Time to Learn?
The Organizational Structure of Schools and Student Achievement

Ozkan Eren

Southern Methodist University

Daniel L. Millimet*

Southern Methodist University

September 2005

Abstract

Utilizing parametric and nonparametric techniques, we asses the impact of a heretofore relatively unexplored ‘input ’in the educational process, time allocation, on the distribution of academic acheivement. Our results indicate that school year length and the number and average duration of classes are salient determinants of student performance. However, the effects are not homogeneous – in terms of both direction and magnitude – across the distribution. We find that students below the median benefit from a shorter school year, while a longer school year benefits students above the median. Furthermore, low-achieving students benefit from fewer, shorter classes per day, while high-achieving students benefit from more and longer classes per day.

JEL: C14, I21, I28
Keywords: student achievement, school quality, school year length, stochastic dominance, distributional analysis

*Corresponding address: Daniel Millimet, Department of Economics, Box 0496, Southern Methodist University, Dallas, TX 75275-0496. Tel: (214) 768-3269. Fax: (214) 768-1821. E-mail: millimet@mail.smu.edu.

1 Introduction

The stagnation of student achievement over the past few decades in the United States is well-documented (e.g., Epple and Romano 1998; Hoxby 1999), despite the fact that per pupil expenditures have increased an average of roughly 3.5% per annum over the period 1890 – 1990 (Hanushek 1999) and that aggregate public expenditures on primary and secondary education total approximately $200 billion (Betts 2001). Given the discontinuity that exists between educational expenditures and student achievement, an important body of research has emerged attempting to discover the primary influences on student learning. However, a potentially important 'input' in the educational process that has been overshadowed is time allocation; specifically, time spent in school and time spent in classes.

To partially address this gap, we assess the impact of several measures of the organizational structure of the learning environment on student achievement. In particular, we focus on the (i) length of the school year, (ii) number of class periods per day, and (iii) average length per class period. There are several reasons a priori to believe that such variables may impact student learning. First, as found in Eren and Millimet (2005), the organizational structure of the school day affects student misbehavior, as measured by the number of instances in which a student is punished for disobeying of school rules, receives an in-school suspension, receives an out-of-school suspension, and skips class. Moreover, Figlio (2003) documents that disruptive student behavior adversely impacts the test performance of peers. Second, there may be advantages to different organizational structures in terms of optimally conveying information and minimizing repetitive teaching activities. Finally, school year length directly affects the amount of time students spend in school, and may impact the curriculum choices of schools. For instance, Pischke (2003) finds that shorter school years increased the probability that students had to repeat a grade level, although he finds no long-run adverse impact.

From a policy perspective, the findings reported herein should be of substantive interest. Since such organizational details are well within the control of school administrators and/or state policymakers, the policy implications are obvious. Moreover, re-organization of the school day is relatively costless, especially relative to other educational policies such as reductions in class size. Altering the length of the school year, on the other hand, is not budget-neutral. For example, the Texas state legislature is currently finalizing legislation that would require all school districts to have a uniform start date for the school year after Labor Day, and end no later than June 7 (*Dallas Morning News*, 13 May 2005, p. 20A). Proponents of such a legislative mandate argue that early school year start dates cost the state of Texas an estimated $332 million annually in foregone tourism revenue, and as much as another $10 million due to the electricity costs from cooling schools during the month of August, not to mention additional teacher salaries (Strayhorn 2000). Advocates of the earlier start date are concerned that student academic achievement would suffer from a shorter school year. Thus, empirical evidence on the link between school year length and student performance will help inform the current political debate, at least in Texas.

To proceed, we use a nationally representative sample of tenth grade public school students from the US National Educational Longitudinal Survey (NELS) – conducted in 1990 – and assess the impact of school organization using both parametric and nonparametric techniques. First, we utilize standard regression analysis to analyze effects on the conditional mean of student test scores, restricting school organization to only an intercept effect. Next, we allow school organization to have both an intercept and slope effect, by altering the efficacy of other inputs in the educational production function. For example, it may be that more qualified teachers are more effective at raising student test performance if there are fewer, but longer, class periods per day.[1] We test for such interaction effects using the familiar Oaxaca-Blinder decomposition of mean differences. Finally, since the preceding parametric analysis may mask heterogeneous effects of school organization by focusing on the (conditional) mean, we extend the analysis to the examination of the conditional distribution of test scores, controlling for a host of student, family, and school attributes and incorporating the notion of the Oaxaca-Blinder decomposition.

Our comparisons of the conditional distributions of test scores across students facing different organizational structures utilize recently developed nonparametric tests for stochastic dominance (SD). Testing for SD is an extremely useful and insightful companion to standard regression analysis for several reasons. First, it allows one to examine effects of the 'treatment' in question at different parts of the distribution. This is particularly useful in assessments of educational policies since many interventions are unlikely to have identical effects on both high- and low-achieving students and policymakers may assign different 'weights' to different parts of the distribution. In this respect, SD analysis offers the same advantage as estimators of Quantile Treatment Effects (QTEs), defined as the difference in the distributions of the outcome of interest at various quantiles across groups exposed to different treatments (see, e.g., Bitler et al. 2005). SD analysis simply advances the analysis a step further by providing a welfare framework for ranking various treatments given nonconstant QTEs.

Second, as just stated, the notion of SD is explicitly couched within a welfare framework. Specifically, a finding of SD (at a particular order) makes it clear that all utility/welfare functions belonging to a particular class would agree that one policy is preferable to another. Thus, individual comparisons based

[1] In a related literature, Maasoumi et al. (2004) find important interaction effects between class size and several other inputs in the determination of student achievement.

on specific indices (e.g., mean comparisons in the case of first order SD) are unnecessary. In addition, the inability to infer a dominance relation is equally as informative, indicating that any (implicit) welfare ordering based on a particular index (such as the conditional mean) is highly subjective; different indices – even within the same general class of utility/welfare functions – will yield different substantive conclusions.

The power of such SD relations, combined with the recently developed theory necessary to conduct rigorous statistical tests for the presence of such relations, has led to their growing application. For example, Maasoumi and Millimet (2005) examine changes in US pollution distributions over time and across regions at a point in time. Maasoumi and Heshmati (2000) analyze changes in the Swedish income distribution over time as well as across different population subgroups. Fisher et al. (1998) compare the distribution of returns to different length US Treasury Bills. Particularly relevant to the analysis at hand are previous applications of SD to the analysis of various treatment effects on the distribution of the outcome of interest. For instance, Amin et al. (2003) analyze the effect of a micro-credit program in Bangladesh on the distribution of consumption of participants versus non-participants. Abadie (2002) analyzes the impact of veteran status on the distribution of civilian earnings. Bishop et al. (2000) compare the distribution of nutrition levels across populations exposed to two different types of food stamp programs. Anderson (1996) compares pre- and post-tax income distributions in Canada over several years.

The results are quite revealing. In particular, we reach five main conclusions. First, a longer school year is associated with higher unconditional, but not conditional, mean test scores. Second, shorter class periods, but more class per day, is associated with higher unconditional and conditional mean test scores. However, none of the mean-based effects are large in magnitude. Third, a longer school year and reorganization of the school day to include shorter, but more, classes are found to improve unconditional test scores across virtually the entire distribution. However, the effects are not uniform; in particular, students in the middle of the distribution seem to gain the most. Fourth, when we examine the conditional test score distributions, we find extremely heterogeneous effects from school organizational structure. Specifically, the academic performance of students below the median benefits from a shorter school year, while a longer school year benefits students above the median. Thus, a uniform start date – of the variety proposed in Texas – does not appear optimal (when considering student achievement only). Furthermore, we find that low-achieving students benefit from fewer, shorter classes per day, while high-achieving students benefit from more and longer classes per day. Finally, while the mean-based effects of school organization are not found to be overly meaningful economically, the distributional analysis indicates that such organizational details are meaningful determinants of test performance for students in the tails of the (conditional) distribution.

The remainder of the paper is organized as follows. Section 2 defines the various dominance relations and describes the tests used to identify such relations in the data. Section 3 discusses the data. Section 4 presents the results. Section 5 concludes.

2 Empirical Methodology

2.1 Regression Approach

To initially examine the data, we utilize standard regression techniques. First, we estimate a linear regression model of the form

$$t_{ij} = \mu + h_{ij}\gamma + z_j\tau + \eta_{ij} \tag{1}$$

where t_{ij} is the test score for individual i in school j, h is a lengthy vector of individual, family, class, teacher, and school attributes, z is a vector of the school organization variables, and η is a mean zero, normally distributed error term. Second, for each set of school organization attributes (i.e., school year length, number of class periods per day, and average length of classes), we split the sample into the appropriate sub-samples, exclude these variables from z, and re-estimate (1) on each sub-sample. The results enable us to perform the traditional Oaxaca-Blinder decompositions of mean test score gaps. Specifically, the mean test score gap between, say, students attending schools with a short versus long school year, may be expressed as

$$\bar{t}_l - \bar{t}_s = \underbrace{(\mu_l - \mu_s)}_{I} + \underbrace{(\overline{h}_l - \overline{h}_s)\gamma_s + (\overline{\widetilde{z}}_l - \overline{\widetilde{z}}_s)\tau_s}_{E} + \underbrace{\overline{h}_l(\gamma_l - \gamma_s) + \overline{\widetilde{z}}_l(\tau_l - \tau_s)}_{C} \tag{2}$$

where l and s denote long and short school year, respectively, short school year is treated as the 'dominant' category, and $\widetilde{z}$ is a sub-vector of z which excludes the school year length variables. If the difference in coefficients (term C) in (2) is large in absolute value, then one may infer the presence of important interactions between school organization and the returns to other test score inputs. If the difference in endowments (term E) in (2) is large in absolute value, then one may 'significant' associations between the set of conditioning variables and school organizational structure.

2.2 Distributional Approach

2.2.1 Test Statistics

Our distributional comparisons are based on the notion of SD. Several tests for SD have been proposed in the literature; the approach herein is based on a generalized Kolmogorov-Smirnov test.[2] To begin, let X

[2] Maasoumi and Heshmati (2000) provide a brief review of the development of alternative tests.

and Y denote two outcome (test score) variables being compared (e.g., X (Y) might refer to test scores of students exposed to a short (long) school year). $\{x_i\}_{i=1}^N$ is a vector of N possibly dependent observations of X; $\{y_i\}_{i=1}^M$ is an analogous vector of realizations of Y. In the spirit of the historical development of such two-sample tests, $\{x_i\}_{i=1}^N$ and $\{y_i\}_{i=1}^M$ each constitute one sample. Thus, we refer to dependence between x_i and x_j, $i \neq j$, as *within-sample dependence* (similarly for observations of Y), and dependence between X and Y as *between-sample dependence.*

Assuming general von Neumann-Morgenstern conditions, let $\mathcal{U}_1$ denote the class of (increasing) utility functions u such that utility is increasing in test scores (i.e. $u' \geq 0$), and $\mathcal{U}_2$ the class of social welfare functions in $\mathcal{U}_1$ such that $u'' \leq 0$ (i.e. concavity). Concavity represents an aversion to inequality in the achievement of students; a high concentration of both high- and low-achieving students is undesirable. Let $F(x)$ and $G(y)$ represent the cumulative density functions (CDF) of X and Y, respectively, which are assumed to be continuous and differentiable.

Under this notation, X First Order Stochastically Dominates Y (denoted X FSD Y) *iff* $\mathrm{E}[u(X)] \geq \mathrm{E}[u(Y)]$ for all $u \in \mathcal{U}_1$, with strict inequality for some u.[3] Equivalently,

$$F(z) \leq G(z) \qquad \forall z \in \mathcal{Z}, \text{ with strict inequality for some } z. \tag{3}$$

where $\mathcal{Z}$ denotes the union of the supports of X and Y. If X FSD Y, then the expected welfare from X is at least as great as that from Y for all increasing welfare functions, with strict inequality holding for some utility function(s) in the class. The distribution of X Second Order Stochastically Dominates Y (denoted as X SSD Y) *iff* $\mathrm{E}[u(X)] \geq \mathrm{E}[u(Y)]$ for all $u \in \mathcal{U}_2$, with strict inequality for some u. Equivalently,

$$\int_{-\infty}^{z} F(v)dv \leq \int_{-\infty}^{z} G(v)dv \qquad \forall z \in \mathcal{Z}, \text{ with strict inequality for some } z. \tag{4}$$

If X SSD Y, then the expected social welfare from X is at least as great as that from Y for all increasing and concave utility functions in the class $\mathcal{U}_2$, with strict inequality holding for some utility function(s) in the class. FSD implies SSD and higher orders.

Define the following generalizations of the Kolmogorov-Smirnov test criteria:

[3] Note that SD relations offer insights into which distribution provides greater welfare *considering only the outcome of interest*, without regard for other considerations such as cost. If the ‘treatment’ is costly, a separate cost-benefit analysis is required to decide if the welfare gains exceed the costs.

$$d = \sqrt{\frac{NM}{N+M}} \min \sup_{z \in \mathcal{Z}} [F(z) - G(z)] \tag{5}$$

$$s = \sqrt{\frac{NM}{N+M}} \min \sup_{z \in \mathcal{Z}} \int_{-\infty}^{z} [F(u) - G(u)]\, du \tag{6}$$

where min is taken over $F - G$ and $G - F$, in effect performing two tests in order to leave no ambiguity between the 'equal' and 'unrankable' cases. Our nonparametric tests for FSD and SSD are based on the empirical counterparts of d and s using the empirical CDFs, where the empirical CDF for X is given by

$$\widehat{F}_N(x) = \frac{1}{N} \sum_{i=1}^{N} \mathrm{I}\,(X \leq x) \tag{7}$$

and $\mathrm{I}(\cdot)$ is an indicator function; $\widehat{G}_M(y)$ is defined similarly for Y. If $\widehat{d} \leq 0$ ($\widehat{s} \leq 0$) to a degree of statistical certainty, then the null hypothesis of FSD (SSD) is not rejected (see Appendix A for details).

To this point X and Y have represented two *unconditional* test score variables. However, dependence between school level attributes such as organization structure and other determinants of student achievement may confound estimation of the 'treatment' effect of interest with the impact of other characteristics (Hanushek 1979). In particular, student achievement may be related to family background variables (such as family structure and income), teacher attributes (such as salary and experience), and school characteristics (such as school size and the ability of peers).[4] To control for the myriad of determinants of student performance that may generate a spurious correlation between school organization and student test scores, we follow in the spirit of Dearden et al. (2002) and perform dominance tests on the *conditional* distribution of test scores.

To obtain these conditional distributions, we control for a host of observable attributes and conduct dominance tests on the distributions of test scores purged of the 'average' effects of these attributes (in the 'dominant' group; defined below). The conditioning covariates (discussed below) represent individual (such as race, gender, lagged test scores, and lagged grade point average), family (such as parental education, socioeconomic status, and family composition), class (such as subject and class size), teacher (such as race, gender, experience, and education), and school (such as enrollment, number of teachers, and average teacher salaries) attributes. Because our conditioning set is quite exhaustive, the implicit *selection on observables* assumption required to identify the causal effect of school organization appears reasonable.

To proceed, we estimate *separate* educational production function models for students exposed to each

[4] For a theoretical account of the cognitive development of students, see Todd and Wolpin (2003).

type of organizational structure:

$$t_{ijk} = \alpha_k + \widetilde{h}_{ij}\beta_k + \widetilde{\varepsilon}_{ijk}, \qquad k = 1, ..., K \tag{8}$$

where t_{ijk} is the test score for individual i in school j with a school organizational structure of type k (within a particular set such as school year length), $\widetilde{h} = [h\ \widetilde{z}]$ (defined previously), $\widetilde{\epsilon}$ is the error term, and there are K organizational categories (within the set of time variables being analyzed).[5] Note, that controls for school organization are omitted from (8), thereby allowing the error term to capture the residual effect of school organization not captured by the included regressors. As such, one possibility at this point would be to simply analyze the distributions of $\widehat{\varepsilon}_{ijk} \equiv \widehat{\alpha}_k + \widehat{\widetilde{\varepsilon}}_{ijk}$, which correspond to test scores *net of all observable characteristics* (evaluated at the *type-specific* returns, β_k).[6] In previous work (e.g., Maasoumi et al. 2004), we refer to these as *Partial Residual* (PR) distributions.

However, a richer approach is to compare what we refer to as the *Full Residual* (FR) distributions, which account for differences in the returns to observables as in the usual Oaxaca-Blinder decomposition approach. To show how this is done, we re-write the first-stage regression (8) for type k as

$$\begin{aligned} t_{ijk} &= \alpha_k + \widetilde{h}_{ij}\beta_k + \widetilde{\epsilon}_{ijk} \\ &= \alpha_k + \widetilde{h}_{ij}\beta_k + \widetilde{\epsilon}_{ijk} + \left(\widetilde{h}_{ij}\beta_{k'} - \widetilde{h}_{ij}\beta_{k'}\right) \\ &= \alpha_k + \widetilde{h}_{ij}\beta_{k'} + \widetilde{h}_{ij}(\beta_k - \beta_{k'}) + \widetilde{\epsilon}_{ijk} \end{aligned} \tag{9}$$

where type k' is implicitly treated as the 'dominant' category (Neuman and Oaxaca 2003).In the analysis below, we compare the previous intercept-adjusted residual distribution of $\widehat{\epsilon}_{ijk'}$ with $\widehat{\epsilon}^{FR}_{ijk} \equiv \left(\widehat{\alpha}_k + \widetilde{h}_{ij}(\widehat{\beta}_k - \widehat{\beta}_{k'}) + \widehat{\widetilde{\epsilon}}_{ijk}\right)$, $k \neq k'$.

2.2.2 Inference

The asymptotic null distribution of the test statistics, d and s, depend on the unknown distributions, F and G. In the analysis below, we approximate the empirical distribution of the test statistics using standard bootstrap methods as in Maasoumi and Heshmati (2000) and Maasoumi and Millimet (2005), and report the estimated significance level. Maasoumi and Heshmati (2005) provide a comparison among alternative methods for inference.

[5] To be clear, when analyzing, say, school year length, there are two groups – short and long – implying $K = 2$. In addition, the first-stage regression includes controls for the number of class periods per day and the average length per class in $\widetilde{z}$.

[6] The intercepts are included as part of the residuals, otherwise the conditional distributions will all be mean zero, precluding the possibility of first order dominance. This is also done since we do not wish to claim the models are perfectly specified.

To evaluate the null $\text{H}_o : d \leq 0$, we first report in our tables whether the observed empirical distributions are *seemingly* rankable by FSD or SSD; we present the sample values of $\max\{d_1\}$, $\max\{d_2\}$, $\widehat{d}$, $\max\{s_1\}$, $\max\{s_2\}$, and $\widehat{s}$ (see Appendix A). We then obtain bootstrap estimates of the probability that d lies in the non-positive interval (i.e. $\Pr\{d \leq 0\}$) using the relative frequency of $\{\widehat{d}^* \leq 0\}$, where $\widehat{d}^*$ is the bootstrap estimate of d (500 repetitions are used). If this interval has a large probability, say 0.90 or higher, and $\widehat{d} \leq 0$, we may infer dominance to a desirable degree of confidence. If this interval has a low probability, say 0.10 or smaller, and $\widehat{d} > 0$, we may infer the presence of significant crossings of the empirical CDFs, implying an inability to rank the outcomes. Finally, if the probability lies in the intermediate range, say between 0.10 and 0.90, there is insufficient evidence to distinguish between equal and unrankable distributions. This is a classic confidence interval test; specifically, we are assessing the likelihood that the event $d \leq 0$ has occurred. Similarly, we estimate $\Pr\{s \leq 0\}$ to evaluate the second order dominance proposition given by $\text{H}_o : s \leq 0$.[7]

A final, necessary comment pertains to inference in the FR tests (i.e., those incorporating the Oaxaca-Blinder decomposition). Due to the usage of a common set of coefficient estimates in obtaining both residual distributions being compared, there *necessarily* exists between-sample dependence. For example, the FR test compares the distributions of $\widehat{\epsilon}_{i1}$ and $\widehat{\epsilon}_{i2}^{FR}$. The former depends on $\{t_{i1}, h_{i1}, \beta_1(t_1, h_1)\}$, where t_1 and h_1 represent the full data vector for t and h for the $k = 1$ sample; the latter, $\widehat{\epsilon}_{i2}^{FR} = t_{i2} - h_{i2}\beta_1$, depends on $\{t_{i2}, h_{i2}, \beta_1(t_1, h_1)\}$. This source of dependence is atypical. Between-sample dependence usually arises when the same individuals appear in the two samples being compared (e.g., distributions of pre- and post-tax incomes for a sample of individuals). To handle this more common type of between-sample dependence, pairwise (or 'clustered') bootstrap samples are drawn in order to maintain the dependence in the resampled data (Linton et al. 2005). In the current situation, the between-sample dependence is maintained by re-estimating the first-stage equations (8) and (9) on each bootstrap resample. Specifically, by resampling N observations $\{t_{i1}^*, h_{i1}^*\}$ and M observations $\{t_{i2}^*, h_{i2}^*\}$ nonparametrically and re-estimating (8), we obtain the resampled distributions of $\widehat{\epsilon}_{i1}^*$ and $\widehat{\epsilon}_{i2}^{FR*}$, where the former depends on $\{t_{i1}^*, h_{i1}^*, \beta_1^*(t_1^*, h_1^*)\}$ and the latter depends on $\{t_{i2}^*, h_{i2}^*, \beta_1^*(t_1^*, h_1^*)\}$. Thus, as in the usual pairwise bootstrap case, the source

[7] Note, we do not impose and test the Least Favorable Case (LFC) of equality of the distributions. This could be done by combining the data on X and Y and bootstrapping from the combined sample (e.g., Abadie 2002). Our bootstrap samples still contain N (M) observations from X (Y). As argued in Linton et al. (2005), working under LFC has some undesirable power consequences as it can produce biased tests that are not similar on the boundary of the null. This happens when the boundary of the null itself is composite. The bootstrap methods employed herein, combined with a fixed critical value at zero (the boundary of the null hypothesis), renders our tests 'asymptotically similar' and unbiased on the boundary (Maasoumi and Heshmati 2005).

of between-sample dependence is maintained in the resampling procedure.

3 Data

The data is obtained from the National Education Longitudinal Study of 1988 (NELS:88), a large longitudinal study of eighth grade students conducted by the National Center for Education Statistics (NECS). The NELS:88 sample was chosen in two stages. In the first stage, a total number of 1032 schools were selected from a universe of approximately 40,000 schools. In the second stage, up to 26 students were selected from each of the sample schools based on race and gender. The original sample, therefore, contains approximately 25,000 eighth grade students. Follow-up surveys were administered in 1990, 1992, 1994 and 2000.

To measure academic achievement, students were administered cognitive tests in reading, social science, mathematics and science during the base year (eighth grade), first follow-up (tenth grade), and second follow-up (twelfth grade). Each of the four grade specific tests contain material appropriate for each grade, but included sufficient overlap from previous grades to permit measurement of academic growth.[8] While four test scores are available per student, teacher and class information used in the conditioning set (discussed below) are only available for two subjects per student; thus, our sample is restricted to two observations per student.[9]

We utilize three categorical measures of the organizational structure of the learning environment are: (i) length of the school year, divided into two categories: 180 days or less and 181 or more days (180+ days), (ii) the number of class periods per school day, divided into three categories: six or fewer periods, seven periods, and eight or more periods (8+), and (iii) average length per class (in minutes), divided into four categories: 45 minutes or less, 46-50 minutes, 51-55 minutes and 56 or more minutes (56+).[10]

To construct the final sample, we focus on the student achievement of tenth grade public school students, pooling test scores across all four subjects (and including subject indicators in the conditioning set), as in Boozer and Rouse (2001). We include only students with non-missing test score data and the relevant school structure variables. The final sample contains 11,431 students representing 826 schools (20,123 total

[8]We follow Boozer and Rouse (2001) and Altonji et al. (2003) and utilize the raw item response theory (IRT) scores for each test.

[9]The two subjects vary across students, however, so that all four subjects are represented in the sample.

[10]The NELS:88 includes six categories for length of the school year, four categories for the class periods per day and five categories for average length of the class period. To help manage the number of SD tests, we combine 1-174 days, 175 days, 176-179 days and 180 days into one category; 181-184 days and 185+ days into one other category; eight and 9+ class periods into one category; and, 1-40 minutes and 41-45 minutes into one category.

observations).

In the regression approach and the conditional SD tests, we utilize an extensive set of individual, family, teacher, class, and school characteristics available from the NELS:88. Specifically, the vector h in (1) includes the following (in addition to a constant):

Individual. race, gender, religion, eighth grade test score, eighth grade composite GPA;

Family. father's education, mother's education, family composition, family income, number of siblings, indicators of home reading material (books and newspaper), indicator for a home computer;

Teacher. race, gender, age, education;

Class. subject indicators, class size, number of minority students in the class;

School. urban/rural status, region, total school enrollment, grade-level enrollment, average daily attendance rate, student racial composition, percentage of students receiving free lunch, average dropout rate of tenth graders prior to graduation, number of full-time teachers, number of teachers by race, number of teachers by education, teacher salaries, an indicator for whether teachers have gone on strike in the past four years, percentage of students in remedial reading, remedial math, and bilingual education.

Dummy variables are used to control for missing values of the individual, family, teacher, class and school controls. In the interest of brevity, Table 1 displays the summary statistics only for some of the variables that utilized in the analysis; the remainder are available upon request.

4 Results

4.1 Regression Results

Ordinary Least Squares (OLS) estimates of (1) are displayed in Table 2; heteroskedasticity robust standard errors are given beneath the coefficients. The point estimates suggest a negative impact on student test scores from a longer school year, a positive impact of structuring the school day to include more class periods, and a negative impact of having longer class periods. However, only the final estimates with respect to the average length of a class period are statistically significant. Specifically, we find that the *ceteris paribus* effect of changing from classes that on average last 45 minutes or less to classes that exceed

50 minutes lowers student test scores by roughly 0.6 points. Since the mean test score is approximately 50, this represents a decline of slightly above one percent.

Table 3 presents the Oaxaca-Blinder decomposition results, as well as the unconditional mean test score gaps across the various sub-samples. In Panel A, we split the sample into students exposed to short (180 days or less) and long (181 days or more) school years. On average, students in schools with shorter school years receive lower test scores, by roughly 0.4 points. Differences in average endowments alone generates a difference in mean test scores of 0.6 points (favoring students exposed to longer school years); differences in the coefficients (inclusive of the intercepts) only partially offset this gap. Thus, on the whole, while a shorter school year is associated with lower test scores, this is reflective of its correlation with lower endowments and masks the positive interactions between a shorter school year and other inputs, explaining the negative (albeit statistically insignificant) coefficient on the dummy variable representing a longer school year in Table 2.

In Panel B, we split the sample into three sub-samples based on the number of class periods per day. In terms of the unconditional mean test score gaps, there is a clear monotonic relationship: six or fewer class periods per day yields the lowest mean test score, followed by seven class periods, and eight or more class periods yields the highest mean test score. In all three pairwise comparisons, the majority (if not more) of the unconditional mean gap is attributable to differences in endowments. There is, however, modest evidence of positive interactions between the presence of eight or more class periods (relative to six or fewer) and other inputs, as well as between the presence of seven class periods (relative to eight or more).

Lastly, in Panel C, we split the sample into four sub-samples based on the average length of a class period. In terms of the unconditional mean test score gaps, there is a non-monotonic relationship: students on average perform worst when classes are 51-55 minutes, followed by 56 minutes or longer and 46-50 minutes, and perform best on average when classes are 45 minutes or less. Moreover, at most 40% of the unconditional gaps favoring classes that are 45 minutes or shorter are attributable to differences in endowments. Thus, there is evidence of significant, positive interactions between shorter classes and the efficacy of other educational inputs. As before, this is consistent with the results in Table 2; specifically, the statistically significant, negative effects of longer class periods.

In sum, then, the parametric regression results indicate little effect of school year length on student test scores, but suggest a statistically significant, although perhaps not overly economically significant, positive effect of structuring the school day to include more, but shorter, class periods. For example, maintaining a six hours of instruction, but switching from six hour long class periods to nine 40 minute class periods is expected to raise student test scores by roughly 0.8 points, or nearly two percent. Although not large economically, given that such reorganization is seemingly costless makes this is an important finding. To examine if such effects are heterogeneous across the test score distribution, we now turn to the distributional results.

4.2 Unconditional Stochastic Dominance Results

The unconditional SD results are displayed in Table 4. The left panels in Figures 1-10 plot the differences in the CDFs across the quantiles. This is useful for quantification of the 'treatment' effects, as well as assessing the uniformity of any effects across the distribution.[11] Panel A and Figure 1 present the results for school year length. We observe no SD ranking in either the first- or second-degree sense, despite the fact that the unconditional mean favors students exposed to a longer school year (see Table 3). However, examining Figure 1 shows that the failure to find a ranking occurs only because the CDFs cross in the lower tail. Above the tenth percentile or so, we find that a longer school year is positively associated with student test scores, and that this effect is roughly uniform, hovering around the unconditional mean test score gap of 0.4. In terms of inference, since $\widehat{d} = 0.045 > 0$ and the bootstrap $\Pr(\widehat{d}^* \leq 0) = 0.066 < 0.10$, we can state that the crossings of the CDFs in the lower tail are statistically significant at the standard confidence levels.

Panel B and Figures 2-4 contain the results pertaining to the number of class periods per day. Consonant with the discussion above, we observe a monotonic relationship even at the distributional level. Specifically, we observe that the distribution of test scores for students with eight or more class periods first order dominates the distributions for seven class periods and six or fewer class periods; the distribution for students with seven class periods first order dominates the distribution for six or fewer class periods. However, only the FSD relation of eight or more class periods over six or fewer class periods is statistically significant at conventional levels ($\Pr(\widehat{d}^* \leq 0) = 0.980$); the dominance of eight or more class periods over seven class periods is statistically significant in the second order sense ($\Pr(\widehat{s}^* \leq 0) = 0.994$). This finding is much more informative than simply reporting the mean test score gaps in Table 3. We can state to a degree of statistical certainty that any policymaker with a social welfare function that is increasing in test scores would prefer the (unconditional) distribution associated with eight or more class periods (relative to six or fewer class periods), and any policymaker with a social welfare function that is increasing and concave in test scores would prefer the (unconditional) distribution associated with eight or more class periods (relative to seven class periods).

Examination of Figures 2-4 highlights an additional advantage of the distributional analysis: quantifying

[11] Plots of the actual CDFs or the integrated CDFs (used in the detection of SSD) are available upon request.

the heterogeneous impact of the treatment in question. In particular, all three plots indicate that the largest gains from having more class periods per day are obtained by students in the middle of the distribution (between roughly the 20^{th} and 60^{th} percentiles); students in the tails are less affected.

Lastly, Panel C and Figures 5-10 present the results for average class duration. As in Panel B, we are able to rank distributions in either the first- or second-degree sense in the majority of cases. Consistent with the rankings of the unconditional mean test scores in Table 3, we observe the distribution associated with classes 45 minutes or less first order dominates the other three distributions. Moreover, the ranking over the distribution for classes 51-55 (56 or more) minutes long is statistically significant in the first- (second-) degree sense ($\Pr(\widehat{d}^* \leq 0) = 0.952$ and $\Pr(\widehat{s}^* \leq 0) = 0.942$, respectively). We also find a statistically significant second order ranking for the distribution associated with classes of 46-50 minutes over the distribution for classes of 51-55 minutes ($\Pr(\widehat{s}^* \leq 0) = 0.970$).[12] The remaining distributions are not rankable to a degree of statistical certainty.

Examining the plots, for the cases of statistically significant rankings, we continue to find nonuniform impacts of the various 'treatments.' In particular, in Figures 6-8 we find more severe effects for students in the middle of the distribution. For example, switching from an average class length of 56 or more minutes to an average of 45 minutes or fewer is associated with an increased test score of roughly three points at the 60^{th} percentile, one point below the tenth percentile, and 1.5 points above the 85^{th} percentile. Thus, at least in unconditional terms, we find fairly robust evidence that the structure of the school day impacts student achievement, especially middle-achieving students. To see if these results change after accounting for covariates, we turn to the conditional distributional analysis.

4.3 Conditional Stochastic Dominance Results

The conditional SD results are displayed in Table 5. The right panels in Figures 1-10 plot the differences in the conditional CDFs across the quantiles. Panel A and Figure 1 present the results for school year length. Whereas we failed to observe any ranking in the first- or second-degree sense using the unconditional distributions, we now observe an SSD ranking favoring a shorter school year, consonant with the negative point estimate reported in Table 2. However, this ranking is not statistically significant ($\Pr(\widehat{s}^* \leq 0) = 0.678$); the bootstrap results for the test of FSD indicates that the crossing of the conditional CDFs is

[12]Note, Table 4 reports an observed FSD ranking of 46-50 minutes over 51-55 minutes, despite the left panel of Figure 8 indicating that the CDFs cross briefly at two points. This discrepancy arises due to differences in the points of the support used in the computations to conduct the actual SD tests versus generate the plots. However, the fact that the FSD ranking is statistically insignificant shows that the bootstrap is picking up these crossings and highlights the importants of utilizing statistical inference in distributional comparisons. Our computer code is written in STATA, and is available upon request.

statistically significant ($\Pr(\widehat{d}^* \leq 0) = 0.000$). Despite the insignificant SSD ranking, examination of Figure 1 is still quite informative. The plot indicates that the conditional CDFs cross only once – essentially at the median) – and low-achieving (high-achieving) students benefit from a short (long) school year. Moreover, the magnitude of the effect is quite substantial in the tails of the distribution: test scores in the bottom (top) ten percent would fall (rise) by roughly three (four) points by switching from a short to a long school year. Recall, the sample mean test score is roughly 51. This substantially heterogeneous impact of school year length clearly highlights the advantages of the distributional approach, revealing what is masked by simply reported effects on the conditional mean (as in Table 2).

The results also have important ramifications for the policy debate over school year length, at least in Texas. As stated in the introduction, the Texas state legislature is currently finalizing legislation that would impose a uniform start date across all public school districts in the state. However, because the impact on student achievement is not uniform, a decentralized decision-making process, whereby districts would have the autonomy to vary in their school year length and students could sort themselves into districts according to their preferences, seems preferable (at least from the perspective of maximizing student achievement).

Panel B and Figures 2-4 present the results pertaining to the number of class periods per day. For two of the three pairwise comparisons, we find a statistically insignificant SSD ranking favoring *fewer* class periods per day (0-6 over seven: $\Pr(\widehat{s}^* \leq 0) = 0.534$; seven over 8+: $\Pr(\widehat{s}^* \leq 0) = 0.876$); we do not observe any FSD or SSD ranking in the other case (0-6 versus 8+ class periods). Again, though, examination of the plots reveals a heterogeneous impact of school organization. All three plots suggest that low-achieving (high-achieving) students benefit from fewer (more) class periods per day. Specifically, Figures 2 and 4 indicate that students below (above) the median gain from fewer (more) class periods per day. Figure 3, on the other hand, suggests that students above the approximately tenth percentile gain from having 8+ class periods relative to only 0-6 class periods. However, relative to the effects of school year length, the magnitudes of the various impacts are small.

The final set of results – with respect to average class length – are displayed in Panel C and Figures 5-10. Consonant with the rankings of the unconditional CDFs in Table 4, we observe several rankings favoring shorter classes. However, all observable rankings are now only in the second-degree sense, highlighting the importance of incorporating the impact of school organization on not just the conditional mean of the test score distribution, but also its dispersion. In particular, the distribution associated with classes 45 minutes or less second order dominates the other three distributions. Moreover, the ranking over the distribution for classes 46-50 and 51-55 are statistically significant ($\Pr(\widehat{s}^* \leq 0) = 0.924$ and $\Pr(\widehat{s}^* \leq 0) = 0.906$, respectively). We also find a statistically significant second order ranking for the distribution associated with classes of 51-55 minutes over the distribution for classes of 56+ minutes ($\Pr(\widehat{s}^* \leq 0) = 0.910$). The remaining distributions are not rankable to a degree of statistical certainty, although we also observe that the empirical distribution for 46-50 minutes second order dominates the empirical distribution for 51-55 minutes.

Turning to the plots, we continue to find a similar pattern of heterogeneous impacts. Specifically, in all the figures, we observe that low-achieving (high-achieving) students benefit from shorter (longer) classes. Moreover, in the cases where the SSD ranking is statistically meaningful, we can conclude that the low-achieving students gain more than the high-achieving students lose from short classes, such that any policymaker with a social welfare function in the class $\mathcal{U}_2$ would prefer shorter classes. Again, these results highlight the explicit welfare framework in which SD is couched, as well as the ability of SD testing to yield insights that may enable policymakers to more easily reach a consensus.

In interpreting Figures 2-10, it is important recall that the conditional CDFs in Figures 2-4 control for the average duration of classes, and the conditional CDFs in Figure 5-10 control for the number of class periods per day. As a result, combining the findings indicates that low-achieving students benefit from fewer, shorter classes per day, while high-achieving students benefit from more and longer classes per day. This suggests that different students learn better from different length school days: low-achieving (high-achieving) students benefit from shorter (longer) school days.

5 Conclusion

The stagnation of academic achievement in the United States and elsewhere has given rise to a growing literature seeking to understand the determinants of student learning. Utilizing parametric and nonparametric techniques, we asses the impact of a heretofore relatively unexplored 'input 'in the educational process, time allocation, on the conditional mean and conditional distribution of tenth grade test performance. Our results indicate that the allocation of time to school, as measured by the length of the school year, as well as the allocation of time within the school day, as measured by the number and average duration of classes matters. However, the effects are not homogeneous across students; thus, a narrowly defined focus on the conditional mean masks the 'true' effects of such school organizational details. Moreover, the distributional analysis – based on the notion of stochastic dominance – offers opportunities to reach a consensus in policy debates that can become very politicized, as evidenced by the Texas debate over school year length.

Specifically, our distributional analysis shows that the small effects of school organization on unconditional and conditional mean test scores are extremely misleading: students below the median benefit from a shorter school year, while a longer school year benefits students above the median, and low-achieving students benefit from fewer, shorter classes per day, while high-achieving students benefit from more and longer classes per day. Thus, flexibility both within and across districts in terms of the structure of the school day and year would allow students to sort themselves in their ‘optimal’ learning environment (at least in terms of maximizing test performance).

References

[1] Abadie, A. (2002), "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 97, 284-292.

[2] Altonji, J., T. Elder, and C. Taber (2003), "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, forthcoming.

[3] Amin, S., A.S. Rai, and G. Topa (2003), "Does Microcredit Reach the Poor and Vulnerable? Evidence From Northern Bangladesh," *Journal of Development Economics*, 70, 59-82.

[4] Anderson, G. (1996), "Nonparametric Tests of Stochastic Dominance in Income Distributions," *Econometrica*, 64, 1183-1193.

[5] Barrett, G. and S. Donald (2003), "Consistent Tests for Stochastic Dominance," *Econometrica*, 71, 71-104.

[6] Betts, J.R. (2001), "The Impact of School Resources on Women's Earnings and Educational Attainment: Findings from the National Longitudinal Survey of Young Women," *Journal of Labor Economics*, 19, 635-657.

[7] Bishop, J.A., J.P. Formby, and L.A. Zeager (2000), "The Effect of Food Stamp Cashout on Undernutrition," *Economics Letters*, 67, 75-85.

[8] Bitler, M.P., J.B. Gelbach, and H.H. Hoynes (2005), "Distributional Impacts of the Self-Sufficiency Project," NBER Working Paper No. 11626.

[9] Boozer, M.A. and C. Rouse (2001), "Intraschool Variation in Class Size: Patterns and Implications," *Journal of Urban Economics*, 50, 163-189.

[10] Dearden, L., J. Ferri, and C. Meghir (2002), "The Effect of School Quality on Educational Attainment and Wages," *Review of Economics and Statistics*, 84, 1-20.

[11] Epple, D. and R. Romano (1998), "Competition Between Private and Public Schools, Vouchers, and Peer Group Effects," *American Economic Review*, 88, 33-62.

[12] Eren, O. and D.L. Millimet (2005), "Time to Misbehave?" unpublished manuscript, Department of Economics, Southern Methodist University.

[13] Figlio, D.N. (2003), "Boys Named Sue: Disruptive Children and their Peers," unpublished manuscript, Department of Economics, University of Florida.

[14] Fisher, G., D. Wilson, and K. Xu (1998), "An Empirical Analysis of Term Premiums Using Significance Tests for Stochastic Dominance," *Economics Letters*, 60, 195-203.

[15] Hanushek, E.A. (1979), "Conceptual and Empirical Issues in the Estimation of Educational Production Functions," *Journal of Human Resources*, 14, 351-388.

[16] Hanushek, E.A. (1999), "The Evidence on Class Size," in S.E. Mayer and P. Peterson (eds.) *Earning and Learning: How Schools Matter*, Washington, DC: Brookings Institute, 130-168.

[17] Hoxby, C.M. (1999), "The Productivity of Schools and Other Local Public Goods Producers," *Journal of Public Economics*, 74, 1-30.

[18] Linton, O., E. Maasoumi, and Y.J. Whang (2005), "Consistent Testing for Stochastic Dominance: A Subsampling Approach," *Review of Economic Studies*, 72, 735-765.

[19] Maasoumi, E. and A. Heshmati (2000), "Stochastic Dominance Amongst Swedish Income Distributions," *Econometric Reviews*, 19, 287-320.

[20] Maasoumi, E. and A. Heshmati (2005), "Evaluating Dominance Ranking of PSID Incomes by Various Household Attributes," IZA Discussion Paper No. 1727.

[21] Maasoumi, E. and D.L. Millimet (2005), "Robust Inference Concerning Recent Trends in U.S. Environmental Quality," *Journal of Applied Econometrics*, 20, 55-77.

[22] Maasoumi, E., D.L. Millimet, and V. Rangaprasad (2004), "Class Size and Educational Policy: Who Benefits from Smaller Classes?" unpublished manuscript, Department of Economics, Southern Methodist University.

[23] Neuman, S. and R.L. Oaxaca (2004), "Wage Decompositions with Selectivity Corrected Wage Equations: A Methodological Note," *Journal of Economic Inequality*, 2, 3-10.

[24] Pischke, J.-S. (2003), "The Impact of Length of the School Year on Student Performance and Earnings: Evidence from the German Short School Year," NBER Working Paper No. 9964.

[25] Strayhorn, C.K. (2000), "An Economic Analysis of the Changing School Start Date in Texas," *Window on State Government* Special Report, December.

[26] Todd, P.E. and K.I. Wolpin (2003), "On the Specification and Estimation of the Production Function for Cognitive Achievement," *Economic Journal*, 113, F3-F33.

A Appendix: Technical Details

A.1 Computation of $\widehat{d}$ and $\widehat{s}$

The test for FSD requires:

(i) computing the values of $\widehat{F}(z_j)$ and $\widehat{G}(z_j)$ for z_j, $j = 1, ..., J$, where J denotes the number of points in the support $\mathcal{Z}$ that are utilized ($J = 300$ in the application, where the points are equally spaced beginning at the first percentile and ending at the 99^{th} percentile of the empirical support, $\mathcal{Z}$, to focus attention away from extreme values),

(ii) computing the differences $d_1(z_j) = \widehat{F}(z_j) - \widehat{G}(z_j)$ and $d_2(z_j) = \widehat{G}(z_j) - \widehat{F}(z_j)$, and

(iii) finding $\widehat{d} = \sqrt{\frac{NM}{N+M}} \min\{\max\{d_1\}, \max\{d_2\}\}$.

If $\widehat{d} \leq 0$ (to a degree of statistical certainty), then the null of FSD is not rejected. Furthermore, if $\widehat{d} \leq 0$ and $\max\{d_1\} < 0$, then X FSD Y. On the other hand, if $\widehat{d} \leq 0$ and $\max\{d_2\} < 0$, then Y FSD X. If $\widehat{d} = \max\{d_1\} = \max\{d_2\} = 0$, then the (estimated) distributions of X and Y are identical. The test for SSD requires the following additional steps:

(i) calculating the sums $s_{1j} = \sum_{k=1}^{j} d_1(z_k)$ and $s_{2j} = \sum_{k=1}^{j} d_2(z_k)$, $j = 1, ..., J$, and

(ii) finding $\widehat{s} = \sqrt{\frac{NM}{N+M}} \min\{\max\{s_{1j}\}, \max\{s_{2j}\}\}$.

If $\widehat{s} \leq 0$ (to a degree of statistical certainty), then the null of SSD is not rejected. Moreover, if $\widehat{s} \leq 0$ and $\max\{s_{1j}\} < 0$, then X SSD Y; otherwise, if $\max\{s_{2j}\} < 0$, then Y SSD X.

Table 1. Summary Statistics.

Variable	Mean	SD
10th grade test score	51.029	9.940
School year length		
180 days or less	0.228	0.419
180+ days	0.771	0.419
Class periods		
0-6 periods	0.447	0.497
7 periods	0.345	0.475
8+ periods	0.207	0.405
Class length		
1-45 minutes	0.218	0.413
46-50 minutes	0.297	0.456
51-55 minutes	0.392	0.488
56+ minutes	0.091	0.287
8th grade test scores	51.349	9.749
8th grade composite GPA	2.984	0.725
Mother's education		
High School Dropout	0.126	0.332
High School	0.359	0.479
Junior College	0.123	0.328
College Less Than 4 Years	0.079	0.270
College Graduate	0.127	0.334
Master Degree	0.062	0.241
PH.D, M.D, etc	0.015	0.124
Father's education		
High School Dropout	0.127	0.333
High School	0.305	0.460
Junior College	0.102	0.303
College Less Than 4 Years	0.073	0.261
College Graduate	0.137	0.344
Master Degree	0.072	0.259
PH.D, M.D, etc	0.039	0.194

NOTES: Appropriate panel weights utilized. The variables listed are only a subset of those utilized in the analysis. The remainder are exlcuded in the interest of brevity. The full set of sample statistics are available upon request.

Table 2. OLS Estimates of Time Variables

Variable	Coefficient (Standard Error)
School Year Length	
180+ days	-0.101
	(0.142)
Class Periods	
7 periods	0.251
	(0.166)
8+ periods	0.212
	(0.225)
Class Length	
46-50 minutes	-0.351
	(0.211)
51-55 minutes	-0.654
	(0.254)
56+ minutes	-0.620
	(0.302)

NOTES: Appropriate panel weights utilized. Standard errors are corrected for arbitrary heteroskedasticity. Controls included for race, gender, religion, 8th grade test score, 8th grade composite GPA, father's education, mother's education, home reading material, home computer, family composition, family income, number of siblings, teacher race, teacher gender, teacher age, teacher education, class subject, class size, number of minority students in the class, region, urban and rural status, school enrollment, grade-level enrollment, average daily attendance rate in the school, racial composition of the school, percentage of students receiving free lunch in the school, average dropout rate of 10th graders prior to graduation, number of total full time teachers as well as by race and education in school, teacher salary, indicator for whether teachers have gone on strike in the past four years, percentage of students in the school in remedial reading and remedial math, and percentage of students in school in bilingual education.

Table 3. Oaxaca-Blinder Decompositions of Mean Test Score Gaps.

		Observed	Portion of Observed Gap Due to Differences in:		
		Gap	Intercepts	Endowments	Coefficients
X	***Y***		(I)	(E)	(C)
A. 10th Grade School Days					
180 Days or Less	180+ Days	-0.435	10.557	-0.601	-10.391
B. 10th Grade School Periods					
0-6 Periods	7 Periods	-0.649	-0.534	-0.674	0.559
0-6 Periods	8+ Periods	-2.590	9.853	-1.942	-10.500
7 Periods	8+ Periods	-1.941	10.386	-2.194	-10.134
C. 10th Grade Class Minutes					
1-45 Minutes	46-50 Minutes	0.838	6.508	-0.180	-5.490
1-45 Minutes	51-55 Minutes	2.642	4.367	1.005	-2.730
1-45 Minutes	56+ Minutes	1.919	-16.398	0.104	18.212
46-50 Minutes	51-55 Minutes	1.805	-2.141	1.554	2.391
46-50 Minutes	56+ Minutes	1.081	-22.905	1.307	22.679
51-55 Minutes	56+ Minutes	-0.724	-20.764	-1.140	21.181

NOTES: Appropriate panel weights utilized. Positive numbers in columns 4-6 indicate advantage to X; negative numbers indicate advantage to Y. See Table 2 for set of covariates.

Table 4. Unconditional Stochastic Dominance Tests.

Distributions		Observed	First Order Dominance						Second Order Dominance					
X	***Y***	**Ranking**	$d_{1,MAX}$	$d_{2,MAX}$	d	$Pr\{d_1^* \leq 0\}$	$Pr\{d_2^* \leq 0\}$	$Pr\{d^* \leq 0\}$	$s_{1,MAX}$	$s_{2,MAX}$	s	$Pr\{s_1^* \leq 0\}$	$Pr\{s_2^* \leq 0\}$	$Pr\{s^* \leq 0\}$
A. 10th Grade School Days														
180 Days or Less	180+ Days	None	1.155	0.045	0.045	0.000	0.066	0.066	197.195	0.002	0.002	0.010	0.380	0.390
B. 10th Grade School Periods														
0-6 Periods	7 Periods	Y FSD X	2.368	-0.043	-0.043	0.000	0.210	0.210	303.672	-0.130	-0.130	0.000	0.660	0.660
0-6 Periods	8+ Periods	Y FSD X	6.111	-0.292	-0.292	0.000	0.980	0.980	1021.302	-0.292	-0.292	0.000	1.000	1.000
7 Periods	8+ Periods	Y FSD X	4.325	-0.252	-0.252	0.000	0.782	0.782	724.110	-0.252	-0.252	0.000	0.994	0.994
C. 10th Grade Class Minutes														
1-45 Minutes	46-50 Minutes	X FSD Y	-0.013	2.334	-0.013	0.206	0.000	0.206	-0.013	300.522	-0.013	0.444	0.000	0.444
1-45 Minutes	51-55 Minutes	X FSD Y	-0.225	5.662	-0.225	0.952	0.000	0.952	-0.235	1021.129	-0.235	0.972	0.000	0.972
1-45 Minutes	56+ Minutes	X FSD Y	-0.198	3.299	-0.198	0.848	0.000	0.848	-0.198	518.836	-0.198	0.942	0.000	0.942
46-50 Minutes	51-55 Minutes	X FSD Y	-0.108	4.748	-0.108	0.590	0.000	0.590	-0.286	801.728	-0.286	0.970	0.000	0.970
46-50 Minutes	56+ Minutes	X SSD Y	0.032	2.248	0.032	0.148	0.000	0.148	-0.237	318.286	-0.237	0.868	0.000	0.868
51-55 Minutes	56+ Minutes	None	1.681	0.052	0.052	0.000	0.056	0.056	220.550	0.201	0.201	0.004	0.310	0.314

NOTES: All results use appropriate panel weights. Probabilities are obtained via 500 bootstrap repetitions. No observed ranking implies only that the distributions are not rankable in the first- or second-degree sense. See text for details.

Table 5. Full Residual Stochastic Dominance Tests.

Distributions		Observed	First Order Dominance						Second Order Dominance					
X	***Y***	**Ranking**	$d_{1,MAX}$	$d_{2,MAX}$	d	$Pr\{d_1^* \leq 0\}$	$Pr\{d_2^* \leq 0\}$	$Pr\{d^* \leq 0\}$	$s_{1,MAX}$	$s_{2,MAX}$	s	$Pr\{s_1^* \leq 0\}$	$Pr\{s_2^* \leq 0\}$	$Pr\{s^* \leq 0\}$
A. 10th Grade School Days														
180 Days or Less	180+ Days	X SSD Y	3.078	4.352	3.078	0.000	0.000	0.000	-0.437	311.437	-0.437	0.678	0.000	0.678
B. 10th Grade School Periods														
0-6 Periods	7 Periods	X SSD Y	1.304	1.515	1.304	0.000	0.000	0.000	-0.136	96.337	-0.136	0.534	0.000	0.534
0-6 Periods	8+ Periods	None	3.199	0.202	0.202	0.000	0.036	0.036	328.711	5.682	5.682	0.116	0.068	0.184
7 Periods	8+ Periods	X SSD Y	0.539	2.140	0.539	0.000	0.000	0.000	-0.269	160.081	-0.269	0.876	0.000	0.876
C. 10th Grade Class Minutes														
1-45 Minutes	46-50 Minutes	X SSD Y	1.790	5.087	1.790	0.000	0.000	0.000	-0.645	513.864	-0.645	0.924	0.000	0.924
1-45 Minutes	51-55 Minutes	X SSD Y	0.768	7.155	0.768	0.000	0.000	0.000	-0.609	730.650	-0.609	0.906	0.000	0.906
1-45 Minutes	56+ Minutes	X SSD Y	0.989	5.991	0.989	0.000	0.000	0.000	-0.851	657.962	-0.851	0.888	0.000	0.888
46-50 Minutes	51-55 Minutes	X SSD Y	1.588	3.210	1.588	0.000	0.000	0.000	-0.289	217.963	-0.289	0.794	0.000	0.794
46-50 Minutes	56+ Minutes	None	1.779	1.811	1.779	0.000	0.000	0.000	67.835	117.009	67.835	0.422	0.000	0.422
51-55 Minutes	56+ Minutes	X SSD Y	0.865	2.599	0.865	0.000	0.000	0.000	-0.226	238.753	-0.226	0.910	0.000	0.910

NOTES: All results use appropriate panel weights. Probabilities are obtained via 500 bootstrap repetitions. No observed ranking implies only that the distributions are not rankable in the first- or second-degree sense. See text for details. First-stage regression control for race, gender, religion, 8th grade test score, 8th grade composite GPA, father's education, mother's education, home reading material, home computer, family composition, family income, number of siblings, teacher race, teacher gender, teacher age, teacher education, class subject, class size, number of minority students in the class, region, urban and rural status, school enrollment, grade-level enrollment, average daily attendance rate in the school, racial composition of the school, percentage of students receiving free lunch in the school, average dropout rate of 10th graders prior to graduation, number of total full time teachers as well as by race and education in school, teacher salary, indicator for whether teachers have gone on strike in the past four years, percentage of students in the school in remedial reading and remedial math, percentage of students in school in bilingual education, number of and length of class periods (Panel A only), school year length and class period length (Panel B only), and school year length and number of class periods (Panel C only).

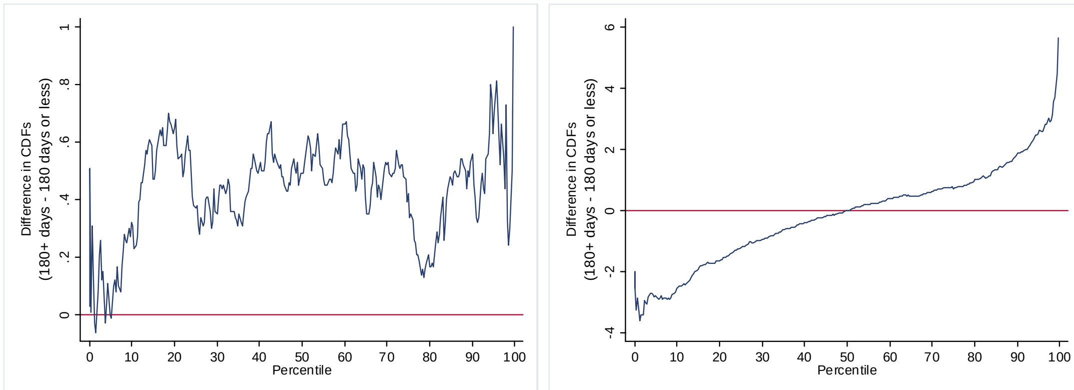

Figure 1: Difference in CDFs: 180 Days or Less & 180+ Days. Plot in the left (right) is for unconditional (full residual) SD tests.

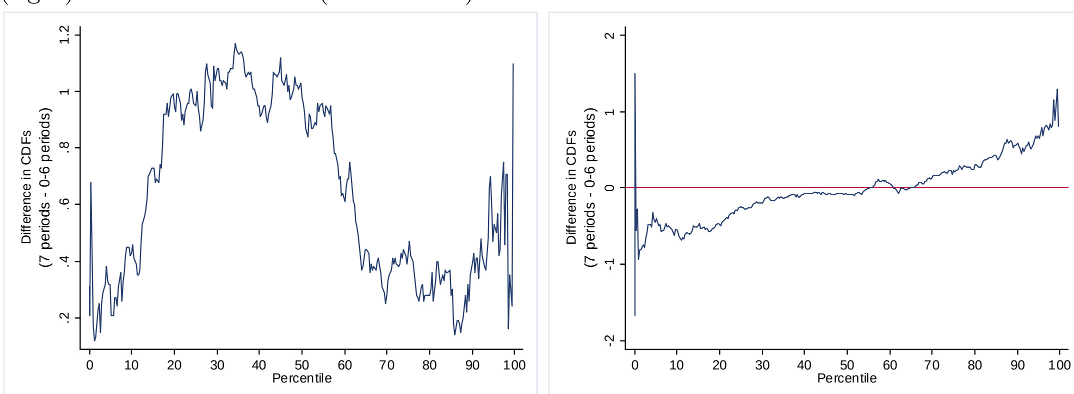

Figure 2: Difference in CDFs: 0-6 Periods & 7 Periods. Plot in the left (right) is for unconditional (full residual) SD tests.

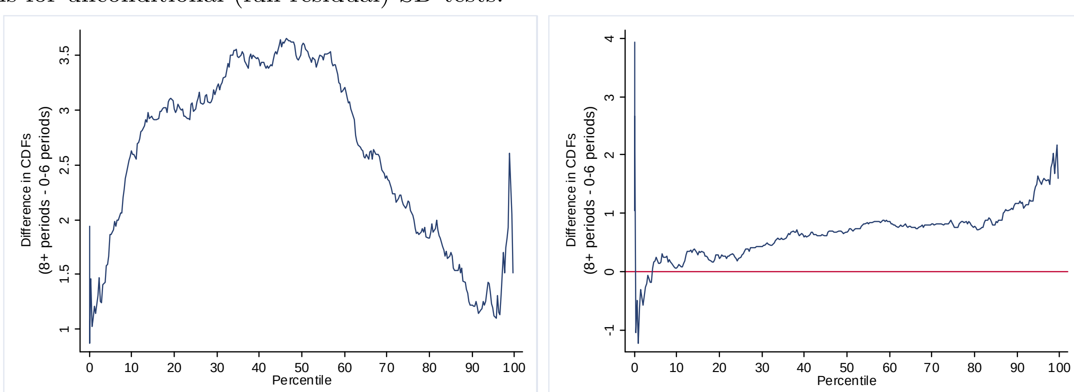

Figure 3: Difference in CDFs: 0-6 Periods & 8+ Periods. Plot in the left (right) is for unconditional (full residual) SD tests.

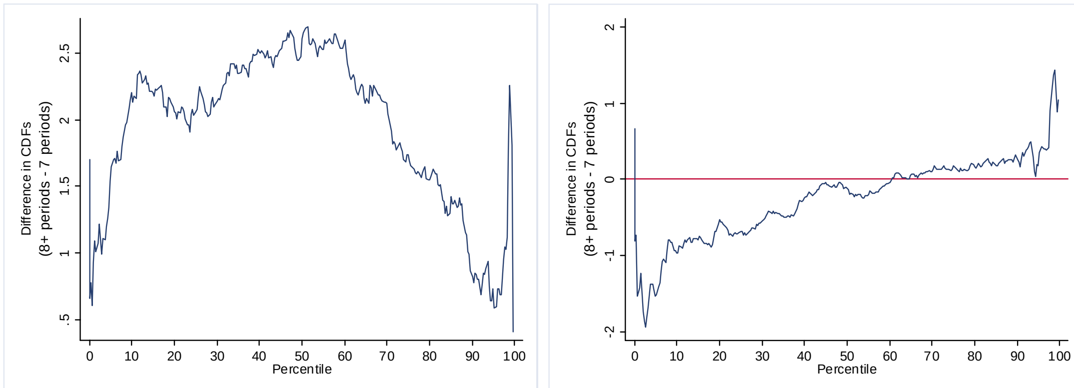

Figure 4: Difference in CDFs: 7 Periods & 8+ Periods. Plot in the left (right) is for unconditional (full residual) SD tests.

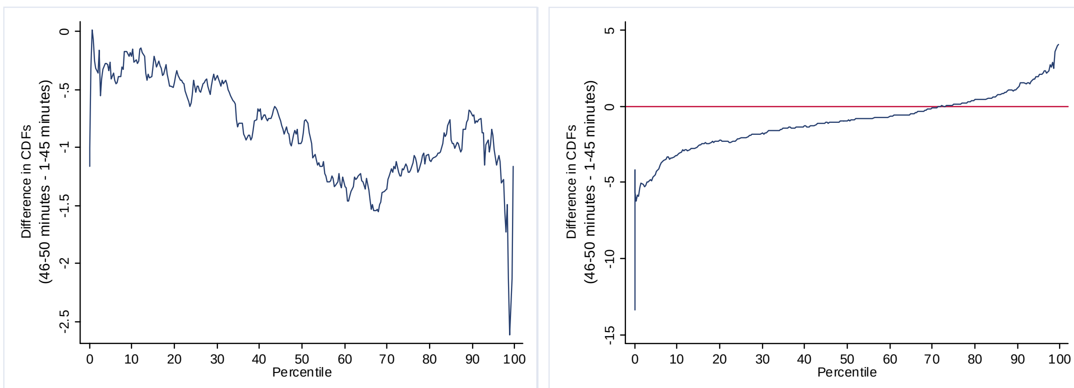

Figure 5: Difference in CDFs: 1-45 Minutes & 46-50 Minutes. Plot in the left (right) is for unconditional (full residual) SD tests.

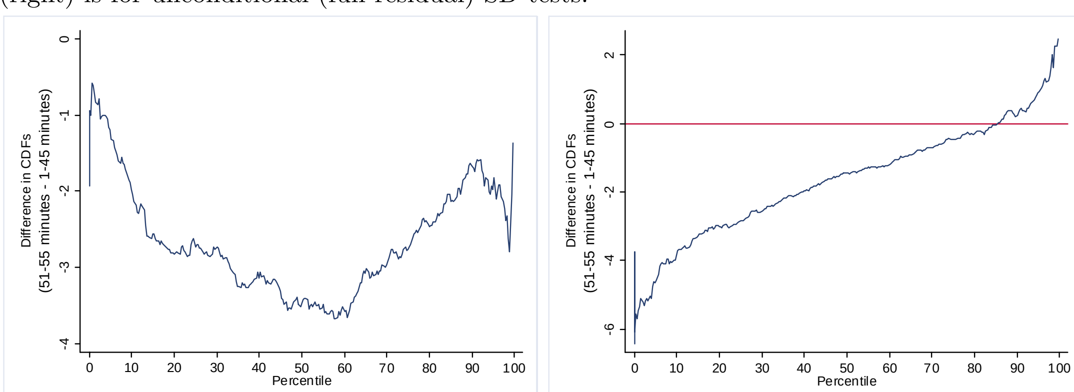

Figure 6: Difference in CDFs: 1-45 Minutes & 51-55 Minutes. Plot in the left (right) is for unconditional (full residual) SD tests.

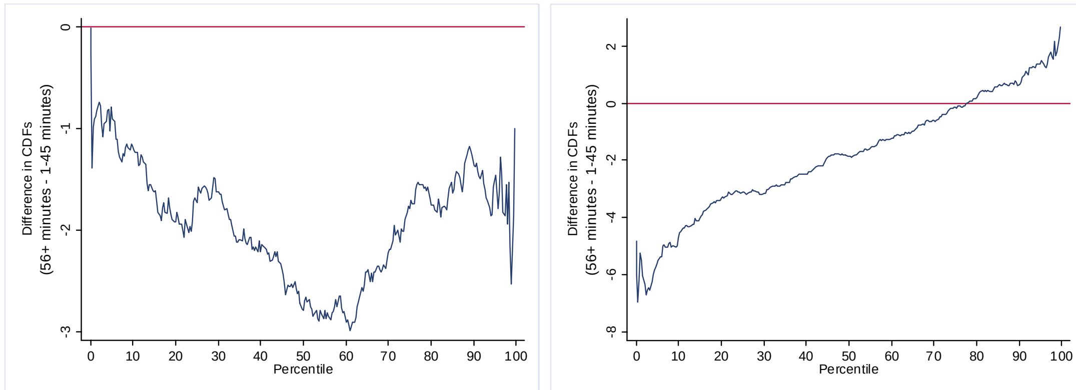

Figure 7: Difference in CDFs: 1-45 Minutes & 56+ Minutes. Plot in the left (right) is for unconditional (full residual) SD tests.

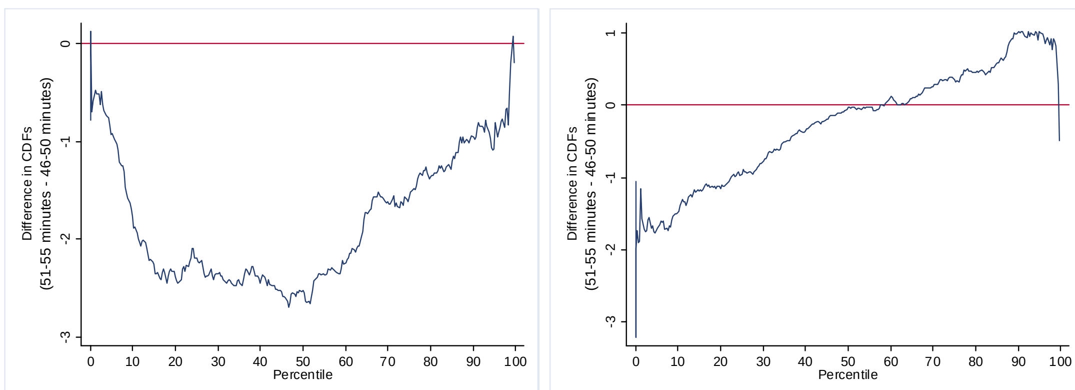

Figure 8: Difference in CDFs: 46-50 Minutes & 51-55 Minutes. Plot in the left (right) is for unconditional (full residual) SD tests.

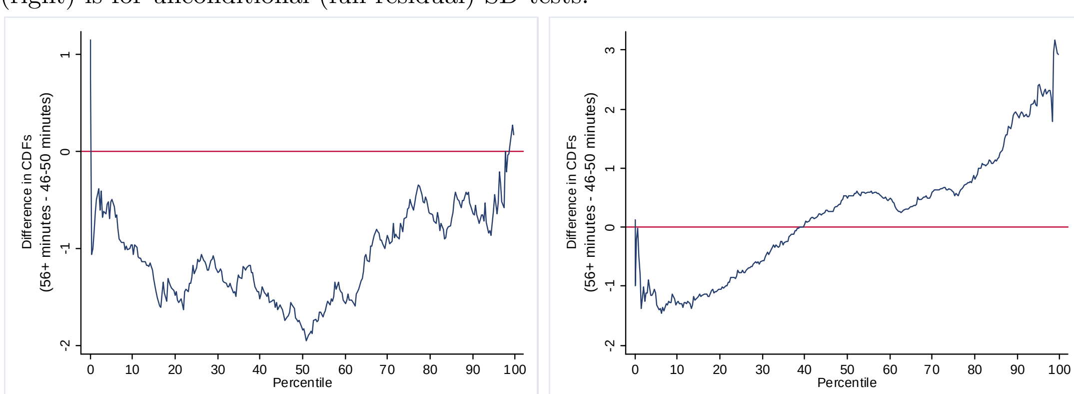

Figure 9: Difference in CDFs: 46-50 Minutes & 56+ Minutes. Plot in the left (right) is for unconditional (full residual) SD tests.

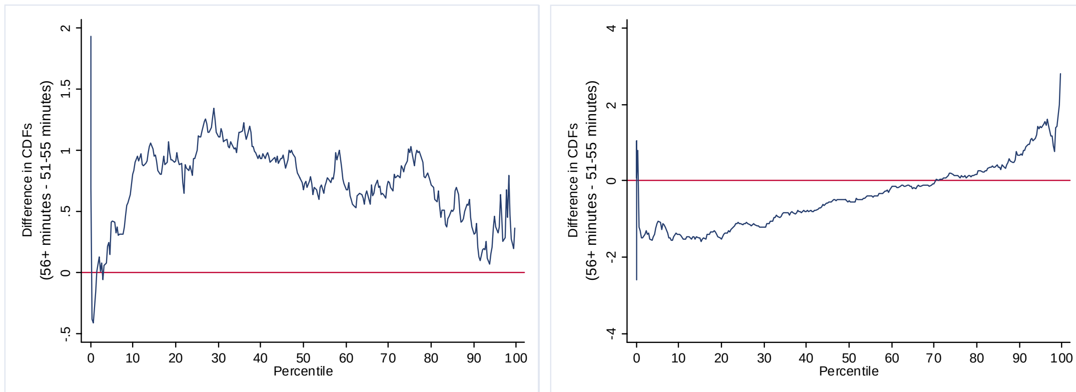

Figure 10: Difference in CDFs: 51-55 Minutes & 56+ Minutes. Plot in the left (right) is for unconditional (full residual) SD tests.