NBER WORKING PAPER SERIES

INSTRUCTION TIME, CLASSROOM QUALITY, AND ACADEMIC ACHIEVEMENT

Steven G. Rivkin
Jeffrey C. Schiman

Working Paper 19464
http://www.nber.org/papers/w19464

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2013

We thank Marcus Casey, Robert Kaestner, Cuiping Long, Darren Lubotsky, Ben Ost, Houston Stokes, Javaeria Qureshi and participants at the UIC economics research lunch and the National Institute for Educational Evaluation in the Ministry of Education of Spain for helpful comments. Rivkin thanks the US Department of Education, Institute for Education Sciences for financial support. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Instruction Time, Classroom Quality, and Academic Achievement
Steven G. Rivkin and Jeffrey C. Schiman
NBER Working Paper No. 19464
September 2013
JEL No. I21,I24,I25,I28

ABSTRACT

Many countries, American jurisdictions and charter schools have recently embraced longer school days or more time devoted to core academic classes. Recent research generally supports the notion that additional time raises achievement, though difficulties isolating an exogenous source of variation raise questions about the strength of much of the evidence. Moreover, it seems likely that the magnitude of any causal link between achievement and instruction time depends upon the quality of instruction, the classroom environment, and the rate at which students translate classroom time into added knowledge. In this paper we use panel data methods to investigate the pattern of instruction time effects in the 2009 Programme for International Student Assessment (PISA) data. The empirical analysis shows that achievement increases with instruction time and that the increase varies by both amount of time and classroom environment. These results indicate that school circumstances are important determinants of the likely benefits and desirability of increased instruction time.

Steven G. Rivkin
Department of Economics
University of Illinois at Chicago
601 South Morgan UH725 M/C144
Chicago, IL 60607
and NBER
sgrivkin@uic.edu

Jeffrey C. Schiman
University of Illinois at Chicago
jschim2@uic.edu

I. Introduction

The belief that increased time on task raises output would go unchallenged in most settings, but public schooling is an exception. Arguments of extensive inefficiencies that dampen the return to additional time or spending are widespread. Nonetheless, many countries and American jurisdictions have recently embraced longer school days or more time devoted to core academic classes. The conceptual appeal is clear: additional time allows teachers to "cover more material and examine topics in greater depth and in greater detail, individualize and differentiate instruction, and answer students' questions" (Farbman 2012).

Many point to KIPP Academy schools for evidence of the benefits of extended time in class. Through a longer school day and Saturday school, instruction time averages around 1,700 hours per year in KIPP schools, roughly 60% more than the US average, and evidence suggests that KIPP students significantly outperform similar students in regular public schools (Farbman 2011).[1] Of course KIPP academy schools differ along other dimensions as well so it is difficult to isolate the specific mechanisms that account for KIPP's apparent success.[2]

Recent research on instruction time generally supports the notion that additional time raises achievement, though difficulties isolating an exogenous source of variation raise questions about the strength of much of the evidence.[3] To illustrate the empirical difficulty consider the difference between academic and vocational secondary schooling.

[1] See http://www.kipp.org/our-approach for more details about how KIPP academy schools operate.

[2] In a recent paper, Angrist et al. (2010) attempt to isolate the contributions of various factors to the educational success of KIPP students.

[3] Recent work includes (Coates 2003; Gijselaers and Schmidt 1995; Kuehn and Landeras 2012; Lavy 2010; Lavy 2012; Mandel and Süssmuth 2011; Marcotte 2007; Marcotte and Hemelt 2008; Roland G. Fryer 2011; and Wiermann 2005). Lavy (2010) emphasizes the identification problem and adopts an empirical approach that provides a foundation for our work in this paper.

Academic schools typically spend more time in mathematics and language arts instruction and boast higher achievement than vocational schools, but instruction time differences across sectors does not provide a valid source of identification because of the positive selection into academic schools. Alternatively, instruction-time variation resulting from the desire to supplement the education of lower achievers would tend to produce downward biased estimates. Thus it is not possible to determine a priori if the simple correlation between achievement and instruction time overstates or understates the causal relationship.

Moreover, it seems likely that the magnitude of any causal link between achievement and instruction time depends upon the quality of instruction, the classroom environment, and the rate at which students translate classroom time into added knowledge. For example, expanded instruction time in response to poor mathematics achievement may have little impact if an ineffective curriculum, inadequate teacher subject matter knowledge, or disruptive behavior led to the low achievement in the first place. Furthermore, even if existing class time is effective, there may be decreasing marginal benefits to additional minutes if the quality of instruction, classroom environment, or student effort diminishes with time.

In this paper we focus on such potential heterogeneity and investigate the pattern of instruction time effects using the 2009 Programme for International Student Assessment (PISA) data. We are particularly interested in the mediating effects of teacher quality and the classroom environment and the character of any diminishing returns. In order to overcome biases introduced by the non-random allocation of instruction time and unobserved differences in school quality, we build on the work of Lavy (2010) and use within-school variation across subjects or grades to identify the effects.

The appeal of a focus on within school variation across subjects is the fact that students taking both mathematics and language arts (for example) bring the same general skills and experience as well as the same school environment for each subject. Therefore neither student heterogeneity in general ability and work habits nor general school quality will contaminate estimates of instruction-time effects identified from average instruction-time differences between subjects. This leaves only subject specific factors related to instruction time as potential confounding factors, and the focus on school average rather than individual instruction time differences mitigates problems introduced by consideration of subject-specific skills when determining course placement.

The possibility that school quality or school average ability differences by subject could exist and either precipitate or result from instruction-time differences remains, and therefore we also consider grade differences within subject and school as an alternative source of variation. This comparison is complicated by the fact that 9^{th} graders have one fewer year of schooling than 10^{th} graders, and clearly it is not the same students in both grades. Nonetheless as we discuss below, as long as skill and instruction-time differences across cohorts are not related, the within-subject variation produces lower bound estimates.

The empirical analysis shows that achievement increases with instruction time and that the increase varies by both amount of time and classroom environment. First, there is evidence of diminishing returns, though the rate of decrease appears to be quite gradual. Second, there is evidence that better classroom environment as indexed by responses to questions about student behavior and student-teacher interactions also appears to raise the benefit of additional instruction time. These results indicate that school circumstances are important determinants of the likely benefits and desirability of increased instruction time.

II. Data

The data come from the 2009 Programme for International Student Assessment (PISA), a survey and assessment administered to fifteen year old students around the world. At least 150 schools are randomly selected in each country based on a stratified sample design, and within each school, 35 students that are 15 years old are sampled at random. Each student is assessed in mathematics, science, and language arts and then answers a set of questions about family background, school environment, home environment, and study habits.[4] The PISA test focuses on knowledge applications and is thought to be highly informative about the quality of preparation for higher education and the labor market. A representative from each school also provides information on staff, environment, and pedagogical and human resource practices.

We focus on mathematics and language arts separately from science and language arts because the quality of mathematics education likely affects performance on the science examination. Because our research design identifies instruction time effects on the basis of between subject differences in test scores and instruction time, this potential

[4] Each student is assigned five achievement measures for each subject called plausible values. To estimate regressions using plausible values, one must estimate separate regressions with each of the five plausible values and then average across the estimates. Estimating separately by plausible value may give different results in smaller samples (e.g. samples less than 6,000), but in samples larger than 6,000, practically speaking, the estimates will be very similar (Adams and Wu 2002). Here we present estimates based on the first plausible value through the estimates are insensitive to choice of plausible value. Upon request, tables are available from the authors.

spillover is especially problematic. We expect there to be little mathematics and language arts spillover and little science to language arts spillover at the high school level. In the appendix, we provide results based on within school and grade science and language arts differences.

The PISA test was administered in 2000, 2003, 2006, and 2009, and we use the 2009 wave because of the richness of information on instruction time and availability of measures about classroom environment and instructional quality. In 2009, students are asked the number of mathematics, science, and language arts classes attended per week and the length in minutes of an average class. This potentially permits us to separately identify the effects of additional classes per week and minutes per class, but in reality there is little variation in average class length across subjects. To calculate average instruction time for each subject, grade, and school, we multiply the average weekly number of classes by the length of an average class.

The variability and range of responses to the instruction time questions does raise concerns about data quality and measurement that may not be fully addressed by aggregation. Non-trivial numbers of students report more than ten classes per week in a subject or class lengths of over two hours. In order to mitigate errors in variables bias, we exclude information on classes per week or average length of classes from the school average calculations if reported number of classes exceeds ten or class length exceeds two hours.[5] These restriction set to missing approximately 1% of student reported information on instruction time.

[5] We test the sensitivity of our estimates to different restrictions on weekly number of classes and average class lengths. In general, the estimates are insensitive to how we restrict the data. Upon request, tables are available from the authors.

By comparison, the 2006 data used in Lavy (2010) (at the time of his writing, 2009 data was not yet available) report instruction time categories only. In 2006, students responded to weekly time spent in each subject in five intervals: "*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; and 6 or more hours a week*." A clear disadvantage of this taxonomy is the absence of detailed information on numbers of classes and minutes. In addition, the taxonomy produces instructional time distributions that differ substantially from those for 2000 and for 2009. While the majority of weekly instruction time would fall in the *2 to < 4 hour* category based on survey responses in 2000 and 2009, the distribution is more evenly split between *2 to < 4 hour* and the *4 to < 6 hour* categories in 2006 (not shown), raising concerns about the accuracy of student responses.

We use factor analysis to generate the indexes of classroom environment and teacher quality based on a series of questions listed in Table 6 along with the factor loadings. The index of the quality of the classroom environment comes from a series of questions to school representatives that ask to "what extent the learning of students is hindered by the following phenomenon." The phenomena include disruption, other aspects of student behavior, student-teacher interactions and other aspects of teacher behavior. Respondents could check "*not at all, some of the time, most of the time, or all of the time.*" A higher value on the index of quality reflects a classroom environment that is more conducive to learning.

The absence of direct measures of instruction quality leads us to focus on measures of the shortage of qualified teachers for subjects and the quality of teacher personnel practices. Such questions are not ideal instruments to measure the quality of instruction, as they are likely to provide fairly noisy information about teacher performance. Moreover, personnel practices are likely to be related to other aspects of school operations such as the processes through which curricula are chosen and budgets determined. Therefore the index of instruction quality may perhaps be interpreted more accurately as a measure of the quality of management practices.

PISA test booklets are designed so that not every student takes both a mathematics and language arts component. Instead, each student is randomly assigned one of twenty-one test booklets, fourteen of which contain both mathematics and language arts components.[6] Those who do not take a math or language arts component have their scores for these subjects imputed by the PISA test makers based on the available assessments.[7] Because our research design relies on information across subjects, prior to aggregation we drop students who do not take both math and language arts. This restriction drops approximately 30 percent of the sample. We also limit the analysis sample to the same set of observations used in all regressions, which drops approximately 3,060 school-by-grade-by-subject observations.

The main sample used in this analysis includes 47,580 school-by-grade-by-subject observations for 9th and 10th grade students in 16,154 schools in 72 countries. We focus on these two grades in order to minimize complications introduced by grade retention and to avoid cells with small numbers of students. Some components of the analysis restrict the sample to only schools with both 9th and 10th grades.

[6] A recent working paper by Borghans and Schils (2012) discusses in more detail the variation in PISA test booklets in the 2006 data.

[7] The variable "bookid" denotes which subjects are contained in each student's test booklet. We drop those with booklet ID 2, 4, 6, 13, 22, 24, and 26 because these booklets do not contain both math and language arts components. The subject clusters and book IDs are described at www.oecd.org/dataoecd/15/31/48580826.xls on Table 2.2

III. Empirical Model

This section describes the empirical framework used to investigate the effects of instructional time on achievement. Conceptually, the empirical framework must address potential biases introduced by confounding student and school factors and the likelihood that the benefits of additional time vary by both the quality of instruction and classroom environment. The association between instruction time on the one hand and both student and school factors results from the fact that instruction time is determined by family selection of schools, assignment of students to schools and courses of study, and systemic rules about school operations. Academically oriented students are much more likely to attend academic high schools that devote more class time to mathematics and language arts. Schools may assign higher achievers in a subject to courses that meet more often, or schools may assign struggling students to additional remedial sections. Governments concerned about poor performance in mathematics may mandate a minimum amount of instructional time, or governments with a strong commitment to mathematics may mandate more class time along with higher salaries and stronger teacher training. Finally, the analysis must consider possible endogenous family responses to realized school quality, as additional instructional time outside of school can substitute for lower or less productive school instruction time (Todd and Wolpin, 2003). Two things quickly become clear: instruction time is likely to be related to a number of factors that may themselves be determinants of achievement, and the direction of those relationships and therefore the direction of any bias from unobserved factors is ambiguous.

By comparison, potential heterogeneity in returns to additional instruction time is likely to be more predictable along at least two important dimensions. First, diminishing returns to additional time are likely to set in at some point due to fatigue. Second, extending the class time taught by an ineffective teacher is likely to yield little return, as is extending time in a classroom plagued by disruption or poor relations between students and teachers. It is these two dimensions that we explore in the empirical analysis.

IIIa. Baseline Empirical Model

Identification of the effect of instruction time on achievement requires exogenous variation that is not related to unobserved differences in students and schools. Existing research shows that available variables explain little of the variation in the quality of instruction and student skill, and therefore it is necessary to account for unobserved student and school factors. Fortunately, as Lavy (2010) points out, the testing of students in multiple subjects enables the use of panel data methods that account for differences in school and teacher quality, school climate, and student ability that span both subjects. The instruction time effect can be identified by the difference in time devoted to mathematics relative to language arts for each student, and all between student and school differences in the allocation of instruction time can be ignored.

A potential problem with comparisons between class time in language arts and mathematics is the purposeful placement of students into courses on the basis of subject specific skills and interest. Weaker mathematics students are more likely to be placed in lower level mathematics courses that could meet less frequently, and the data may not contain information on that could be used to control for underlying math skill. Therefore following Lavy (2010) we aggregate instruction time and test scores for each student to the school-by-grade-by-subject level. Such aggregation eliminates the potential confounding influence of within school variation in the mathematics-reading skill differential.

In order to highlight the key assumptions underlying the various fixed effects structures, we begin with a simple specification that ignores heterogeneity in the return to class time. Equation (1) models achievement A in subject k in grade g in school s in country c as a function of minutes per week of instruction M and a series of error components that capture interactions among country, school, grade and subject. Note that country is fully subsumed by school and is included as an interaction with grade and subject to highlight the potential importance of country policies and practices regarding curriculum, accountability, funding, and other factors.

$$A_{kgsc} = M_{kgsc}\lambda + \eta_s + \phi_g + \sigma_k + \zeta_c + \theta_{kgc} + \tau_{ks} + \omega_{gs} + v_{kgsc} \quad (1)$$

The school-by-grade fixed effect (ω) accounts for differences in average ability, level of disruption, and school quality that are common across subjects for students in a particular cohort, grade, school, and country. Therefore all subject invariant differences in academic skills and school quality at each grade level are removed which controls for the primary confounding factors. Only within school and grade instruction-time differences among subjects remain for identification. Note, importantly, that the school-by-grade fixed effects fully account for a range of subject invariant influences including national minimum school starting and leaving ages, school funding and governance structures, and family background.

The school-by-grade fixed effect does not account for differences among subjects in either instruction time or various other factors that could influence achievement. These include, but are not limited to, national curricula, the quality of instruction in one subject in relation to the other, and subject-specific student skills. The country-by-grade-by-subject fixed effect (θ) captures some such influences including national curricula, but it does not capture subject-specific abilities or instructional quality specific to schools that are related to instruction time. Our estimate of instruction time effects would be biased upward if the difference between school-by-grade average mathematics and language arts instruction time is positively related to the difference in average abilities in mathematics and language arts, as would be the case if analytically skilled students attended schools that devoted more time to mathematics instruction. A similar upward bias would arise if the instruction difference was positively related to the difference in the quality of mathematics versus language arts instruction. Of course, a negative relationship between instruction time in a subject on the one hand and ability or instructional quality on the other would introduce a negative bias.

It is not clear whether confounding subject-specific factors introduce bias. Nonetheless, the availability of multiple grades per school enables an alternative approach that accounts directly for school-by-subject factors. Rather than identifying effects on the basis of within school-by-grade instruction-time differences across subjects, effects can be identified on the basis of instruction-time differences across grades for the same subjects. Essentially this amounts to including a school-by-subject fixed effect into Equation (1) and excluding the school-by-grade fixed effect.

If instruction time is significantly related to achievement, a larger instruction time difference in mathematics courses between 9th and 10th grade should be associated with a larger difference in test scores. As opposed to the school-by-grade fixed effect specifications, the school-by-subject specification accounts for subject-specific differences among schools in both school quality and average student skills. However, this advantage is potentially offset by the fact that 9th and 10th grade mathematics and language arts scores are produced by different students who are not in the same grade cohort. Importantly, the strict exogeneity assumption does not require equality in average ability or the quality of instruction across grades but only that any differences are not related to differences in grade-average instruction time. As long as the course schedule for a subject and grade is not responsive to grade differences in student or teacher skills one would not expect performance-induced changes in the course schedule to occur and introduce bias.

However, inadequate treatment of learning dynamics can introduce correlation between lagged instruction time and the error which violates the strict exogeneity assumption. Unless learning fully depreciates each year, a better 9th grade education will raise achievement in 10th grade as well as 9th grade. Therefore additional 9th grade instruction time will tend to increase 10th grade achievement. As Meghir and Rivkin (2011) illustrate, fixed effect estimates that are based on achievement differences across grades will tend to introduce a downward bias in models that compare achievement in the respective grades and do not account for prior achievement. Even though our analysis does not compare achievement of the same student in successive grades, persistence in the structure of instruction time across cohorts would still attenuate estimates based on instructional time differences between 9th and 10th grade. Importantly, the direction of bias introduced by this specification error is unambiguously toward zero, meaning that the school-subject fixed effect estimates are likely to provide a lower bound of the true average instruction time effect.

Persistence in the structure of instruction time across grades also complicates the interpretation of the estimates from the school by grade specifications discussed above. Specifically, identification based on mathematics and language arts instruction time differences in 10^{th} grade, for example, does not produce an unbiased estimate of the effects of instruction time on achievement in 10^{th} grade if the difference in 10^{th} grade is correlated with the difference in 9^{th} grade. Rather the estimate would capture the effect of instruction time in 10^{th} grade plus persistent effects from differences in previous grades. The magnitude and direction of the bias would depend upon the correlation between instruction time differences in 9^{th} and 10^{th} grades. In this sample the correlation equals 0.42, indicating that the school by grade fixed effect estimates will tend to overstate the effect of instruction time in a grade. Thus the true effect likely lies between the estimates produced by the school by grade and those produced by the school by subject specifications.

A final complication arises from the possibility that parents respond to school inputs when determining family education inputs (Todd and Wolpin 2003). The direction of bias that would arise from an endogenous family response is unclear. For example, if parents judge the school to lack instruction time in a particular subject, they may compensate by studying more with their child at home. Assuming that more parental help is positively related to student achievement and negatively related to classroom instruction time, failing to account for the endogenous parental response would tend to bias downward the estimated effect of instructional time.

As an informal specification test, we include subject-specific measures of out-of-school instruction time. The 2009 wave of PISA asks "How many hours do you typically spend per week attending <out-of-school-time lessons> in the following subjects (at school, at home or somewhere else)?" The student can respond "*do not attend; less than 2 hours; 2 to 4 hours; 4 to 6 hours; or 6 or more hours*." We aggregate student responses to these questions to the school-by-grade-by-subject level for the same reason we aggregate instruction time. Appendix Table A3 shows that the inclusion of this variable has little effect on the in-school instruction time estimate, providing evidence that any such parental behavior may not introduce bias in this framework.

IIIb. Dimensions of Heterogeneity

We explore the possibility that there are diminishing returns to instruction time and that the effect varies by classroom environment and the quality of instruction. Instruction time varies by total number of minutes per week and by the division of that time into classes, and it may matter if the 180 minutes per week are divided into four 45-minute classes or three 60-minute classes. The information on number of classes per week and minutes per class potentially enables the identification of diminishing returns along each of these dimensions. In reality, there is very little variation across subjects in the length of classes at a school, and therefore we focus on total minutes and use quadratic and higher-order terms to investigate the presence and magnitude of diminishing returns.

Identification of heterogeneity by classroom environment and the quality of instruction requires measures of each. The surveys lack direct measures of student behavior, and a growing body of evidence highlights the weakness of observed characteristics as measures of teacher quality. School administrators do answer a series of questions about student behavior and student-teacher relations, and we use the responses to construct an index of classroom environment. Administrators also respond to questions about teacher evaluation and support and whether there are teacher shortages in specific subjects, and we use these responses to construct an index of instruction quality. Specifically, the indexes come from separate factor analyses that take he ordinal character of the responses into consideration. These indexes do not vary within schools, but they can be interacted with instruction time to produce information on heterogeneity in the return to additional instruction time by classroom environment and the quality of instruction.

Table 6 lists the variables used to generate each factor and the factor loadings. Single combinations of factors explain 89% and 99% of the variance in instructional and classroom quality, respectively. The factor weightings illustrate the importance of the teacher shortage indicators in the construction of the quality of instruction factor and the high correlation of all student behavior and student-teacher interaction variables in the construction of the classroom environment index. Thus it is not possible to separately identify the effects of disruption, the quality of student-teacher interactions, student attendance, or disrespectful behavior toward teachers or peers.

IV. Results

We report a series of estimates that characterize the relationship between achievement and instruction time using the fixed effect specification described in the previous section. Because a school's class length tends not to vary across subject or grade, we present results for both the number of classes and total minutes per week in most tables. The initial set of results report the average effect of instructional time on achievement. Subsequently we explore the existence of non-linear effects of both minutes and classes, and this section concludes with the results of the investigation of potential heterogeneity in the effects of instructional time by classroom environment and the quality of instruction. Prior to presenting the fixed effect results we describe the within-school variation across subjects in class time and achievement used to identify the estimates.

IVa. Instructional Time Differences Between Subjects

In Table 1, we describe the joint-distribution of instruction time in mathematics and language arts for both total weekly minutes and the number of classes. Although the diagonal elements have the highest frequencies a substantial share of schools report instructional time disparities between subjects. Consider first the top panel on weekly minutes. Among students reporting language arts minutes between 200 and 219, only slightly more than half report mathematics minutes that fall in the same category. Among those with other than 200 to 219 minutes of mathematics instruction time, the majority spends more time in mathematics than language arts classes. Not surprisingly, at higher levels of language arts instructional time a larger share of students spend less as opposed to more time in mathematics classes.

A similar pattern holds for classes per week, the primary source of within-school instructional time variation. Students that attend four language arts classes per week are more likely to attend five or more mathematics classes than fewer than four. However, students that attend five language arts classes per week are less likely to attend six or more mathematics classes than fewer than five.

Table 1 documents the existence of adequate within-school instructional time variation to identify effects, and we now describe patterns of test score differences to examine whether the raw test score data are consistent with the belief that longer classes raise achievement. Table 2 reports differences in average test score (mathematics minus language arts) by the joint distribution of mathematics and language arts instructional time based on both minutes and classes per week. This table has the same structure as Table 1 but replaces the cell shares with the average score differences.

A finding that entries above the diagonal (where instructional time for math exceeds instructional time for language arts) tend to be more positive than entries along the diagonal (where there is little or no difference between subjects) which in turn tend to be more positive than entries below the diagonal (where instructional time for language arts exceeds that for mathematics) would be consistent with a positive effect of instructional time, and the pattern in Table 2 provides support for such an effect. In the top panel there are only three negative entries above the diagonal, and Table 1 shows that these are three of the smallest of the above-diagonal cells. In contrast, there are ten negative entries below the diagonal including three of the six largest entries. Finally, entries along the diagonal tend to fall in between those above and those below.

IVb. Baseline Estimates

This section begins with results from the basic models that estimate the average effect of instructional time and then moves to results from models with a more flexible parameterization of the relationship between achievement and time. All tables report coefficients from specifications with school-by-grade fixed effects and specifications with school-by-subject fixed effects as well as robust standard errors clustered by school. The main sample includes 47,580 school-grade-subject cells, and roughly two thirds of the sample contains schools with both 9th and 10th grade. Therefore the remaining one third does not contribute to the identification of the estimates based on the school-by-subject fixed effect specification.

Table 3 reports estimates of the relationship between achievement and instructional time as measured by both weekly minutes and the number of classes for specifications without fixed effects, with school-by-grade fixed effects, and with school-by-subject fixed effects. The two panels share a similar pattern of highly significant estimates that decline by more than 60 percent with the inclusion of school-by-grade fixed effects and another 30 percent when school-by-subject effects replace school-by-grade effects.

The smaller estimates from the specifications with school by subject fixed effects are consistent with the issues raised in the previous section. Factors that could contribute to the observed pattern include subject specific skills that are positively related to instruction time and not accounted for in the specifications with school by grade fixed effects, correlation between instruction-time differences in the current and prior grades that inflate estimates from the school by grade fixed effect specifications, attenuation bias in the specifications with school by subject fixed effects introduced by violation of the assumption that 9th grade instructional time has no effect on 10th grade achievement, or larger measurement error-induced attenuation bias in the models with school by subject fixed effects.

Although there is little direct evidence exists on subject-specific skills, available information suggests that the contribution of the other factors likely varies. On the one hand, an analysis of residual variances in Appendix Table A2 find little or no evidence in support of larger measurement error-induce attenuation bias in the school by subject fixed effect specifications. On the other hand, available evidence does support the belief that the effects of instruction time in prior years contribute to the observed pattern. First, Jacob, Lefgren, and Sims (2008), Rothstein (2010), and Kain and Staiger (2008) find that at least a portion of the knowledge acquired in a grade persists into the future. Second, the correlation between school average instruction-time differences in ninth and tenth grades equals 0.42 in the PISA data. Such persistence in effects and correlation in instruction time differences leads the effects of instruction time in prior years to inflate the school by grade fixed effect estimates and to attenuate the school by subject fixed effect estimates as discussed in Section III.

Note that the instruction-time coefficient remains positive in the fully saturated specification with both school by grade and school by subject fixed effects (not reported), though the estimate is much smaller and quite imprecise. Unfortunately, the final column in Appendix Table A2 shows that less than three percent of the variation in the school-average instruction-time difference remains, consistent with the notion that there is inadequate variation in actual instruction time to generate a precise, unbiased estimate not attenuated by measurement error.

The instructional time measure provides another dimension over which differences in magnitudes arise, as the magnitude of the effect is generally larger in the regressions based on classes as opposed to weekly minutes. Consider the average class length of roughly 50 minutes. The school-by-grade fixed-effect coefficient indicates that the addition of one class per week would raise achievement by roughly 2 points on average, while the addition of 50 minutes per week would raise achievement by roughly 1.25 points on average. Note, however, that this difference becomes much smaller in the school-by-subject specifications. One interpretation is that the return to additional time diminishes more quickly when classes are lengthened than when schools increase the number of classes per week, though the lack of within school variation in class length precludes a direct test of this hypothesis.

In Figure 1, we present additional evidence on the relationship between instruction time and achievement. Each figure scatters the mean residuals from two separate regressions where instruction time and achievement, respectively, are regressed on country-by-school-by-grade effects and school-by-grade (left panels) or school-by-subject (right panels) fixed effects. In all the figures, the relationship between study time and achievement is positive and strong, though as expected the slope is less steep in regressions with school-by-subject fixed effects.

We now investigate the possibility of diminishing returns to additional minutes. Table 4 reports results from the three specifications with weekly minutes entered as a quadratic, and the results in both fixed effect specifications strongly support the hypothesis of diminishing returns. Importantly, the return to additional minutes diminishes quite slowly, becoming negative at over 500 minutes per week in both specifications, a number that exceeds the 95th percentile.

Table 5 reports results from fixed effect specifications that group weekly minutes and classes into seven and five categories respectively. Although both specifications produce a generally positive relationship between achievement and minutes, there are some inconsistencies. For example, in column (2) with school-by-subject fixed effects, the highest category in both the minutes and classes specification the estimate is negative. However, this category likely contains substantial error in measurement. Attendance in greater than six classes per week may reflect efforts to remediate low performance or may result from reporting error, and a number of observations that report weekly minutes above 280 (80 minutes per day if students attend school six days per week) may also suffer from reporting error.

IVc. Heterogeneity by classroom environment and the quality of instruction

The notion that the return to additional time depends crucially on the quality of the learning environment fits with the emphasis on the role of disruption in education production presented in Lazear (2001) and more general consideration of the quality of teachers and schools. In this section we investigate the possibility of variation in the return to instruction time by reported student and teacher behavior.

Because the instructional and classroom quality indexes do not vary within schools, the direct effects on achievement cannot be identified. However, we can interact these measures with the instructional time variables in order to investigate heterogeneity in the returns to instruction time along these dimensions. The instructional quality index ranges from 1.1 to 4.2 with a mean of 3.51, and the classroom environment measure ranges from 1.2 to 4.7 with a mean of 3.45.

The results in Table 7 provide some support for the hypothesis that the return to additional instructional time increases with the quality of the classroom environment and no support for the hypothesis that the return increases with the quality of instruction. In the school by grade fixed effect specifications the coefficients on the classroom environment interactions are positive and significant at the 1 percent level. As expected, the magnitude and significance is smaller in the school by subject specifications, where the still positive coefficients are slightly less than half as large as those from the school by grade fixed effect specifications. The small and insignificant coefficients on the interaction with the quality of instruction is consistent with the notion that the quality of instruction does not affect the return to additional class time, but we believe that the more compelling interpretation is that our indirect measures fail to capture salient differences in teacher effectiveness.

We evaluate the return to instruction time at the 25th, 50th, and 75th percentiles of classroom quality using the school by grade fixed effect estimates and the distribution of the classroom environment index. These suggest that an additional hour of weekly instruction time raises achievement by more than twice as much at the 75th percentile (0.025 standard deviations) than at the 25th percentile (0.011 standard deviations). Schools with behaviors that place them in the lower tail in terms of classroom environment therefore are likely to realize little or no benefit from increases in instructional time. The fact that these survey questions provide noisy information about the quality of the classroom environment including the degree of disruption raise the possibility of much greater heterogeneity along this dimension as well as by the quality of instruction.

The model of education production in Lazear (2001) suggests the possibility of a nonlinear relationship between the level of disruption and learning, and we now examine a more flexible specification that includes interactions with indicators for quartile of the instructional quality and classroom environment distributions (the lowest quartile interactions are excluded). Again there is little or no evidence of heterogeneity along our measure of the quality of instruction. The school by grade fixed effect specification produces a monotonically increasing return to additional instructional time as the quality of the classroom environment increases, while the school by subject fixed effect specification suggests that only the bottom quartile schools fail to receive the benefit accruing to all others. An additional school by subject fixed effect regression that interacts instruction time with an indicator for not being in the bottom classroom environment quartile produces an interaction coefficient of 0.027 that approaches significance at the 10 percent level (the standard error equals 0.018). This provides additional evidence that it is the schools with poor classroom environments that realize little or no benefit from additional instruction time.

V. Conclusions and Policy Implications

Instructional time has become an important element in school reform discussions, as many advocate for increases in time devoted to mathematics and reading instruction. A shortage of compelling empirical evidence has hindered the decision-making process, and a primary goal of this paper is to build on the contributions of recent work and provide additional information. The analysis uses panel data methods made possible by the richness of the PISA data, and the fixed effects models accounted for student and school heterogeneity including differences by subject in some specifications.

The empirical analysis provides strong evidence in favor of the notion that additional time raises achievement using a series of specifications and measures of instructional time. Given the character of the deficiencies of the two fixed effects models, the results suggest that the effect is positive and modest in magnitude on average. Although instructional time is found to exhibit diminishing returns, the rate of decrease appears to be quite gradual.

Perhaps most important, the benefit of additional instructional time appears to vary with the quality of the classroom environment. The results produced by both specifications show that schools with low quality classroom environments likely realize little or no benefit from additional instruction time. On the one hand, it does not appear that schools can compensate for poor environments with additional time. If anything, additional time might be expected to degrade further the quality of the classroom environment as it becomes more difficult for students to sit and listen. On the other hand, there would appear to be substantial complementarities between policies that improve the classroom environment such as the strict discipline demanded in KIPP Academy schools and those that expand instruction time. Thus these results are consistent with the large benefits found for attendance at KIPP Academy charter schools.

In contrast, the estimates provide little or no evidence of a relationship between the return to additional instruction time and the quality of instruction. Yet given the absence of direct measures of teacher quality, class size, and other established determinants of the quality of instruction, this finding may simply reflect the weakness of the quality of instruction measure. Additional research is called for to gain a better understanding of heterogeneity by the quality of instruction.

Bibliography

Adams, Ray, and Margaret Wu. 2002. "PISA 2000 Technical Report". OECD. http://www.oecd.org/edu/preschoolandschool/programmeforinternationalstudentassessmentpisa/33688233.pdf.

Angrist, Joshua D., Susan M. Dynarski, Thomas J. Kane, Parag A. Pathak, and Christopher R. Walters. 2010. "Inputs and Impacts in Charter Schools: KIPP Lynn." American Economic Review 100 (2): 239–43.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2011. "The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood". Working Paper 17699. National Bureau of Economic Research. http://www.nber.org/papers/w17699.

Coates, Dennis. 2003. "Education Production Functions Using Instructional Time as an Input." Education Economics 11 (3): 273–292.

Farbman, David. 2011. Learning Time in America: Trends to Reform the American School Calendar. http://www.eric.ed.gov/PDFS/ED521518.pdf.

Farbman, David. 2012. "The Case for Improving and Expanding Time in School: A Review of Key Research and Practice." http://www.timeandlearning.org/files/CaseforMoreTime_1.pdf.

Gijselaers, Wim H., and Henk G. Schmidt. 1995. "Effects of Quantity of Instruction on Time Spent on Learning and Achievement." Educational Research and Evaluation 1 (2): 183–201. doi:10.1080/1380361950010204.

Jacob, Brian A., Lars Lefgren, and David Sims. 2008. "The Persistence of Teacher-Induced Learning Gains". NBER Working Paper 14065. National Bureau of Economic Research, Inc. http://ideas.repec.org/p/nbr/nberwo/14065.html.

Kane, Thomas J., and Douglas O. Staiger. 2008. "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation". NBER Working Paper 14607. National Bureau of Economic Research, Inc. http://ideas.repec.org/p/nbr/nberwo/14607.html.

Herbst, Mikolaj, Daniel Munich, Steven Rivkin, and Jeffrey Schiman. 2012. Understanding the Divergent Trends in PISA Test Results for Poland and the Czech Republic. Working Paper.

Kuehn, Zoe, and Pedro Landeras. 2012. Study Time and Scholarly Achievement in PISA. Working Paper. FEDEA. http://ideas.repec.org/p/fda/fdaddt/2012-02.html.

Lavy, Victor. 2010. Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries. NBER Working Paper. National Bureau of Economic Research, Inc. http://ideas.repec.org/p/nbr/nberwo/16227.html.

Lavy, Victor. 2012. "Expanding School Resources and Increasing Time on Task: Effects of a Policy Experiment in Israel on Student Academic Achievement and Behavior". Working Paper 18369. National Bureau of Economic Research. http://www.nber.org/papers/w18369.

Lazear, Edward P. 2001. "Educational Production." The Quarterly Journal of Economics 116 (3) (August 1): 777–803. doi:10.1162/00335530152466232.

Mandel, Philipp, and Bernd Süssmuth. 2011. Total Instructional Time Exposure and Student Achievement: An Extreme Bounds Analysis Based on German State-Level Variation. CESifo Working Paper Series. CESifo Group Munich. http://ideas.repec.org/p/ces/ceswps/_3580.html.

Marcotte, Dave E. 2007. "Schooling and Test Scores: A Mother-natural Experiment." Economics of Education Review 26 (5): 629–640.

Marcotte, Dave E., and Steven W. Hemelt. 2008. "Unscheduled School Closings and Student Performance." Education Finance and Policy 3 (3) (July 1): 316–338. doi:10.1162/edfp.2008.3.3.316.

Meghir, Costas, and Steven Rivkin. 2011. "Chapter 1 - Econometric Methods for Research in Education." In Handbook of the Economics of Education, ed. Stephen Machin and Ludger Woessmann Eric A. Hanushek, Volume 3:1–87. Elsevier. http://www.sciencedirect.com/science/article/pii/B9780444534293000016.

National Center on Time & Learning. Why Time Matters. http://timeandlearning.org/?q=why-time-matters.

Roland G. Fryer, Jr. 2011. Injecting Successful Charter School Strategies into Traditional Public Schools: Early Results from an Experiment in Houston. NBER Working Paper. National Bureau of Economic Research, Inc. http://ideas.repec.org/p/nbr/nberwo/17494.html.

Rothstein, Jesse. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." The Quarterly Journal of Economics 125 (1): 175–214.

Todd, Petra E., and Kenneth I. Wolpin. 2003. "On the Specification and Estimation of the Production Function for Cognitive Achievement*." The Economic Journal 113 (485): F3–F33. doi:10.1111/1468-0297.00097.

Wiermann, Christian. 2005. Class Size, Instruction Time and Central Exit Examinations : Disentangling the Relative Contributions to Scholastic Achievement. Working Papers of the Research Group Heterogenous Labor. Research Group Heterogeneous Labor, University of Konstanz/ZEW Mannheim. http://ideas.repec.org/p/knz/hetero/0504.html.

Table 1: Joint Distribution of Mathematics and Language Arts Instructional Minutes and Classes Based on Student Level Data

1. Minutes per week (proportion of math total)

	Mathematics							
Language Arts	*0-99*	*100-179*	*180-199*	*200-219*	*220-239*	*240-279*	*280+*	*Total*
0-99	0.41	0.04	0.02	0.00	0.02	0.01	0.02	8,220
100-179	0.32	0.58	0.13	0.18	0.08	0.08	0.06	50,051
180-199	0.15	0.11	0.48	0.01	0.14	0.08	0.04	37,238
200-219	0.02	0.14	0.01	0.57	0.01	0.16	0.05	36,695
220-239	0.03	0.05	0.18	0.01	0.47	0.06	0.03	30,976
240-279	0.03	0.05	0.14	0.16	0.23	0.49	0.16	53,403
280+	0.05	0.03	0.04	0.07	0.06	0.13	0.64	39,445
Total	9,102	45,251	42,250	34,082	35,384	51,455	38,504	256,028

2. Classes per week (proportion of math total)

	Mathematics					
Language Arts	*0-2*	*3*	*4*	*5*	*6+*	*Total*
0-2	0.51	0.12	0.03	0.02	0.03	21,347
3	0.22	0.49	0.15	0.06	0.04	45,877
4	0.18	0.31	0.53	0.21	0.10	84,279
5	0.06	0.07	0.21	0.56	0.20	73,743
6+	0.03	0.02	0.08	0.15	0.63	44,135
Total	20,867	44,914	88,115	77,017	38,468	269,381

Notes: The joint distributions are calculated at the student level. The sample starts with only students that have a math and language arts component in their PISA exam (304,070 students). We then merge the student sample by country, school, grade, and subject to the analysis sample and drop observations not used in the regressions.

Table 2: Mathematics minus language arts score difference by instructional time in mathematics and language arts

1. Minutes per week

	Mathematics						
Language Arts	*0-99*	*100-179*	*180-199*	*200-219*	*220-239*	*240-279*	*280+*
0-99	1.4	7.4	-6.3	-2.8	15.4	7.3	7.1
100-179	2.8	0	8.2	2.1	7.2	0.1	0.4
180-199	-1.7	1.2	0.8	-1.4	4.1	2.5	2.4
200-219	3.7	-6.4	-5.8	1.4	4.7	1.3	13.3
220-239	15.9	6.3	1.6	2	2.8	6	10.9
240-279	2.8	-9	-1.3	-8.7	-17.4	-1.1	7.9
280+	-5.6	-11.9	-3.6	-1.1	3.8	5.6	4.2

2. Classes per week

	Mathematics				
Language Arts	*0-2*	*3*	*4*	*5*	*6+*
0-2	-2.8	3.8	-3.4	2.4	9.4
3	1	-0.9	4.7	3.1	5.7
4	-3.9	-1.4	0.7	0.7	9.3
5	-10	-5	-8.8	-1	5.1
6+	-16.2	1.6	-0.8	0.8	10.7

Table 3: Estimated Effects of Weekly Instructional Minutes and Classes per Week on Achievement

	(1)	(2)	(3)
Panel A:			
Weekly Minutes of Instruction	0.068***	0.025***	0.017**
	(0.008)	(0.006)	(0.008)
Panel B:			
Weekly Number of Class Periods	4.946***	2.078***	1.153**
	(0.497)	(0.346)	(0.541)
School-by-grade fixed effect	N	Y	N
School-by-subject fixed effect	N	N	Y
Sample Size	47,580	47,580	47,580
# of Schools	16,154	16,154	16,154

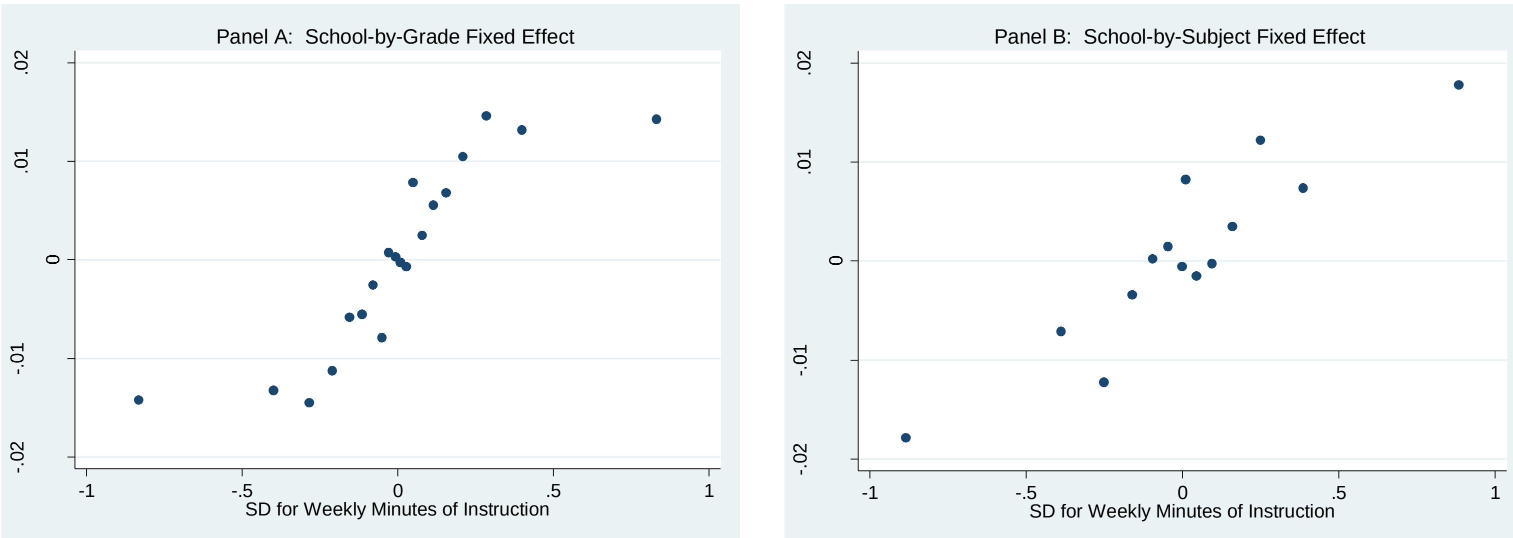
Notes: The dependent variable in all regressions is stacked school-by-grade-by-subject average test scores based on PV1MATH and PV1READ. Estimates are insensitive to choice of plausible value. A consistent sample of schools is used for all regressions in the following tables. All regressions also include a country-by-grade-by-subject effect. Prior to aggregation to the country-school-grade-subject level, the sample is limited to students who had both math and language arts components in their 2009 PISA exam. The sample consists of 47,580 school-by-grade-by-subject observations from 16,154 schools.

Robust Standard errors clustered by school are in parentheses.

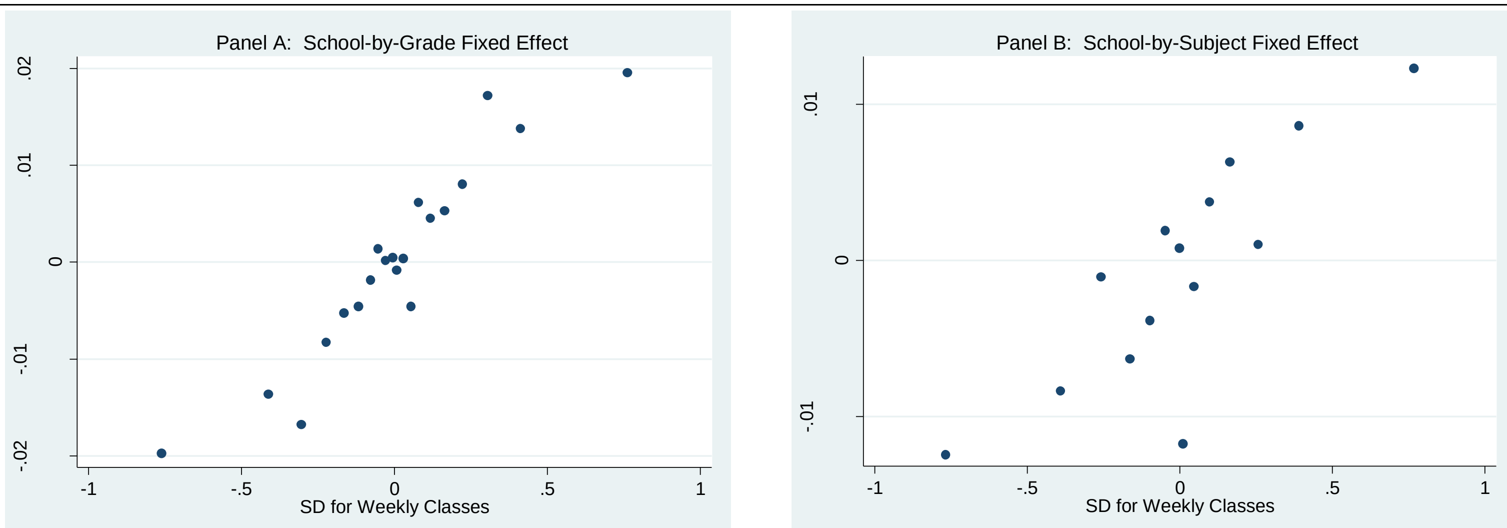
*** Significant at the 1 percent level.; ** Significant at the 5 percent level; * Significant at the 10 percent level.

Figure 1: Estimated Effect of Instructional Minutes and Classes Per Week on Standard Deviations of Achievement

Panel A: Weekly Minutes of Instruction



Panel B: Weekly Number of Classes



Notes: Similar to Chetty, Friedman, and Rockoff (2011), this figure presents the regression estimates non-parametrically. To plot each figure, we regress both instruction time and test scores in standard deviations on country-by-grade-by-subject fixed effects as well as school-by-grade or school-by-subject fixed effects. After both regressions, we calculate residuals, group them based on the instruction time residuals, and scatter the grouped residuals against each other.

Table 4: Estimated Effects of Weekly Minutes Using a Quadratic Specification

	(1)	(2)	(3)
Weekly Minutes of Instruction	0.3410***	0.0701***	0.0973***
	(0.0197)	(0.0131)	(0.0209)
Weekly Minutes of Instruction Squared	-0.00043***	-0.00007***	-0.00012***
	(0.00003)	(0.00002)	(0.00003)
School-by-grade fixed effect	N	Y	N
School-by-subject fixed effect	N	N	Y
Sample Size	47580	47580	47580
# of Schools	16154	16154	16154

Notes: Robust Standard errors clustered by school are in parentheses.

Table 5. Estimated Effects of Weekly Instructional Minutes and Classes per Week on Achievement from Regressions Using Instructional Time Categories

	(1)	(2)
Panel A: Average Shares of Students in each Minutes Per Week Category (relative to 200-219)		
0 to 99	-3.5	-20.6***
	(2.5)	(3.9)
100 to 179	-2.2	-6.6***
	(1.3)	(2.1)
180 to 199	4.5***	-5.2*
	(1.6)	(2.8)
220 to 239	1.7	0.8
	(1.7)	(2.9)
240 to 279	5.9***	0.7
	(1.3)	(2.4)
280+	4.4***	-2.7
	(1.6)	(2.7)
Panel B: Average Shares in each Weekly Classes Category (relative to 4)		
0 to 2	-8.2***	-16.7***
	(1.7)	(2.8)
3	-2.9***	-2.9*
	(1.0)	(1.6)
5	2.1**	1.6
	(0.9)	(1.8)
6 to 10	3.9***	-4.1*
	(1.3)	(2.3)
School-by-grade fixed effect	Y	N
School-by-subject fixed effect	N	Y
Sample Size	47580	47580
# of Schools	16154	16154

Notes: Robust Standard errors clustered by school are in parentheses.

Table 6. Factor Analysis of Questions on Student and Teacher Behavior

	Instructional Quality	
	Factor Loadings	Scoring Coefficients
Principal Discusses Job with Teachers	0.0114	-0.0006
Principal Makes Suggestions to Teachers	-0.0629	-0.0150
Teachers are Observed by an Authority Figure	-0.0958	-0.0146
No Lack of Science Teachers	0.8811	0.2962
No Lack of Mathematics Teachers	0.8929	0.3272
No Lack of Language Arts Teachers	0.8618	0.2599
No Lack of Teachers in other subjects	0.7863	0.1649
	Classroom Quality	
	Factor Loadings	Scoring Coefficients
Lack of Student Absences	0.7056	0.1633
Lack of Student Disruption	0.7571	0.1824
Lack of Student Skipping	0.7870	0. 2510
Students Respect Teachers	0.8021	0.2397
Lack of Student Drug Use	0.7011	0.1527
Lack of Student Bullying	0.7284	0.1898

Notes: Each school representative responds to a series of questions about the school and classroom climate (Q11, Q17, Q23, and Q26). We use responses to these questions in our factor analysis.

The eigenvalue for the teacher quality factor is 2.948 and the proportion of variance it explains is 83%. We predict only the first factor and call it classroom hindrances. The eigenvalue for the classroom quality factor is 3.356 and the proportion of variance it explains is 99%. Higher values on both scales reflect higher quality. Given the ordered categorical nature of the variables, we use a Polychoric correlation matrix to conduct the factor analysis.

Table 7. Estimated Effects of Instructional Time, by Teacher and Classroom Quality

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A:						
Weekly Minutes of Instruction	-0.001	-0.000	-0.049*	-0.013	-0.054*	-0.021
	(0.024)	(0.033)	(0.025)	(0.037)	(0.031)	(0.043)
Weekly Minutes*Instruction Quality	0.008	0.005	-	-	0.002	0.003
	(0.007)	(0.010)	-	-	(0.007)	(0.010)
Weekly Minutes*Classroom Quality	-	-	0.021***	0.009	0.021***	0.008
	-	-	(0.007)	(0.011)	(0.007)	(0.011)
Panel B:						
Weekly Number of Class Periods	0.806	1.952	-2.050	-0.228	-2.051	0.701
	(1.442)	(2.055)	(1.395)	(2.362)	(1.707)	(2.644)
Weekly Classes*Instruction Quality	0.364	-0.230	-	-	0.001	-0.373
	(0.396)	(0.588)	-	-	(0.423)	(0.632)
Weekly Classes*Classroom Quality	-	-	1.178***	0.407	1.178***	0.514
	-	-	(0.393)	(0.674)	(0.419)	(0.719)
School-by-grade fixed effect	Y	N	Y	N	Y	N
School-by-subject fixed effect	N	Y	N	Y	N	Y
Sample Size	47580	47580	47580	47580	47580	47580
# of Schools	16154	16154	16154	16154	16154	16154

Table 8. Estimated Effects of Instructional Time, by quartile of Teacher and Classroom Quality

	Instructional Quality				Classroom Quality			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Weekly Minutes of Instruction	0.023**	0.023	0.020**	0.016	-0.004	-0.002	-0.004	-0.002
	(0.010)	(0.015)	(0.010)	(0.014)	(0.010)	(0.016)	(0.010)	(0.016)
Weekly Minutes*2nd Quality Quartile	-0.004	-0.011	-	-	0.021	0.027	-	-
	(0.013)	(0.021)	-	-	(0.014)	(0.022)	-	-
Weekly Minutes*3rd Quality Quartile	0.004	-0.012	-	-	0.039***	0.030	-	-
	(0.013)	(0.021)	-	-	(0.013)	(0.022)	-	-
Weekly Minutes*4th Quality Quartile	0.012	0.008	-	-	0.049***	0.021	-	-
	(0.015)	(0.023)	-	-	(0.014)	(0.023)	-	-
Weekly Minutes*(2nd through 4th Quality Quartile)	-	-	0.007	0.002	-	-	0.037	0.027
	-	-	(0.011)	(0.017)	-	-	(0.011)***	(0.018)
School-by-grade fixed effect	Y	N	Y	N	Y	N	Y	N
School-by-subject fixed effect	N	Y	N	Y	N	Y	N	Y
Sample Size	47,580	47,580	47,580	47,580	47,580	47,580	47,580	47,580
# of Schools	16,154	16,154	16,154	16,154	16,154	16,154	16,154	16,154

Notes: The regression includes the main effects of instruction time, the quality quartiles, and interactions of instruction time and quality quartiles. We omit the lowest quality category as the base category. Because the quality quartile main effects do not vary within a school, they are captured perfectly by the fixed effects.

Appendix Table A1: Descriptive Statistics

	Math		Language Arts	
	Mean	SD	Mean	SD
Average Test Score	457.14	83.82	457.28	79.34
Average Weekly Number of Classes	4.35	1.16	4.43	1.22
Average Length in Minutes of an Average Class	51.76	12.15	51.46	11.99
Average minutes per week	221.55	69.23	223.83	69.71
Average Instruction Quality	3.51	0.76	3.51	0.76
Average Classroom Quality	3.45	0.73	3.45	0.73
# of Schools		16,154		

Notes: To calculate weekly minutes of instruction, we multiply the school-by-grade-by-subject average number of weekly classes attended by the length of an average class (ST28Q01*ST29Q01 and ST28Q02*ST29Q02). Prior to aggregation to the grade-by-school-by-subject level, students who reported having more than 10 classes per week or average class lengths greater than 120 minutes were set to missing.

Total number of observations is 48,528 and each represents a country-by-school-by-grade-by-subject average value. In all analyses that follow, standard errors will be clustered on school of which there are 16,452.

Appendix Table A2: Percent of Variation in Instruction Time Measures Explained by the Fixed Effects

Average weekly minutes	0%	43%	88%	87%	97%
Average Weekly Classes	0%	50%	90%	91%	98%
School-by-grade fixed effects	N	N	Y	N	Y
School-by-subject fixed effects	N	N	N	Y	Y
Subject-by-grade-by-country effects	N	Y	Y	Y	Y

Notes: Average weekly minutes, average weekly classes, and average minutes per class are used as dependent variables. The independent variables used in each regression are indicated in the table. The percent indicates r^2 from each regression.

Appendix Table A3: Out-of-school study

		(1)	(2)		(3)	(4)
Weekly Minutes of Instruction	0.025***	0.030***	0.014*			
	(0.006)	(0.006)	(0.008)			
Weekly Number of Class Periods				2.078***	2.231***	0.846
				(0.346)	(0.348)	(0.540)
Out-of-school Lessons (rel. to 2 to < 4 Hours)						
None		0.525	13.726***		0.604	13.723***
		(1.361)	(1.848)		(1.362)	(1.848)
<2 Hours per week		-3.270**	-3.053		-3.182**	-3.044
		(1.438)	(2.126)		(1.437)	(2.127)
4 to <6 Hours per week		1.091	-2.468		0.983	-2.423
		(1.752)	(2.522)		(1.752)	(2.523)
6+ Hours per week		0.921	-5.696**		0.607	-5.642*
		(2.163)	(2.887)		(2.164)	(2.888)
School-by-grade fixed effect	Y	Y	N	Y	Y	N
School-by-subject fixed effect	N	N	Y	N	N	Y
Sample Size	47,580	46,136	46,136	47,580	46,136	46,136
# of Schools	16,154	16,038	16,038	16,154	16,038	16,038

* p<0.10, ** p<0.05, *** p<0.01

Appendix Table A4: Estimated Effects of Weekly Instructional Minutes and Classes per Week on Achievement in Science and Language Arts

	(1)	(2)	(3)
Panel A:			
Weekly Minutes of Instruction	0.091***	0.013***	0.034***
	(0.006)	(0.003)	(0.007)
Panel B:			
Weekly Number of Class Periods	5.917***	0.778***	2.337***
	(0.341)	(0.158)	(0.388)
School-by-grade fixed effect	N	Y	N
School-by-subject fixed effect	N	N	Y
Sample Size	47,786	47,786	47,786
# of Schools	16,301	16,301	16,301

Appendix Table A5. Estimated Effects of Instructional Time on Achievement in Science and Language Arts, by Teacher and Classroom Quality

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A:						
Weekly Minutes of Instruction	0.023**	0.028	-0.007	0.017	0.005	0.016
	(0.011)	(0.028)	(0.011)	(0.027)	(0.014)	(0.034)
Weekly Minutes*Teacher Quality	-0.003	0.002			-0.004	0.000
	(0.003)	(0.008)			(0.003)	(0.008)
Weekly Minutes*Classroom Quality			0.006*	0.005	0.007**	0.005
			(0.003)	(0.008)	(0.003)	(0.008)
Panel B:						
Weekly Number of Class Periods	1.347**	2.313	-0.238	1.689	0.424	1.819
	(0.534)	(1.423)	(0.559)	(1.513)	(0.668)	(1.723)
Weekly Classes*Teacher Quality	-0.160	0.007			-0.253*	-0.054
	(0.146)	(0.397)			(0.153)	(0.431)
Weekly Classes*Classroom Quality			0.289*	0.188	0.357**	0.205
			(0.153)	(0.427)	(0.161)	(0.463)
School-by-grade fixed effect	Y	N	Y	N	Y	N
School-by-subject fixed effect	N	Y	N	Y	N	Y
Sample Size	47786	47786	47786	47786	47786	47786
# of Schools	16301	16301	16301	16301	16301	16301

Notes: Robust Standard errors clustered by school are in parentheses.
* p<0.10, ** p<0.05, *** p<0.01