DO DIFFERENCES IN SCHOOLS' INSTRUCTION TIME EXPLAIN INTERNATIONAL
ACHIEVEMENT GAPS? EVIDENCE FROM DEVELOPED AND DEVELOPING
COUNTRIES

Victor Lavy

Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence
from Developed and Developing Countries
Victor Lavy
NBER Working Paper No. 16227
July 2010, Revised August 2012
JEL No. I21,I28

ABSTRACT

The time children spend in school varies across countries. Do these differences explain international
gaps in pupils' academic achievements? In this paper, I estimate the effects of instructional time on
students' achievement. I use the PISA 2006 data with samples from more than 50 countries. I find
that instructional time has a positive and significant effect on test scores bit it is much lower in developing
countries. Evidence suggests that the productivity of instructional time is higher in countries that implemented
school accountability measures, and in countries that give schools autonomy in budgetary decisions
and in hiring and firing teachers.

Victor Lavy
Department of Economics
Hebrew University of Jerusalem
Mount Scopus
Jerusalem 91905
ISRAEL
and University of Warwick, England
and also NBER
msvictor@mscc.huji.ac.il

I. Introduction

The amount of time students spend in public schools varies widely from one country to another. For example, among European countries such as Belgium, France and Greece, 15-year-old pupils receive an average of more than 1,000 hours per year of total compulsory classroom instruction, while in England, Luxembourg and Sweden the average is only 750 hours per year.[1] Children aged 7-8 in England, Greece, France and Portugal, receive an average instructional time of more than 800 hours per year, while in Finland and Norway, they receive less than 600 hours. Similar differences among countries exist in the number of classroom lessons per week in different subjects as evident from data from the 2006 Program for International Student Assessment (PISA), a unique international education survey of 15-year-old students conducted by the Organization for Economic Co-operation and Development (OECD) and designed to allow for cross-country comparisons. Pupils aged 15 in Denmark, for example, receive 4 hours of instruction per week in math and 4.7 hours in language, while pupils of the same age in Austria receive only 2.7 hours of weekly classroom lessons in math and 2.4 hours in language, the data show. Overall, total weekly hours of instruction in math, language and science is 55 percent higher in Denmark (11.5 hours) than in Austria (7.4 hours). Similar magnitudes of disparities in instructional time appear among the Eastern European and developing countries included in PISA 2006. Can these large differences explain some of the differences across countries in pupils' achievements in different subjects? What is the likely impact of changes in instructional time? This latter question is of policy relevance in many countries. For example, U.S. President Barack Obama has argued in recent years that American children should go to school longer – either staying later in the day or attending later into the summer or both. He has spoken of this goal, of extending the school week and year, as a central element in his proposed education reform.[2] An example of implementation of these ideas is the forty schools across Chicago that signed on to Mayor Rahm Emanuel's longer school day initiative earlier this year. Students in these schools will have an extra 90 minutes of instruction every day.[3]

---

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

[2] President Barack Obama said (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 — to have a chance to compete for jobs and paychecks against foreign children. He said "Our children spend over a month less in school than children in South Korea…. The challenges of a new century demand more time in the classroom. 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.

[3]   See   http://www.suntimes.com/news/education/7757540-418/chicago-public-schools-system-launches-longer-school-day-website.html.

While research in recent years provides convincing evidence about the effect of several inputs in the education production function[4], there is limited evidence on the effect of classroom instructional time. This evidence ought to be important for policymakers because increasing instructional time is a relatively simple option to implement, provided that resources are available, and, moreover, such an increase would be feasible in many countries.

The simple correlations and the simple OLS regression relationship between classroom instructional time per week and test scores of pupils aged 15 in the 2006 PISA exams in math, science and language are positive and highly significant. Of course, these correlations do not represent causal relationships because of potential selection bias and endogeneity issues. In this paper, I investigate the causal relationship between instructional time and pupils' knowledge in these subjects. 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. In this analysis of productivity, I focus on structural features such as accountability, autonomy and governance of schools.

I use 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 percent of the world economy took part in PISA 2006. The study focused on science, but the assessment also included language and mathematics. In addition, it collected data on student, family and institutional factors, including information about the amount of instructional time per week in each of the subjects tested. I exploit these features of the data to identify the effect of instructional time on academic achievement. First, having test scores in multiple subjects for each student, and relatively large variation in instructional time across subjects allows me to use within student estimation of the effect of instructional time. At the same time, I am able to control 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. Second, the considerable within student variation in instructional time allows for additional analysis. 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 hours per week in Iceland. The respective figures in science are 1.8 in the Netherlands and 3.6 in New Zealand, while in language they are 2.4 in Austria and 4.7 in Germany. I use this significant variation to

---

[4] For example, studies on class size (e.g. Kruger 1999, Hoxby 2000), teachers' training and certification (Kane Rockoff and Staiger, 2007), remedial education (Jacob and Lefgren, 2004), teacher effect (Rockoff, 2004, Rivkin, Hanushek, and Kain, 2005), and computer aided instruction (Barrow, Markman and Rouse, 2009).

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 subjects together.

Numerous studies have evaluated 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 United States and found insignificant effects, perhaps due to limited variation in the length of the school year there, or perhaps also due to the omission of other variables. Card and Krueger (1992), and Betts and Johnson (1998) used state-level data in the United States to examine the same effect and found a positive significant effect on earnings, perhaps because they studied earlier periods when length of the school year varied more widely and because state-level data may be of sufficient size to contend with the potential effect of unobserved heterogeneity Card and Krueger also presented results controlling for state effects, showing that the positive effect of a longer school year length vanishes within states and conditional on other school quality variables. Lee and Barro (2001) examined 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 found no effects of the length of the school year on internationally comparable test scores. A more recent study by Wöessmann (2003), which also analyzed cross-country test score data, corroborated this finding. Additional instructional time had an effect that is significant but of negligible size, he reported. However, these two studies attempted 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), overcame potential selection and endogeneity problems by using a West-German short school year in 1966-67 which reduced those school years by about two-thirds for some students. This change increased grade repetition in primary school, and lowered enrollment in higher secondary school tracks, but it had no adverse effect on earnings and employment later in life. Hansen (2008) reported that more instructional time due to fewer snow-related school day cancelations and delayed testing increased student performance in Colorado and Maryland. Marcotte and Hemelt (2008) reported that substantial snowfall (leading to fewer days spent at school) is associated with lower pupils' performance in Maryland.

The results I present in this paper show that instructional time has a positive and significant effect on the academic achievements of pupils. However, the estimates are much lower than the "naïve" OLS estimates which overstate the extent to which countries like the United States might "catch up" to

other developed nations by increasing instructional time. The size of the estimated effects is modest to large. On average, a one-hour increase per week in math, science or language instruction raises the test score in these subjects by 0.15 of a standard deviation of the within student distribution of test scores. The effect is larger for girls, immigrants and pupils from low socioeconomic status families. Estimates based on the sample of the former Soviet Eastern European countries are very similar to the average effect obtained from the sample of OECD developed countries. However, the evidence based on a sample of developing countries suggests a much lower effect of additional 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. Echoing the OECD results, the effect is much larger for girls, immigrants and pupils from disadvantaged backgrounds. Overall, the main results presented in this paper are very robust to a variety of robustness checks with respect to the identification assumptions and to threats to their validity. In addition, this evidence is almost identical to related results I obtained with data from Israel using two identification strategies (Lavy 2012). In that paper, the first method was identical to the one used in this paper, and the second stemmed from a natural experiment that provided an opportunity to exploit a sharp change in the method of funding schools in Israel. Remarkably, both methods led to identical results about the effect of instruction time on test scores in core subjects. The results from the research in Israel, and the results from the research presented here in an international context, are, remarkably, almost identical.

In the second part of this 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. However, I am able to estimate the effect of their interactions with instructional time in each subject. The evidence suggests 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. These findings emphasize the importance of quality and quantity of instructional time in bridging the gaps in student achievement across countries.

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

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 receive more instructional time than others. Of course, it could be that at the school level, such variation is not purely random. However, 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,

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

Where A_{ijk} is the achievement of the ith student, in the jth school, in the kth subject, H_{kj} is instructional time in the kth subject in the jth school, X is a vector of characteristics of the ith student in the jth school and S_j is a vector of characteristics of the jth 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 one to control for a number of sources of potential biases related to unobserved characteristics of the school, the student, or their interaction. In some situations, students might be placed or be sorted according to their ability across schools that provide more (or less) instructional time in some subjects. If, for example, more able students attend better schools that provide more instructional hours overall in each subject, it would cause γ to be biased downward unless one also were to account for student and school fixed effects. The bias would have an opposite sign if the less able students are exposed to more instructional time. However, identification of the effect of instructional time based on a comparison of the performance of the same student in different subjects is immune to these potential biases.

Three remarks about this identification strategy merit mention. First, the necessary assumption for this identification strategy is that the effect of instructional time is the same for all subjects, implying

5

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 language 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 have high ability in math, for example, may select or be placed in a school that specializes in math, and they may receive more instructional time in math as a result. However, this issue may be less of a concern in the sample I use here. Such tracking is largely conducted 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. In addition, 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 whether the school sorts and selects students based on subject-specific considerations. For example, I observe in the PISA data information about whether admission to a school takes into account the student's academic record, or the student's need or desire for a special program. I also note whether a school uses tracking in forming classes, and whether it is a public or a private school. I assume that a school that does not use academic ability as an admission criterion, or does not take into account student's desire for a particular program, will most likely not select students on subject-specific considerations. Such selection is also less likely to take place in schools that do not use any form of tracking by ability, or in public schools. Indeed, the results that I present below are very similar across the various stratified samples based on schools' admission and tracking policies and on whether they are private or public.

III. Data

The Program for International Student Assessment, known by its acronym PISA, provides regular data on the OECD country education systems and the knowledge and skills of their students. 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. PISA does not examine students' mastery of specific school curricula. Instead, it evaluates students' ability to apply knowledge and skills in key subject areas and their ability to analyze reason and communicate effectively as they examine, interpret and solve problems. PISA measures student performance in language, 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 that 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 participating OECD country has a PISA Governing Board representative, 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 PISA surveys. For each survey, an international contractor (usually a testing and assessment agency) is 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 in each participating country. PISA has Subject Matter Expert Groups for its three key areas of testing – language, 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 international contractor and participating countries review the questions and carefully check them for cultural bias. PISA uses only those questions that have been approved unanimously.

Students answer a 20- to 30-minute background questionnaire, providing information about themselves, their homes, and their attitudes to learning. 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 the following question (see online appendix for the exact format of this question in the PISA 2006 student's questionnaire): "How much time do you typically spend per week studying the following subjects in regular lessons at school?" The student could choose from the following options: *No time, Less than 2 hours a week, 2 or more but less than 4 hours a week, 4 or more but less than 6 hours a week, or more hours a week.* I merged the first two options into one as *less than two hours a week* and used the other categories as is. I computed the school average in each subject using the mid values of each range. Because the PISA data allow for overall instruction time per week in all subjects to be combined, I have aggregated and averaged this information at the country annual level for all OECD countries. A high correlation emerges between these country means and the administrative data on total annual hours of instruction in secondary schooling as reported in the OECD report, "Education at Glance."

The means of instructional time in the developed OECD countries in math, science and language are 3.53, 3.06, and 3.54 hours per week, respectively. In the Eastern European sample, mean instructional time is lower than in the OECD countries in all subjects (3.30 for math, 2.77 for science and 3.08 in language). These Eastern European figures are similar to the mean instructional time in the developing countries sample (3.48 for math, 2.97 for science and 3.24 for language). Tables A1-A3 in the online appendix present the mean instructional time in each of the subjects for each of the countries included in the three samples of the 2006 PISA. As discussed in the introduction, large variation exists on many levels. Variation emerges among the countries that make up the three broad multi-country groups, and within individual countries, among different schools and among the various subjects. As a result, a significant proportion of the overall variation in instruction time is due to the variation within countries among schools. (Evidence is not shown here, but is available from the author.) The PISA data - which include instruction time per week for the three subjects studied here, all other subjects, and for the overall school week - explain some of the reasons behind the variations. As a result, the data show that a large part of the variation across subjects in instruction time is explained by variation across schools in total instruction time per week. Of course, the variation in total weekly instruction time across schools originates from many sources, for example differential government resource allocation to schools by their socio-economic background, differential contribution from parents and the community of local

school authority and more. Many of these factors vary at the school or country level, and our identification methodology, using the within school and pupil estimation strategies, will account for them.

Certain additional issues regarding the data and my methodologies bear mention. Different countries vary in the length of the school year (number of weeks of study), but is not a limitation in this case since the within school and pupil estimation strategies control for the length of the school year. Regarding instruction time, the measurements reflect the average in the school, not the hours reported by each student. And, it is worth underscoring that the variation in subject instruction hours across students in the same school is not due to an endogenous decision of the students. This is because the PISA sample in each country includes students in 9th and 10th grades, levels at which the curriculum of study in most or all countries is completely compulsory. Students may choose some courses only in the latter part of high school, perhaps at 11th and 12th grades. Variation in subject instruction hours could stem from tracking practices in schools, however, and I address this issue by directly examining the sensitivity of my results to differences in tracking practices across schools.

Tables A1-A3 in the online appendix present the country means of the PISA test scores in each of the subjects for each of the countries. Table A4 in the online appendix present the means of the PISA test scores 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 of the same pupil to explain. The average weekly 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 in the online appendix 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 score regressions based on the sample of the OECD countries. For each subject, I report estimates from three specifications: 1) without any controls, 2) with country fixed effects, and 3) 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 of measures 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 on PISA test scores 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 language 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 an 5.76 points on the PISA test. 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.

It is important to note here that in a related paper (Lavy, 2012) I report results based on the same method I use in this paper and also based on an alternative identification strategy that relies on a natural experiment that exploit a sharp change in the funding method of schools in Israel. Remarkably, both of these two methods lead to identical point estimates for the effect of instructional time on core subjects test scores, and these two sets of estimates from the previous paper are almost identical to the results presented in this paper in Table 3. It is also important to emphasize that the natural experiment method that I used in the Israeli context allowed me to test some of the central assumptions I make while using the pupil fixed effect identification method. These tests support the assumption that the effect of additional hours of instruction on each of the subjects is very similar and that there are no cross-subjects spillover effects. These previous findings strengthen the overall credibility of the results I present in this paper and also their causal interpretation.

10

The estimates of the indicators of instructional time presented in lower part of panel A of Table 2 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 percent higher than the respective estimate obtained from pooling all three subjects together. However, pooling math and language test scores yields an estimate of 7.42 and pooling science and language yields an estimate of 4.27. The pattern emerging from the OECD country group does not permit me to conclude whether the average productivity of instructional time is lower for any of the three subjects. 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 and checks for threats of identification

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 based on a sample of schools instead of pupils. I present these results in online appendix Table A5. I obtain the variables used by collapsing the pupils' data to the school-level respective means. The pattern in this table is very similar to that of 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 student micro data. The estimate based on all three subjects is 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 by including interactions between the subject dummies and pupil's characteristics. The estimates from this more flexible specification approach (not shown here, but available from the author) are very similar to those presented in Table 3, though overall they are about 10 percent lower.

---

[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.

The first robustness check concerns whether the evidence in column 2 Table 3 reflects some subject-specific selection and sorting in some schools. It is based on the data available from the PISA school questionnaire about how much consideration is given in the admission decisions to student's academic record and whether placements tests are used in this process. (See the exact wording of the question 19 in the online appendix.) I expect that the validity of the identification strategy will be underscored here. As a result, sample of schools that do not pay any attention to applicants' previous academic records and do not use any admission exams should not be sensitive to endogenous sorting and selection. 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 in which past academic achievements are irrelevant for admission is the largest, accounting for about two-thirds of the whole sample. The estimates from this sample are only marginally different from those obtained from full sample: the OLS estimate 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. I also estimated the effect of hours of instruction based on a sample of schools that do not use a student's desire for a particular program as a criterion for admission. This information is also based on the PISA school questionnaire question 19. The results based on this sample are very similar those obtained from the full sample.

Another potential source of selection bias is tracking pupils into classes within schools according to their ability. One can expect that schools practicing 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, a bias in the estimated effect of hours of instruction will occur. In Table 4, I present results for three different samples distinguished by schools' tracking policies. (See the exact wording of the question in the online appendix.) Columns 1-2 show the estimates for a sample of schools that practice tracking at the class level – those that group students in classes according to their ability, that is. Columns 3-4 show the results based on a sample of schools that sort pupils into different ability study groups within classes. Columns 5-6 show the results based on a sample of schools that do not practice any form of pupil tracking. 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 ability classes 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 significantly different from the point estimates obtained from the full sample; 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 private schools are more likely to base admission on previous academic record and on additional exams, and to track pupils by ability. To address 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 language test scores is 6.09 (sd=0.428), just barely higher than the estimate from a sample that included also the private schools.

Overall, it is apparent that the results are not driven by potential selection and sorting of students based on subject-specific considerations related to selective admission or tracking pupils in classes by abilities, as underscored by the findings presented in Tables 3 and 4 and the public schools-only sample. This is an important result because schools that admit pupils based on academic record or that track students by ability 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. For example, certain schools may come to be known as "math-oriented" or "literature-oriented." The more effective teachers in each of these areas may gravitate to the schools that emphasize these subjects – perhaps because they like teaching students motivated in their subjects, or because they believe being the math teacher in the math-oriented school confers higher prestige than being the math teacher in a the literature-oriented school. If these schools add more hours of instructions to their favorite subject of specialization, then teachers' quality will confound the effect of instruction time. Such specialized-focus schools almost always select their students based on subject-specific ability and motivation and often also use tracking. As a result, the robustness of the evidence presented above regarding patterns of admission and student selection criteria is especially reassuring; the effect of instruction time that I estimate is not capturing other effects such as teachers' quality that might skew the results.

The lack of any large, discernable differences in the effect of hours of instruction by admission or tracking policies of schools suggests that our estimates are not biased by unobservables correlated with sorting or selection of pupils based on subject-specific hours of instruction. Table 5 provides further

evidence to support this conclusion. First, I add to the regressions certain control variables, indicators of whether the school offers a special study program in science or math that may attract students with special interest and ability in those subjects. The first set of controls is based on question 20 in the PISA school questionnaire. It consists of indicators for school activities that promote student engagement with science (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, 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 columns 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. This suggests that the fact that many schools offer special programs and activities in science and math is not source of concern for bias.

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 specialized schools with a particular strength in a given subject are less likely to struggle to find and hire qualified teachers. The OLS and pupil fixed effect estimates are presented in columns 3 and 4, respectively, of Table 5. They are almost identical to those presented in columns 1-2 of Table 3. I also estimate the potential effect in other ways, through various samples stratified by the extent of lack of qualified teachers. For example I test the matter by including only schools that report lack of qualified teachers in at least two subjects, or through a sample including 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 subgroups. 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 percent) 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 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 language classes on boys 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 certain family characteristics: the average years of schooling of both parents, and immigrants, first generation and second generation. The productivity of instructional time is clearly higher (35 percent) for pupils from low-education families. The effect of additional instructional time for immigrants reveals a striking pattern. As compared to children of native parents, the benefits of additional instruction time are marginally higher (12 percent) for children of first-generation immigrants, but they are much higher (30 percent) for second-generation immigrants.

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 countries. The mean test scores for this sample are: 472.4 for math, 480.4 for science, 458.3 for language.(See Table A2 in the online appendix.). The standard deviations in the pupil-level distribution of test scores are 97.8 in math, 97.9 in science and 105.0 in language – similar to those in the OECD sample.

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 is much more important.

The estimate for girls is again higher (26 percent) than for boys, and it is much higher (33 percent) 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 is also among the highest in the overall sample. Their mean test scores are 21 percent lower than in the OECD countries. The mean test scores for this sample of developing countries are: 398.5 in math, 403.4 in science, and 397.1 in language. The standard deviation in the pupil level distribution of test scores is around 100 in the three subjects.

The estimates show much lower instructional time productivity 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- pupil 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 percent higher than for boys) and of immigrants (X percent). However, in this sample the effect of additional instructional time is lower by 26 percent 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 can be explained or eliminated by bridging the gap in instructional time and in its productivity in the different subjects. The mean instructional times in math, science and language in the rich countries are 3.5, 3.1 and 3.5 hours per week; in the poor countries they are 3.5, 3.0 and 3.2 hours per week, respectively. The gaps in instructional time, where they exist, are relatively small. The two distinct groups of countries spend the same amount of time on math, and wealthy countries spend only marginally more time on science (14 percent) and language (9 percent). At the same time, the mean test scores in the developing countries sample are much lower: 398.5 in math, 403.4 in science, and 397.1 in language. Therefore, the gap in mean test scores between the developing and the OECD developed countries is very large, more than 20 percent, 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

---

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

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.

V. Correlates of Productivity Differences of Instructional Time across Countries

The productivity of instructional time is endogenous, and it can be affected by a wide variety of factors - 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 teachers' and school principals' effort and efficiency, and, thus, influence the productivity of instructional time. Examples include accountability measures, such as publishing school league tables based on national tests, and 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.

Certain aspects and characteristics of this dimension of the educational landscape were a part of the 2006 PISA, through a survey of school head masters and the gathering of data on various indicators comparable manner for all the participating countries. Here I use 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.

A variety of PISA data relating to these aspects of the educational environment and its potential to affect productivity. For instance, PISA 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. Another index ranks the school's quality of educational resources, 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 salaries and potential raises, and determining and allocating the budget. The second index measures the school's responsibility for curriculum and assessment, including school independence in deciding the courses offered and their content, the textbooks used, and the method of assessing pupil performance.

17

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 9 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 percent) or teachers (29 percent). The means of the other indices are less interpretable.

In columns 2 and 3 of Table 9, 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 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 does not change much when I include all the interactions simultaneously in the regression. These results, shown in column 4 of Table 9, 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, more than 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

18

similarly evaluate their teachers. However, the 2006 PISA questionnaire data do not provide enough details to allow an understanding of how exactly such evaluations are done and whether they are used to reward school staff or affect their wages. As a result, 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 assess 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 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 influences staffing and the budget leads to higher productivity of instructional time. Second, having a board that influences instructional content and assessment has no measurable effect on the productivity of instruction in school. This evidence, presented in column 5 of Table 9, strengthens the overall findings that school autonomy in personal and budgetary issues is conducive to enhancing pupils' learning and achievement; again, there is no parallel evidence with respect to school pedagogic autonomy.

VI. Conclusions

In this paper, I empirically measure the effects of instructional time on students' academic achievement. The evidence from a sample of 15-year-old students from more than 50 countries consistently shows that additional 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. 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 average test scores by about 0.15 of the within-pupil standard deviation and by about 0.07 of the between-pupil standard deviation. From a policy perspective, any evaluation of the merits of adding instructional time should take into account its cost relative to other potentially beneficial inputs or interventions. Policymakers would be advised to consider that adding instructional time in a given subject may be associated with beneficial spill-over effects by leading to more demanding and advanced coursework. For example, if high school students from a different country/school spend twice as much class time on math than students from a different country/school, they are much more likely to cover algebra rather than just geometry. Such an increase in the level of challenge in coursework may have increased performance in PISA along with the effect

realized solely through more time devoted to subjects. The PISA data do now allow disentangling these two channels of effect of change in instruction time.[7]

The estimated effect of additional instructional time is strikingly lower for the sample of developing countries, the 2006 PISA results show. The gains in developing countries were only half of the size of those in developed countries. The results are all the more worrying because the developing countries included in the PISA sample (Chile, Argentina and Thailand, for example) 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, 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 PISA sample. The significant association between instructional time productivity and the structure and working environments of educational systems in OECD countries points toward a path for improvements in all nations. For developing countries in particular, one avenue for enhanced educational and economic progress clearly lies in closing this gap in productivity of instructional time.

VII. References

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.

Goodman Joshua, "The Labor of Division: Returns to Compulsory Math Coursework", Harvard Kennedy School, November 2009.

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.

Hansen Ben "School Year Length and Student Performance: Quasi-Experimental Evidence", University of California Santa Barbara, 2008.

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.

---

[7] See Goodman (2009) for a discussion of the literature on the effects of coursework on achievement and earnings.

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.

Lavy Victor, "School Resources, Time on Task and Students' Performance and Behavior", NBER Working paper Number ???, 2012.

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

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.

Marcotte Dave E. and Steven Hemelt "Unscheduled Closings and Student Performance," with. 2008. *Education Finance and Policy*. v. 3(3), pp. 316-38.

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.

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.

I. Introduction

      The amount of time students spend in public schools varies widely from one country to another. For example, among European countries such as Belgium, France and Greece, 15-year-old pupils receive an average of more than 1,000 hours per year of total compulsory classroom instruction, while in England, Luxembourg and Sweden the average is only 750 hours per year.[1] Children aged 7-8 in England, Greece, France and Portugal, receive an average instructional time of more than 800 hours per year, while in Finland and Norway, they receive less than 600 hours. Similar differences among countries exist in the number of classroom lessons per week in different subjects as evident from data from the 2006 Program for International Student Assessment (PISA), a unique international education survey of 15-year-old students conducted by the Organization for Economic Co-operation and Development (OECD) and designed to allow for cross-country comparisons. Pupils aged 15 in Denmark, for example, receive 4 hours of instruction per week in math and 4.7 hours in language, while pupils of the same age in Austria receive only 2.7 hours of weekly classroom lessons in math and 2.4 hours in language, the data show. Overall, total weekly hours of instruction in math, language and science is 55 percent higher in Denmark (11.5 hours) than in Austria (7.4 hours). Similar magnitudes of disparities in instructional time appear among the Eastern European and developing countries included in PISA 2006. Can these large differences explain some of the differences across countries in pupils' achievements in different subjects? What is the likely impact of changes in instructional time? This latter question is of policy relevance in many countries. For example, U.S. President Barack Obama has argued in recent years that American children should go to school longer – either staying later in the day or attending later into the summer or both. He has spoken of this goal, of extending the school week and year, as a central element in his proposed education reform.[2] An example of implementation of these ideas is the forty schools across Chicago that signed on to Mayor Rahm Emanuel's longer school day initiative earlier this year. Students in these schools will have an extra 90 minutes of instruction every day.[3]

---

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

[2] President Barack Obama said (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 — to have a chance to compete for jobs and paychecks against foreign children. He said "Our children spend over a month less in school than children in South Korea…. The challenges of a new century demand more time in the classroom. 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.

[3] See http://www.suntimes.com/news/education/7757540-418/chicago-public-schools-system-launches-longer-school-day-website.html.

While research in recent years provides convincing evidence about the effect of several inputs in the education production function[4], there is limited evidence on the effect of classroom instructional time. This evidence ought to be important for policymakers because increasing instructional time is a relatively simple option to implement, provided that resources are available, and, moreover, such an increase would be feasible in many countries.

The simple correlations and the simple OLS regression relationship between classroom instructional time per week and test scores of pupils aged 15 in the 2006 PISA exams in math, science and language are positive and highly significant. Of course, these correlations do not represent causal relationships because of potential selection bias and endogeneity issues. In this paper, I investigate the causal relationship between instructional time and pupils' knowledge in these subjects. 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. In this analysis of productivity, I focus on structural features such as accountability, autonomy and governance of schools.

I use 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 percent of the world economy took part in PISA 2006. The study focused on science, but the assessment also included language and mathematics. In addition, it collected data on student, family and institutional factors, including information about the amount of instructional time per week in each of the subjects tested. I exploit these features of the data to identify the effect of instructional time on academic achievement. First, having test scores in multiple subjects for each student, and relatively large variation in instructional time across subjects allows me to use within student estimation of the effect of instructional time. At the same time, I am able to control 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. Second, the considerable within student variation in instructional time allows for additional analysis. 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 hours per week in Iceland. The respective figures in science are 1.8 in the Netherlands and 3.6 in New Zealand, while in language they are 2.4 in Austria and 4.7 in Germany. I use this significant variation to

---

[4] For example, studies on class size (e.g. Kruger 1999, Hoxby 2000), teachers' training and certification (Kane Rockoff and Staiger, 2007), remedial education (Jacob and Lefgren, 2004), teacher effect (Rockoff, 2004, Rivkin, Hanushek, and Kain, 2005), and computer aided instruction (Barrow, Markman and Rouse, 2009).

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 subjects together.

Numerous studies have evaluated 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 United States and found insignificant effects, perhaps due to limited variation in the length of the school year there, or perhaps also due to the omission of other variables. Card and Krueger (1992), and Betts and Johnson (1998) used state-level data in the United States to examine the same effect and found a positive significant effect on earnings, perhaps because they studied earlier periods when length of the school year varied more widely and because state-level data may be of sufficient size to contend with the potential effect of unobserved heterogeneity Card and Krueger also presented results controlling for state effects, showing that the positive effect of a longer school year length vanishes within states and conditional on other school quality variables. Lee and Barro (2001) examined 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 found no effects of the length of the school year on internationally comparable test scores. A more recent study by Wöessmann (2003), which also analyzed cross-country test score data, corroborated this finding. Additional instructional time had an effect that is significant but of negligible size, he reported. However, these two studies attempted 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), overcame potential selection and endogeneity problems by using a West-German short school year in 1966-67 which reduced those school years by about two-thirds for some students. This change increased grade repetition in primary school, and lowered enrollment in higher secondary school tracks, but it had no adverse effect on earnings and employment later in life. Hansen (2008) reported that more instructional time due to fewer snow-related school day cancelations and delayed testing increased student performance in Colorado and Maryland. Marcotte and Hemelt (2008) reported that substantial snowfall (leading to fewer days spent at school) is associated with lower pupils' performance in Maryland.

The results I present in this paper show that instructional time has a positive and significant effect on the academic achievements of pupils. However, the estimates are much lower than the "naïve" OLS estimates which overstate the extent to which countries like the United States might "catch up" to

other developed nations by increasing instructional time. The size of the estimated effects is modest to large. On average, a one-hour increase per week in math, science or language instruction raises the test score in these subjects by 0.15 of a standard deviation of the within student distribution of test scores. The effect is larger for girls, immigrants and pupils from low socioeconomic status families. Estimates based on the sample of the former Soviet Eastern European countries are very similar to the average effect obtained from the sample of OECD developed countries. However, the evidence based on a sample of developing countries suggests a much lower effect of additional 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. Echoing the OECD results, the effect is much larger for girls, immigrants and pupils from disadvantaged backgrounds. Overall, the main results presented in this paper are very robust to a variety of robustness checks with respect to the identification assumptions and to threats to their validity. In addition, this evidence is almost identical to related results I obtained with data from Israel using two identification strategies (Lavy 2012). In that paper, the first method was identical to the one used in this paper, and the second stemmed from a natural experiment that provided an opportunity to exploit a sharp change in the method of funding schools in Israel. Remarkably, both methods led to identical results about the effect of instruction time on test scores in core subjects. The results from the research in Israel, and the results from the research presented here in an international context, are, remarkably, almost identical.

In the second part of this 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. However, I am able to estimate the effect of their interactions with instructional time in each subject. The evidence suggests 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. These findings emphasize the importance of quality and quantity of instructional time in bridging the gaps in student achievement across countries.

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

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 receive more instructional time than others. Of course, it could be that at the school level, such variation is not purely random. However, 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,

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

Where A_{ijk} is the achievement of the ith student, in the jth school, in the kth subject, H_{kj} is instructional time in the kth subject in the jth school, X is a vector of characteristics of the ith student in the jth school and S_j is a vector of characteristics of the jth 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 one to control for a number of sources of potential biases related to unobserved characteristics of the school, the student, or their interaction. In some situations, students might be placed or be sorted according to their ability across schools that provide more (or less) instructional time in some subjects. If, for example, more able students attend better schools that provide more instructional hours overall in each subject, it would cause γ to be biased downward unless one also were to account for student and school fixed effects. The bias would have an opposite sign if the less able students are exposed to more instructional time. However, identification of the effect of instructional time based on a comparison of the performance of the same student in different subjects is immune to these potential biases.

Three remarks about this identification strategy merit mention. First, the necessary assumption for this identification strategy is that the effect of instructional time is the same for all subjects, implying

5

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 language 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 have high ability in math, for example, may select or be placed in a school that specializes in math, and they may receive more instructional time in math as a result. However, this issue may be less of a concern in the sample I use here. Such tracking is largely conducted 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. In addition, 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 whether the school sorts and selects students based on subject-specific considerations. For example, I observe in the PISA data information about whether admission to a school takes into account the student's academic record, or the student's need or desire for a special program. I also note whether a school uses tracking in forming classes, and whether it is a public or a private school. I assume that a school that does not use academic ability as an admission criterion, or does not take into account student's desire for a particular program, will most likely not select students on subject-specific considerations. Such selection is also less likely to take place in schools that do not use any form of tracking by ability, or in public schools. Indeed, the results that I present below are very similar across the various stratified samples based on schools' admission and tracking policies and on whether they are private or public.

III. Data

The Program for International Student Assessment, known by its acronym PISA, provides regular data on the OECD country education systems and the knowledge and skills of their students. 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. PISA does not examine students' mastery of specific school curricula. Instead, it evaluates students' ability to apply knowledge and skills in key subject areas and their ability to analyze reason and communicate effectively as they examine, interpret and solve problems. PISA measures student performance in language, 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 that 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 participating OECD country has a PISA Governing Board representative, 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 PISA surveys. For each survey, an international contractor (usually a testing and assessment agency) is 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 in each participating country. PISA has Subject Matter Expert Groups for its three key areas of testing – language, 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 international contractor and participating countries review the questions and carefully check them for cultural bias. PISA uses only those questions that have been approved unanimously.

Students answer a 20- to 30-minute background questionnaire, providing information about themselves, their homes, and their attitudes to learning. 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 the following question (see online appendix for the exact format of this question in the PISA 2006 student's questionnaire): "How much time do you typically spend per week studying the following subjects in regular lessons at school?" The student could choose from the following options: *No time, Less than 2 hours a week, 2 or more but less than 4 hours a week, 4 or more but less than 6 hours a week, or more hours a week.* I merged the first two options into one as *less than two hours a week* and used the other categories as is. I computed the school average in each subject using the mid values of each range. Because the PISA data allow for overall instruction time per week in all subjects to be combined, I have aggregated and averaged this information at the country annual level for all OECD countries. A high correlation emerges between these country means and the administrative data on total annual hours of instruction in secondary schooling as reported in the OECD report, "Education at Glance."

The means of instructional time in the developed OECD countries in math, science and language are 3.53, 3.06, and 3.54 hours per week, respectively. In the Eastern European sample, mean instructional time is lower than in the OECD countries in all subjects (3.30 for math, 2.77 for science and 3.08 in language). These Eastern European figures are similar to the mean instructional time in the developing countries sample (3.48 for math, 2.97 for science and 3.24 for language). Tables A1-A3 in the online appendix present the mean instructional time in each of the subjects for each of the countries included in the three samples of the 2006 PISA. As discussed in the introduction, large variation exists on many levels. Variation emerges among the countries that make up the three broad multi-country groups, and within individual countries, among different schools and among the various subjects. As a result, a significant proportion of the overall variation in instruction time is due to the variation within countries among schools. (Evidence is not shown here, but is available from the author.) The PISA data - which include instruction time per week for the three subjects studied here, all other subjects, and for the overall school week - explain some of the reasons behind the variations. As a result, the data show that a large part of the variation across subjects in instruction time is explained by variation across schools in total instruction time per week. Of course, the variation in total weekly instruction time across schools originates from many sources, for example differential government resource allocation to schools by their socio-economic background, differential contribution from parents and the community of local

school authority and more. Many of these factors vary at the school or country level, and our identification methodology, using the within school and pupil estimation strategies, will account for them.

Certain additional issues regarding the data and my methodologies bear mention. Different countries vary in the length of the school year (number of weeks of study), but is not a limitation in this case since the within school and pupil estimation strategies control for the length of the school year. Regarding instruction time, the measurements reflect the average in the school, not the hours reported by each student. And, it is worth underscoring that the variation in subject instruction hours across students in the same school is not due to an endogenous decision of the students. This is because the PISA sample in each country includes students in 9th and 10th grades, levels at which the curriculum of study in most or all countries is completely compulsory. Students may choose some courses only in the latter part of high school, perhaps at 11th and 12th grades. Variation in subject instruction hours could stem from tracking practices in schools, however, and I address this issue by directly examining the sensitivity of my results to differences in tracking practices across schools.

Tables A1-A3 in the online appendix present the country means of the PISA test scores in each of the subjects for each of the countries. Table A4 in the online appendix present the means of the PISA test scores 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 of the same pupil to explain. The average weekly 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 in the online appendix 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 score regressions based on the sample of the OECD countries. For each subject, I report estimates from three specifications: 1) without any controls, 2) with country fixed effects, and 3) 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 of measures 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 on PISA test scores 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 language 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 an 5.76 points on the PISA test. 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.

It is important to note here that in a related paper (Lavy, 2012) I report results based on the same method I use in this paper and also based on an alternative identification strategy that relies on a natural experiment that exploit a sharp change in the funding method of schools in Israel. Remarkably, both of these two methods lead to identical point estimates for the effect of instructional time on core subjects test scores, and these two sets of estimates from the previous paper are almost identical to the results presented in this paper in Table 3. It is also important to emphasize that the natural experiment method that I used in the Israeli context allowed me to test some of the central assumptions I make while using the pupil fixed effect identification method. These tests support the assumption that the effect of additional hours of instruction on each of the subjects is very similar and that there are no cross-subjects spillover effects. These previous findings strengthen the overall credibility of the results I present in this paper and also their causal interpretation.

10

The estimates of the indicators of instructional time presented in lower part of panel A of Table 2 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 percent higher than the respective estimate obtained from pooling all three subjects together. However, pooling math and language test scores yields an estimate of 7.42 and pooling science and language yields an estimate of 4.27. The pattern emerging from the OECD country group does not permit me to conclude whether the average productivity of instructional time is lower for any of the three subjects. 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 and checks for threats of identification

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 based on a sample of schools instead of pupils. I present these results in online appendix Table A5. I obtain the variables used by collapsing the pupils' data to the school-level respective means. The pattern in this table is very similar to that of 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 student micro data. The estimate based on all three subjects is 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 by including interactions between the subject dummies and pupil's characteristics. The estimates from this more flexible specification approach (not shown here, but available from the author) are very similar to those presented in Table 3, though overall they are about 10 percent lower.

---

[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.

The first robustness check concerns whether the evidence in column 2 Table 3 reflects some subject-specific selection and sorting in some schools. It is based on the data available from the PISA school questionnaire about how much consideration is given in the admission decisions to student's academic record and whether placements tests are used in this process. (See the exact wording of the question 19 in the online appendix.) I expect that the validity of the identification strategy will be underscored here. As a result, sample of schools that do not pay any attention to applicants' previous academic records and do not use any admission exams should not be sensitive to endogenous sorting and selection. 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 in which past academic achievements are irrelevant for admission is the largest, accounting for about two-thirds of the whole sample. The estimates from this sample are only marginally different from those obtained from full sample: the OLS estimate 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. I also estimated the effect of hours of instruction based on a sample of schools that do not use a student's desire for a particular program as a criterion for admission. This information is also based on the PISA school questionnaire question 19. The results based on this sample are very similar those obtained from the full sample.

Another potential source of selection bias is tracking pupils into classes within schools according to their ability. One can expect that schools practicing 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, a bias in the estimated effect of hours of instruction will occur. In Table 4, I present results for three different samples distinguished by schools' tracking policies. (See the exact wording of the question in the online appendix.) Columns 1-2 show the estimates for a sample of schools that practice tracking at the class level – those that group students in classes according to their ability, that is. Columns 3-4 show the results based on a sample of schools that sort pupils into different ability study groups within classes. Columns 5-6 show the results based on a sample of schools that do not practice any form of pupil tracking. 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 ability classes 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 significantly different from the point estimates obtained from the full sample; 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 private schools are more likely to base admission on previous academic record and on additional exams, and to track pupils by ability. To address 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 language test scores is 6.09 (sd=0.428), just barely higher than the estimate from a sample that included also the private schools.

Overall, it is apparent that the results are not driven by potential selection and sorting of students based on subject-specific considerations related to selective admission or tracking pupils in classes by abilities, as underscored by the findings presented in Tables 3 and 4 and the public schools-only sample. This is an important result because schools that admit pupils based on academic record or that track students by ability 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. For example, certain schools may come to be known as "math-oriented" or "literature-oriented." The more effective teachers in each of these areas may gravitate to the schools that emphasize these subjects – perhaps because they like teaching students motivated in their subjects, or because they believe being the math teacher in the math-oriented school confers higher prestige than being the math teacher in a the literature-oriented school. If these schools add more hours of instructions to their favorite subject of specialization, then teachers' quality will confound the effect of instruction time. Such specialized-focus schools almost always select their students based on subject-specific ability and motivation and often also use tracking. As a result, the robustness of the evidence presented above regarding patterns of admission and student selection criteria is especially reassuring; the effect of instruction time that I estimate is not capturing other effects such as teachers' quality that might skew the results.

The lack of any large, discernable differences in the effect of hours of instruction by admission or tracking policies of schools suggests that our estimates are not biased by unobservables correlated with sorting or selection of pupils based on subject-specific hours of instruction. Table 5 provides further

evidence to support this conclusion. First, I add to the regressions certain control variables, indicators of whether the school offers a special study program in science or math that may attract students with special interest and ability in those subjects. The first set of controls is based on question 20 in the PISA school questionnaire. It consists of indicators for school activities that promote student engagement with science (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, 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 columns 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. This suggests that the fact that many schools offer special programs and activities in science and math is not source of concern for bias.

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 specialized schools with a particular strength in a given subject are less likely to struggle to find and hire qualified teachers. The OLS and pupil fixed effect estimates are presented in columns 3 and 4, respectively, of Table 5. They are almost identical to those presented in columns 1-2 of Table 3. I also estimate the potential effect in other ways, through various samples stratified by the extent of lack of qualified teachers. For example I test the matter by including only schools that report lack of qualified teachers in at least two subjects, or through a sample including 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 subgroups. 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 percent) 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 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 language classes on boys 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 certain family characteristics: the average years of schooling of both parents, and immigrants, first generation and second generation. The productivity of instructional time is clearly higher (35 percent) for pupils from low-education families. The effect of additional instructional time for immigrants reveals a striking pattern. As compared to children of native parents, the benefits of additional instruction time are marginally higher (12 percent) for children of first-generation immigrants, but they are much higher (30 percent) for second-generation immigrants.

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 countries. The mean test scores for this sample are: 472.4 for math, 480.4 for science, 458.3 for language.(See Table A2 in the online appendix.). The standard deviations in the pupil-level distribution of test scores are 97.8 in math, 97.9 in science and 105.0 in language – similar to those in the OECD sample.

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 is much more important.

The estimate for girls is again higher (26 percent) than for boys, and it is much higher (33 percent) 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 is also among the highest in the overall sample. Their mean test scores are 21 percent lower than in the OECD countries. The mean test scores for this sample of developing countries are: 398.5 in math, 403.4 in science, and 397.1 in language. The standard deviation in the pupil level distribution of test scores is around 100 in the three subjects.

The estimates show much lower instructional time productivity 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- pupil 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 percent higher than for boys) and of immigrants (X percent). However, in this sample the effect of additional instructional time is lower by 26 percent 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 can be explained or eliminated by bridging the gap in instructional time and in its productivity in the different subjects. The mean instructional times in math, science and language in the rich countries are 3.5, 3.1 and 3.5 hours per week; in the poor countries they are 3.5, 3.0 and 3.2 hours per week, respectively. The gaps in instructional time, where they exist, are relatively small. The two distinct groups of countries spend the same amount of time on math, and wealthy countries spend only marginally more time on science (14 percent) and language (9 percent). At the same time, the mean test scores in the developing countries sample are much lower: 398.5 in math, 403.4 in science, and 397.1 in language. Therefore, the gap in mean test scores between the developing and the OECD developed countries is very large, more than 20 percent, 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

---

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

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.

V. Correlates of Productivity Differences of Instructional Time across Countries

The productivity of instructional time is endogenous, and it can be affected by a wide variety of factors - 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 teachers' and school principals' effort and efficiency, and, thus, influence the productivity of instructional time. Examples include accountability measures, such as publishing school league tables based on national tests, and 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.

Certain aspects and characteristics of this dimension of the educational landscape were a part of the 2006 PISA, through a survey of school head masters and the gathering of data on various indicators comparable manner for all the participating countries. Here I use 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.

A variety of PISA data relating to these aspects of the educational environment and its potential to affect productivity. For instance, PISA 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. Another index ranks the school's quality of educational resources, 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 salaries and potential raises, and determining and allocating the budget. The second index measures the school's responsibility for curriculum and assessment, including school independence in deciding the courses offered and their content, the textbooks used, and the method of assessing pupil performance.

17

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 9 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 percent) or teachers (29 percent). The means of the other indices are less interpretable.

In columns 2 and 3 of Table 9, 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 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 does not change much when I include all the interactions simultaneously in the regression. These results, shown in column 4 of Table 9, 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, more than 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 do not provide enough details to allow an understanding of how exactly such evaluations are done and whether they are used to reward school staff or affect their wages. As a result, 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 assess 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 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 influences staffing and the budget leads to higher productivity of instructional time. Second, having a board that influences instructional content and assessment has no measurable effect on the productivity of instruction in school. This evidence, presented in column 5 of Table 9, strengthens the overall findings that school autonomy in personal and budgetary issues is conducive to enhancing pupils' learning and achievement; again, there is no parallel evidence with respect to school pedagogic autonomy.

VI. Conclusions

In this paper, I empirically measure the effects of instructional time on students' academic achievement. The evidence from a sample of 15-year-old students from more than 50 countries consistently shows that additional 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. 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 average test scores by about 0.15 of the within-pupil standard deviation and by about 0.07 of the between-pupil standard deviation. From a policy perspective, any evaluation of the merits of adding instructional time should take into account its cost relative to other potentially beneficial inputs or interventions. Policymakers would be advised to consider that adding instructional time in a given subject may be associated with beneficial spill-over effects by leading to more demanding and advanced coursework. For example, if high school students from a different country/school spend twice as much class time on math than students from a different country/school, they are much more likely to cover algebra rather than just geometry. Such an increase in the level of challenge in coursework may have increased performance in PISA along with the effect

realized solely through more time devoted to subjects. The PISA data do now allow disentangling these two channels of effect of change in instruction time. [7]

The estimated effect of additional instructional time is strikingly lower for the sample of developing countries, the 2006 PISA results show. The gains in developing countries were only half of the size of those in developed countries. The results are all the more worrying because the developing countries included in the PISA sample (Chile, Argentina and Thailand, for example) 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, 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 PISA sample. The significant association between instructional time productivity and the structure and working environments of educational systems in OECD countries points toward a path for improvements in all nations. For developing countries in particular, one avenue for enhanced educational and economic progress clearly lies in closing this gap in productivity of instructional time.

VII. References

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.

Goodman Joshua, "The Labor of Division: Returns to Compulsory Math Coursework", Harvard Kennedy School, November 2009.

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.

Hansen Ben "School Year Length and Student Performance: Quasi-Experimental Evidence", University of California Santa Barbara, 2008.

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.

---

[7] See Goodman (2009) for a discussion of the literature on the effects of coursework on achievement and earnings.

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.

Lavy Victor, "School Resources, Time on Task and Students' Performance and Behavior", NBER Working paper Number ???, 2012.

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

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.

Marcotte Dave E. and Steven Hemelt "Unscheduled Closings and Student Performance," with. 2008. *Education Finance and Policy*. v. 3(3), pp. 316-38.

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.

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	
		Student		Student		Student
	OLS	FE	OLS	FE	OLS	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 (all three subjects pooled)

	Tracking By Class		Track In Class		No Tracking	
		Student		Student		Student
	OLS	FE	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)
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)
Number of students	212,169		201,138		160,188	

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. 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 (all three subjects pooled)

	Control Added For			
	Special Science School Activities		Scarcity of Teachers in Each Subject	
		Student		Student
	OLS	FE	OLS	FE
	(1)	(2)	(3)	(4)
Hours of instruction	18.37	5.59	19.58	5.75
	(0.73)	(0.39)	(0.72)	(0.37)
Number of students	460,734		224,508	

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. 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

	Boys		Girls	
		Student FE		Student FE
	OLS		OLS	
	(1)	(2)	(3)	(4)
Hours of instruction	20.25	4.99	18.62	5.62
	(0.86)	(0.40)	(0.77)	(0.41)
Number of students	224,508		236,226	

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. 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

	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)
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)
Number of students	235,539		225,195		23,103		22,092	

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

Index	Index's Mean (1)	Separate Spec.		Joint Spec.	
		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 Students
			Mathematics	Science	Reading	All (sum)	Mathematics	Science	Reading	All (average)	
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 Students
			Mathematics	Science	Reading	All (sum)	Mathematics	Science	Reading	All (average)	
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

#	Country	Code	Hours of Instruction per week				Pisa Score				Number of Students
			Mathematics	Science	Reading	All (sum)	Mathematics	Science	Reading	All (average)	
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	
		School		School		School
	OLS	FE	OLS	FE	OLS	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.