Instruction Time, Classroom Quality, and Academic Achievement

Steven G. Rivkin and Jeffrey C. Schiman*

February 2013

PRELIMINARY DRAFT

DO NOT CITE

Abstract:.

* Rivkin: University of Illinois at Chicago, University of Texas at Dallas, and NBER; Schiman: University of Illinois at Chicago. We thank Marcus Casey, Robert Kaestner, Cuiping Long, Darren Lubotsky, Ben Ost, and participants of the UIC economics research lunch for many helpful comments.

I. Introduction

The statement that additional time on task raises output seems uncontroversial in most settings, even if a longer workday might raise the error or injury rates as output increased. However, this belief is not accepted as simple fact in the area of education, where debate also persists over whether additional dollars raise school quality. Inefficiencies potentially dampen the return to additional instruction time as well as increased spending in other areas. In addition, students may exhibit diminishing ability to remain engaged and well behaved as instruction time rises.

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 in greater detail, individualize and differentiate instruction, and answer students' questions" (National Center on Time & Learning). Many point to KIPP Academy schools for evidence of the benefits of extended time in class. Instructional 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, and it is difficult to isolate the specific mechanisms that account for KIPP's apparent success.[2]

Recent work focusing on instructional time generally supports the notion that additional time raises achievement, though difficulties isolating an exogenous source of

[1] http://www.kipp.org/our-approach

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

variation raise questions about the strength of much of the evidence.[3] Perhaps the simplest example is the difference between academic and vocational secondary schooling. Academic schools boast higher achievement and more instructional time, but this comparison certainly does not constitute causal evidence. Alternatively, lower achieving schools may boost instructional time in order raise test scores, or schools may require additional instruction only for struggling students. These latter cases would introduce bias in the opposite direction than that resulting from the academic-vocational split, thus it is not even possible to determine a priori if the simple correlation between achievement and instructional time overstates or understates the causal relationship.

Moreover, it seems highly likely that the magnitude of any causal link between achievement and instructional time depends upon the quality of instruction and the classroom environment. Expanded instructional time in response to poor mathematics achievement may be unlikely to have much impact if an ineffective curriculum, inadequate teacher subject matter knowledge, or disruptive behavior led to the low achievement in the first place. Even if existing class time is effective, there may be decreasing marginal benefit to additional minutes if the quality of instruction, classroom environment or student effort diminishes with time.

In this paper we investigate the relationship between instructional time and achievement using 2009 Programme for International Student Assessment (PISA) data. We are particularly interested in the mediating effects of teacher quality and the classroom environment on that relationship. In order to overcome biases introduced by

---

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.

the non-random allocation of instruction time, we build on the work of Lavy (2010) that uses variation across subjects within schools in order to account for unobserved ability and school quality. Importantly, we also make use of other aspects of the panel data structure to control for variation among schools in the quality of instruction in specific subjects.

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, though some inconsistencies in the findings highlight the need for additional research. Although it appears that the instructional time does exhibit diminishing returns, the rate of decrease appears to be quite gradual. Education quality appears to introduce another source of heterogeneity in the benefit of instructional time, as the findings show that hindrances to learning reduce the benefit of additional minutes and classes per week.

The next section develops the conceptual framework and empirical model, and Section 3 describes the data. Section 4 presents the results, focusing on the inter-relationship between instruction time, classroom environment and the quality of instruction. The final section summarizes the analysis and considers implications for education policy.

II. Empirical Model

This section develops an empirical framework with which to investigate the effects of instructional time on achievement. It first develops a conceptual model that incorporates the inter-relationship among instructional time, classroom environment, quality of instruction and ability. This highlights both the difficulty in identifying the effect of instructional time and likely heterogeneity in that effect resulting from variation in the other variables.

The second part of this section describes the panel data methods used to identify the effects of instruction time and potential interactions with classroom environment and the quality of instruction. The analysis must account for student and school heterogeneity as well as complications introduced by error in the measurement of class time, classroom environment, and other relevant factors. In addition, the estimation 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).

IIa. Conceptual Framework

Equation (1) models average achievement (*A*) in subject *s*, grade *g*, and school *j* as a function of instruction time (*T*), school value added per-unit of time (*V*), and ability (α). Because subject-specific student skill and interest influence classroom placement, within subject and school variation in instruction time is likely correlated with unobserved subject-specific ability that cannot be accounted for. Therefore, in the basic model we aggregate to the school-by-subject level and focus on school average differences in achievement and instruction time across subjects. When we allow for nonlinearities in the relationship between achievement and time, we focus on the shares of students in different time categories.

(1) $A_{sgj} = f_A(T_{sgj}, V_{sgj}, \alpha_{sgj})$

Equation (2) models value-added (*V*) as a function of the quality of instruction (*Q*), classroom environment (*D*), and instruction time (*T*). Teacher effectiveness and the curriculum are assumed to be the primary determinants of the quality of instruction, and skill at classroom management and the underlying student propensity to disrupt are assumed to be the primary determinants of the classroom environment. Importantly, both the quality of instruction and classroom environment may vary with total instruction time. As time in class increases student behavior and teacher quality may worsen, potentially leading to diminishing returns to longer classes.

$$(2)\, V_{sgj} = f_V(T_{sgj}, Q_{sgj}, D_{sgj})$$

Equation (3) represents instruction time (*T*) as a function of the quality of instruction, the classroom environment, and average ability. Aggregation to the school-by-grade-by-subject level eliminates within-school variation correlated with ability, but average ability may still be related to average instruction time through student sorting among schools. The quality of instruction likely also influences *T*; schools may attempt to compensate for low quality or devote additional time to stronger subjects. Finally, the sensitivity of student behavior and the quality of instruction to the class length likely influences the choice of *T*. Disruptive behavior may increase with the number of instructional minutes, and the quality of instruction may deteriorate. All in all, the relationship illustrated in Equation (3) illustrates that the failure to account fully for ability, classroom environment and the quality of instruction would almost certainly lead to biased estimates of the effect of instructional time on achievement.

$$(3)\, T_{sgj} = f_T(\alpha_{sgj}, Q_{sgj}, D_{sgj})$$

The substitution of Equation (1) into Equation (3) highlights potentially important sources of heterogeneous effects including the level of disruption and the quality of instruction. Many of the same processes that potentially introduce bias also affect the return from additional instructional time. In addition, the relationship between these determinants of learning per unit of time and the number of minutes may also depend on the organization of minutes during the week. Specifically, an increase in the number of classes per week in a subject may have a different effect than an increase in class length.

$$(4)\ A_{sgj} = f_A(T_{sgj}, f_V(T_{sgj}, Q_{sgj}, D_{sgj}), \alpha_{sgj})$$

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, 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 estimate of instructional time.

As an informal specification test, we include subject-specific measures of out-of-school study 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. If after accounting for these subject-specific measures the estimate of instruction time does not change, this provides some evidence in support of the notion that any parental response introduces little bias. In fact, this is exactly what we find (not shown), suggesting that any such parental behavior either has little effect on achievement or is not strongly related to instruction time.

IIb. Fixed Effect Specifications

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, the structure of the PISA data enables the use of panel data methods to account for differences in school quality, school climate, and student ability. First, the availability of tests in multiple subjects allows for the inclusion of student fixed effects and identification of the estimates on the basis of within-student instruction time variation across subjects. Not only do these effects account for student heterogeneity, they also account for between school differences in school climate and the quality of instruction. Clearly the student fixed effects, used by Lavy (2010), account for the primary confounding factors. Because we aggregate to the school-by-grade-by-subject level, the school-by-grade fixed effect rather than the student fixed effect accounts for student heterogeneity and school average differences in the quality of instruction and climate that are common across subjects.

Importantly, neither the student nor school-by-grade fixed effects account for subject differences in ability or school quality, and such differences could be related to instruction time and bias the estimates. Above we discussed the possibility that prior student performance could influence classroom placement and instruction on a student level, but decisions about school curricula could also be influenced by the distributions of mathematics and language arts skills for students and for teachers. Additional instruction time in the subject where students, teachers or curricula are relatively stronger would introduce an upward bias in the estimated effect of time, while additional time in the weaker subject would introduce a downward bias.

In order to account for subject-specific factors we take advantage of the availability of data for multiple grades in many schools and include school-by-subject fixed effects. In models with only school-by-subject fixed effects, the estimates are identified by instructional time differences between grades but within subjects. Therefore unobserved differences across subjects do not introduce bias. These models rely on the assumption that average instruction time differences between 9th and 10th grade in each subject are not related to differences in average ability, behavior or the quality of instruction. There are almost certainly cohort differences in average ability and behavior and grade differences in average teacher quality, but these will only introduce bias if they are systematically related to instructional time.

A model that accounts for both grade and subject specific differences would include both school by grade and school by year fixed effects, and we estimate such a model. This can be viewed as a difference-in-difference-in-difference model, where the difference between mathematics and language arts scores for 10th grade minus the difference in 9th grade is related to the difference between mathematics and language arts instruction time for 10th grade minus the difference in 9th grade. If the difference in mathematics and language arts test scores is larger for the grade in which there is a larger instructional time difference the regression will yield a positive instructional time estimate.

A final issue introduced by comparisons across grades is the treatment of learning dynamics. Unless learning fully depreciates each year, a better 9th grade education will raise achievement in 10th grade as well as 9th grade. Therefore longer 9th grade instruction time will tend to increase 10th grade achievement if additional time is valuable. As Meghir and Rivkin (2011) illustrate, fixed effect estimates 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. Our analysis does not compare achievement of the same student in successive grades. Nonetheless, persistence in the effect of instruction time in 9th grade could still attenuate estimates based on instructional time differences between 9th and 10th grade.

Importantly, the form of PISA instruction time questions enables the specification of a flexible relationship between achievement and instruction time. Students are asked to report the number of classes taken each week and the length of classes in minutes for both mathematics and reading. Therefore we are able to examine the possibility that the effect of an increase in the number of classes differs from the effect of longer classes as well as decreasing returns to longer classes.

The baseline linear model shown in Equation (5) models average achievement in subject s in grade g in school j in country c as a function of the number of classes per week (C), minutes per class (M), and a set of error components (ε). An inspection of the error illustrates the factors accounted for by the various fixed effects and the maintained assumptions of the respective models.

(5) $A_{sgjc} = C_{sgjc}\gamma + M_{sgjc}\lambda + \varepsilon_{sgjc}$

Equation (6) decomposes ε into subject, grade, school, and country components as well as their interactions. Notice that all country interactions with the exception of the country-subject-grade term are omitted, as these are subsumed by the corresponding terms that replace country with school. School-by-subject-by-grade fixed effects cannot be included, because they would capture all variation in time use. Therefore we include a country-by-subject-by-grade term that captures national differences in curriculum and other factors that contribute to performance.

(6) $\varepsilon_{sgjc} = \eta_s + \phi_g + \sigma_j + \varsigma_c + \theta_{sgc} + \tau_{sj} + \omega_{gj} + \upsilon_{sgj}$

Each student contributes a mathematics and a reading test score to the grade averages, and the school-by-grade fixed effect accounts for those differences in average ability, level of disruption and school quality that are common across subjects. This removes all instruction time variation between schools and grades, meaning that the instruction time effects are identified by within school and grade differences.

Note that the school-by-grade fixed effects do not account for subject-specific abilities or instructional quality that is related to instruction time, and these could introduce bias. For example, consider the possibility that 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. Alternatively, the instruction difference could be positively related to the difference in the quality of mathematics versus language arts instruction. In either case the school-by-grade fixed effects model would generate estimates that are biased upward. Of course negative relationships between instruction time on the one hand and ability or instructional quality differences on the other would introduce a negative bias. Note that the use of student fixed effects in place of school-by-grade fixed effects would generate a similar pattern of estimates.

It is not clear whether confounding school-by-subject factors introduce bias. Nonetheless, the availability of multiple grades per school enables the inclusion of school-by-subject fixed effects. These effects account for systematic differences across subjects that could be related to instruction time differences and introduce bias. In models with these fixed effects, identification comes from within subject, between-grade variation in school average instruction time.

It is possible to include both school-by-grade and school-by-subject fixed effects in the same specification, though this pushes the data quite hard. There would be four observations per school (two grades and two subjects) and three dummy variables. Nonetheless, we do report estimates from fully saturated models as a further sensitivity check.

A final potential source of bias comes from school factors specific to a subject and grade. For example, a school may emphasize and devote more instruction time and higher teacher quality to mathematics in 9th as opposed to 10th grade. We are not able to account directly for any such confounding variation in teacher quality, but we are not aware of any evidence that this is the case in practice. Therefore we believe that within school and subject differences constitute an exogenous source of variation in instructional time is unlikely to be violated.

III. Data

The data come from the Programme for International Student Assessment (PISA), a survey and assessment test of mainly 15 year old students in over 65 countries. At least 150 schools are randomly selected in each country based on a stratified sample design, and, within each school, 35 students are sampled at random. Each student is assessed in math, science, and reading and then answers a changing set of questions on family background, school environment including classes per subject, home environment, and study habits. A representative from each school also provides information on the school's staff, environment, and pedagogical and human resource practices including the length of classes.

The PISA tests focus on knowledge application and are thought to be highly informative about the quality of preparation for higher education and the labor market. We focus exclusively on mathematics and language arts because the quality of mathematics education likely affects performance on the science examination. This is a problem, because our method identifies instruction time effects on the basis of the relationship between subject differences in test scores and instruction time. We expect there to be little spillover between mathematics and language arts at the high school level.[4]

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. In 2009 students are asked the number of math, science, and reading classes attended per week, and

[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). In this draft of the paper we present estimates based on the first plausible value.

schools are asked the length in minutes of classes. This permits us to investigate whether the effects of additional classes vary with the length of those classes.

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 both 2000 and 2009. While the majority of students would fall in the *2 to < 4 hour* category 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). As it is much more demanding on students to report total hours per week as opposed to number of classes, we believe that the 2006 data likely contain substantial error.

Perhaps even more important, preliminary work showed that the estimated relationship between achievement and instructional time is quite sensitive to the method used to impute minutes to each category. It turns out that Lavy's midpoint method generates much larger and significant estimates than those generated from imputations based on the 2000 or 2009 empirical distributions within each category.[5]

The measures of classroom environment are derived from school responses to a series of questions concerning disruption, other aspects of student behavior, student-

[5] Herbst, Munich, Rivkin, and Schiman (2012) show that for both Poland and the Czech Republic the estimates are quite sensitive to the imputation method. For Poland the student fixed effect instructional time coefficient equals 0.14 using Lavy's midpoint method and 0.052 using the 2009 empirical distribution for Poland based on reported minutes, and the corresponding estimates for the Czech Republic are 0.27 and 0.099.

teacher interactions and other aspects of teacher behavior (see Appendix Table A1 for the complete list). Respondents could check not at all, some of the time, most of the time, or all of the time. The responses are used in a factor analysis to produce a composite measure of the classroom environment.

The main sample used in this analysis includes 419,651 students in 9th and 10th grade in 16,586 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 ninth and tenth grades, and one set of regressions excludes schools that report very short or very long class lengths (shorter than 37.5 minutes or longer than 62.5 minutes).

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 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 sets 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. Prior to presenting the estimates we describe the within-school variation across subjects in class time and achievement used to identify the estimates.

IVa. Instructional Time Differences Between Subjects

Table 1 reports the joint distribution of instructional 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 reporting divergent minutes spent in mathematics classes, the majority spend 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 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 instructional time variation to identify effects based on within-school differences, and we now examine 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 reading) by instructional time categories based on both minutes and classes per week. This table has the same structure as Table 1 but replaces the frequency with the average score difference.

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 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. We turn now to the results from the regression analysis of instructional time effects.

IVb. Estimated Effects of Instructional Time

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 49,745 school-grade-subject cells, and roughly two thirds of the sample contains schools with both 9^{th} and 10^{th} grade. Therefore the remaining one third does not contribute to the identification of the estimates based on the school-by-subject 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 50 percent with the inclusion of school by grade fixed effects and another 40 to 60 percent when school by subject effects replace school by grade effects. The estimate remains positive in the fully saturated specification reported in the final column, though the estimate is much smaller and statistically insignificant. Unfortunately there does not appear to be enough variation remaining to generate a precise estimate.

A number of factors could contribute to smaller estimates in the school by subject specifications including confounding school-by-subject specific factors in the school by grade specifications, attenuation bias in the school by subject specifications introduced by the structure of learning dynamics, or larger measurement error-induced attenuation bias in the models with school by subject fixed effects due to the fact that only a fraction of the schools contain 9^{th} and 10^{th} grades. Little direct evidence exists on the importance of the first two components, but information on instructional time residual variance following the inclusion of the respective fixed effects that is reported in Appendix Table A2 suggests that the magnitude of any fixed effect amplification of measurement bias is likely to be similar in the two specifications. The third and fourth columns show that each of these fixed effects removes roughly 90 percent of the variance from each of the instructional time measures.

The instructional time measure provides another dimension over which differences in magnitudes arise, as the magnitude of the effect is generally larger in the regression based on classes as opposed to weekly minutes. Consider the average class length of roughly 50 minutes. The school-by-grade coefficient indicates that the addition of one additional class would raise achievement by roughly 2.4 points on average, while the addition of 50 minutes would raise achievement by roughly 1.5 points on average. Note, however, that this difference becomes much smaller in the school-by-subject specifications. One possible source of the difference could be that additional classes raise achievement by more than an equivalent increase in the number of minutes generated by lengthening the existing number of classes.

We attempt to learn more about such nonlinearities below, but first we 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 400 minutes per week in both specifications, values above 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, the pattern tends is more consistent and diminishing returns more pronounced in the school-by-subject specifications. In these the only exception is the estimate for the highest category in both the minutes and classes specification, and this category likely contains the most error in measurement. It is unclear what it means for students to attend more than 5 of 6 classes per week in a subject, and weekly minutes above 280 places students in the upper tail of the distribution. In contrast, the school-by-grade fixed effect estimates suggest that attending class 180-199 minutes per week raises achievement relative to attending 200-219 minutes per week.

Table 6 presents additional results on potential nonlinearities in the return to instructional time by classifying classes on the basis of class length. In order to mitigate problems introduced by measurement error and small cells we focus on the vast majority of classes between 37.5 and 62.5 minutes in length. Once again the two specifications offer quite different findings, with the school-by-subject specifications producing a more consistent pattern of estimates. In this specification the benefits of an additional class are roughly half as large for classes with 40 or 45 minutes as they are for classes of 50 to 55 minutes. This finding is not consistent with a simple story of diminishing returns to instructional time, as the return to a 20 percent shorter class is roughly 50 percent smaller. Note both that the highest and likely noisiest category of minutes produces a smaller and less precisely estimated coefficient, and the school-by-grade specifications show little or no systematic relationship between class length and the return to additional classes.

IVc. Learning Hindrances and the Return to Instructional Time

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 test the hypothesis that hindrances to learning reduce the benefit of instructional time.

Our approach applies factor analysis to the set of survey questions on the behaviors of students and teachers filled out by the school representative in order to generate one or more measures related to the quality of education. The results of the factor analysis reported in Table 7 show that a single factor explained 96 percent of the variance, and we therefore use only that factor in the regressions. The high explanatory power of the single factor reflects the strong correlation among all the variables, and the factor weightings illustrate that negative teacher and student behaviors each increase the value of the factor.

Because the school hindrance measure does not vary within schools the direct effect on achievement cannot be identified. However, we can interact this measure with the instructional time variables in order to test the hypothesis that school hindrances reduce the benefit of additional time.

The results in Table 8 support the belief that hindrances to learning (lower education quality) reduce the return to instructional time, though the interaction term coefficients for specifications with school-by-subject fixed effects are not significant. A one standard deviation increase in the learning hindrance index reduces the estimated return to additional weekly minutes by roughly 20 percent and to additional classes by roughly 15 percent. Given the absence of direct measures of teacher effectiveness, class size, and the quality of the curriculum, the findings provide fairly strong evidence that the ultimate value of additional time depends on the many factors that influence the learning environment.

V. Conclusions and Policy Implications

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

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, though some inconsistencies in the findings highlight the need for additional research. Although it appears that the instructional time does exhibit diminishing returns, the rate of decrease appears to be quite gradual. Education quality appears to introduce another source of heterogeneity in the benefit of instructional time, as the findings show that hindrances to learning reduce the benefit of additional minutes and classes per week.

In terms of school policy, the results highlight both the potential value of additional time and the need to consider carefully local circumstances and conditions including the initial value of instructional time. In the presence of a poor learning environment the benefit of additional time might be quite low, and it is important to examine the effect of additional time on the learning environment. The opportunity cost of additional time must be weighed carefully against the expected benefits, and the return is likely to justify the investment in some circumstances but not in others.

Bibliography (INCOMPLETE)

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.

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.

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.

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.

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.

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 Reading Instructional Minutes and Classes Based on Student Level Data

1. Minutes per week (frequencies)

	Mathematics							
Language Arts	*0-99*	*100-179*	*180-199*	*200-219*	*220-239*	*240-279*	*280+*	*Total*
0-99	5,538	2,598	1,351	271	820	652	952	12,182
100-179	4,210	38,413	8,197	8,918	4,250	5,724	3,602	73,314
180-199	1,969	7,172	29,660	362	6,944	6,084	2,083	54,274
200-219	223	9,436	316	28,376	436	12,085	2,932	53,804
220-239	340	3,375	10,830	275	24,309	4,355	1,952	45,436
240-279	338	3,516	8,797	8,175	11,819	36,722	8,970	78,337
280+	690	2,265	2,304	3,522	3,273	9,812	36,077	57,943
Total	13,308	66,775	61,455	49,899	51,851	75,434	56,568	375,290

2. Classes per week (frequencies)

	Mathematics					
Language Arts	*0-2*	*3*	*4*	*5*	*6+*	*Total*
0-2	15,776	7,690	3,943	2,580	1,687	31,676
3	6,565	32,226	19,054	6,942	2,385	67,172
4	5,433	20,260	68,412	23,464	5,623	123,192
5	1,930	4,410	27,530	62,870	11,384	108,124
6+	1,081	1,586	9,746	16,800	36,011	65,224
Total	30,785	66,172	128,685	112,656	57,090	395,388

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

1. Minutes per week

	Mathematics						
Language Arts	*0-99*	*100-179*	*180-199*	*200-219*	*220-239*	*240-279*	*280+*
0-99	2.5	8.4	-3.7	-4.1	20.4	6.7	11.9
100-179	2.9	0.0	7.6	1.9	7.4	0.0	1.9
180-199	-2.1	1.6	0.3	-0.5	3.8	2.3	1.3
200-219	2.2	-6.6	-3.5	0.9	2.2	0.6	12.2
220-239	12.5	7.0	0.4	-1.3	2.9	6.0	9.8
240-279	3.6	-8.7	-1.3	-8.9	-17.9	-1.9	8.3
280+	-5.3	-11.3	-2.6	-1.2	4.2	5.4	4.3

2. Classes per week

	Mathematics				
Language Arts	*0-2*	*3*	*4*	*5*	*6+*
0-2	-2.3	3.5	-2.2	5.0	11.5
3	1.4	-1.0	4.6	3.2	6.6
4	-4.3	-1.4	0.4	0.0	9.2
5	-12.2	-3.7	-9.6	-1.3	5.6
6+	-14.7	3.1	-0.2	0.2	10.5

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

Panel A:				
Weekly Minutes of Instruction	0.072***	0.030***	0.018***	0.006
	(0.008)	(0.005)	(0.007)	(0.009)
	[0.061]	[0.025]	[0.015]	
Panel B:				
Weekly Number of Classes	5.597***	2.426***	1.142***	0.270
	(0.495)	(0.335)	(0.482)	(0.482)
	[0.082]	[0.035]	[0.017]	
# of schools		16,586		
School-by-grade fixed effects	N	Y	N	Y
School-by-subject fixed effects	N	N	Y	Y

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. Regressions in *Panel B* condition on the length of an average class.

Robust Standard errors clustered by school are in parentheses. In square brackets are effect sizes interpreted as the standard deviation increase in test scores from a one standard deviation increase in the instruction measure.

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

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

Weekly Minutes	0.347***	0.069***	0.108***
	(0.023)	(0.014)	(0.018)
	[0.292]	[0.058]	[0.090]
Weekly Minutes Squared	-0.00043***	-0.00006***	-0.00013***
	(0.00004)	(0.00002)	(0.00002)
	[-0.232]	[-0.034]	[-0.071]
# of Schools		16,586	
School-by-grade fixed effects	N	Y	N
School-by-subject fixed effects	N	N	Y

Notes: Robust Standard errors clustered by school are in parentheses. In square brackets are effect sizes interpreted as the standard deviation increase in test scores from a one standard deviation increase in the instruction measure.

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

Average Shares of Students in each Minutes Per Week Category (relative to 200-219)		
0-99	-5.8**	-20.1***
	(2.5)	(3.7)
100-179	-3.0**	-5.5***
	(1.3)	(2.0)
180-199	3.4**	-4.2
	(1.6)	(2.7)
220-239	0.4	1.1
	(1.7)	(2.8)
240-279	5.8***	3.1
	(1.3)	(2.3)
280+	4.5***	-1.4
	(1.6)	(2.7)
Weekly Classes (relative to 4)		
0-2	-9.3***	-13.5***
	(1.7)	(2.6)
3	-3.9***	-1.0
	(0.9)	(1.5)
5	2.4***	2.6
	(0.9)	(1.6)
6+	4.3***	-1.5
	(1.3)	(2.1)
School-by-grade fixed effects	Y	N
School-by-subject fixed effects	N	Y

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

Table 6. Estimated Effects of Classes per Week on Achievement, By Class Length

$37.5 < minutes \leq 42.5$	3.251***	0.905
	(0.821)	(0.644)
$42.5 < minutes \leq 47.5$	1.850***	1.071*
	(0.531)	(0.564)
$47.5 < minutes \leq 52.5$	3.263***	1.910***
	(0.601)	(0.583)
$52.5 < minutes \leq 57.5$	4.017***	1.988***
	(0.936)	(0.638)
$57.5 < minutes \leq 62.5$	3.250***	0.946
	(1.111)	(0.736)
# of Schools	14,629	14,629
School-by-grade fixed effects	Y	N
School-by-subject fixed effects	N	Y

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

Table 7. Factor Analysis of Questions on Student and Teacher Behavior

	Factor 1 – Learning Hindrances	
	Factor Loadings	Scoring Coefficients
Teachers' low expectations of students	0.603	0.108
Student absenteeism	0.725	0.163
Poor student-teacher-relations	0.748	0.176
Disruption of classes by students	0.755	0.163
Teachers not meeting individual student needs	0.651	0.137
Teacher absenteeism	0.659	0.115
Students skipping classes	0.759	0.120
Students lacking respect for teachers	0.776	0.186

Notes: Each school representative responds to a series of questions about to "what extent the learning of students is hindered by the following phenomenon." The school representative could check not at all (1), very little (2), to some extent (3), or a lot (4). Thus, higher values indicate that learning is *more* hindered.

The eigenvalue for the first factor is 4.057 and the proportion of variance it explains is 96%. We predict only the first factor and call it classroom hindrances. Higher values on the classroom hindrance scale reflect a lower quality environment. Given the ordered categorical nature of the variables, we use a Polychoric correlation matrix conduct the factor analysis.

Table 8. Estimated Effects of Instructional Time, by Classroom Hindrances

Panel A:			
Weekly Minutes of Instruction	0.080***	0.051*	0.038
	(0.017)	(0.028)	(0.026)
	[0.067]	[0.043]	
Weekly minutes-by-Classroom Hindrances	-0.020***	-0.013	-0.012
	(0.006)	(0.010)	(0.009)
	[-0.059]	[-0.039]	
Panel B:			
Weekly Number of Classes	4.560***	3.338**	1.050
	(1.033)	(1.692)	(1.562)
	[0.067]	[0.049]	
Weekly Classes-by- Classroom Hindrances	-0.824**	-0.819	-0.295
	(0.369)	(0.607)	(0.542)
	[-0.047]	[-0.047]	
School-by-grade fixed effects	Y	N	Y
School-by-subject fixed effects	N	Y	Y

Notes: Robust Standard errors clustered by school are in parentheses. Estimates in *Panel B* are conditional on length of an average class and its interaction with classroom hindrances.

Appendix Table A1: Descriptive Statistics

	Math		Reading	
	Mean	SD	Mean	SD
Average Weekly Number of Classes	4.35	1.17	4.43	1.21
Average Length in Minutes of an Average Class	51.66	11.97	51.37	11.86
Average minutes per week	221.42	67.88	223.74	68.24
Average Test Score	454.03	84.00	453.76	80.94
# of Schools		16,586		

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 49,745 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 16,586.

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

Average weekly minutes	0%	44%	88%	88%
Average Weekly Classes	0%	50%	90%	91%
School-by-grade fixed effects	N	N	Y	N
School-by-subject fixed effects	N	N	N	Y
Subject-by-grade-by-country effects	N	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.