in the OECD sample is 3.38 hours, and the within pupil standard deviation in instructional time is 1.02 — comparable in magnitude to the standard deviation in instructional time between students, 1.08. The rest of Table A4 presents the evidence for the Eastern Europe and developing countries samples. No dramatic differences are observed in the within and between pupil standard deviations of these two samples in comparison to the OECD sample.

IV. Results

A. Estimates of the effects of instructional time in OECD countries

Table 2 reports the estimated coefficients of instructional time from subject specific test scores regressions based on the sample of the OECD countries. For each subject I report estimates from three specifications: firstly without any controls, secondly with country fixed effects and thirdly with country fixed effects and pupil characteristics. The first row presents the OLS estimates when instructional time is measured in hours per week. The following two rows report estimates when three indicators, one for each for the following groups, measure instructional time: less than 2 hours per week, 2-3 hours per week, and 4+ hours. The first group indicator is the omitted group in the regression.

The estimated effects of instructional time presented in Table 2 are all positive, very large, always significantly different from zero and not dramatically sensitive to the addition of controls to the regression. For example, the estimate for total instructional time in mathematics is 21.69 with no controls, 27.98 with country fixed effects and 24.45 with the addition of student's controls. In science, the respective estimates are about 25-30 percent higher than in math. The reading estimates are much lower than in math and science. The estimates of the instructional time' indicators presented in panel II show that the largest marginal effect of one additional hour of instruction is when classroom hours are increased from less than 2 hours to 2-3 hours. In math for example, such a change is associated with an increase in test scores of about 0.5 of the standard deviation (s.d) of between pupils test score distribution and more than a standard deviation of the within pupils test score distribution. We will see below that the OLS estimates are highly biased upward.

Column 1 in Table 3 presents estimates from regressions based on a pooled sample of all three subjects (with subject fixed effects included as controls) while column 2 presents estimates when student fixed effects are included. The OLS estimates in column 1 are very similar to the estimates presented in Table 2. The within-student estimates in column 2 are all positive and much smaller than the OLS estimates in column 1 but they are still very precisely measured. Assuming a constant linear effect of instructional time, the effect of one additional hour of classroom instruction in the within student regression is 5.76 points. The effect amounts to 0.15 of the standard deviation within pupil and 0.07 standard deviation of the between pupil test score distribution. However, the more relevant scale for the effect size is the within pupil standard deviation as this is the variation that we use to estimate the effect of instructional time in the within pupil regression.

The estimates of the instructional time' indicators suggest some non-linearity in the effect of instructional time, with a larger effect in the range of 1-2 hours than at higher levels. The marginal effect of an hour in the 2-3 hour range is 4.20 [= (6.3 points/1.5 hours)], while in the range of 4+ hours the effect is only 2.48 [= (12.42 points/5 hours)], both of which are lower than the average effect of 5.76 which suggests that the first two hours of instruction have the highest effect.[5]

The productivity of classroom-hours might be different for different subjects. In order to check for such variation I estimate models based on the three possible samples that include only two of three subjects. The lower panel of Table 3 presents estimates based on the sample that pools the math and science test scores. The estimated effects of classroom-hours obtained from this sample is higher, 7.14, about 24% higher than the respective estimate obtained from pooling all three subjects together. However, pooling math and reading test scores yields an estimate of 7.42 and pooling science and reading yields an estimate of 4.27. This pattern does not permit me to conclude in which of the three subjects there is lower average productivity of instructional time in the OECD countries. However, the average (6.28) of the three estimates obtained from three samples that include only two of the three subjects is very close to the estimate (5.76) obtained by pooling all three subjects.

B. Robustness of main results

In this section, I present a set of robustness checks and alternative specifications that support the causal interpretation of the findings reported in column 2 of Table 3. Since the variation in hours of instruction is at the school level, the first check of robustness is estimates based on a sample of schools instead of pupils. I present these results in appendix Table A5. I obtain the variables that I use in this estimation by collapsing the pupils' data to the school level respective means. The pattern of estimates in this table is very similar to those presented in column 2 of Table 3. The OLS estimates in the two tables are practically identical while the school fixed effect estimates based on the school level sample are slightly lower than the estimates based on students micro data. The estimate based on all three subjects is

[5] For the range of 4+, it is impossible to compute the exact effect per hour because it is an open-ended range and the mean is not known, therefore, I assume arbitrarily a mean of 5 hours.

lower by 17 percent than the respective estimate in Table 3, the estimate based on math, and science only is lower by 9 percent than the respective estimate in Table 3.

Next I also examined how sensitive the treatment estimates are to including interactions between the subject dummies and pupil's characteristics. The treatments estimates from this more flexible specification (not shown here and available from the author) are very similar to those presented in Table 3 though overall they are about 10 percent lower.

The first robustness check of whether the evidence in column 2 Table 3 is reflecting some subject specific selection and sorting in some schools is based on the data available from the PISA school questionnaire question 19 about how much consideration is given in the admission decisions to student's academic record and whether placements tests are used in this process. I expect that the validity of the identification strategy not be sensitive to endogenous sorting and selection in a sample of schools that do not pay any attention to previous academic records of its applicants and that do not use any admission exams. In columns 3-4 of Table 3, I report results from such a sample of schools and in columns 5-6, I report estimates based on a sample of schools that consider student's academic record for admission. The sample of students in schools where past academic achievements are irrelevant for admission is the largest and it includes about two thirds of the whole sample. The estimates from this sample are only marginally different from those obtained from full sample: the OLS estimates is lower, 16.97 versus 19.58, and pupil fixed effect estimate is higher, 6.008 versus 5.76. The OLS and the pupil fixed effect estimates in columns 5-6 are also only marginally higher than the estimates obtained from the full sample and reported in columns 1-2 of Table 3. The estimate based on the sample of schools that admit pupils based on their academic record yields lower estimates, but they are not statistically different from the respective estimates presented in any of the columns of Table 3.

Another potential source of selection bias is tracking pupils to classes within schools according to their ability because one can expect that schools that practice such tracking will also tend to select and admit pupils based on subject specific strengths. If the strengths or specializations of schools are correlated with hours of instruction in different subjects, it will lead to a bias in the estimated effect of hours of instruction. In Table 4, I present results for three different samples distinguished by schools' tracking policy. Columns 1-2 present estimates for a sample of schools that practice tracking at the class level, namely they group their students in classes according to their ability. In columns 3-4, I report results based on a sample of schools that track pupils to different ability study groups within classes. In columns 5-6 I report results based on a sample of schools that do not practice any form of pupil's tacking. The OLS and the pupil fixed effect estimates in the first row in columns 1-2 are quite similar to the respective estimates presented in columns 3-4. They are marginally higher than the estimates obtained from the full sample and reported in columns 1-2 of Table 3. The sample of schools that track pupils into different classes by ability yields estimates that are higher by 15 percent than the respective estimates in Table 3 and the sample of schools that practice within class tracking yields estimates that are higher by 7 percent of the estimates in Table 3. However, in both cases these estimates are not different significantly from the point estimates obtained from the full sample as the confidence intervals of the latter estimates include the point estimates obtained from each of these samples. Finally, the effect of instruction hours on test scores in schools that practice no tracking at all is 5.17, not significantly different from the estimate from the full sample (5.76) but significantly lower than the estimates obtained for schools which practice tracking between classes.

Another potential source of bias can originate from the inclusion of private schools in the PISA sample. For example, 18 percents of the schools in the OECD sample are classified as private. This could be of concern because admission based on previous academic record and on additional exams as well as tracking pupils to study groups by ability is much more prevalent among private schools. To assess these concerns I therefore estimated the effect of instruction hours based on a sample that included only the public schools in the PISA sample. The estimated effect of instruction school hours based on pooling together the math, science and English test scores is 6.09 (sd=0.428), just barely higher than the estimate from a sample that included also the private schools. The estimate based on just math and science is 7.501 (sd=0.643), only marginally higher than the 7.14 (sd=0.55) obtained from the full sample.

Overall, based on the evidence presented in Table 3 and 4 and the results from a sample that includes only public schools, it is apparent that potential selection and sorting of students based on subject specific considerations related to selective admission or tracking pupils in classes by abilities is not driving the results. This is an important result because it is expected that schools that admit pupils based on academic record or that track students by ability will also tend to select and admit pupils based on subject specific strengths. If the strength or specialization of schools is also correlated with hours of instruction in different subjects, it might bias the estimates of the effect of hours of instruction. The lack of any large discernable differences in the effect of hours of instruction by admission or tracking policies of schools suggests that unobservables that are correlated with sorting or selection of pupils based on subject specific hours of instruction consideration are not biasing our estimates. Table 5 provides further evidence that support this conclusion. First, I add to the regressions as control variables indicators of whether the school offers a special study program in science or math which may attract students with special interest and ability in science and math. The first set of controls is based on question 20 in the PISA school questionnaire. It consists of indicators for school activities that promote engagement with science among students (science clubs, science fairs, science competitions, extracurricular science projects and excursions and field trips). The second set is based on question 22 in the PISA school questionnaire. It consists of indicators of school programs such as trips to museums, trips to science and technology centers, and extracurricular environmental projects and lectures and seminars with guest speakers, all of which provide opportunities for students to learn about science and environmental topics. The motivation for including these control variables is that they most likely will eliminate a potential bias in the estimated effect of hours of instruction due to selection or sorting of students to schools based on special abilities and interest in science and math. These results are presented in column 1 and 2 of Table 5. Note that the OLS (column 1) and fixed effects (column 2) estimates, are almost identical to the respective estimates presented in columns 1-2 of Table 3, suggesting that the fact that many schools offer special programs and activities in science and math a are not source of concern that our estimates are biased due to subject specific sorting and selection.

Another robustness check of our evidence is based on the data available in PISA (school questionnaire question 14) about lack of qualified teachers for each of the following subjects: science, mathematics, language, and other subjects. I have added a control variable for whether the school's capacity to provide instruction in a given subject is hindered by a lack of qualified teachers in that subject. The rationale for adding this control is that schools that specialize and have a particular strength in a given subject will be less likely to have difficulties in hiring qualified teachers. The OLS and pupil fixed effect estimates are presented in columns 3 and 4, respectively, of Table 5 and they are almost identical to those presented in columns 1-2 of Table 3. I also estimate the various models in various samples stratified by the extent of lack of qualified teachers, for example including only schools that report lack of qualified teachers in at least two subjects or a sample that includes only schools without lack of qualified teachers in any subject. The results obtained from these samples are practically identical.

C. Heterogeneous treatment effects

To gain further insights into the effect of instructional time, I explore heterogeneous effects of classroom hours for different sub-samples. In Table 6, I report separate estimates for boys and for girls. The estimates show a positive impact of instructional time for both genders but the effect is marginally higher (by 13%) for girls than for boys but this difference is not significantly different from zero. A somewhat lower gender difference in the effect of instructional time is observed when only math and science test scores are pooled jointly in the estimation. This pattern may suggest that the gender related difference in marginal productivity of instructional time is due to the marginally lower effect of reading classes on boys reading proficiency than on girls. The pattern of the non-linear effect further suggests that most of this gender difference comes from the higher effect of 4+ hours on girls than on boys.

In Table 7, I report results for two sub-samples stratified by the average years of schooling of both parents and for two sub-samples of immigrants, first generation and second generation. The productivity of instructional time is clearly higher (35%) for pupils from low education families.

However, again this differential productivity does not exist when only math and science are used in the estimation, which suggests that pupils from low education families benefit significantly more than other pupils from additional classroom instruction in reading.

Finally, an interesting pattern is seen in the estimated effect of instructional time for immigrants. Firstly, the estimates are marginally higher for first generation immigrants but they are much higher (30%) for second-generation immigrants. Secondly, these differences are even larger when these estimates are based on pooling in the estimation only math and science test scores. Instruction time in school is 69 percent more productive for second-generation immigrants in comparison to natives (an estimate of 11.99 versus 7.11). This suggests that the relative gain for an hour of instruction in reading is much lower for second-generation immigrants than for natives.

D. Evidence from middle and low income countries

The first row in Table 8 presents evidence based on a sample of the following middle-income countries, all former Soviet block: Bulgaria, Czech Republic, Estonia, Croatia, Hungary, Lithuania, Latvia, Montenegro, Poland, Romania, Russian Federation, Serbia, Slovak Republic and Slovenia. The mean test scores of the three subjects in this sample are all lower than the respective means of the OECD developed countries: math - 472.4, science - 480.4, reading - 458.3 (see Table A2). The standard deviations in the pupil level distribution of test scores are similar to those in the OECD sample, 97.8 in math, 97.9 in science and 105.0 in reading.

The OLS estimates of the effect of instructional time are much higher in this sample than in OECD developed countries. The OLS estimate of the continuous hours of instruction variable is 38.2 versus 19.58 in the OECD sample. However, the within pupil estimate is 6.07, almost identical to the respective OECD estimate. This suggests that the selection or endogeneity in school resources in the Eastern European countries are much more important.

The estimate for girls is again higher (26%) than for boys and it is much higher (by 33%) for pupils from low education families. The higher effect of hours of instruction on second-generation immigrants is again evident as in the OECD sample.

The lower panel in Table 8 presents estimates based on a sample of developing countries (Argentina, Azerbaijan, Brazil, Chile, Colombia, Indonesia, Jordan, Kyrgyzstan, Mexico, Thailand, and Tunisia).[6] These four countries are among the best performing countries among all participants in PISA 2006 and their mean instructional time in all three subjects are also among the highest in the overall sample. The mean test scores in this sample of developing countries are 21% lower than in the OECD countries: math – 398.5, science - 403.4, reading – 397.1. The standard deviation in the pupil level distribution of test scores is around 100 in the three subjects.

The estimates show a much lower productivity of instructional time than the estimates of the OECD or the middle-income Eastern European countries. The effect of a change of one classroom hour is only 2.99 points which is equal to 0.06 of the within pupil standard deviation and 0.04 of the between standard deviation. This effect size is about half the effect size estimated for the OECD developed economies and for the Eastern Europe sample. The gap is even larger based on a comparison of the estimates derived from pooling only math and science test scores. The largest difference in terms of the non-linear specification of instructional hours between the two groups of countries is in the effect of changing from less than two hours to 2-3 hours of instruction per week.

Overall, instructional time in the sample of the developing countries is much more effective in improving test scores of girls (38% higher than for boys) and of immigrants. However, in this sample the effect of instructional time is lower by 26% for pupils from low education families than for pupils from educated families.

The results from the samples of rich and poor countries can be used to compute what proportion of the gap in knowledge between these sets of countries can be explained or eliminated by bridging the

[6] I do not include in this sample the new industrialized countries of Korea, Honk Kong, Macau and Chinese Taipei because their income per capita is much higher than the developing countries and in the PISA classification they are not included in the sample of the developing countries.

gap in instructional time and in its productivity in the different subjects. The mean instructional time in math, science and reading in the rich countries are 3.5, 3.1 and 3.5 while in the poor countries they are 3.5, 3.0 and 3.2. The gap in instructional time is relatively small, 0% in math, 14% in science and 9% in reading. The mean test scores in the developing countries sample are much lower: 398.5 in math, 403.4 in science, and 397.1 in reading. Therefore, the gap in mean test scores between the developing and the OECD developed countries is very large, over 20% and its size is about one standard deviation in each of the subjects. Obviously, equalizing the instructional time in the poor countries to the level in rich countries will not significantly eliminate the test score gap between these two parts of the world. However, the poor countries can reduce this gap by raising the marginal productivity of instructional time to the level in rich countries. The average instructional time in the three subjects in the developing countries sample is 3.2. Converging to the productivity of instructional time in the OECD countries will therefore raise achievements in each of the three subjects by 0.10 of a standard deviation. In section V, I will explore what structural changes in the education system in developing countries can lead to convergence of the productivity of instructional time to the level in the OECD countries.

E. Evidence from Primary and Middle Schools in Israel

Using Within Student between Subject Variation in Instructional time

Data for elementary and middle schools is based on the GEMS (Growth and Effectiveness Measures for Schools - *Meizav* in Hebrew) datasets for the years 2002-2005. The GEMS includes a series of tests and questionnaires administered by the Division of Evaluation and Measurement of the Ministry of Education.[7] The GEMS is administered at the midterm of each school year to a representative 1-in-2 sample of all elementary and middle schools in Israel, so that each school participates in GEMS once every two years. The GEMS student data include test scores of fifth and eighth graders in math, science, Hebrew, and English. In principle, all students except those in special education classes are

[7] The GEMS are not administered for school accountability purposes and only aggregated results at the district level are published. For more information on the GEMS see the Division of Evaluation and Measurement website (in Hebrew): http://cms.education.gov.il/educationcms/units/rama/odotrama/odot.htm.

tested and the proportion of students who are tested is above 90 percent. The raw test scores used a 1-to-100 scale that I transform into z-scores to facilitate interpretation of the results.

The test scores for the years 2002-2005 are linked to student administrative records collected by the Israel Ministry of Education. The administrative records include student demographics that I use to construct all measures of students' background characteristics. Using the linked datasets, I build a panel for elementary schools and a panel for middle schools. I drop any schools with an annual enrollment lower than 10 students from the panel. The elementary school panel includes data for 5th grade student test scores for the years 2002-2005. The sample is restricted to Jewish public schools that have mixed-gender classes. There are 939 elementary schools with test score data. As every school is sampled once in two years, we have two observations of the same school and grade for more than 90 percent of the schools. The middle school panel includes 8th-grade student test scores for the years 2002-2005. The sample is restricted to Jewish middle schools. There are 475 schools in the sample, of which 85 percent appear in two years. As there are multiple years for each school, I pool all years and exploit within student variation in instructional time across years.

The GEMS also includes interviews with all teachers and the school principal. The questionnaire for home teachers of all classes included questions about classroom instruction time in each subject and the total per week. I use teachers' responses to these items to compute the school average for 5th and 8th grade instructional time in each subject. The mean per grade is preferred over the class level measure to avoid selection due to within school and grade endogenous allocation of instructional time to various subjects. However, the mean at the grade is very highly correlated with the class level figure for classroom-hours of instruction.

Table 9 presents the estimates for instructional time in fifth grade (columns 1-3) and eighth grade (columns 4-6). Three different specifications are used. The first includes only year fixed effects, the second adds pupil demographic controls and the third adds school fixed effects. All the 5th grade estimates are positive and most are significantly different from zero. Some of the 8th grade estimates are negative and many are not significantly different from zero.

Table 10 presents estimates of the effect of instructional time from a sample that pools all or sub sets of the three subjects, for 5^{th} and 8^{th} grades. The first row of the table presents the OLS estimates (with controls for year effects, student's characteristics and school fixed effects) and the third row presents the estimates based on student fixed effects. The estimate based on within student variation across all three subjects (column 4) is 0.058 for fifth and 0.029 for 8^{th} grade (column 8). The OLS estimates are larger for both grades but these differences are not nearly as large as we saw in the PISA sample, most likely because there is not much selection in allocation of instruction time to primary and middle schools in Israel.

The other columns in the Table 10 present estimates based on samples that pool at a time only two of the subjects. In 5^{th} grade, all of these estimates are about equal to the estimate based on a sample that pooled test scores of all three subjects. The estimates from the 8^{th} grade sample are smaller, being similar in the two set of pairs (math and science and math and English) but much lower in the third pair science and English.[8]

In Table 10 I also report estimates based on sub groups, first for male and female, and secondly for pupils from low and high education families. Unlike the evidence from the PISA sample, there is no systematic pattern of differences in the estimated effect of instructional time between boys and girls. Also unlike the PISA OECD estimates, in the Israeli results there is some evidence that pupils from higher education families have a higher productivity of school instructional time.

B. Using Pupil's Longitudinal Data and within Subject Variation in Instructional Time

The structure of the GEMS allows me to follow a sample of students from elementary schools at 5^{th} grade in 2002 to middle schools at 8^{th} grade in 2005).[9] I take advantage of this feature and construct a longitudinal dataset at the student level to examine how changes in students' achievement in the three

[8] The lower productivity of instruction time in middle school in Israel is consistent with the view that this part of the schooling cycle is the weakest link in the school system and there are discussions and recommendations to abolish it and make the 6-8 grades an integral part of secondary schools.

[9] I link only a fourth of the students because except the large cities almost all other localities were sampled once every two years.

subjects are associated with changes in their instructional time (due to their transition from elementary school to middle school). I first estimate the following first difference equation by differencing out two relationships like equation (1) for each student (one for middle school and one for elementary school):

$$A_{ijmk} - A_{ijpk} = \mu_i + \gamma\,(H_{ijmk} - H_{ijpk}) + \beta X_{ij} + \delta S_j + (\varepsilon_j + \eta_k) + u_{ijk} \qquad (2)$$

where *p* denotes primary school and *m* denotes middle school. A student fixed effect is differenced out from this equation. However, I attempted specifications that included as controls the students' background characteristics, the average characteristics of their cohort in elementary and middle school, a grade fixed effect, a fixed effect for all students who attended the same primary school and a fixed effect for all students who attend the same middle school.[10] I therefore base the identification on contrasting the change in hours of instruction in each subject across elementary and middle school, and within students.

Table 11 presents the longitudinal student fixed effect estimates of the effect of instructional time on test scores, for each subject separately and for the pool of all test scores. The estimated effect of instructional time is positive and significant in math, science and it is positive in English as well but it is not precisely measured. The highest estimate is in math and the lowest in English. The estimated effect of instructional time obtained from a sample of all subjects together is 0.036, which is larger than the estimate that is reported in Table 10 (0.029) for 8th grade and smaller than the estimate for 5th grade (0.058) reported in Table 9. It is also higher than the average of the two, which is 0.043 and also lower than the estimate that I obtained from the sample of all the developed OECD countries but we have to note that the later estimate is for ninth and tenth grade students.

As an alternative to the difference specification (equation 2), I also estimated the following value added model with a very flexible specification:

$$A_{ijmk} = \mu_i + \gamma\, H_{ijmk} + \beta X_{ij} + \delta S_j + \lambda\, a_{iq} + \theta\, a_{iq(1)} + \sigma\, a_{iq(2)} + u_{ijk} \qquad (3)$$

where now a_{iq} represents *same*-subject test score in 5th grade, $a_{iq(1}$ and $a_{iq(2)}$ are the two cross-subjects test scores in 5th grade, and λ, θ and σ are (vectors of) subject-specific parameters that capture the effects of

[10] The results are virtually identical when these controls are omitted from the regression. They are also qualitatively unchanged when I simply include a separate fixed effect for each primary and each middle school.

5th grade test scores in the same- and cross-subjects. The parameter estimates of the effect of hours of instruction on test score in 8th grade are very similar to those reported above and therefore are not reported here.

VI. Correlates of Productivity Differences of Instructional Time across Countries

The productivity of instructional time is endogenous and a variety of factors can affect it. For example, the quantity and quality of other school inputs, teachers' education and training, class size, computers, science labs and so on. All of these inputs might interact with learning hours and shape the productivity of instructional time in school. Similarly, various structural features of the education system may affect the productivity of instructional time by affecting teachers and school principals' effort and efficiency. For example, accountability measures, such as publishing school league tables based on national tests or using pupils' performance measures to determine school staff compensation. Another relevant structural characteristic of the education system is the degree of autonomy that schools have in hiring and dismissing teachers. We can presume that more flexibility in staffing decision might lead to a better match between teachers and schools and create an environment that induces more effort and responsibility among school staff. The survey of school head masters in PISA 2006 provides information on a few aspects and characteristics of the education system of the dimensions discussed above. In this section, I use several indices or indicators of these characteristics that PISA 2006 produced in a comparable manner for all the countries in the sample. I use here the OECD sample because it is the largest in terms of number of countries and schools in the sample and because it exhibit relatively large variation in structure and characteristics of schools.

The first set of characteristics includes three binary indicators of school accountability measures: whether achievements data are posted publicly, whether achievements data are used in evaluation of school principal performance, and whether achievements data are used in evaluation of teachers' performance. Next is a PISA index that ranks the school's quality of educational resources, which is based on teachers' qualifications, class size and the quality of other school inputs. Two additional indices measure the degree of school autonomy. The first measures the school's autonomy in resource allocation: hiring and firing teachers, determining teachers' starting and change in salaries, determining and allocating the budget. The second index measures the school's responsibility for curriculum and assessment: school independence in deciding on the courses offered and their content, textbook used, and method of assessing pupils' performance.

Because these indices are the same in each school for all subjects, their main effect cannot be included as covariates in a regression that includes a school fixed effect. However, the interactions of these indices with instructional time can be included in the within pupil regression of achievement. Note that the pupil fixed effect absorbs the school fixed effect and therefore it also controls for any school level factor that is correlated with or determines these indices. In other words, even if the distribution of these indices across schools is not random, the school fixed effect will control for such heterogeneity. Therefore, the identifying assumption for the effect of the interaction between the indices and the hours of instruction is that the heterogeneity in these indices across schools is not subject specific.

Table 12 presents the estimated coefficients from these regressions. The first column presents the means of the indicators or indices. Accountability is not widespread among OECD countries as only 33.5 percent of the schools post their mean achievement publicly, and even fewer use them to evaluate school principals (22%) or teachers (29%). The means of the other indices are less interpretable.

In column 2 and 3 of Table 12, I present the estimates of the main effect of instructional hours and the estimates of the interaction of instructional hours and each of the school level indices. I include the interactions one at a time so each pair of estimates comes from a different regression. The estimated main effect of instruction hours is always positive and significant and it does not vary very much across the different regressions and from the estimate presented in Table 3. Three of the six estimated effects of the interaction terms are significantly different from zero. The same three remain significantly different from zero and their point estimate did not change much when I included all the interactions simultaneously in the regression. These results, shown in column 4 of Table 12, suggest the multi-co-linearity among the various indices does not prevent the estimation of the unique effect of each index.

The overall pattern is that the productivity of instructional time is higher in schools that implement school accountability measures, and in schools that have a degree of independence in allocating their resources. The index of quality of educational resources has a positive coefficient but it is not precisely measured. On the other hand, school flexibility in determining its curriculum and pupils' assessment measures do not have a significant effect on the productivity of instruction hours. Note that this index has no significant effect even when entered the regression as the sole interaction with hours of instruction. But I should emphasize that the main effect on pupils' achievement of school pedagogical autonomy, can still be positive even though it does not vary with hours of instruction across the three subjects measured in PISA.

The main effect of instructional time in the regression when all indices are included simultaneously is 4.676. In schools that post the achievements of their students publicly, this estimate is 6.64, over 40 percent higher. A similar large effect is evident in schools that evaluate school principals according to their students' performance though no such effect is evident in schools that similarly evaluate their teachers. However, the 2006 PISA questionnaire data does not provide enough details to allow an understanding of how exactly such an evaluation is done and whether it is used to reward school staff or affect their wages so we should be cautious in interpreting these results.

Another interesting feature of the school structure in PISA 2006 is governance, in particular the role of the school governing board. Four questions allow the measurement of the role of the governing board in influencing staffing, the budget, and instructional content and assessment. Adding to the regression interactions terms between these four measures (indicators) and instructional hours did not change at all the point estimates of the already included interaction terms. However, the pattern of the estimates of these new interaction terms is interesting since it is consistent with the evidence of the other interaction terms. First, having a board that affect staffing and the budget leads to higher productivity of instructional time. Second, having a board that influence instructional content and assessment has no measurable effect on the productivity of instruction in school. This evidence, which is presented in column 5 of Table 12, strengthens the overall findings that school autonomy in personal and budgetary issues is conducive to enhance pupils' learning and achievement while there is no parallel evidence with respect to school pedagogic autonomy.

VII. Conclusions

In this paper, I measure empirically the effects of instructional time on students' academic achievement. The evidence from a sample of 15 year olds from over fifty countries and from a sample of 10 and 13 year olds in Israel consistently show that instructional time has a positive and significant effect on test scores. The OLS results are highly biased upward but the within student estimates are very similar across groups of developed and middle-income countries and age groups. The effect of instructional time can be considered moderate or even large relative to other school level interventions for which we have reliable evidence. In the OECD sample, one additional hour of instruction increases on average test scores by about 0.15 of the within pupil standard deviation in test scores and by about 0.07 standard deviation of the between pupil standard deviation. Of course, a judgment on the merit of enhancing instructional time should also take into account the cost of adding instructional time relative to the cost of increasing the level of other inputs or of other interventions.

The estimated effect of instructional time is much lower in the sample of developing countries that participated in PISA 2006. The estimated effect of instructional time in this sample is only half of the effect size in the developed countries. The developing countries included in the PISA sample, for example Chile, Argentina or Thailand, are much more developed than the 'typical' developing country. Given the recent evidence from India, Kenya and other very poor developing countries about the high rate of absenteeism of teachers from work, we can expect that the productivity of instructional time in the poorest developing counties in Africa and in South East Asia is even lower than in our sample of developing countries. In these countries, we can expect to have much more scope for improvement by closing the gap in productivity of instructional time relative to the OECD.

The significant association between structural characteristics of the education system and the work environment of teachers in OECD countries and the average productivity of instructional time points to directions of how productivity can be improved in developed and in poorer countries.

VIII. References

Angrist, J. and V. Lavy (1999). “Using Maimonides’ Rule to Estimate the Effect of Class Size on Scholastic Achievement”, *Quarterly Journal of Economics*, 114: 533-75.

Angrist, J. and V. Lavy (2001). "The Effect of Teachers' Training on Student Achievements.”) *Journal of Labor Economics*, volume 19, no. 2, pp. 343-369.

Angrist, J. and V. Lavy (2002). “New Evidence on Classroom Computers and Pupil Learning”, *The Economic Journal*,. Volume 112, pp: 735-765.

Angrist, J. and V. Lavy, (2009) “The Effect of High-Stakes High School Achievement Awards: Evidence from a Randomized Trial”, *American Economic Review,* Volume 99, No. 1384-1414.

Angrist, J. P. Lang, and P. Oreopoulos, “Incentives and Services for College Achievement: Evidence from a Randomized Trial”, *American Economic Journal: Applied Economics, 2009*.

Barrow, L. L. Markman and C. E. Rouse, “Technology's Edge: The Educational Benefits of Computer-Aided Instruction”, *American Economic Journal: Economic Policy*, Vol. 1, No. 1, February, 52–74. 2009.

Betts, J. R. and Johnson, E. (1998). “A Test of Diminishing Returns to School Spending”, mimeographed, University of California San Diego.

Card, D. and Krueger, A. (1992). “Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States.”, *Journal of Political Economy*, vol. 100, pp. 1-40.

Eide, E. and Showalter, M.H. (1998). “The Effect of School Quality on Student Performance: A Quantile Regression Approach,” *Economics Letters*, vol. 58, pp. 345-50.

Grogger, J. (1996). “Does School Quality Explain the Recent Black/White Wage Trend?” *Journal of Labor Economics*, vol. 14, pp. 231-53.

Hanushek. E. A (2003) “The Failure of Input Based Schooling Policies”, Economic Journal 113: F64-98.

Jacob, B. and Lefgren L. (2004). “Remedial education and student achievement: A regression-discontinuity analysis”, *Review of Economics and Statistics*, vol. 86, pp. 226-44.

Kane.T. J, J. E. Rockoff and D. O. Staiger (2007) What Does Certification Tell Us About Teacher Effectiveness? Evidence From New York City, NBER Working Paper 12155.

Krueger, A. (1999) “Experimental Estimates of Education Production Functions”, *Quarterly Journal of Economics* 114: 497-532.

Lazear, E., (2001) “Educational Production”, *Quarterly Journal of Economics* 116: 777-803.

Lavy, V. (2009) “Performance Pay and Teachers’ Effort, Productivity and Grading Ethics”, *American Economic Review,* Volume 99, No. 5, pp. 979-2011.

Lavy, V. and A. Schlosser, (2005).”Targeted Remedial Education for Under-Performing Teenagers: Costs and Benefits”, *Journal of Labor Economics*, vol. 23, pp. 839-74.

Lee, J.-W. and Barro, R. (2001), “School Quality in a Cross-Section of Countries.”, *Economica,* vol. 68, 465-88.

Machin, S. and McNally, S. (2004), .The Literacy Hour., IZA Discussion Paper 1005.

Pischke, J.-S. “The Impact of Length of the School Year on Student Performance and Earnings: Evidence from the German Short School Years,” *Economic Journal* 117, October 2007, 1216-1242.

Rivkin. S. G, E. A. Hanushek, and J. F. Kain (2005) “Teachers, Schools, and Academic Achievement”, *Econometrica* 73: 417-59.

Rockoff. J. E (2004) “The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data”, *American Economic Review* 94: 247-52.

Rizzuto, R. and Wachtel, P. (1980). “Further Evidence on the Returns to School Quality” *Journal of Human Resources*, vol. 15, pp. 240-54.

Rose, H., and Betts, J.R. (2004). “The Effect of High School Courses on Earnings”, *Review of Economics and Statistics*, vol. 86, pp. 497-513.

Wöessmann, L. (2003). “Schooling Resources, Educational Institutions and Student Performance: The International Evidence". *Oxford Bulletin of Economics and Statistics*, vol. 65, pp. 117-70.

Table 1 - Means and Standard Deviations of Instructional Time in OECD, Eastern European, and Developing Countries

Subject	Mean Value	Std. Dev	Proportion of pupils by weekly instruction time: < 2 Hours	2-3 Hours	4-5 Hours	6 Hours +
Panel A: 22 OECD Countries						
All Subjects	3.38	(1.48)	13.16	40.43	36.45	9.97
Math	3.53	(1.38)	8.72	39.54	43.14	8.60
Science	3.06	(1.57)	21.14	42.72	25.53	10.61
Reading	3.54	(1.44)	9.61	39.02	40.66	10.71
Panel B: 14 Eastern European Countries						
All Subjects	3.05	(1.56)	22.51	39.59	29.29	8.61
Math	3.30	(1.48)	15.36	38.97	37.59	8.08
Science	2.77	(1.68)	33.38	37.21	17.53	11.88
Reading	3.08	(1.45)	18.79	42.59	32.75	5.86
Panel C: 13 Developing Countries						
All Subjects	3.23	(1.71)	22.86	34.72	27.51	14.90
Math	3.48	(1.69)	18.72	30.73	34.06	16.50
Science	2.97	(1.74)	29.03	37.17	18.53	15.27
Reading	3.24	(1.65)	20.85	36.27	29.94	12.95

Notes: The first column shows the mean of instruction time per week and the second column presents the respective standard deviations. The thrid to sixth columns presents the proportion of pupils by the amount of weekly hours of instruction time. The sample in panel A includes 22 OECD developed countries: Australia, Austria, Belgium, Canada, Germany, Denmark, Spain, Finland, France, Greece, Ireland, Iceland, Italy, Japan, Luxembourg, Netherlands, Norway, New Zealand, Portugal, Sweden, Switzerland, United Kingdom. Panel B includes 14 countries of Eastern Europe: Bulgaria, Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Montenegro, Poland, Romania, Russian Federation, Serbia, Slovak Republic, Slovenia. Panel C includes 13 developing countries: Argentina, Azerbaijan, Brazil, Chile, Colombia, Indonesia, Jordan, Kyrgyzstan, Mexico, Thailand, Tunisia, Turkey, Uruguay. Standard errors are reported in parentheses.

Table 2 - OLS Regressions of Test Scores on Instructional Time, OECD Sample

	Mathematics			Science			Reading		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
I. Continuous Hours:									
Hours	21.69	27.98	24.45	26.24	38.36	33.92	4.56	15.43	12.48
	(1.03)	(1.19)	(1.10)	(0.80)	(0.90)	(0.85)	(1.00)	(1.32)	(1.19)
II. Categorical Hours:									
2-3 Hours	40.92	47.97	43.03	44.67	53.70	48.48	49.25	50.73	42.20
	(8.16)	(7.32)	(6.67)	(2.63)	(2.82)	(2.54)	(10.38)	(8.54)	(7.66)
4 Hours +	63.73	70.11	61.89	77.11	90.48	80.40	55.69	64.88	53.41
	(8.21)	(7.41)	(6.76)	(2.98)	(3.24)	(2.96)	(10.42)	(8.67)	(7.79)
Country dummies		✓	✓		✓	✓		✓	✓
Individual characteristics			✓			✓			✓

Notes: The table shows OLS regression estimates of student test scores on hours of school instruction in a particular subject. In the first regression hours of instruction is a continuous variable. In the second regression hours enters the regression as binary variables for a particular number of hours learned per subject per week. The base (omitted) category is 1 hour. Controls on individual characteristics include binary variables for gender, fathers' and mothers' education and immigrant status. The sample includes 22 OECD developed countries (see notes to Table 1). Standard errors in parentheses are clustered at the school level. Each regression contains 137 083 observations.

Table 3 - Estimated Effect of Instructional Time on Test Scores, OECD Sample

	Whole Sample		Sample Divided by School Admission Policy			
			Academic Record is Irrelevant		Academic Record Taken into Account	
	OLS	Student FE	OLS	Student FE	OLS	Student FE
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Mathematics + Science + Reading					
Regression I.						
Hours of instruction	19.58	5.76	16.97	6.01	21.08	6.21
	(0.72)	(0.37)	(0.86)	(0.50)	(1.73)	(0.89)
Regression II.						
2-3 Hours	46.90	6.30	43.18	7.53	55.59	6.54
	(2.65)	(1.09)	(3.78)	(1.62)	(5.66)	(2.38)
4 Hours +	67.88	12.42	62.71	14.08	73.69	13.10
	(2.88)	(1.28)	(3.99)	(1.78)	(6.03)	(2.83)
Number of students	460,734		266,769		86,370	
	B. Mathematics + Science					
Regression I.						
Hours of instruction	25.48	7.14	21.84	8.60	27.56	7.57
	(0.73)	(0.55)	(0.88)	(0.75)	(1.79)	(1.33)
Regression II.						
2-3 Hours	45.65	9.38	40.13	10.81	55.19	12.17
	(2.58)	(1.52)	(3.49)	(2.57)	(5.06)	(3.43)
4 Hours +	73.87	16.96	65.72	20.71	80.47	18.66
	(2.82)	(1.81)	(3.70)	(2.84)	(5.58)	(4.07)
Number of students	307,156		177,846		57,580	

Notes: The table shows OLS and FE regressions of student scores on hours of instruction in a particular subject. Fixed effects are at the student level. Each regression also includes subject fixed effects. In the first regression hours of instruction is measured a continuous variable. In the second regression hours enters the regression as binary variables for a particular number of hours learned per subject per week. The base (omitted) category is 1 hour. The sample includes 22 OECD developed countries (see notes to Table 1). Standard errors in parentheses are clustered at the school level.

Table 4 - Estimated Effect of Instructional Time on Test Scores by School Tracking Policy

	Tracking By Class		Track In Class		No Tracking	
	OLS	Student FE	OLS	Student FE	OLS	Student FE
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Mathematics + Science + Reading					
Regression I.						
Hours of instruction	19.88	6.61	19.01	6.17	20.02	5.17
	(1.05)	(0.53)	(1.00)	(0.56)	(1.36)	(0.68)
Regression II.						
2-3 Hours	41.26	10.99	39.23	6.88	58.78	3.93
	(4.15)	(1.63)	(3.87)	(1.76)	(4.80)	(1.76)
4 Hours +	63.62	16.12	59.48	13.45	78.40	9.82
	(4.41)	(1.78)	(4.08)	(1.98)	(5.35)	(2.23)
Number of students	212,169		201,138		160,188	
	B. Mathematics + Science					
Regression I.						
Hours of instruction	22.01	10.13	24.06	8.58	30.14	3.36
	(1.01)	(0.73)	(1.02)	(0.82)	(1.45)	(1.05)
Regression II.						
2-3 Hours	39.21	17.67	37.31	10.81	58.11	3.30
	(3.91)	(2.23)	(3.62)	(2.50)	(4.85)	(2.29)
4 Hours +	64.40	26.93	64.41	20.11	89.27	5.86
	(4.15)	(2.44)	(3.84)	(2.86)	(5.48)	(3.01)
Number of students	141,446		134,092		106,792	

Notes: Table 4 replicates Table 3 in samples defined by tracking status - whether the school tracks students by classes, within classes, or not at all. The table shows OLS and FE regressions of student scores on hours of instruction in a particular subject. Fixed effects are at the student level. Each regression also includes subject fixed effects. In the first regression hours of learning is a continuous variable. In the second regression hours enters the regression as binary variables for a particular number of hours learned per subject per week. The base (omitted) category is 1 hour. The sample includes 22 OECD developed countries (see notes to Table 1). Standard errors in parentheses are clustered at the school level.

Table 5 - Estimated Effects of Instruction Time on Test Scores, with Controls Included in the Regressions for Special Science Activities in School and for Scarcity of Teachers in Each Subject

	Control Added For			
	Special Science School Activities		Scarcity of Teachers in Each Subject	
	OLS	Student FE	OLS	Student FE
	(1)	(2)	(3)	(4)
	A. Mathematics + Science + Reading			
Regression I.				
Hours of instruction	18.37	5.59	19.58	5.75
	(0.73)	(0.39)	(0.72)	(0.37)
Regression II.				
2-3 Hours	42.67	5.94	46.79	6.27
	(2.67)	(1.09)	(2.65)	(1.09)
4 Hours +	62.59	14.73	67.70	12.38
	(2.91)	(1.29)	(2.87)	(1.27)
Number of students	460,734		224,508	
	B. Mathematics + Science			
Regression I.				
Hours of instruction	24.10	6.65	25.47	7.08
	(0.75)	(0.55)	(0.73)	(0.55)
Regression II.				
2-3 Hours	41.31	8.28	45.54	9.21
	(2.58)	(1.51)	(2.58)	(1.51)
4 Hours +	67.87	15.19	73.72	16.75
	(2.82)	(1.80)	(2.81)	(1.79)
Number of students	307,156		149,672	

Notes: The table shows OLS and FE regressions of student scores on hours of instruction in a particular subject. Fixed effects are at the student level. Each regression also includes subject fixed effects. In the first regression hours of instruction is a continuous variable. In the second regression hours enters the regression as binary variables for a particular number of hours learned per subject per week. The base (omitted) category is 1 hour. The sample includes 22 OECD developed countries (see notes to Table 1). Standard errors in parentheses are clustered at the school level.

Table 6 - Estimated Effect of Instructional Time on Test Scores, by Gender, OECD Sample

	Boys		Girls	
	OLS	Student FE	OLS	Student FE
	(1)	(2)	(3)	(4)
	A. Mathematics + Science + Reading			
Regression I.				
Hours of instruction	20.25	4.99	18.62	5.62
	(0.86)	(0.40)	(0.77)	(0.41)
Regression II.				
2-3 Hours	46.82	6.22	46.66	5.91
	(3.09)	(1.19)	(2.85)	(1.22)
4 Hours +	67.86	11.20	67.16	12.13
	(3.39)	(1.37)	(3.09)	(1.40)
Number of students	224,508		236,226	
	B. Mathematics + Science			
Regression I.				
Hours of instruction	26.35	6.90	24.75	7.25
	(0.86)	(0.60)	(0.80)	(0.63)
Regression II.				
2-3 Hours	45.66	9.51	45.81	8.73
	(2.87)	(1.65)	(2.87)	(1.76)
4 Hours +	74.42	16.48	73.73	16.92
	(3.19)	(1.93)	(3.12)	(2.09)
Number of students	149,672		157,484	

Notes: The table shows OLS and FE regressions of student scores on hours of instruction in a particular subject. Fixed effects are at the student level. Each regression also includes subject fixed effects. In the first regression hours of instruction is a continuous variable. In the second regression hours enters the regression as binary variables for a particular number of hours learned per subject per week. The base (omitted) category is 1 hour. The sample includes 22 OECD developed countries (see notes to Table 1). Standard errors in parentheses are clustered at the school level.

Table 7 - Heterogeneity in Estimated Effect of Instructional Time on Test Scores, OECD Sample.

	High Parental Education		Low Parental Education		Immigrants - First Generation		Immigrants - Second Generation	
	OLS	Stud.FE	OLS	Stud.FE	OLS	Stud.FE	OLS	Stud.FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. Mathematics + Science + Reading							
Regression I.								
Hours of instruction	19.64	4.83	17.85	6.54	39.90	6.37	37.62	7.62
	(0.86)	(0.42)	(0.74)	(0.44)	(1.95)	(0.88)	(2.03)	(0.95)
Regression II.								
2-3 Hours	47.70	5.44	43.65	6.77	61.42	9.44	60.64	6.97
	(3.48)	(1.28)	(2.50)	(1.22)	(4.65)	(2.14)	(4.99)	(2.31)
4 Hours +	69.89	10.30	60.62	14.05	105.66	12.89	101.45	10.69
	(3.69)	(1.45)	(2.79)	(1.47)	(5.83)	(2.58)	(6.04)	(2.76)
Number of students	235,539		225,195		23,103		22,092	
	B. Mathematics + Science							
Regression I.								
Hours of instruction	24.67	7.11	24.06	7.14	47.17	8.76	42.05	11.99
	(0.86)	(0.63)	(0.78)	(0.61)	(1.96)	(1.26)	(1.98)	(1.38)
Regression II.								
2-3 Hours	47.44	9.49	41.79	9.16	60.23	10.15	61.20	13.56
	(3.40)	(1.82)	(2.41)	(1.64)	(4.70)	(3.12)	(4.97)	(3.33)
4-5 Hours	75.01	17.31	66.62	16.48	116.27	17.80	107.12	20.40
	(3.60)	(2.13)	(2.73)	(1.94)	(6.09)	(3.89)	(6.07)	(4.11)
Number of students	157,026		150,130		15,402		14,728	

Notes: The table reports estimates of the effect of instruction time on test scores for the following sub-samples: pupils from high edcuation families, pupils from low education families, first generation immigrants, and second generation immigrants. Fixed effects are at the student level. Each regression also includes subject fixed effects. In the first regression hours of instruction is a continuous variable. In the second regression hours enters the regression as binary variables for a particular number of hours learned per subject per week. The base (omitted) category is 1 hour. The sample includes 22 OECD developed countries (see notes in Table 1). Standard errors in parentheses are clustered at the school level.

Table 8 - Estimates of Effect of Instructional Time on Test Scores, Samples of Eastern European and Developing and Countries

	All	Boys	Girls	High Parental Education	Low Parental Education	Immigrant 1st Gen.	Immigrant 2nd Gen.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Eastern European Countries						
OLS	38.20	38.89	37.25	41.20	33.37	26.35	35.68
	(1.28)	(1.42)	(1.38)	(1.56)	(1.25)	(3.32)	(2.70)
Fixed Effects	6.07	5.15	6.49	5.03	6.67	5.53	7.26
	(0.56)	(0.59)	(0.59)	(0.66)	(0.62)	(2.07)	(1.88)
Number of students	177,015	84,612	92,403	78,006	99,009	3,525	5,604
	Developing Countries						
OLS	36.60	38.17	35.24	43.27	29.64	58.13	51.54
	(1.20)	(1.36)	(1.24)	(1.38)	(1.23)	(5.34)	(4.15)
Fixed Effects	2.99	2.39	3.29	3.41	2.60	18.59	11.11
	(0.80)	(0.87)	(0.90)	(0.94)	(0.88)	(4.65)	(3.91)
Number of students	238,938	108,927	130,011	76,970	82,322	1,642	2,210

Notes: The table shows OLS and fixed effect regressions of scores on hours of instructional time for two samples. The first sample includes the following 14 Eastern European countries: Bulgaria, Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Montenegro, Poland, Romania, Russian Federation, Serbia, Slovak Republic, Slovenia. The second sample includes the following 13 developing countries: Argentina, Azerbaijan, Brazil, Chile, Colombia, Indonesia, Jordan, Kyrgyzstan, Mexico, Thailand, Tunisia, Turkey, Uruguay.

Table 9 - Estimates of the Effect of Instructional Time on Test Scores in Israel

	5th Grade			8th Grade		
	OLS	Controls	School FE	OLS	Controls	School FE
Subject	(1)	(2)	(3)	(4)	(5)	(6)
Math	0.075	0.104	0.037	0.099	0.129	0.030
	(0.014)	(0.014)	(0.018)	(0.023)	(0.022)	(0.026)
Science	0.041	0.065	0.043	-0.018	0.004	-0.010
	(0.010)	(0.009)	(0.016)	(0.012)	(0.011)	(0.022)
English	0.029	0.053	0.058	-0.014	0.026	-0.001
	(0.018)	(0.016)	(0.020)	(0.029)	(0.023)	(0.024)
Included Controls:						
Year Fixed-Effects	✓	✓	✓	✓	✓	✓
Individual Pupil Controls		✓	✓		✓	✓
School Fixed Effects			✓			✓
Number of schools		939			475	
Number of students		110,544			104,729	

Notes: The table shows estimates of the effects of hours of instructional time on student scores, using Israeli data from 2002 and 2005. Standard errors are clustered at the school level. Individual controls include: a sex dummy, both parents' years of schooling, number of siblings, immigration status and ethnic origin.

Table 10 - OLS and Pupil Fixed Effects in Israel Using Various Combinations of Pooled Subjects

		5th Grade				8th Grade			
Sample		Math & Science (1)	Math & English (2)	Science & English (3)	All 3 Subjects (4)	Math & Science (5)	Math & English (6)	Science & English (7)	All 3 Subjects (8)
All	OLS	0.075	0.082	0.058	0.071	0.037	0.090	0.010	0.036
		(0.008)	(0.011)	(0.008)	(0.007)	(0.010)	(0.017)	(0.010)	(0.010)
	FE	0.055	0.060	0.060	0.058	0.041	0.036	0.015	0.029
		(0.010)	(0.016)	(0.012)	(0.007)	(0.012)	(0.024)	(0.015)	(0.009)
Boys	OLS	0.076	0.085	0.061	0.073	0.037	0.086	0.008	0.034
		(0.009)	(0.012)	(0.009)	(0.008)	(0.011)	(0.019)	(0.011)	(0.011)
	FE	0.055	0.059	0.062	0.059	0.038	0.035	0.013	0.026
		(0.010)	(0.017)	(0.014)	(0.008)	(0.014)	(0.026)	(0.017)	(0.011)
Girls	OLS	0.074	0.080	0.054	0.068	0.038	0.091	0.012	0.037
		(0.008)	(0.011)	(0.009)	(0.008)	(0.011)	(0.018)	(0.011)	(0.011)
	FE	0.056	0.059	0.057	0.057	0.044	0.034	0.017	0.031
		(0.011)	(0.018)	(0.012)	(0.008)	(0.014)	(0.027)	(0.015)	(0.010)
Higher Parental Education	OLS	0.090	0.093	0.069	0.083	0.049	0.088	0.018	0.044
		(0.008)	(0.011)	(0.008)	(0.007)	(0.011)	(0.017)	(0.010)	(0.010)
	FE	0.054	0.034	0.049	0.047	0.039	0.024	0.004	0.022
		(0.011)	(0.017)	(0.015)	(0.008)	(0.013)	(0.029)	(0.015)	(0.010)
Lower Parental Education	OLS	0.075	0.084	0.056	0.070	0.030	0.096	0.005	0.032
		(0.009)	(0.013)	(0.010)	(0.009)	(0.013)	(0.020)	(0.012)	(0.012)
	FE	0.057	0.078	0.068	0.066	0.040	0.047	0.022	0.033
		(0.012)	(0.019)	(0.014)	(0.009)	(0.015)	(0.026)	(0.016)	(0.010)

Notes: The table shows OLS and fixed effect regressions of scores on continuous hours of instructional time for the Israeli data, using different subject combinations, for 5th and 8th grade. Estimates include subject and year fixed effects. Standard errors are clustered at the school level. Individual controls include: a sex dummy, both parents' years of schooling, number of siblings, immigration status and ethnic origin.

Table 11 - Pupil Fixed Effect Estimates of Instructional Time on Test Scores in Israel based on a Panel Data of Pupils Observed Both at 5th and 8th Grade in Israel.

Sample	Math (1)	Science (2)	English (3)	All Subjects - OLS	All Subjects - FE
All	0.086	0.074	0.013	0.026	0.036
	(0.019)	(0.015)	(0.024)	(0.004)	(0.004)
Boys	0.095	0.050	0.031	0.032	0.041
	(0.030)	(0.023)	(0.037)	(0.006)	(0.006)
Girls	0.080	0.097	0.000	0.021	0.032
	(0.025)	(0.020)	(0.032)	(0.005)	(0.005)
Pupils with Higher Parental education	0.067	0.046	-0.019	0.028	0.026
	(0.033)	(0.024)	(0.034)	(0.005)	(0.006)
Pupils with Lower Parental education	0.093	0.080	0.040	0.028	0.043
	(0.029)	(0.024)	(0.041)	(0.006)	(0.006)
Descriptive Statistics:					
Mean change in hours	1.156	3.765	1.749	-0.690	-0.690
SD of change in hours	(2.176)	(2.715)	(1.512)	(3.438)	(3.438)
Number of schools	686	686	686	686	686
Number of students	4822	4822	4822	4822	4822

Notes: This table estimates the effect of continuous hours on scores, for each subject separately, and for all subjects pooled together, using OLS and Student Fixed Effects. Standard errors are clustered at the school level. Individual controls include: a sex dummy, both parents' years of schooling, number of siblings, immigration status and ethnic origin. Column 4 pools the samples from columns 1-3, and includes a subject fixed effect. Column 5 does the same, but includes a student fixed effect.

Table 12 - Estimated Effects of School Characteristics Interacted with Instructional Hours, OECD Countries.

		Separate Spec.		Joint Spec.	
Index	Index's Mean (1)	Hours Main Effect. (2)	Hours interact-ed with Index (3)	Hours interact-ed with Index (4)	Hours interact-ed ith Index (5)
Achievement data are posted publicly (e.g. in the media). (Binary Variable)	.335 (.472)	5.017 (.447)	2.744 (.840)	1.962 (.903)	2.452 (.912)
Achievement data are used in evaluation of the principal's performance (Binary Variable)	.216 (.411)	5.153 (.432)	2.106 (.889)	2.158 (1.135)	2.317 (1.134)
Achievement data are used in evaluation of teachers' performance (Binary Variable)	.294 (.456)	5.501 (.458)	.345 (.819)	-1.230 (1.015)	-.934 (1.010)
Quality of Educational Resources: Index, (Range -3.45 to 2.1)	.150 (.989)	5.834 (.395)	.099 (.393)	.435 (.399)	.442 (.400)
School Responsibility for Resource Allocation: Index, (Range -1.1 to 2.0)	-.058 (.946)	5.925 (.380)	1.224 (.398)	.842 (.433)	.938 (.435)
School Responsibility for Curriculum & Assessment: Index (Range -1.4 to 1.3)	.052 (.964)	5.830 (.386)	-.247 (.399)	-.451 (.427)	-.561 (.429)
School Governing Board Influences Staffing (Binary Variable)	.363 (.481)	4.981 (.523)	2.599 (.763)		1.199 (.883)
School Governing Board Influences Budget (Binary Variable)	.706 (.455)	3.759 (.711)	2.974 (.843)		1.834 (.925)
School Governing Board Influences Instructional Content (Binary Variable)	.162 (.368)	5.973 (.429)	-.588 (.968)		-.199 (1.069)
School Governing Board Influences Assessment (Binary Variable)	.219 (.413)	6.018 (.464)	-.837 (.831)		-.802 (.922)
Hours Main Effect				4.676 (.713)	3.255 (.964)

Notes: This table looks into the effect of hours when it is interacted with various school characteristics (means shown in column 1). The estimates presented in columns 2 and 3 are based on regressions when each characteristic enters the regression separately. In columns 4 and 5 all characteristics are jointly included. Regressions include hours, interaction between hours and the school characteristic, subject dummies, subject dummies interacted with school characteristics, and pupil fixed effects. The sample includes 22 OECD developed countries that are listed in the notes of Table 3.

Table A1 - Average Hours of Instructional Time and Pisa Scores, for OECD Countries

#	Country	Code	Hours of Instruction per week				Pisa Score				Number of
			Mathematics	Science	Reading	All (sum)	Mathematics	Science	Reading	All (average)	Students
1	Australia	AUS	3.5	2.8	3.5	9.8	516.2	523.0	508.3	515.8	14,170
2	Austria	AUT	2.8	2.2	2.4	7.4	509.3	513.8	494.0	505.7	4,927
3	Belgium	BEL	3.2	2.3	3.1	8.6	526.9	516.2	506.9	516.6	8,857
4	Canada	CAN	3.9	3.5	3.9	11.3	517.4	522.5	512.4	517.5	22,646
5	Switzerland	CHE	3.5	2.0	3.4	8.9	527.8	507.6	496.2	510.5	12,192
6	Germany	DEU	3.4	2.7	3.2	9.3	503.7	516.0	496.2	505.3	4,891
7	Denmark	DNK	3.9	2.8	4.8	11.5	512.4	495.1	494.1	500.5	4,532
8	Spain	ESP	3.1	2.8	3.2	9.1	501.4	504.4	479.7	495.2	19,604
9	Finland	FIN	3.0	2.7	2.7	8.4	549.9	563.7	547.2	553.6	4,714
10	France	FRA	3.4	2.5	3.6	9.5	497.0	496.1	488.6	493.9	4,716
11	United Kingdom	GBR	3.4	3.7	3.4	10.5	497.5	514.3	496.0	502.6	13,152
12	Greece	GRC	3.0	2.8	2.8	8.6	461.9	476.8	462.1	466.9	4,873
13	Ireland	IRL	3.2	2.2	3.1	8.5	502.2	509.4	518.8	510.1	4,585
14	Iceland	ISL	4.2	2.6	4.0	10.7	505.2	490.8	484.3	493.4	3,789
15	Italy	ITA	3.2	2.5	3.9	9.6	473.8	487.2	477.4	479.4	21,773
16	Japan	JPN	3.7	2.3	3.3	9.4	525.8	534.1	500.1	520.0	5,952
17	Luxembourg	LUX	3.4	2.1	3.1	8.5	490.5	487.0	480.5	486.0	4,567
18	Netherlands	NLD	2.5	2.0	2.5	7.1	537.2	530.4	513.8	527.1	4,871
19	Norway	NOR	2.9	2.3	3.1	8.4	489.9	486.4	484.5	486.9	4,692
20	New Zealand	NZL	3.9	3.6	3.9	11.4	523.0	532.3	523.3	526.2	4,823
21	Portugal	PRT	3.2	2.9	2.8	8.9	470.2	478.7	476.6	475.2	5,109
22	Sweden	SWE	2.6	2.4	2.7	7.7	503.3	504.3	508.5	505.4	4,443
	Average		3.3	2.6	3.3	9.2	506.5	508.6	497.7	504.3	8358.1
	Standard Deviation		0.4	0.5	0.6	1.3	21.5	21.1	19.0	19.6	6089.3
	Total										183,878

Notes: The table shows, for each OECD country, average hours of instruction per week, for Mathematics, Science and Reading, and the total for all three subjects. Average Scores are also shown for these categories.

Table A2 - Average Hours of Instructional Time and Pisa Scores, for Eastern European Countries

#	Country	Code	Hours of Instruction per week				Pisa Score				Number of
			Mathematics	Science	Reading	All (sum)	Mathematics	Science	Reading	All (average)	Students
1	Bulgaria	BGR	2.6	2.3	2.6	7.5	417.2	439.4	407.2	421.3	4,498
2	Czech Republic	CZE	3.5	3.0	3.2	9.7	536.0	537.7	510.0	527.9	5,932
3	Estonia	EST	3.7	2.9	3.1	9.7	517.2	534.5	502.9	518.2	4,865
4	Croatia	HRV	2.7	1.8	2.9	7.3	467.3	493.3	477.1	479.2	5,213
5	Hungary	HUN	2.9	2.2	2.8	7.9	496.7	508.9	488.4	498.0	4,490
6	Lithuania	LTU	3.1	2.4	3.2	8.7	485.3	486.5	468.7	480.1	4,744
7	Latvia	LVA	3.9	2.5	3.2	9.7	491.1	493.7	484.6	489.8	4,719
8	Montenegro	MNE	2.7	2.5	2.6	7.8	395.2	408.8	387.8	397.3	4,455
9	Poland	POL	3.9	2.4	4.1	10.4	500.3	503.0	512.7	505.3	5,547
10	Romania	ROU	2.5	1.9	2.8	7.3	415.0	416.3	392.0	407.7	5,118
11	Russian Federation	RUS	3.2	3.3	1.8	8.3	478.6	481.4	442.3	467.4	5,799
12	Serbia	SRB	2.8	2.5	2.8	8.1	436.1	436.8	403.0	425.3	4,798
13	Slovak Republic	SVK	2.9	2.2	2.7	7.8	494.7	491.1	470.2	485.3	4,731
14	Slovenia	SVN	2.8	2.2	2.7	7.7	482.3	494.3	468.9	481.8	6,595
	Average		3.1	2.4	2.9	8.4	472.4	480.4	458.3	470.3	5107.4
	Standard Deviation		0.5	0.4	0.5	1.0	41.3	40.2	44.0	41.3	640.5
	Total										71,504

Notes: The table shows, for 14 Eastern European countries, average hours of instruction per week, for Mathematics, Science and Reading, and the total for all three subjects. Average Scores are also shown for these categories. The sample includes 14 countries of Eastern Europe: Bulgaria, Croatia, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Montenegro, Poland, Romania, Russian Federation, Serbia, Slovak Republic, Slovenia.

Table A3 - Average Hours of Instructional Time and Pisa Scores, for Developing Countries

			Hours of Instruction per week				Pisa Score				Number of
#	Country	Code	Mathematics	Science	Reading	All (sum)	Mathematics	Science	Reading	All (average)	Students
1	Argentina	ARG	2.6	2.0	2.1	6.8	388.3	398.9	384.4	390.5	4,339
2	Azerbaijan	AZE	3.3	2.5	3.2	9.0	476.6	385.3	355.2	405.7	5,184
3	Brazil	BRA	2.7	2.0	2.6	7.3	365.8	385.5	389.9	380.4	9,295
4	Chile	CHL	3.1	2.1	3.1	8.3	417.5	442.6	447.8	435.9	5,233
5	Columbia	COL	3.7	3.0	3.4	10.2	373.5	391.5	390.9	385.3	4,478
6	Indonesia	IDN	3.5	2.7	3.2	9.5	380.7	384.8	383.6	383.0	10,647
7	Jordan	JOR	3.1	2.9	3.2	9.2	388.9	427.0	409.4	408.4	6,509
8	Kyrgyzstan	KGZ	2.6	1.9	2.5	7.0	316.0	326.4	290.9	311.1	5,904
9	Mexico	MEX	3.5	2.7	3.3	9.5	420.8	422.5	427.6	423.6	30,971
10	Thailand	THA	3.4	3.4	2.7	9.5	425.2	430.0	425.3	426.8	6,192
11	Tunisia	TUN	3.0	2.3	2.8	8.0	363.5	384.3	378.5	375.4	4,640
12	Turkey	TUR	3.4	2.6	3.5	9.5	428.0	427.9	453.4	436.5	4,942
13	Uruguay	URY	3.0	2.2	2.4	7.6	435.2	438.1	425.0	432.8	4,839
	Average		3.1	2.5	2.9	8.6	398.5	403.4	397.1	399.7	7936.4
	Standard Deviation		0.4	0.5	0.4	1.1	41.0	32.2	43.0	34.8	7177.6
	Total										103,173

Notes: The table shows, for 13 Developing Countries, average hours of instruction per week, for Mathematics, Science and Reading, and the total for all three subjects. Average Scores are also shown for these categories. The sample includes 13 developing countries: Argentina, Azerbaijan, Brazil, Chile, Colombia, Indonesia, Jordan, Kyrgyzstan, Mexico, Thailand, Tunisia, Turkey, Uruguay.

Table A4 - Descriptive Statistics - Test Score and Instructional Time

	Test scores			Instructional time		
	OECD Developed	Eastern Europe	Developing countries	OECD Developed	Eastern Europe	Developing countries
Mean	513.4	485.6	413.5	3.38	3.05	3.23
Standard Deviation between pupils	84.4	86.9	75.1	1.02	0.88	1.22
Standard Deviation within pupils	38.8	40.9	46.7	1.08	1.28	1.19

Notes: The table contains means, and the standard deviation within and between pupils, for 3 different samples: OECD countries, Eastern Europe, and Developing Countries.

Table A5 - Regressions of Test Scores on Instructional Time using School Level Means

	Whole Sample		Boys		Girls	
	OLS	School FE	OLS	School FE	OLS	School FE
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Mathematics + Science + Reading					
Hours of instruction	20.28	4.80	19.45	4.18	20.13	5.11
	(0.77)	(0.48)	(0.86)	(0.52)	(0.78)	(0.45)
Number of Students	19,731		18,894		18,792	
	B. Mathematics + Science					
Hours of instruction	25.89	6.47	25.38	6.19	26.15	7.11
	(0.80)	(0.60)	(0.88)	(0.67)	(0.81)	(0.64)
Number of Students	13,154		12,596		12,528	

Notes: These regressions are run using collapsed school level data. For example, hours refers to the mean of continuous hours of learning, averaged to the school level. Fixed effects are at the student level. Hours of learning is a continuous variable. The sample includes 22 OECD developed countries: Australia, Austria, Belgium, Canada, Germany, Denmark, Spain, Finland, France, Greece, Ireland, Iceland, Italy, Japan, Luxembourg, Netherlands, Norway, New Zealand, Portugal, Sweden, Switzerland, United Kingdom. Each regression includes subject dummies, and school fixed effects. Standard errors in parentheses are clustered at the school level.

Table A6 - Subject Combinations Estimates using OLS and Pupil Fixed Effects in Israel: Pooled 5th and 8th Grades

Sample		Math & Science (1)	Math & English (2)	Science & English (3)	All 3 Subjects (4)
All	OLS	0.047	0.063	0.038	0.048
		(0.005)	(0.008)	(0.006)	(0.005)
	FE	0.034	0.036	0.039	0.036
		(0.005)	(0.008)	(0.007)	(0.004)
Boys	OLS	0.053	0.072	0.037	0.052
		(0.006)	(0.009)	(0.007)	(0.006)
	FE	0.039	0.041	0.040	0.040
		(0.005)	(0.008)	(0.008)	(0.005)
Girls	OLS	0.042	0.053	0.039	0.043
		(0.006)	(0.008)	(0.007)	(0.006)
	FE	0.029	0.030	0.036	0.031
		(0.005)	(0.008)	(0.007)	(0.004)
Higher Parental Education	OLS	0.052	0.057	0.053	0.053
		(0.006)	(0.008)	(0.007)	(0.005)
	FE	0.026	0.026	0.023	0.025
		(0.005)	(0.008)	(0.008)	(0.005)
Lower Parental Education	OLS	0.050	0.072	0.032	0.049
		(0.007)	(0.009)	(0.008)	(0.006)
	FE	0.041	0.043	0.049	0.044
		(0.005)	(0.009)	(0.007)	(0.005)
Top Deciles	OLS	0.062	0.087	0.041	0.060
		(0.008)	(0.010)	(0.009)	(0.008)
	FE	0.041	0.040	0.024	0.035
		(0.007)	(0.011)	(0.010)	(0.006)
Bottom Deciles	OLS	0.039	0.046	0.046	0.043
		(0.008)	(0.011)	(0.009)	(0.007)
	FE	0.028	0.033	0.046	0.035
		(0.006)	(0.010)	(0.009)	(0.005)

Notes: This table is a version of table 8, however using 5th and 8th grade pooled together. The table shows OLS and fixed effect regressions of scores on continuous hours of instructional time for the Israeli data, using different subject combinations. Estimates include subject and year fixed effects. Standard errors are clustered at the school level. Individual controls include: a sex dummy, both parents' years of schooling, number of siblings, immigration status and ethnic origin.