University of Hawai`i at Mānoa
Department of Economics
Working Paper Series

Saunders Hall 542, 2424 Maile Way,
Honolulu, HI 96822
Phone: (808) 956 -8496
www.economics.hawaii.edu

Working Paper No. 17-4

Sleep and Student Success: The Role of Regularity vs. Duration

By
Phuc Luong
Lester Lusher
Vasil Yasenov

October 2017

Sleep and Student Success: The Role of Regularity vs. Duration*

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

[a]University of California, Davis

[b]University of Hawaii at Manoa

[c]Goldman School of Public Policy, University of California, Berkeley

[d]Institute for the Study of Labor (IZA)

Abstract

Recent correlational studies and media reports have suggested that sleep regularity - the variation in the amount of sleep one gets across days - is a stronger determinant of student success than sleep duration - the total amount of sleep one receives. We identify the causal impacts of sleep regularity and sleep duration on student success by leveraging over 165,000 student-classroom observations from a large university in Vietnam where incoming freshmen were randomly assigned into course schedules. These schedules varied significantly: some had the same daily start time across the week, while others experienced extreme shifts. Across a multitude of specifications and samples, we precisely estimate no discernible differences in achievement between students with highly varying start times versus students with consistent schedules. Moreover, we find much smaller gains to delayed school start times compared to previous studies.

Keywords: School Start Time, Sleep Regularity, Education Policy

JEL codes: I20, I21, I23

*We would like to thank Scott Carrell, Jennifer Heissel, Paco Martorell, Matthew Naven, and Giovanni Peri 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: pthluong@ucdavis.edu; Vasil Yasenov, Email: yasenov@berkeley.edu

1 Introduction

Researchers and policymakers have increasingly been interested in the link between sleep and student success. Early studies documented a positive correlation between hours of sleep 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). 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.[1] Despite increased discussions on the issue of student sleep deprivation and movements to delay school start times, 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."[2] 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).

While the debate around the merits of additional sleep continues, recent scientific studies have suggested that *sleep regularity* may be a stronger correlate than *sleep duration* in determining student success (Phillips et al., 2017; Smarr, 2015; Trockel et al., 2000). That is, students who experience a steady level of daily sleep with consistent sleep and wake times (*sleep regularity*) tend to perform better academically than students who simply get a higher total number of hours of daily sleep (*sleep duration*).[3] This discussion has also expanded to major media outlets, including CNN, USA Today, and the Independent.[4] Though the authors

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

[2]Specifically, 39% of students reported feeling tired. The next two most common responses were "stressed" at 29% and "bored" at 26%. Source: *https://www.usatoday.com/story/news/nation/2015/10/23/survey-students-tired-stressed-bored/74412782/*

[3]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.

[4]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.

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.

This finding, if proven to be causal, would have critical implications for policies related to school schedules. At the post-secondary level, college students are frequently left with heavily imbalanced schedules, with some weekdays starting lecture in the early morning and others beginning in the afternoon. 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 sleep 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. Finally, other studies have identified correlations between sleep regularity and other outcomes, including obesity (Spruyt et al., 2011), behavioral and emotional stability (Pesonen et al., 2010), and worker productivity (Åkerstedt, 2003). Given the biological basis for the theory of how sleep influences outcomes, identifying a causal link between sleep regularity and student performance would likely have implications for these other settings as well.

This paper is, to our knowledge, the first to identify the causal effects of sleep regularity versus sleep duration 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 consistent 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 uniquely capable of identifying the causal effect of sleep regularity on student performance since previous studies, though occasionally equipped with random assignment, had no variation in school start times both *within* (across weekdays) and *across* students.

Our analysis utilizes over 165,000 student-classroom observations across a four year period (2011-2015) from a 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).[5] In our sample of freshmen, the average age was 18.3 years and 43.0% were female.

We find no evidence in support of a causal link between sleep regularity and student achievement. Students who are randomly assigned to school schedules with relatively consistent daily start times do not outperform students 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, multidimensional clustering of standard errors, and various measures of spread of daily start times. In our preferred specification, we can rule out detrimental effects larger than 2% 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.

Moreover, we estimate only small 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 1.7 percentage point reduction in the probability a student receives a zero in the course and a 0.021 standard deviation increase in class grade for classes taught in the morning.[6] When we consider the full sample of classes, these treatment effects shrink closer to zero, suggesting that later start times had no effect on student performance in afternoon

[5]More precisely, both lack of total sleep or an inconsistent sleep schedule cause daytime fatigue since they interfere with a student's circadian rhythm, an internal body clock that cycles the body between alertness and sleepiness through the day. 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. See the National Sleep Foundation at *https://sleepfoundation.org/* for related information.

[6]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.

classes. These “sleep duration” effects are relatively small when juxtaposed to previous studies: Carrell et al. (2011) 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) estimates a two percentage point gain in math test scores in response to relatively small delays in school start time, and Heissel and Norris (2017) estimate effects between 0.06 and 0.08 standard deviations in response to delaying start times by one hour.[7] Overall, while our findings suggest some evidence of sleep impacting student success, we find relatively small returns to delayed school start times with little importance for consistency in start times across the week.

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 Data

2.1 Institutional 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 two-year Associates’ and a four-year Bachelor’s, across three fields: Technology, Business, and Design. Although two-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.

All incoming students have a single declared major, and each major is embedded within a single de-

[7] 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 (forthcoming), who find that boys experience larger gains in the classroom in response to delayed start times.

partment. Regardless of their major, all students take certain common core subjects, such as Math and English. First-year students also take some major-specific courses. For instance, Marketing majors have to take a Web Design course in their first semester. 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.[8]

Most courses are broken down into multiple sections (i.e., meeting times), with each section taught by a single 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 randomized into their meeting times conditional on their declared major and the first letter of their first name. In Section 3 we present several tests for this randomization.

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 class 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 lecture

[8]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.

attendance and performance on midterms and a final examination. Typically, final examinations 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.[9] 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 end up conceding a final grade of zero for the course.

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 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.

2.2 Summary Statistics

Table 2 presents summary statistics for our full sample. Our data contain over 10,000 admitted freshmen from 2011 to 2015. 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

[9] 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.

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 9:16am. Panel C presents student-classroom level summary statistics for the outcomes of interest. 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.[10]

3 Econometric Specification and Identification

Our primary analysis estimates the following specification:

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

where y_{icst} is an 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 Tuesdays a student was randomly assigned a first classroom at 8:15am, then this variable would equal 8.25 for all the student's classrooms that were taught on Tuesdays.[11] 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. In our primary specification, we calculate the standard

[10]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.

[11]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). Given that students with different later start times are perhaps more likely to wake up at similar times (e.g., starts times of 12:30pm vs. 1:20pm), discussed in further detail later, we also consider a "top-coding" for $StartTime_{icst}$ as a robustness check.

deviation in start times for student i in term t across all class days.[12] 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. Figure 2 presents the distribution of these covariates, in which we observe large variation in both variables.

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

Consequently, our analyses hinge on comparing the academic outcomes of students with varying (spread of) start times within the same classroom by subjecting them to the same classroom-level shocks. These shocks may include the instructor's characteristics (e.g., ability, effectiveness, grading practices), the time at which the classroom was taught, and its size. 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

[12]That is, $sd[StartTime]_{it} = \sqrt{\frac{\sum_{d=1}^{6}(DailyStartTime_{dit} - \overline{DailyStartTime_{it}})^2}{5}}$ where $DailyStartTime_{dit}$ denotes the start time of student i's classes in term t on school day d (Monday/Tuesday/Wednesday/Thursday/Friday/Saturday).

their school schedule during their first two semesters of enrollment.[13]

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. Assuming student sleep schedules roughly correspond to their class start times, 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.

3.1 Balance Test

As is customary with any study claiming 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 $sd[StartTime]_{it}$. Second, we examine the sensitivity of our estimated coefficients from equation (1) to the inclusion of various fixed effects. If unobserved student or classroom characteristics are associated with (the spread of) start times, and these characteristics have a meaningful relationship with student performance,

[13] 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.

then we would expect our estimates for β and γ to shift in response to the inclusion of fixed effects.

Table 3 presents the results from the regressions of our four student characteristics on the two covariates of interest. The only statistically significant coefficients we find are for student gender. However, these coefficients of -0.003 and -0.009 are economically tiny. They say that a one standard deviation increase in the spread of daily start times (a one hour increase in average daily start time) is associated with a 0.9 (0.3) percentage point decrease in the probability the student is male. The remaining coefficients are, in addition to being statistically insignificant, economically insignificant. 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 coefficients on gender likely achieved significance by natural variation, as opposed to actual sorting of students into start times by gender. 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 remain relatively unchanged in response to student fixed effects.

4 Results

4.1 Main Results

Table 4 presents our main results. Each of the four panels considers a different student achievement outcome variable denoted in the top left corner. In column (1), we only include our two regressors of interest, while column (2) includes student and classroom fixed effects. Column (3) two-way clusters the standard errors from (2) to account for plausible autocorrelation across observations by student and by classroom (Cameron et al., 2011). Columns (4) through (6) parallel the first three columns while instead considering a "top-coding" for $StartTime_{icst}$. Specifically, all observations with afternoon start times (12:30pm, 1:20pm, 2:15pm, 3:05pm) were coded with a start time of 12:30pm and $sd[StartTime]_{it}$ was recalculated accordingly. This top-coding accounts for the likely fact that the variation of afternoon school start times is unlikely to generate variation of sleep across students.

Starting with column (1) in the first panel, we find a very small, yet precisely estimated, change in the probability the student received a zero in the course in response to a higher spread of start times and to delayed start times. Recall that this outcome variable can roughly be interpreted as "dropping" the course. Our point estimates remain almost completely unchanged when we include student and classroom fixed effects, lending credence to the notion of student randomization into course schedules. In column (3), after accounting for potential serial correlation by student and by classroom, estimates remain precisely estimated. For instance, with 95% confidence we can rule out a 0.4 percentage point increase in the probability the student received a zero in the course in response to a one standard deviation increase in the student's spread of daily school start time. These results remain largely unchanged when we top-code the start time.

The remaining three panels of Table 4 consider different outcome variables. Overall, we find very little effects of increased spread of daily start times and delayed start times on student achievement. In our full specifications, a one standard deviation increase in the spread of start times decreases a student's grades by 0.1% of a standard deviation, while a one-hour increase in start time is associated with a 0.7% of a standard deviation increase in class grade. Even after clustering our standard errors two-way, we can rule out detrimental effects as large as 2.2% 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. Similarly, we can rule out performance boosts as large as 1.7% of a standard deviation in class grade in response to a one hour delay in school start time. These results remain roughly stable regardless of whether we consider raw grades, condition on receiving a non-zero grade in the course, include student and classroom fixed effects, or top-code daily start times.

4.2 Subsamples and Alternate Measure of Spread

In this subsection, we consider several subsamples for which start times would plausibly have an especially impactful effect, as well as test the sensitivity of the results to measuring the spread of daily start times using the inter-quartile range (IQR). These results are presented in Table 5.

The most obvious subsample of interest is the group of classes that were taught in the morning. These are classes for which the consequences of irregular sleep (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 as measured by the IQR has no impact on student performance in morning classes. Second, the estimated effects of delayed school start time relative to the full sample results shift in the anticipated direction. We estimate a 1.7 percentage point reduction in the probability the student received a zero in the class, and a 2.1% of a standard deviation increase in grade, in response to a one hour delay in school start time. For the latter result, we can rule out effects larger than 7% of a standard deviation at the 5% significance level. While these effects suggest a positive response for students to delayed start times, it's important to note that these estimates are significantly smaller than those found in prior studies. Most notably, Carrell et al. (2011) also utilize random assignment at the college level to find decreases in student performance between 11.6% and 13.9% of a standard deviation in response to a 7am first period class. Not only are these estimates significantly larger than those in our study, but they also apply to their full sample of both morning and afternoon classes.

Next, we consider subsamples by gender. One study by Lusher and Yasenov (forthcoming) 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 (-0.016) is notably larger than that from the full sample (-0.001). Still, this estimate is rather small in economic terms: a one standard deviation increase in the spread of start times causes a 1.6% 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. [14]

5 Conclusions

Recent correlational studies and media reports have suggested that sleep regularity - the variation in the amount of sleep one gets across days - is a stronger determinant of student success than sleep duration - the total amount of sleep one receives. In this paper we estimate the causal effect of sleep regularity versus sleep duration on student performance by using over 165,000 student-classroom observations from a large university in Vietnam. Crucially, in this setting, all incoming 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. Across a variety of specifications and classifications, we do not find any evidence that higher spreads of start times adversely affected student performance. Moreover, we find only small boosts in student performance in response to delayed school start times.

These findings have several 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. Additionally, it may be the case that in our setting, students

[14] Finally, 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.

who were randomly assigned volatile school schedules were able to adjust their sleep schedules such that their sleep regularity was similar to students who had constant daily start times. In other words, our analysis assumes that student sleep schedules roughly correspond to their daily start times, which we believe to be likely since classes started as early as 6:30am. Finally, similar to previous studies, our setting grants us the opportunity to estimate the causal effect of later school start times overall on student achievement. Though we find improvements in student performance in response to later school start times, our estimates suggest that the returns to delayed school start times are significantly smaller than those reported previously.

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.

CAMERON, A. C., J. B. GELBACH, AND D. L. MILLER (2011): "Robust inference with multiway clustering," *Journal of Business & Economic Statistics*, 29, 238–249.

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.

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

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.

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

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.

Table 2: Summary Statistics

	Mean	SD	Obsevations
Panel A. Student level statistics			10,130
Male	0.570	0.495	
Age	18.323	0.804	
Native Ethnicity	0.926	0.263	
Pre-collegiate Grade	14.755	4.487	
Panel B. Course-section-term (classroom) level statistics			3,386
Number of Students	49.344	28.145	
STEM Course	.494	.500	
Grade (conditional on non-zero)	6.269	1.227	
Start Time	09:16 am	1hr 14min	
Panel C. Student-classroom level outcomes			167,078
Received Zero	0.094	0.291	
Grade (conditional on non-zero)	5.964	1.912	

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

Table 3: Balance Test

	Dependent Variable - Student Characteristics			
	Indicator for Male	Age	Native Ethnicity	Pre-Collegiate Score
Start Time	-0.003**	-0.002	-0.000	0.004
	(0.002)	(0.003)	(0.002)	(0.014)
	[0.296]	[1.000]	[1.000]	[1.000]
σ(Start Time)	-0.009*	0.008	0.005	0.023
	(0.005)	(0.012)	(0.004)	(0.050)
	[0.609]	[1.000]	[1.000]	[1.000]
Observations	167078	167078	167078	167078

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. 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: Main Results

	Full Model			Top-coded Start Time		
	(1)	(2)	(3)	(4)	(5)	(6)
Outcome: Received Zero						
Start Time	-0.001***	0.000	0.000	-0.002***	0.000	0.000
	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
σ(Start Time)	-0.004***	-0.004**	-0.004	-0.002***	-0.006***	-0.006
	(0.001)	(0.002)	(0.004)	(0.001)	(0.002)	(0.005)
Observations	167078	167078	167078	167078	167078	167078
Outcome: Raw Grade [1,10]						
Start Time	-0.018***	0.009*	0.009	-0.022***	0.011**	0.011
	(0.002)	(0.005)	(0.007)	(0.002)	(0.006)	(0.008)
σ(Start Time)	0.007	0.005	0.005	-0.003	0.002	0.002
	(0.005)	(0.012)	(0.018)	(0.005)	(0.014)	(0.023)
Observations	151459	151367	151367	151459	151367	151367
Outcome: Standardized Grade						
Start Time	0.002*	0.003	0.003	0.002**	0.004	0.004
	(0.001)	(0.003)	(0.004)	(0.001)	(0.003)	(0.004)
σ(Start Time)	0.008***	0.002	0.002	0.007***	0.002	0.002
	(0.002)	(0.006)	(0.011)	(0.002)	(0.007)	(0.013)
Observations	167070	167070	167070	167070	167070	167070
Outcome: Standardized Grade $\| > 0$						
Start Time	0.001	0.007*	0.007	0.001	0.008**	0.008
	(0.001)	(0.003)	(0.005)	(0.001)	(0.004)	(0.005)
σ(Start Time)	0.004	-0.001	-0.001	0.003	-0.004	-0.004
	(0.002)	(0.008)	(0.011)	(0.003)	(0.009)	(0.014)
Observations	151452	151360	151360	151452	151360	151360
Class FE		X	X		X	X
Student FE		X	X		X	X
Twoway Cluster SE			X			X

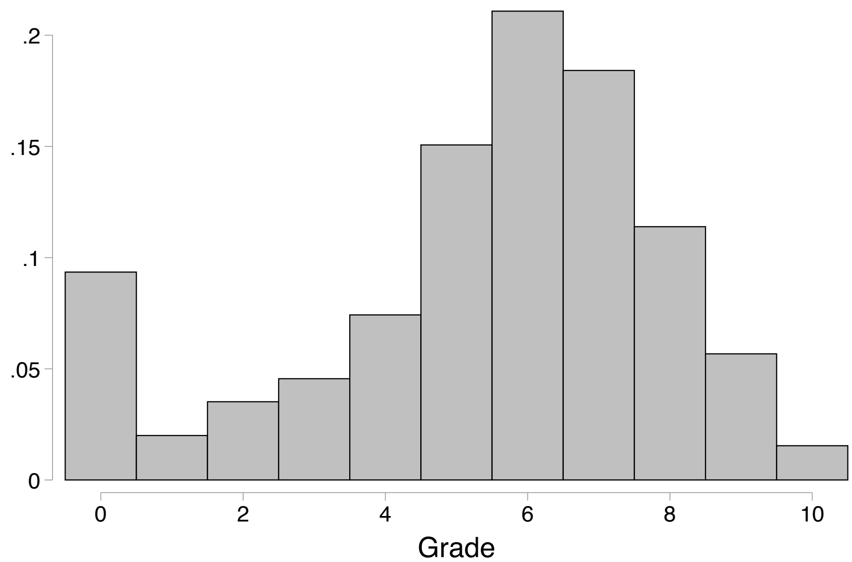
Notes: The unit of observation is student-classroom. "Start Time" is measured in hours. "σ(Start Time)" is the standard deviation of a student's school start times across days of the week. Robust standard errors presented in parentheses. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Table 5: Main Results - Subsamples and Other Measures of Spread

	Full Model					Top-coded Start Time				
	Full Sample	Morning Classes	Male Students	Low Achievers	STEM Courses	Full Sample	Morning Classes	Male Students	Low Achievers	STEM Courses
Outcome: Received Zero										
Start Time	0.000	-0.017***	0.000	-0.001	-0.000	0.000	-0.017***	0.000	-0.000	-0.000
	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.001)	(0.001)	(0.002)
σ(Start Time)	-0.004	-0.000	-0.005	-0.017**	-0.006	-0.006	-0.003	-0.007	-0.015*	-0.010
	(0.004)	(0.005)	(0.005)	(0.008)	(0.005)	(0.005)	(0.006)	(0.006)	(0.009)	(0.006)
Observations	167078	84354	99296	60082	74112	167078	84354	99296	60082	74112
									est9	est10
Outcome: Received Zero										
Start Time	0.000	-0.017***	0.000	-0.000	-0.000	0.000	-0.017***	0.000	0.000	-0.000
	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.001)	(0.001)	(0.002)
IQR(Start Time)	-0.001	0.000	-0.002	-0.005	-0.002	-0.000	0.001	-0.001	-0.002	-0.002
	(0.002)	(0.002)	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)	(0.003)	(0.005)	(0.003)
Observations	167078	84354	99296	60082	74112	167078	84354	99296	60082	74112
									est9	est10
Outcome: Standardized Grade $\mid > 0$										
Start Time	0.007	0.021	0.005	0.005	0.010	0.008	0.022	0.007	0.006	0.011
	(0.005)	(0.025)	(0.006)	(0.008)	(0.008)	(0.005)	(0.025)	(0.006)	(0.008)	(0.008)
σ(Start Time)	-0.001	0.009	-0.016	0.009	-0.008	-0.004	0.015	-0.022	0.006	-0.006
	(0.011)	(0.015)	(0.014)	(0.021)	(0.016)	(0.014)	(0.019)	(0.017)	(0.023)	(0.019)
Observations	151360	76412	87710	54893	67212	151360	76412	87710	54893	67212
									est9	est10
Outcome: Standardized Grade $\mid > 0$										
Start Time	0.007	0.020	0.006	0.004	0.010	0.008	0.020	0.007	0.006	0.011
	(0.005)	(0.025)	(0.006)	(0.008)	(0.008)	(0.005)	(0.025)	(0.006)	(0.008)	(0.008)
IQR(Start Time)	-0.003	0.008	-0.009	-0.012	0.003	-0.004	0.010	-0.014*	-0.020	0.005
	(0.006)	(0.008)	(0.007)	(0.012)	(0.008)	(0.006)	(0.008)	(0.007)	(0.014)	(0.008)
Observations	151360	76412	87710	54893	67212	151360	76412	87710	54893	67212
Class FE	X	X	X	X	X	X	X	X	X	X
Student FE	X	X	X	X	X	X	X	X	X	X
Twoway Cluster SE	X	X	X	X	X	X	X	X	X	X

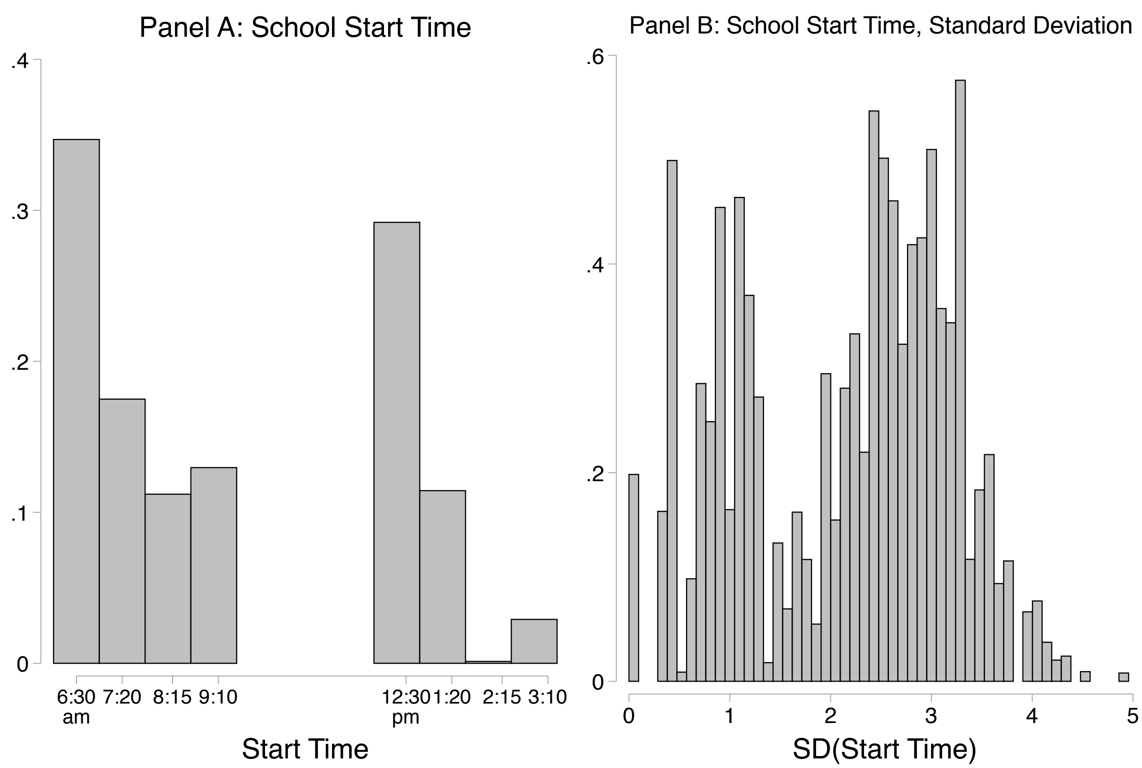
Notes: The unit of observation is student-classroom. "Start Time" is measured in hours. "σ(Start Time)" is the standard deviation of a student's school start times across days of the week. "IQR(Start Time)" is the interquartile range of a student's school start times across days of the week. Robust standard errors presented in parentheses. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Figure 1: Histograms of Raw Grades



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

Figure 2: Histograms of Sleep Consistency 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.