Does Schedule Irregularity Affect Productivity? Evidence from Random Assignment into College Classes*

Lester Lusher[a]†, Vasil Yasenov[b], Phuc Luong[c]

[a]University of Hawaii at Manoa, IZA

[b]Stanford University, IZA

[c]University of Virginia

Abstract

Workers with irregular or on-call work schedules constitute up to 17% of the workforce in the US. We identify the causal impact of schedule regularity on productivity by leveraging data from a Vietnamese university where freshmen were randomly assigned into highly-varying course schedules. Some schedules had consistent start times across the week, while others had extreme shifts in daily start times. Though we find a robust relationship between schedules and self-reported sleep, we precisely estimate no discernible differences in achievement across students with differing start time variability. Like prior studies, we find gains in achievement to delayed start times.

Keywords: School Start Time, Irregular Schedules, Productivity, Education Policy

JEL codes: I20, I21, I23

*We would like to thank Scott Carrell, Stefanie Fischer, Jennifer Heissel, Paco Martorell, Matthew Naven, Giovanni Peri and four anonymous referees for valuable comments. All remaining errors are our own.

†Corresponding author: Lester Lusher, University of Hawaii at Manoa, Saunders Hall 542, 2424 Maile Way, Honolulu, HI 96822 United States. Email: lrlusher@hawaii.edu; Phuc Luong, Email: ptly2n@virginia.edu; Vasil Yasenov, Email: yasenov@stanford.edu

1 Introduction

Workers with irregular or on-call work schedules constitute up to 17% of the workforce in the US. Such irregular schedules are common throughout a wide spectrum of occupations.[1] For instance, in health care, nurses and doctors may work day shifts in the beginning of the week, followed by night shifts at the end of the week. Moreover, many other jobs include "on-call" responsibilities, where employees check into work when summoned, typically due to fluctuating special events or peak hours, and often times these workers are given little notice in advance.[2] Students in higher education also frequently face irregular starts times across weekdays since courses are often divided into Monday/Wednesday/Friday and Tuesday/Thursday blocks.

A likely consequence of an irregular work or school schedule is an irregular sleep schedule. While many studies have investigated the impacts of *sleep duration* on various outcomes,[3] more recent studies have considered the impacts of *sleep regularity*. Sleep duration is simply the total amount of daily sleep one receives, while sleep regularity is the variation in the amount of sleep one receives across days, as determined by the consistency of the individual's sleep and wake times.[4] Prior studies, mostly from the physiology literature, have documented correlations between sleep regularity and numerous outcomes, including worker productivity (Åkerstedt, 2003; Ferri et al., 2016), job satisfaction, chronic fatigue, and cardiovascular symptoms (Ferri et al., 2016), obesity (Spruyt et al., 2011), and behavioral and emotional stability (Pesonen et al., 2010). Overall, individuals who have irregular sleep schedules tend to experience negative health outcomes and to be less productive.

Irregular work and school schedules may adversely impact individuals through channels other than sleep as well. For example, the Econonomic Policy Institute (2015) documents correlations between irregular work schedules and work-family conflict and stress. Irregular schedules likely also interact with "time-of-day" effects, where worker and student productivity are tied to the interaction between a) the type of task being done and b) what time-of-day the task is done. For instance, Pope (2015) documents how, holding start time constant, students tend to be more productive in the morning, and schools can improve learning by moving math (a time-of-day-sensitive course) to the morning and other classes, like English, into the afternoon. Relatedly, a popular literature has investigated the circadian rhythm, a 24-hour internal clock that cycles the body between sleepiness and alertness at regular intervals throughout the day; for instance, most adults experience a dip in energy between 1pm and 3pm. Given sleep irregularity adversely impacts the

[1] Economic Policy Institute (2015). Link: *https://www.epi.org/publication/irregular-work-scheduling-and-its-consequences/*

[2] Some studies have estimated worker willingness to pay for alternative work arrangements. For instance, in Mas and Pallais (2017) an average worker was willing to give up 20% of wagers to avoid a schedule set by an employer on short notice.

[3] For instance, Gibson and Shrader (forthcoming) estimate the causal impact of sleep duration on worker wages.

[4] Consider two individuals, one who sleeps six hours every day, and another who alternates between five and seven hours of sleep every day - these individuals attain the same amount of total sleep, but the latter individual has a more irregular sleep schedule.

circadian rhythm, schedule regularity (conditional on sleep) may interact with circadian cycles as well.[5]

A notable shortcoming from the prior literature is the absence of an identification of a causal channel between schedule regularity and outcomes. It is inherently difficult to generate exogenous variation in individual schedule regularity, as it would require variation in start times both i) within individuals and across days, and ii) across individuals. Namely, one would need to compare individuals who (exogenously) experience inconsistent start times against those who start their days at the same time and (presumably) sleep the same amount every day. Prior studies have done these comparisons by using naturally occurring data i.e. by comparing individuals across ex-post various (sleep) schedules. An obvious likely pitfall from these comparisons is that schedule and sleep regularity are related to unobservable characteristics; for instance, it may be that individuals who opt into irregular schedules and engage in irregular sleep patterns tend to live unhealthier lifestyles, which itself is the true channel for their negative health outcomes.

In this paper, we focus on the causal role of schedule regularity in an educational setting. In general, researchers and policymakers have increasingly been interested in the link between school start times and student success. Early studies documented a positive correlation between sleep duration and academic achievement (e.g., Wolfson and Carskadon, 1998; Wahlstrom, 2002; Pagel et al., 2007), while more recent studies have identified a causal link from delayed school start times to improved student performance (Carrell et al., 2011; Edwards, 2012; Heissel and Norris, 2017). With regard to sleep regularity, studies have suggested it may be a stronger predictor than sleep duration in determining student success (Phillips et al., 2017; Smarr, 2015; Trockel et al., 2000). In fact, in the Phillips et al. (2017) study, students who slept relatively little, but otherwise had a consistent sleep schedule, outperformed those who experienced, on average, higher levels of total sleep but with a higher variance in their amount of daily sleep. The discussions around these correlations have expanded to major media outlets, including CNN, USA Today, and the Independent.[6,7] Primary and secondary schools spanning 44 states have taken measures to delay start times. On a national level, the "ZZZ's to A's Act" was introduced in April 2017 to the US House of Representatives which calls on the US Secretary of Education to make policy recommendations related to the link between school start times and adolescent health, well-being, and performance.[8,9]

[5] See the National Sleep Foundation for more information at *https://sleepfoundation.org/articles/what-circadian-rhythm*

[6] Links: *http://www.cnn.com/2017/06/12/health/student-sleep-grades-study/index.html*, *http://college.usatoday.com/2017/06/12/this-simple-sleep-habit-might-be-the-secret-to-better-grades/*, and *http://www.independent.co.uk/life-style/health-and-families/regular-bedtime-linked-success-work-careers-brigham-women-s-hospital-study-research-a7787886.html*, respectively.

[7] Though the authors of these scientific studies are careful to say that they do not necessarily identify a causal link, the media coverage has thus far implicitly implied causality by prescribing suggestions for students to improve their sleep regularity.

[8] Source: *http://www.startschoollater.net.*

[9] Still, many students continue to report feeling fatigued in the classroom. In a 2015 national well-being survey with over 20,000 high schoolers, the most popular answer to the question "How do you currently feel in school?" was "Tired." 39% of students reported feeling tired, while the next two most common responses were "stressed" at 29% and "bored" at 26%. Source:

A causal relationship between schedule regularity and student performance, if shown to exist, would have critical implications for policies related to school schedules. At the post-secondary level, students could improve their outcomes by selecting classes that grant them a consistent start time. At the secondary school level, "zero-periods", where students attend an additional class before their regular school starts on particular weekdays, would have an especially detrimental effect on student performance since they harm both sleep duration and schedule regularity. Additionally, to help improve total hours of sleep, secondary schools often grant a "free-period" during a student's first-period. While this could improve sleep duration, it may also impair sleep regularity, and thus have an overall adverse effect on student performance.[10] Lastly, and perhaps most importantly, given the physiological basis for the theory of how sleep influences outcomes, identifying a causal link would likely have implications for the aforementioned correlational studies on how sleep influences worker productivity and individual health outcomes.

This paper is, to our knowledge, the first to identify the causal effects of schedule regularity on student performance. To do this, we utilize a setting where college students are randomly assigned to highly-varying course schedules during their first two semesters of enrollment. Some schedules have a consistently early/late start time across the week, while others have a high spread; start times could be as early as 6:30am or as late as 3:05pm. Our analysis, at its core, simply compares students randomly assigned to schedules with highly fluctuating daily start times against students who have relatively consistent late/early start times across the week. Our setting is particularly capable of identifying the causal effect of schedule regularity on student performance since previous studies, though occasionally equipped with random assignment, had little to no variation in school start times both *within* students (across weekdays) and *across* students.

Our analyses utilize data from a single large university in Vietnam. All incoming freshmen are randomly assigned to a schedule of courses for their first year based on their declared major. We have no reason to believe that students in this setting would be any more or less affected by school schedules than the typical student in the general population. This is especially true since the literature suggest the issues with student daytime fatigue are largely biological and are only conditional on the student's gender and progression through adolescence (Heissel and Norris, 2017).[11] In our primary sample of freshmen, the average age was 18.3 years and 43.0% were female.

https://www.usatoday.com/story/news/nation/2015/10/23/survey-students-tired-stressed-bored/74412782/. Fatigue is perhaps an even larger issue at the post-secondary level, where 50% of college students report experiencing daytime sleepiness (Hershner and Chervin, 2014).

[10] Staggered start times have also been discussed as a potential means to reduce casualties from gun violence by reducing logjams at doorways. Link: *https://edition.cnn.com/2018/05/19/us/texas-school-shooting-exits-trnd/index.html*

[11] More precisely, both a lack of total sleep or an inconsistent sleep schedule cause daytime fatigue since they interfere with a student's circadian rhythm. Students experience subtle changes to their circadian rhythm as they progress through adolescence. Thus, if an adolescent is endowed with a schedule that adversely affects their sleep cycle, then the effects are unlikely to differ by, for example, the student's ethnic or cultural background.

Our primary dataset utilizes the entire population of over 165,000 student-classroom observations across a four year period (2011-2015) from the university. Like prior studies, we estimate boosts in students performance in response to a later school start time. In response to a one hour delay in start time, we precisely estimate a 1.4 percentage point reduction in the probability a student receives a zero in the course and a 0.022 standard deviation increase in class grade for morning classes.[12] When we consider the full sample of classes, these treatment effects shrink closer to zero, suggesting that later start times had less effect on student performance in afternoon classes.[13] We do not find, however, evidence in support of a causal link between schedule regularity and student achievement. Students who are randomly assigned to school schedules with relatively consistent daily start times do not outperform their peers who are assigned highly-varying daily start times. These precisely estimated null treatment effects are consistent across a variety of specifications, including estimation of models with classroom and student fixed effects and various measures of spread of daily start times. In our preferred specification, we can rule out detrimental effects larger than 2.5% of a standard deviation on student performance in response to a one standard deviation increase in the spread of a student's daily start times across the week.

Following our initial results, in the spring of 2018, we conducted a large survey at the university asking students to report their course schedules and their average number of hours of sleep by weekday. Though results from this survey can merely provide suggestive evidence,[14] the purpose of the survey was to investigate whether and how randomly assigned schedules correlate with student sleep. In the context of null results for schedule regularity, the survey can help to understand whether schedules were actually impacting sleep in our setting. We find that on days with later school start times, students sleep more, and students with highly fluctuating start times across weekdays experience greater sleep irregularity (have a higher standard deviation of sleep hours across weekdays). We also find that students are more likely to attend class the later it is, and that a greater variation in schedule and sleep volatility is correlated with more absenteeism. Student academic satisfaction, however, is unrelated to improved sleep duration or sleep regularity.

Altogether, while our findings reconfirm earlier studies that delayed start times positively influence student success, we find little importance for consistency in start times across the week. Our results directly suggest that schedule consistency is a relatively unimportant input in the human capital production process.

[12] Described in more detail later, the indicator for receiving a zero in the course is effectively an indicator for whether the student "dropped" the class.

[13] Studies looking at the impacts of school start times include Carrell et al. (2011), Edwards (2012), Heissel and Norris (2017), and Hinrichs (2011). Other related papers looking at the impact of school schedules include Pope (2015), Lusher and Yasenov (2016), and Lusher and Yasenov (2018).

[14] Discussed in additional detail later, there are several notable shortcomings with relying on such a survey to accurately reflect actual sleep behavior. Namely, students may not accurately recall their actual average sleep hours, and students may report hours in a manner that they believe reflects the survey-taker's hypotheses.

Assuming the channel between schedules and sleep is causal, our results further imply that correlations identified by the prior literature in sleep regularity and outcomes may be driven by selection - individuals who choose irregular sleep schedules tend to be less productive or healthy.

The remainder of this paper proceeds as follows. Section 2 discusses the institutional features, data, and summary statistics. Section 3 introduces our econometric specification and tests for randomization. Section 4 presents our results. Section 5 concludes.

2 Institution Background and Data

2.1 University Background

Our data come from a private university located in Vietnam. It is mostly regarded as a relatively average university within the country. It follows a semester system similar to the one in the U.S. where school starts in September and ends in June. Furthermore, the university offers two degrees, a three-year Associates' and a four-year Bachelor's, across three fields: Technology, Business, and Design. Although three-year students may take some Bachelor's level courses, most of their core courses are independent and less demanding. Our data contain information on enrolled students in both degrees. The majority of incoming freshmen at this university live on or nearby campus.

All incoming students have a single declared major, and each major is embedded within a single department. Regardless of their major and department, all students take certain common core subjects, such as Math and English. Additionally, there are some department-specific core subjects that students have to take regardless of their majors within a department; for example, the required courses for first-year students in the Technology Department are Math A1, Math A2, Physics, Chemistry, and Foreign Language.[15] On very rare instances, first-year students take some major-specific courses; for instance, Marketing majors have to take a Web Design course in their first semester. The same course can have both major-specific and department-specific designations across multiple majors and departments such that no first-year course is strictly composed of the same major; so, for instance, though Web Design is a required course for Marketing majors, Web Design courses will also have students from other majors (due to major-specification) and departments (due to department-specification) as well. Generally, nearly all first-year courses are part of a university requirement or a department-specific requirement.

Courses are broken down into multiple sections (i.e., meeting times), with each section taught by a single

[15]The Technology department itself has different majors: e.g., Software Engineering, Information Technology, and Web Development. So, all first-year students in these three majors must take the above-listed courses.

professor. Different from conventional universities, admitted freshmen strictly follow a rigorous academic curriculum established by the staff and faculty. That is, first-year students cannot freely choose their courses and sections. Instead, all first-year students within a major take the same courses, and are allocated into their sections through a multi-stage randomization mechanism; first, within each major, an incoming freshman is designated to one of several potential "homerooms" for the school year. For instance, suppose electrical engineering majors are divided into seven homerooms (EE_1 to EE_7), each with 10 students. Students are randomly assigned to their homerooms, stratified by the first letter of their first name. For example, if there are three electrical engineering majors whose first name starts with the letter P, these students would be distributed across three different homerooms, randomly; one motivation the university stratifies by first name is to ensure that no one homeroom has too many similar names so as to inconvenience instructors. Then, the university randomly assigns homerooms into sections based on course-requirements for the major. For example, the first section for Math A1 may be randomly assigned the EE_5 homeroom from electrical engineering, the M_2 homeroom from marketing, and so on.[16] In short, students by major are randomized into small groups ("homerooms") stratified by the first letter of their first name, and these homerooms are then randomly allocated to course-sections. In Section 3, we present several tests for randomization into course-sections.[17]

Table 1 shows the breakdown of potential meeting times for course-sections. For incoming freshmen, sections can start as early as 6:30am (Period 1) and as late as 3:05pm (Period 10). While each period lasts 45 minutes, sections themselves typically constitute three to four periods. For instance, a section that starts at 6:30am will usually last until 9am or 9:50am. A five to ten minute break is designated between each period. Unlike the U.S., each section only meets once a week and on the same day of the week for the entire semester. Furthermore, meeting times could land on Saturdays such that students are only guaranteed to have no classes on Sundays.

The university follows an 11 point integer grading system ranging from zero to ten. A grade of five or higher is regarded as a passing grade and any grade below that is considered a failing one. A grade of nine or ten is equivalent to an A+ in the U.S. system, an eight to an A, a seven to a B+, a six to a B, and a five to a C. For the majority of courses, students' final grades are determined by a combination of section attendance and performance on midterms and a final examination.[18] Typically, final examinations

[16] Aside from randomization into sections, these homerooms serve no other administrative purpose. Still, students are guaranteed to have classmates from their homeroom across all their course-sections, and so these homerooms may serve as a "study group" for students.

[17] Our de-identified data do not include the first letter of the students' first name.

[18] Anecdotally, students rarely "make up" missed sections by attending a different section, primarily because instructors either did not allow it, or because the instructor did not have an alternative course-section the student could attend. Unfortunately, our data do not contain information on course-section absences.

determine more than 50% of a student's final grade in the course. Both midterms and final exams are taken in the students' assigned section times. The exam questions on the midterms may differ across sections for the same course, while the final exams are identical within a course-semester.[19] Since first-year students follow a strict curricula conditional on their major, students do not have the opportunity to drop their assigned courses and subsequently alter their school schedule. Thus, students who perhaps want to drop a course and consequently put little to no effort in the course end up conceding a final grade of zero.

2.2 Student-classroom performance data

Our primary dataset centers on over 10,000 admitted freshmen at the university from 2011 to 2015. The purpose of this dataset is to look at how student grades are impacted by their schedules. Table 2 presents summary statistics for this sample. In Panel A we show that 57% of these students are male, and the average incoming age was 18.3 years old. Nearly 93% of the students are of the predominant race in Vietnam. We also show the statistics for college entrance exam scores. In Panel B we focus on variables at the "classroom" level, which we define as a combination of course, section, and term. The data contain 3,386 total classrooms, and the average class size was nearly 50 students. The average grade received across classes (not including students who received zeroes) was 6.3, which is roughly equivalent to between a B and B+ in the U.S. system. About half of the classes in our sample were in a STEM field. The average start time of a classroom was 9:08am. With a sample size of over 165,000 observations, Panel C presents summary statistics at the unit of observation for our analyses (student-classroom) for the outcomes of interest.[20] Nearly 10% of our observations correspond to a zero grade, which is mostly comprised of students who essentially "dropped" the class. The average raw grade (conditional on the student receiving a non-zero grade) at the student-classroom level was just under 6.0, which is equivalent to a B in the U.S. system.[21] Lastly, Panel D shows the average standard deviation of school start time was 1 hour and 51 minutes.

Figure 1 presents the distribution of raw scores received across our sample of student-class combinations. Nearly 10% of our observations are marked with a grade of zero, which is an amalgamation of students who essentially stopped participating and "dropped" the course, and students who perhaps significantly failed it despite some effort. In our analysis, we examine several outcome variables, the first being an indicator for

[19]Discussed in more detail later, our models will be able to control for any material or grading differences across course-section-terms with course-section-term (i.e. classroom) fixed effects.

[20]A very small portion of our data contain students who attended the school as freshmen for more than two semesters. We drop the 674 observations which pertain to a student's attendance beyond their first two semesters.

[21]Note that the averages of raw grades need not equal at the classroom and student-classroom levels since classroom level means do not weight by class size. The differences in our statistics imply that larger classes tended to give out lower grades.

whether the student received a zero in the class. We also analyze their raw score, conditional on not receiving a zero. Finally, we consider a normalization of the raw grades to a mean of zero and a standard deviation of one by course-section-term, which allows our estimates to be interpreted in standard deviation units and thus be easily compared to estimates from previous studies. Moreover, in conjunction with classroom fixed effects, it accounts for differential grading practices across classrooms.

2.3 Sleep survey data

Our second dataset centers on a survey conducted in the spring of 2018 on enrolled students at the university. The purpose of this survey was to ask students to report their hours of sleep, by weekday, and link these responses to their school start times and other self-assessments. The survey was advertised as a short survey to be conducted online through Google Forms, and was distributed by university faculty and staff. Survey-takers received no tangible benefits, and were told that their responses would be important for evaluation purposes.

The primary questions of interest on this survey collected the student's school start time across the six school days, and the number of hours they slept on average the night before those school days. With these two series of questions, we can directly test a "first-stage" effect of whether the randomly assigned schedules influenced the students' self-reported sleep. We also asked students to report their satisfaction with their academic performance for the spring 2018 semester, and whether they felt that early school start times (and a higher spread of school start times) were harming their performance. The full list of questions on the survey and corresponding summary graphs can be found in Appendix A.1. A total of 588 students completed the survey. We then took the student-level responses and reshaped the data to the student-weekday level (n=2,972). Importantly, each observation contains information on the time of the student's first class for the school day and how many hours they slept on average the night prior.

There are several shortcomings worth noting when it comes to relying on survey information for sleep behavior. Namely, it is unlikely that students can perfectly recall their average number of sleep hours by day. An estimate of the relationship between start times and sleep hours would be biased then if this measurement error is associated with start time. For instance, it may be that students assume that they sleep fewer (more) hours than they actually sleep on early (late) start days. Exacerbating this potential issue would be a type of "experimenter's bias" effect, where the students may be completing the survey in a manner which conforms to the experimenters' hypotheses. In this case, if students believe that they are "supposed to" sleep more (less) on later (early) start days, then they may report sleeping more (fewer) hours on the days that they have later (earlier) school start times. These shortcomings are insurmountable absent observation of actual sleep behavior, and so later results from our sleep survey are interpreted accordingly with caution.

3 Econometric Specifications and Identification

3.1 Specification using student-classroom performance data

Our primary analyses utilize our student performance data to estimate the following specification:

$$y_{icst} = \alpha + \beta \times StartTime_{icst} + \gamma \times sd[StartTime]_{it} + \lambda_{cst} + \lambda_i + \mu X_{icst} + u_{icst} \tag{1}$$

where y_{icst} is a grade-based outcome for student i taking course c (e.g., Principles of Microeconomics) in section s in school term t (e.g., Fall semester of year 2016). We define a "classroom" as a course-section-term combination cst. $StartTime_{icst}$ is the time (in hours) of student i's earliest classroom on the weekday for which classroom cst held lecture. For instance, if on Saturdays a student was randomly assigned a first classroom at 12:30pm, then this variable would equal 12.50 for all the student's classrooms that were taught on Saturdays.[22,23] Next, $sd[StartTime_{it}]$ is a student-term level variable that captures the spread of the student's school start times across all of his/her class days for each of the student's two freshman terms t. In our primary specification, we calculate the standard deviation in start times for student i across all class days.[24] Note that if a student was randomly assigned the same start time across all class days, then this variable would equal zero. To test the sensitivity of our results, we also consider the inter-quartile range of start times.[25] Figure 2 presents the distribution of these two covariates, in which we observe large variation in both variables: some students have consistently early/late school start times, while others have volatile school start times.

Since our primary regressors of interest are determined at the student-classroom and student-term levels,

[22]Any start times that did not correspond to a whole hour were coded with appropriate fractions. Any start times after noon were coded using "military time" (e.g., 2:15pm as 14.25).

[23]Given that students with different late start times are perhaps more likely to wake up at similar times (e.g., students with starts times of 12:30pm vs. 1:20pm are likely to wake up at the same time), our default analyses "top-codes" for $StartTime_{icst}$ at 12.50. Additional analyses consider various top-codes, and do not significantly impact the main results.

[24]That is, $sd[StartTime]_{it} = \sqrt{\frac{\sum_{d=1}^{6}(StartTime_{id} - \overline{StartTime_{it}})^2}{5}}$ where $StartTime_{idt}$ denotes the start time of student i's classes on school day d (Monday/Tuesday/Wednesday/Thursday/Friday/Saturday), and $\overline{StartTime_{it}}$ is the average daily start time for student i in term t. If a student had a day with no classes, then $sd[StartTime]_{it}$ was calculated using the student's five school days i.e. we make no assumptions about their "start time" during their non-school day.

[25]One limitation of our primary calculations for the spread of start times is that we do not have information on "start times" on Sundays or on days in which the student does not have any classes. Accounting for both factors will require making assumptions about student habits on those days (and perhaps their habits during the night prior to their off-day). For robustness checks, in Appendix Table A3, Table A4, Table A5, we consider our main results while making several assumptions about the student's start time on their off days.

we can simultaneously estimate classroom fixed effects λ_{cst} and student fixed effects λ_i. The former focus on using variation within classrooms and across students, and are estimable since students enrolled in the same classroom may have very different daily start times. Further, classroom fixed effects control for any unobserved factors that vary at the classroom level and affect student performance. For instance, they control for instructor fixed effects since each classroom is taught by exactly one instructor. These, in turn, control for the possibility that particular types of students, by chance, were randomly assigned to classrooms with instructors who were systematically different than other ones. Classroom fixed effects also alleviate the need to have a setting with standardized grading or testing procedures across classrooms since all students within a classroom are assigned the exact same homework and exams. Next, we can control for any student characteristics which would influence the student's academic outcomes (e.g., inherent ability, major choice, family income, living proximity from campus) with student fixed effects. They are estimated by using variation in a student's start times, and spread of start times, across their two semesters of enrollment. As robustness checks, in specifications without student fixed effects we are able to control for gender, age, ethnicity, pre-collegiate score and major fixed effects. Finally, X_{icst} linearly controls for the classroom order for student i in the day the student's classroom cst meets (e.g. the second class of the day for the student).

Our setting grants us the unique capability of estimating both β and γ without concerns of endogeneity since students, conditional on major, were randomly assigned to their school schedule during their first two semesters of enrollment.[26] We can interpret γ as the causal effect of a one standard deviation increase in the spread of daily start times across the week on student success. Similarly, β is the causal effect of a one-hour increase in daily start time. Under the hypothesis that sleep duration improves student success, we should expect a positive coefficient for β: later start times let students sleep more, which leads to improved outcomes. Under the hypothesis that sleep regularity impacts student success, we would expect a negative estimate for γ: students experience an increasingly irregular sleep pattern the higher the spread of start times across the week, which negatively influences their performance.

In addition to this mean regression approach we estimate unconditional quantile regressions following Firpo et al. (2009). This method allows for estimating heterogeneous impacts of school start time and its standard deviation on various points of the unconditional standardized grade distribution. These could be important if the mean overall and mean subgroup effects we estimate mask underlying heterogeneity. For

[26]In a typical college setting with natural variation in daily start times, researchers would face a major issue of selection bias - students who opt into schedules with lower spread of start times (or later start times) may be systematically different in unobservable ways than students a higher spread of start times (or earlier start times). These unobserved factors could then correlate with student outcomes, introducing endogeneity bias.

instance, it may be the case that higher performing students are less affected by course schedules than lower performing students and these groups are not clearly identified by the covariates in the dataset. The quantile regressions, therefore, paint a more complete picture of the statistical relationship between the variables of interest. In practice, we estimate a linear RIF-OLS model using Borgen et al. (2016)'s approach to include fixed effects.[27]

3.2 Specification using sleep survey data

Lastly, to test whether the randomly assigned schedules influence sleep, we use our sleep survey to estimate the following equation:

$$y_{id} = \alpha + \beta \times StartTime_{id} + \gamma \times sd[StartTime]_i + u_{id} \tag{2}$$

where y_{id} is a sleep-related outcome for student i on school day d (Monday through Saturday). Much like our primary specification, $StartTime_{id}$ is the time (in hours) of student i's earliest class on school day d. For instance, if on Tuesdays a student was randomly assigned a first class at 8:15am, then this variable would equal 8.25. Next, $sd[StartTime_i]$ is a student level variable that captures the spread of the student's school start times across all of his/her class days, as measured using the standard deviation.

The primary outcomes of interest with this dataset are 1) student-day level average number of sleep hours (*sleep duration*) and 2) student level spread of average number of sleep hours across days (*sleep regularity*). For sleep duration, we would expect that later start times would coincide with more sleep, and so β can be interpreted as the predicted increase in sleep hours in response to a one hour increase in school start time. For sleep regularity, we would also expect that a greater variance of school start times across days would be associated with a higher variance of sleep hours across days. So, γ would be interpreted as the predicted increase in the standard deviation of a student's sleep hours across days in response to a one standard deviation increase in the spread of his/her daily school start time. In short, positive and statistically significant coefficients β for sleep duration and γ for sleep regularity can be interpreted as evidence of a "first-stage" effect of randomly assigned schedules on sleep.

[27] An alternative estimation method we have considered is the two sample instrumental variable. However, this strategy does not fit our setting well because the two datasets we have are in different format (student-class versus student-day), we do not have individual demographics in the survey data (except for gender) and would not know how to adjust the standard errors for within student correlation. Moreover, our two datasets have wildly different sample sizes which will further undermine the validity of our standard errors. To the best of our knowledge, this problem is yet to be solved in the applied econometrics literature.

3.3 Balance Test

As is customary with any study assuming exogeneity, we can test for whether the schedule assignment policy was truly random. We undertake two distinct approaches. First, we examine whether student characteristics are associated with start times or the spread of start times, conditional on major. That is, we regress each of the student characteristics on our two covariates of interest, $StartTime_{icst}$ and $sd[StartTime]_{it}$, while controlling for major and classroom fixed effects. Results from this analysis will tell us whether conditional on major and classroom, particular types of students were more likely to have a higher spread of start times (or earlier start times). If assignment to schedules was in fact random (conditional on major), then observable student characteristics should have no statistically meaningful association with $StartTime_{icst}$ and $sd[StartTime]_{it}$. Our second approach, presented in the results section, is to examine the sensitivity of our estimated coefficients from equation (1) to the inclusion of student fixed effects. If unobserved student characteristics are associated with (the spread of) start times, and these characteristics have a meaningful relationship with student performance, then we would expect our estimates for β and γ to shift in response to the inclusion of student fixed effects.

Table 3 presents the results from the regressions of our four student characteristics on the two covariates of interest.[28] The only statistically significant coefficients we find are for start time on student age and ethnicity, which are significant at the 10% level. However, both coefficients (of 0.006 and -0.002) are economically tiny; for instance, the former suggests that a one hour increase in start time is associated with a 0.006 year (or 2.19 days) increase in student age. The remaining coefficients are, in addition to being statistically insignificant, economically insignificant. Furthermore, the covariate we would perhaps be most concerned about, pre-collegiate test scores, is completely unassociated with start times and the spread of start times.

For completeness, we present in brackets the p-values from testing the null hypothesis of the true coefficients equaling zero while correcting for testing multiple hypotheses using the Holm-Bonferroni method (Holm, 1979). The lack of small p-values suggests that the coefficient on age likely achieved significance by natural variation, as opposed to actual sorting of students into start times by age. Additionally, presented in the Appendix Table A1, the results from the balance test remain unchanged when we remove major fixed effects, suggesting that even though scheduling was done conditional on major, start times were roughly balanced across majors. Overall, this analysis supports the notion that student schedules were randomly assigned. Moreover, as discussed in further detail in the next section, we find that our estimates for β and γ

[28] Due to data limitations, the university was only able to provide four student-level characteristics for the student performance data. Only gender was collected for the sleep survey data, and thus we do not utilize the survey for a balance test.

remain relatively unchanged in response to student controls and fixed effects.

4 Results

4.1 Student performance results

The impact of delayed school start time

Table 4 presents our main student performance results, where we estimate equation (1) separately across our two main covariates $StartTime_{icst}$ and $sd[StartTime]_{it}$. We first focus on Panel A, which much like previous studies, strictly investigates the impacts of having a delayed school start time on student outcomes. Assuming delayed start times lead to increased sleep duration, and assuming increased sleep improves student performance, we would expect later start times to improve student outcomes. Each column considers a separate regression, and each set of three columns considers our three outcome variables of interest: whether the student received a zero in the course, their raw grade (between 1 and 10), and their standardized grade for the course. Note that the latter two outcomes are conditional on the student not receiving a zero in the course; results for these outcomes can then be interpreted as estimating a treatment effect for a subsample of students who have already been "selected" partially based on the treatment.[29] For each of the three outcomes, we test the sensitivity of the results to the inclusion of student controls (the four student controls from the balance test and major fixed effects) and student fixed effects. Robust standard errors are clustered at the student level.

While we find that delayed school start times had no impact on the probability a student received a zero in the course, we do find statistically significant improvements in course grade (both raw and standardized). Results remain relatively consistent in response to the inclusion of student controls or fixed effects, highlighting the exogenous variation generated by random assignment of students into schedules. From our fully specified models in columns (6) and (9), we predict a 0.011 increase in course grade (0.008 standard deviation increase) in response to a one hour delay in school start time.[30]

Other studies looking at the impacts of sleep duration on student success include Carrell et al. (2011), who find decreases in students performance between 0.12 and 0.14 standard deviations across all classes through the day in response to being randomly assigned a 7am first period class, Edwards (2012) who estimates a two percentage point gain in math test scores in response to relatively small delays in school

[29] More precisely, estimates for the impact of start time on these latter two outcomes would be understated if there is a strong "first-stage" impact on receiving a zero in the course, assuming students induced toward receiving a zero were lower ability and would have been more responsive to the impacts of delayed start times had receiving a zero not been possible.

[30] Later results highlight larger impacts of delayed school start time on morning classes.

start time, and Heissel and Norris (2017) who estimate effects between 0.06 and 0.08 standard deviations in response to delaying start times by one hour. On the other hand, Hinrichs (2011) finds no evidence of student improvement in response to delayed start times. Other related papers include Pope (2015), who identifies varying time-of-day effects on student achievement, Lusher and Yasenov (2016), who find no effects of "double shift schooling" policies on student achievement, and Lusher and Yasenov (2018), who find that boys experience larger gains in the classroom in response to delayed start times.

The impact of start time volatility

In Panel B of Table 4, we turn our focus to the impacts of school start time volatility on student outcomes. Here, start time volatility is measured as the standard deviation of the student's start times across his/her six days of school. If a student has the same start time everyday, then the standard deviation of start times equals zero; this measure increases the more fluctuating a student's start times across the six days are. Assuming that randomly assigned start time volatility harms sleep regularity (the consistency of one's sleep across days), we may expect more volatile start times to be associated with negative student outcomes.

We do not find any evidence in support of this hypothesis: students who were randomly assigned consistent daily start times do not significantly outperform students with highly fluctuating daily start times. We do not estimate a statistically significant relationship for each of our fully specified models across the three outcome variables of interest.[31] From column (9), we predict a 0.004 standard deviation decrease in course grade in response to a one standard deviation increase in the spread of a student's daily school start times. This coefficient is precisely estimated as well: we can rule out detrimental effects as large as 2.5% of a standard deviation in student grades in response to a one standard deviation increase in the spread of a student's daily school start times. To put this bound in perspective, Carrell and West (2010) find that a one standard deviation increase in professor quality is associated with a 5% increase in student scores, or roughly double the size of our upper bound. From Panel C of Table 4, we see that these results are robust to measuring the spread of a student's daily school start times using interquartile range instead of standard deviation.

4.2 Heterogeneity

In this subsection, we consider several subsamples for which start times would plausibly have an especially impactful effect. These results are presented in Table 5. The most obvious subsample of interest is

[31] From columns (4) and (5), we actually estimate an *increase* in a student's raw grade in response to an increase in the spread of daily school start times, but these estimates lose statistical significance (at the 5% level) after including student fixed effects.

the group of classes that were taught in the morning. These are classes for which the consequences of an irregular schedule (and lack of sleep) are likely to have an especially detrimental effect on student achievement. First, we still find that the spread of daily start times has no impact on student performance in morning classes. Second, the estimated effects of delayed school start time relative to the full sample results increase. We estimate a statistically significant 1.4 percentage point reduction in the probability the student received a zero in the class, and a 2.4% of a standard deviation increase in grade, in response to a one hour delay in school start time.

Next, we consider subsamples by gender. One study by Lusher and Yasenov (2018) suggests that primarily due to physiological reasons, male students tend to have a stronger negative response to earlier start times than females. Our results partially support this hypothesis. On one hand, male students do appear to be relatively sensitive to the spread of start times; though our estimate is statistically insignificantly different from zero, the estimate for standardized grade (-0.021) is notably larger than that from the full sample (-0.004). Still, this estimate is rather small in economic terms: a one standard deviation increase in the spread of start times causes a 2.1% of a standard deviation decrease in grade for male students. On the other hand, delayed start times do not appear to be especially beneficial for male students.

For completeness, we find no notable differences when we focus on low-achieving students, as proxied by having a below-median pre-collegiate exam score, or when focusing on courses in STEM fields. We also observe no notable differences when we focus on student-semesters with above-median $sd[StartTime]_{it}$ i.e. those students who had especially volatile schedules. We also find that our results remain relatively unchanged when we consider a "top-coding" for school start time. Top-coding accounts for the plausibility that the variation in later school start times is unlikely to generate variation of sleep across students relative to earlier school start times. For example, it may be that switching from a 6:30am start to an 8:15am start will result in more sleep when juxtaposed with switching from 9:05am to 10:50am. Specifically, under the column titled "9am top coding" in Table 5, all observations with start times after 9am were coded with a start time of 9am and $sd[StartTime]_{it}$ was recalculated accordingly. Lastly, we show a specification in which we cluster the standard errors two-way by student and classroom; though standard errors slightly increase, our conclusions remain essentially unaffected, as we are still able to estimate a precise zero effect of schedule volatility.

Finally, a potential explanation of the observed null effects is that the overall sample and each subgroup consist of similar numbers of "losers" and "gainers" to varying start times and therefore the mean effect is close to zero. To analyze this possibility we estimated a series of unconditional quantile regressions with the standardized grade as an outcome while controlling for student fixed effects. Each coefficient is interpreted as the effect of irregular schedules on the respective quantile of the outcome. Figure 3 shows the estimated coefficients of both measures of sleep irregularity. The gray shaded regions correspond to 95% confidence intervals. This analysis does not uncover any masked heterogeneity. The coefficients of "σ(Start Time)" (left panel) and IQR(Start Time)" (right panel) fluctuate around zero and are largely statistically insignificant. Overall, Figure 3 further confirms the earlier results of no statistically and economically significant impacts of sleep regularity on students' grades.

4.3 Sleep survey results

An underlying assumption for the posited hypotheses is that the randomly assigned start times influenced student sleep. Table 6 presents results from our student sleep survey. Standard errors are clustered at the student level. In the first cell, we regress sleep duration on start time, and find that a one-hour delay in start time is associated with a statistically significant 4.3 minute increase in sleep. In the second column we regress sleep duration on both start time and start time volatility, measured as the standard deviation of the student's start times across the six school days, and find that more volatile schedules are associated with less sleep (though this estimate is statistically insignificant). In the next column, we find that start time volatility is statistically significantly positively related to the standard deviation of sleep hours across school days. In other words, the more varying a student's daily start time is, the more they experience an irregular sleep schedule. In the third column, we find that students with later school start times are not any less dissatisfied with their academic performance; we also estimate no statistically significant relationship between a student's start time volatility and their academic satisfaction. Finally, we find that delayed school start times are statistically significantly associated with decreases in absenteeism, and that students with highly volatile school schedules tend to miss more classes overall. Overall, we find strong evidence that schedules impact self-reported sleep, with mixed evidence on schedules impacting self-reported student performance.[32]

5 Conclusions

Individuals across numerous work and school environments face irregular start times in their schedules. In this paper we estimate the causal effect of schedule regularity and start times on student performance by utilizing student-course level data from a large university in Vietnam. Crucially, in this setting, all incoming

[32]Though our survey elicited major, these regressions do not control for major fixed effects due to a lack of remaining variation. Concern for this omission is mitigated by the fact that balance test results remain very similar with and without including major fixed effects.

freshmen are randomly assigned to a schedule of courses. These schedules varied greatly, where some had the same early/late daily start time across the week, while others had highly-fluctuating daily start times across the week. On top of random assignment, the setting also allows us to control for any unobservable differences across classrooms and students by estimating models with both classroom and student fixed effects.

Like prior studies, we find that delayed school start times positively influences student grades. Delayed school start times have an especially beneficial effect on morning classes, with only a minimal effect on afternoon classes. However, across a variety of specifications and classifications, we do not find any evidence that higher spreads of start times adversely affected student performance. In our preferred specification, we can rule out detrimental effects larger than 2.5% of a standard deviation in course grade in response to a one standard deviation increase in the spread of daily start times. Subsequent survey evidence indeed suggests that randomly assigned school start times influences both sleep duration and sleep regularity in our setting: students with later start times tend to sleep more, and students with volatile daily start times tend to have fluctuating sleep hours across weekdays. Thus, our null results for schedule volatility perhaps cannot be attributed to randomly assigned schedules not impacting sleep regularity.

These findings have several important implications. With respect to the correlational studies on sleep regularity and student achievement (Phillips et al., 2017; Smarr, 2015; Trockel et al., 2000), our findings suggest that the link between sleep regularity and student achievement is likely largely driven by unobserved omitted variables. That is, students who self-select into irregular sleep schedules are low achieving due to reasons beyond getting an irregular amount of sleep. For example, it may be that lower achieving students tend to procrastinate (Beattie et al., 2016), and thus these students cram their studying until the night before an exam, which then leads to an irregular sleep schedule. Given the biological basis for how sleep theoretical impacts productivity and health, our results also suggest that prior correlations outside of education may also be driven by a selection bias - individuals who sort into occupations or shifts with irregular hours may tend to be individuals who tend to experience adverse productivity or health outcomes irrespective of the regularity of their sleep schedule.

References

ÅKERSTEDT, T. (2003): "Shift work and disturbed sleep/wakefulness," *Occupational medicine*, 53, 89–94.

BEATTIE, G., J.-W. P. LALIBERTÉ, AND P. OREOPOULOS (2016): "Thrivers and Divers: Using Non-Academic Measures to Predict College Success and Failure," Tech. rep., National Bureau of Economic Research.

BORGEN, N. T. ET AL. (2016): "Fixed effects in unconditional quantile regression," *Stata Journal*, 16, 403–415.

CARRELL, S. E., T. MAGHAKIAN, AND J. E. WEST (2011): "A's from Zzzz's? The causal effect of school start time on the academic achievement of adolescents," *American Economic Journal: Economic Policy*, 3, 62–81.

CARRELL, S. E. AND J. E. WEST (2010): "Does professor quality matter? Evidence from random assignment of students to professors," *Journal of Political Economy*, 118, 409–432.

EDWARDS, F. (2012): "Early to rise? The effect of daily start times on academic performance," *Economics of Education Review*, 31, 970–983.

FERRI, P., M. GUADI, L. MARCHESELLI, S. BALDUZZI, D. MAGNANI, AND R. DI LORENZO (2016): "The impact of shift work on the psychological and physical health of nurses in a general hospital: a comparison between rotating night shifts and day shifts," *Risk management and healthcare policy*, 9, 203.

FIRPO, S., N. M. FORTIN, AND T. LEMIEUX (2009): "Unconditional quantile regressions," *Econometrica*, 77, 953–973.

GIBSON, M. AND J. SHRADER (forthcoming): "Time use and productivity: The wage returns to sleep," *The Review of Economics and Statistics*.

HEISSEL, J. AND S. NORRIS (2017): "Rise and shine: The effect of school start times on academic performance from childhood through puberty," *Journal of Human Resources*, 0815–7346R1.

HERSHNER, S. D. AND R. D. CHERVIN (2014): "Causes and consequences of sleepiness among college students," *Nat Sci Sleep*, 6, 73–84.

HINRICHS, P. (2011): "When the Bell Tolls: The Effects of School Starting Times on Academic Achievement," *Education Finance and Policy*, 6, 486–507.

HOLM, S. (1979): "A simple sequentially rejective multiple test procedure," *Scandinavian journal of statistics*, 65–70.

LEE, K. A., G. MCENANY, AND D. WEEKES (1999): "Gender differences in sleep patterns for early adolescents," *Journal of adolescent health*, 24, 16–20.

LUSHER, L. AND V. YASENOV (2016): "Double-shift schooling and student success: Quasi-experimental evidence from Europe," *Economics Letters*, 139, 36–39.

——— (2018): "Gender performance gaps: Quasi-experimental evidence on the role of gender differences in sleep cycles," *Economic Inquiry*.

MAS, A. AND A. PALLAIS (2017): "Valuing alternative work arrangements," *American Economic Review*, 107, 3722–59.

PAGEL, J. F., N. FORISTER, AND C. KWIATKOWKI (2007): "Adolescent sleep disturbance and school performance: the confounding variable of socioeconomics." *Journal of clinical sleep medicine: JCSM: official publication of the American Academy of Sleep Medicine*, 3, 19–23.

PESONEN, A.-K., K. RÄIKKÖNEN, E. J. PAAVONEN, K. HEINONEN, N. KOMSI, J. LAHTI, E. KAJANTIE, A.-L. JÄRVENPÄÄ, AND T. STRANDBERG (2010): "Sleep duration and regularity are associated with behavioral problems in 8-year-old children," *International journal of behavioral medicine*, 17, 298–305.

PHILLIPS, A. J. K., W. M. CLERX, C. S. OBRIEN, A. SANO, L. K. BARGER, R. W. PICARD, S. W. LOCKLEY, E. B. KLERMAN, AND C. A. CZEISLER (2017): "Irregular sleep/wake patterns are associated with poorer academic performance and delayed circadian and sleep/wake timing," *Scientific Reports*.

POPE, N. G. (2015): "How the Time of Day Affects Productivity: Evidence from School Schedules," *Review of Economics and Statistics*, 98, 1–11.

SMARR, B. L. (2015): "Digital sleep logs reveal potential impacts of modern temporal structure on class performance in different chronotypes," *Journal of biological rhythms*, 30, 61–67.

SPRUYT, K., D. L. MOLFESE, AND D. GOZAL (2011): "Sleep duration, sleep regularity, body weight, and metabolic homeostasis in school-aged children," *Pediatrics*, 127, e345–e352.

TROCKEL, M. T., M. D. BARNES, AND D. L. EGGET (2000): "Health-related variables and academic performance among first-year college students: implications for sleep and other behaviors," *Journal of American college health*, 49, 125–131.

WAHLSTROM, K. (2002): "Changing times: Findings from the first longitudinal study of later high school start times," *Nassp Bulletin*, 86, 3–21.

WOLFSON, A. R. AND M. A. CARSKADON (1998): "Sleep schedules and daytime functioning in adolescents," *Child development*, 69, 875–887.

6 Tables and Figures

Table 1: Class Schedules

	Period #	Start Time - End Time
Morning	1	06 : 30 am - 07 : 15 am
	2	07 : 20 am - 08 : 05 am
	3	08 : 15 am - 09 : 00 am
	4	09 : 05 am - 09 : 50 am
	5	10 : 00 am - 10 : 45 am
	6	10 : 50 am - 11 : 35 am
		Lunch Break
Afternoon	7	12 : 30pm - 1 : 15 pm
	8	1 : 20 pm - 2 : 05 pm
	9	2 : 15 pm - 3 : 00 pm
	10	3 : 05 pm - 3 : 50 pm
	11	4 : 00 pm - 4 : 45 pm
	12	4 : 50 pm - 5 : 35 pm

Notes: A list of potential meeting times for course-sections. A single class typically lasts three periods.

Table 2: Summary Statistics

	Mean	SD	Obsevations
Panel A. Student level characteristics			10,130
Male	0.569	0.495	
Age	18.322	0.802	
Native Ethnicity	0.926	0.262	
Pre-collegiate Grade	14.755	4.486	
Panel B. Course-section-term (classroom) level characteristics			3,386
Number of Students	49.145	28.074	
STEM Course	0.494	0.500	
Grade (conditional on non-zero)	6.271	1.226	
Start Time	09:08 am	2hr 27min	
Panel C. Student-classroom level outcomes			166,404
Received Zero	0.092	0.289	
Raw Grade (conditional on non-zero)	5.967	1.910	
Standardized Grade (conditional on non-zero)	0.000	0.989	
Panel D. Student-term characteristics			20,080
σ(Start Time)	1 hr 51 min	1 hr 0 min	
IQR(Start Time)	2 hrs 24 min	2 hrs 4 min	

Notes: Each panel shows summary statistics for variables determined at different aggregation levels.

Table 3: Balance Test

	Indicator for Male	Age	Native Ethnicity	Pre-Collegiate Score
Start Time	-0.001	0.006*	-0.002*	0.000
	(0.001)	(0.003)	(0.001)	(0.013)
	[1.00]	[0.511]	[0.512]	[1.00]
σ(Start Time)	-0.004	0.021	-0.000	-0.029
	(0.006)	(0.014)	(0.004)	(0.057)
	[1.00]	[0.864]	[1.00]	[1.00]
Observations	166404	166404	166404	166404

Notes: Each entry on the first and fourth rows present the estimated coefficient of a regression of a student characteristic (shown in the columns) on the respective school scheduling variable while controlling for major and classroom fixed effects. The parenthesis (second and fifth rows) display the estimated standard error clustered by student. The brackets [third and sixth rows] present the corresponding p-values from testing a null hypothesis of statistical significance of the true coefficient while correcting for multiple testing with the Holm-Bonferroni method (Holm, 1979).

Table 4: Student Performance Results

	Received zero			Raw grade [1,10]			Standardized grade		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Start Time									
Start Time	0.001	0.001	-0.000	0.017**	0.014**	0.011*	0.011**	0.009**	0.008**
	(0.001)	(0.001)	(0.001)	(0.007)	(0.007)	(0.006)	(0.004)	(0.004)	(0.004)
Observations	166404	166404	166404	151155	151155	151062	151148	151148	151055
Panel B: Schedule irregularity									
σ(Start Time)	-0.006	-0.007*	-0.004	0.036*	0.041**	0.002	0.020	0.022*	-0.004
	(0.004)	(0.004)	(0.005)	(0.020)	(0.020)	(0.017)	(0.013)	(0.013)	(0.011)
Observations	166404	166404	166404	151155	151155	151062	151148	151148	151055
Panel C: Schedule irregularity									
IQR(Start Time)	-0.000	-0.000	-0.001	0.010	0.009	-0.002	0.006	0.004	-0.004
	(0.002)	(0.002)	(0.002)	(0.009)	(0.009)	(0.007)	(0.006)	(0.006)	(0.005)
Observations	166404	166404	166404	151155	151155	151062	151148	151148	151055
Classroom FE	X	X	X	X	X	X	X	X	X
Student controls		X			X			X	
Student FE			X			X			X

Notes: The unit of observation is student-classroom. "Start Time" is measured in hours. "σ(Start Time)" is the standard deviation of the student's school start times across days of the week. "IQR(Start Time)" is the interquartile range of the student's school start times across days of the week. Student controls include gender, age, ethnicity, pre-college test score and major fixed effects. In addition, all regressions control for student-day class order. Standard errors are clustered at the student level. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Table 5: Student Performance Results by Subsamples

	Subsample							
	Full sample	Morning classes	Male students	Low achievers	High Schedule Volatility	STEM courses	9am top coding	Two-way Clustering
Outcome: Received zero								
Start Time	0.001	-0.014***	0.001	0.001	0.001	0.001	0.001	0.001
	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
σ(Start Time)	-0.004	0.001	-0.004	-0.007	0.008	-0.006	-0.020*	-0.004
	(0.005)	(0.006)	(0.006)	(0.006)	(0.012)	(0.006)	(0.010)	(0.005)
Observations	166404	84020	98730	103078	84543	73812	166404	166404
Outcome: Raw grade [1,10]								
Start Time	0.011	0.015	0.012	0.006	0.010	0.008	0.011	0.011
	(0.007)	(0.026)	(0.009)	(0.009)	(0.009)	(0.015)	(0.007)	(0.009)
σ(Start Time)	0.002	0.027	-0.026	0.016	0.052	0.005	0.002	0.002
	(0.017)	(0.024)	(0.022)	(0.023)	(0.042)	(0.025)	(0.038)	(0.023)
Observations	151062	76274	87460	94534	76319	67082	151062	151062
Outcome: Standardized grade								
Start Time	0.009*	0.024	0.008	0.003	0.010	0.006	0.008*	0.009
	(0.005)	(0.020)	(0.006)	(0.006)	(0.006)	(0.010)	(0.005)	(0.006)
σ(Start Time)	-0.004	0.014	-0.021	0.002	0.019	-0.005	-0.003	-0.004
	(0.011)	(0.016)	(0.013)	(0.014)	(0.026)	(0.014)	(0.023)	(0.014)
Observations	151055	76267	87453	94527	76319	67075	151055	151055
Classroom FE	X	X	X	X	X	X	X	X
Student FE	X	X	X	X	X	X	X	X

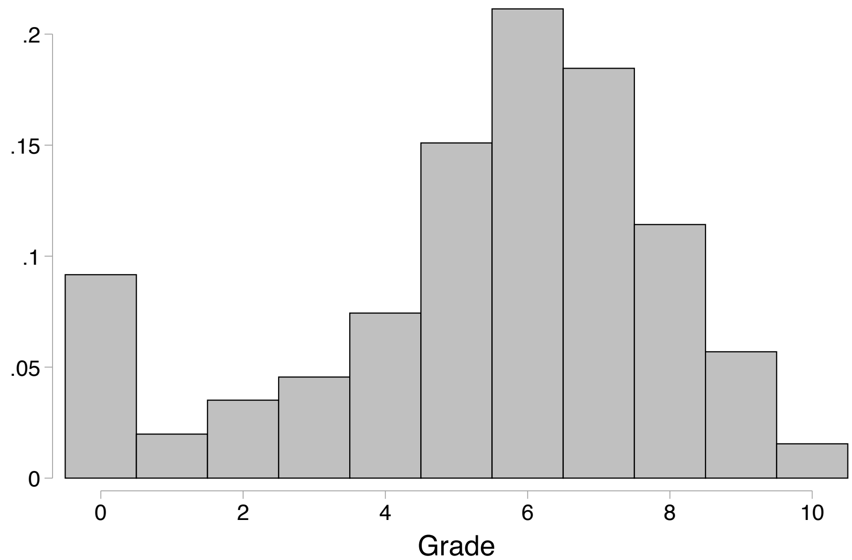
Notes: The unit of observation is student-classroom. "Start Time" is measured in hours. "σ(Start Time)" is the standard deviation of the student's school start times across days of the week. All regressions control for student-day class order. Standard errors are clustered at the student level except under "Two-way clustering", where standard errors are clustered at the student and classroom level. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Table 6: Sleep Survey Results

	Sleep duration: Minutes of sleep	Sleep duration: Minutes of sleep	Sleep irregularity: σ(Minutes of sleep)	(Very) Dissatisfied with academic performance?	Always/often/sometimes misses class?
Start Time	4.332***	4.960***		-0.002	-0.027***
	(0.706)	(0.594)		(0.003)	(0.004)
σ(Start Time)		-3.553	4.633***	-0.020	0.072***
		(3.017)	(1.527)	(0.020)	(0.015)
N (Student-day level)	2972	2972	2972	2972	2972
Mean of outcome	410.249	410.249	55.277	0.315	0.403

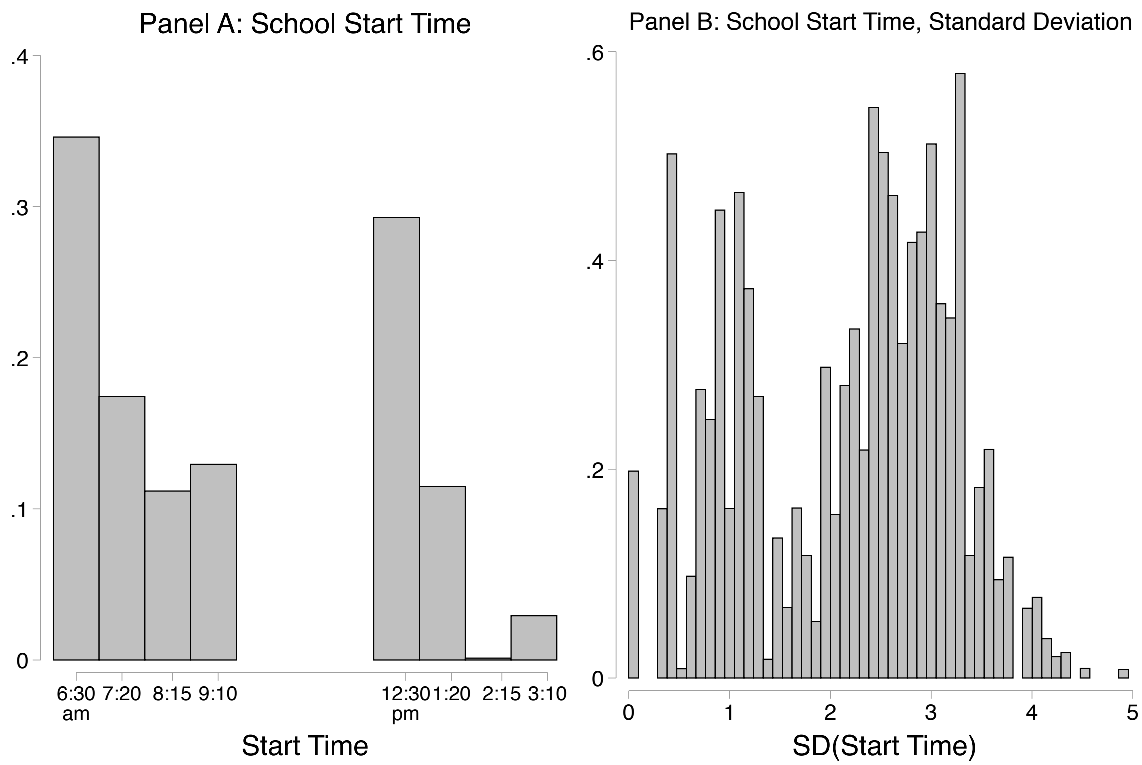
Notes: The unit of observation is at the student-day level. Each column pertains to a single regression. "Start Time" is measured in hours. "σ(Start Time)" is calculated at the student level, and is the standard deviation of a student's school start times across the six school days of the week. The first two columns pertains to the amount of sleep (in minutes) the student reported receiving by day, on average. The outcome variable for the third column is at the student level, calculated as the standard deviation of the student's hours of sleep across the six school days. The fourth column pertains to the outcome variable for whether the student was either dissatisfied or very dissatisfied with their academic performance, while the fourth column considers whether the student reported either sometimes, often, or always missing their class (as opposed to never missing their class). Standard errors are clustered at the student level. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Figure 1: The Distribution of Raw Grades



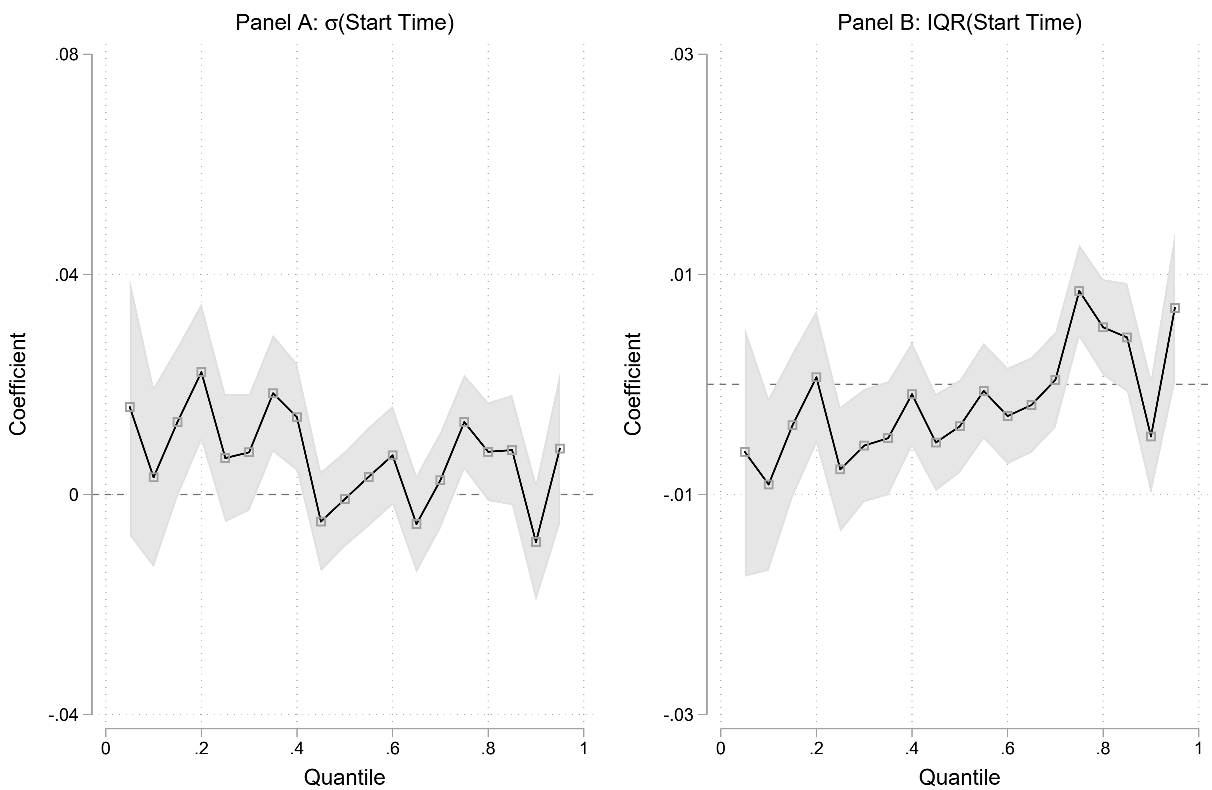
Notes: Grade denotes the raw (non-standardized) final grade received for each student-classroom combination.

Figure 2: The Distributions of Our Two Primary Start Time Measures



Notes: The left panel visualizes the variation of daily school start times across student-classroom observations. The right panel displays the distribution of the standard deviation of a student's daily school start times across all student-term pairs.

Figure 3: Quantile Regression Results



Notes: Panel A (B) shows the estimated coefficients of "σ(Start Time)" ("IQR(Start Time)") on various quantile regressions with standardized grade as the outcome variable. These are estimated via RIF regressions. All regressions have student fixed effects and control for the concurrent number of classes taken by the student within the same day. Standard errors are clustered at the student level.

Appendix A.1 - Questions from student sleep survey

Personal Information

1. **What is your current class standing?**
   - ○ First-year
   - ○ Second-year
   - ○ Third-year
   - ○ Fourth-year
   - ○ Other
2. **What is your major?**

   ____

   ____
3. **What is your gender?**
   - ○ Male
   - ○ Female

Survey Questions

4. **Do you think that a volatile school start schedule affects your sleep?**
   - ○ Yes
   - ○ It depends
   - ○ No
5. **Do you believe that you sleep less, or get worse quality sleep, when your school start schedule is more volatile?**
   - ○ Yes
   - ○ It depends
   - ○ No
6. **How satisfied are you with your performance in school so far this semester?**
   - ○ Very unsatisfied
   - ○ Unsatisfied
   - ○ Apathetic
   - ○ Satisfied
   - ○ Very satisfied

7. Do you believe that you sleep more, or get better quality sleep, when you do not have a class early in the morning?

○ Yes

○ It depends

○ No

8. Do you feel you are more likely to miss a class the earlier it starts?

○ Yes

○ It depends

○ No

Class Schedule

9. What time does your first class start on the following days?

- Monday:
- Tuesday:
- Wednesday:
- Thursday:
- Friday:
- Saturday:

Sleep Schedule

10. On average approximately how many hours of sleep do you get on the following nights?

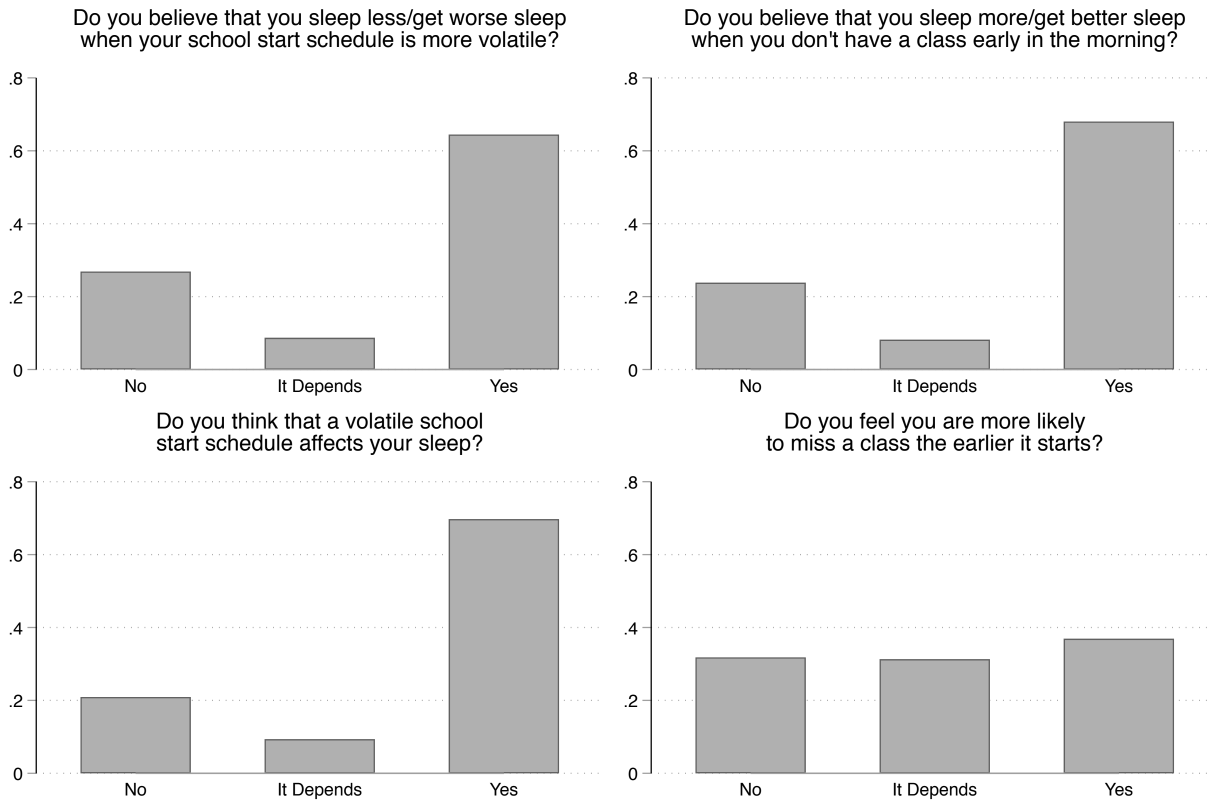
- Monday night:
- Tuesday night:
- Wednesday night:
- Thursday night:
- Friday night:
- Saturday night:

Class Attendance

11. How often do you miss or skip your first class on the following days:

11a. Monday?	◯ Never	◯ Sometimes	◯ Often	◯ Always	◯ No class
11b. Tuesday?	◯ Never	◯ Sometimes	◯ Often	◯ Always	◯ No class
11c. Wednesday?	◯ Never	◯ Sometimes	◯ Often	◯ Always	◯ No class
11d. Thursday?	◯ Never	◯ Sometimes	◯ Often	◯ Always	◯ No class
11e. Friday?	◯ Never	◯ Sometimes	◯ Often	◯ Always	◯ No class
11f. Saturday?	◯ Never	◯ Sometimes	◯ Often	◯ Always	◯ No class

Answers to Sleep-related Survey Questions



Notes: Each bar graph displays the distribution of answers to a separate sleep-related question in our survey. The question is written in the bar headline and the units are fractions.

Appendix A.2 - Additional Tables

Table A1: Balance Test - Removing Major Fixed Effects

	Indicator for Male	Age	Native Ethnicity	Pre-Collegiate Score
Start Time	-0.000	0.007*	-0.002*	-0.007
	(0.001)	(0.003)	(0.001)	(0.013)
	[1.00]	[0.416]	[0.574]	[1.00]
σ(Start Time)	-0.008	0.015	-0.001	0.005
	(0.006)	(0.014)	(0.004)	(0.056)
	[1.00]	[1.00]	[1.00]	[1.00]
Observations	166404	166404	166404	166404

Notes: Each entry on the first and fourth rows present the estimated coefficient of a regression of a student characteristic (shown in the columns) on the respective school scheduling variable while controlling for classroom fixed effects. The parenthesis (second and fifth rows) display the estimated standard error clustered by student. The brackets [third and sixth rows] present the corresponding p-values from testing a null hypothesis of statistical significance of the true coefficient while correcting for multiple testing with the Holm-Bonferroni method (Holm, 1979).

Table A2: Alternate Specifications for School Start Time Effect

	(1)	(2)	(3)	(4)
Panel A: Received Zero				
6:30 am	0.003			
	(0.006)			
7:20 am or earlier		0.004		
		(0.005)		
8:15 am or earlier			0.004	
			(0.005)	
9:05 am or earlier				-0.002
				(0.004)
Observations	166404	166404	166404	166404
Panel B: Raw Grade [1,10]				
6:30 am	-0.040			
	(0.039)			
7:20 am or earlier		-0.048		
		(0.029)		
8:15 am or earlier			-0.039	
			(0.029)	
9:05 am or earlier				-0.055*
				(0.030)
Observations	151062	151062	151062	151062
Panel C: Standardized Grade				
6:30 am	-0.034			
	(0.026)			
7:20 am or earlier		-0.040**		
		(0.020)		
8:15 am or earlier			-0.035*	
			(0.020)	
9:05 am or earlier				-0.036*
				(0.019)
Observations	151055	151055	151055	151055

Notes: The unit of observation is student-classroom. All regressions have student and classroom fixed effects. Standard errors are clustered at the student level. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Table A3: Additional schedule irregularity results - Assume start time on off days is same as average of start time on school days

	Received zero			Raw grade [1,10]			Standardized grade		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
σ(Start Time)	-0.006	-0.007	-0.004	0.044**	0.049**	0.001	0.024*	0.026*	-0.005
	(0.005)	(0.005)	(0.005)	(0.022)	(0.022)	(0.019)	(0.014)	(0.014)	(0.012)
Observations	166404	166404	166404	151155	151155	151062	151148	151148	151055
Classroom FE	X	X	X	X	X	X	X	X	X
Student controls		X			X			X	
Student FE			X			X			X

Notes: The unit of observation is student-classroom. "Start Time" is measured in hours. "σ(Start Time)" is the standard deviation of the student's school start times across days of the week. Student controls include gender, age, ethnicity, pre-college test score and major fixed effects. In addition, all regressions control for student-day class order. Standard errors are clustered at the student level. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Table A4: Additional schedule irregularity results - Assume start times of 6:30am and 3:05pm on off days for low-achieving students

	Received zero			Raw grade [1,10]			Standardized grade		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Assigning 6:30 AM Start Time									
σ(Start Time)	-0.007	-0.007	-0.002	0.010	0.012	0.002	0.004	0.003	-0.007
	(0.006)	(0.007)	(0.007)	(0.027)	(0.028)	(0.027)	(0.018)	(0.018)	(0.017)
Observations	103078	103078	103078	94583	94583	94534	94576	94576	94527
Panel B: Assigning 3:05 PM Start Time									
σ(Start Time)	-0.010*	-0.010	-0.001	0.014	0.022	0.000	0.007	0.009	-0.010
	(0.006)	(0.006)	(0.007)	(0.028)	(0.028)	(0.025)	(0.018)	(0.018)	(0.016)
Observations	103078	103078	103078	94583	94583	94534	94576	94576	94527
Classroom FE	X	X	X	X	X	X	X	X	X
Student controls		X			X			X	
Student FE			X			X			X

Notes: Sample includes students who scored at or below median in pre-collegiate test score. The unit of observation is student-classroom. "Start Time" is measured in hours. "σ(Start Time)" is the standard deviation of the student's school start times across days of the week. Student controls include gender, age, ethnicity, pre-college test score and major fixed effects. In addition, all regressions control for student-day class order. Standard errors are clustered at the student level. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Table A5: Additional schedule irregularity results - Indicator for whether the student had both AM and PM start times across the week

	Received zero			Raw grade [1,10]			Standardized grade		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\mathbb{1}$(Start Time)	-0.003	-0.004	-0.001	0.071*	0.068*	0.000	0.041	0.036	-0.008
	(0.009)	(0.010)	(0.009)	(0.041)	(0.040)	(0.033)	(0.025)	(0.025)	(0.021)
Observations	166404	166404	166404	151155	151155	151062	151148	151148	151055
Classroom FE	X	X	X	X	X	X	X	X	X
Student controls		X			X			X	
Student FE			X			X			X

Notes: The unit of observation is student-classroom. "Start Time" is measured in hours. "$\mathbb{1}$(Start Time)" is an indicator for whether the students daily start times fluctuate between AM or PM (i.e. an indicator equal to one if the student has start times in both the morning and the afternoon, or equal to zero if the student has start time entirely in the morning or in the afternoon). Student controls include gender, age, ethnicity, pre-college test score and major fixed effects. In addition, all regressions control for student-day class order. Standard errors are clustered at the student level. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively.