Economics of Education Review 67 (2018) 158–170

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Academic achievement across the day: Evidence from randomized class schedules

Kevin M. Williams[*,a], Teny Maghakian Shapiro[b]

[a] Department of Economics, Occidental College, 1600 Campus Road, Los Angeles, CA, United States
[b] Slack, Research & Analytics, San Francisco, CA, United States

ARTICLE INFO

Keywords:
Human capital
Educational economics
School schedules
Fatigue
Learning
Start time
Higher education
Adolescents

JEL classification:
I20
J24

ABSTRACT

This study expands our understanding of how school day schedules affect achievement. We focus on three aspects related to scheduling: student fatigue, time of instruction, and instructor schedules. Data cover five academic years at the United States Air Force Academy, where schedules are randomized, grading is standardized, and there is substantial variance in schedule structure. Analyzing over 180,000 student-course outcomes, we find causal evidence of cognitive fatigue brought on by scheduling multiple courses in a row. The expected performance of two students in the same class may differ by as much as 0.15 standard deviations simply owing to their prior schedules. All else equal, students perform better in the afternoon than in the early morning. We also note that instruction improves with repetition. Heterogeneous effects by ability suggest that overall gains are possible. Prioritizing certain schedules would be equivalent to improving teacher quality by one-third of a standard deviation. A reorganization of students' daily school schedules is a promising and potentially low-cost educational intervention.

1. Introduction

Teachers, administrators, and policymakers go to great lengths to improve student achievement: searching for the best educators, employing the newest pedagogical practices, and carefully crafting assignments, all in the hope that students will better understand the material they are presented. Recent research has shown, however, that much of an individual's ability to learn is determined by their mental state (Persson, Welsh, Jonides, & Reuter-Lorenz, 2007) and their daily biological rhythms (Schmidt, Collette, Cajochen, & Peigneux, 2007). Students perform worse when they are mentally taxed or when classes are scheduled at times asynchronous with their internal clocks.

The best evidence on the role of time of day comes from studies on school start times and morning versus afternoon classes (Carrell, Maghakian, & West, 2011; Diette & Raghav, 2017; Edwards, 2012; Heissel & Norris, 2018; Pope, 2015). However, none of these studies take fatigue into account and thus are unable to abstract the effect of time of day from the effect of fatigue–an important distinction to make if findings are being used to reorganize school schedules.

The goal of this study is to expand on the link between students' academic achievement and their daily schedules to causally determine precisely how the organization of courses throughout the school day influences performance. We explore the independent roles of three aspects of the school-day schedule: student fatigue due to classes earlier in the day; the time of day a class is held; and the instructors' schedules. By better understanding how each of these scheduling components affects student achievement, we can offer schools and individuals recommendations on how to improve academic outcomes.

This is the first paper to separately identify the effects of student fatigue, the time a class is held, and the instructors' schedule. This would be difficult, if not impossible, to do in most school settings. Selection by students into specific courses or instructors and a lack of common grading standards across sections are common in secondary and higher education. Further, in schools where students and/or teachers are assigned a class during each period, the effect of time of day cannot be separately identified from the effect of fatigue.

We are able to overcome typical identification hurdles by utilizing data from the United States Air Force Academy (USAFA). The school day at USAFA is split into seven class periods, four before lunch and three after, a daily structure very similar to that of the typical U.S. high school. While the daily schedule at USAFA is standard, there are a number of distinct institutional characteristics that allow for causal assessment of the role of schedules on academic achievement. Student's schedule assignment is random and during the first two years of instruction, students primarily take required core courses. Grading and instruction are standardized across all sections of a course and exams

* Corresponding author.
E-mail addresses: KevinW@Oxy.edu (K.M. Williams), tshapiro@slack-corp.com (T.M. Shapiro).

https://doi.org/10.1016/j.econedurev.2018.10.007
Received 23 August 2017; Received in revised form 6 October 2018; Accepted 14 October 2018
Available online 19 October 2018
0272-7757/ © 2018 Elsevier Ltd. All rights reserved.

are taken during a common testing session. Instructors regularly teach multiple sections of the same course. Relatively small class sizes mean that certain core courses (e.g. Calculus, Introductory Chemistry), routinely have over twenty different sections taught during a single semester, meeting at all times of the school day. Also, students alternate daily between two class schedules within the same semester. Students have a similar academic course load, but the alternating schedule creates variation in how much time students spend in class on a given schedule-day. This allows us to assess how a student performs with one schedule relative to their own performance with a different schedule.

Our data cover five academic years at USAFA, which includes nearly 7000 students and over 200,000 student-course observations in core academic fields. We recognize that USAFA students are not the average student; they were high-achievers in high school and chose to attend a military service academy. Although we do not know for certain if school schedules affect high-achievers or military-types differently than the average student, we have little reason to believe that the students in our sample would be *more* adversely affected by components of their daily schedule than the average teen or young adult.

In addition to our full sample analysis, we focus on the subset of fall-semester freshmen. These students are still in their teens, are new to the academy, and are enrolled almost entirely in required core courses. Much of the focus of changes in school start times and schedules is on teens because of their distinct time preferences and its misalignment with traditional school schedules (Crowley, Acebo, & Carskadon, 2007). To the extent that our findings can be generalized beyond higher education, first-semester students at USAFA provide the closest comparison with secondary student populations.

We find that, all else equal, the afternoon is the best time of day for student learning. Gains from having a class in the afternoon relative to the morning are partially offset by fatigue. Specifically, if a student took her first class of the day at 2:00 pm rather than 7:30 am, she would perform about 0.16 standard deviations better. However, when fatigue is factored in, a student in a 2:00 pm class that follows a full schedule of classes in the morning is predicted to perform only 0.08 standard deviations better than in the 7:30 am class. Even two students sitting in the same section of a class may have different expected grades as large as 0.15 standard deviations simply due to variation in fatigue from their prior schedules. Our results are consistent across full-sample and first-semester freshman analyses.

There is also evidence that instructors improve as the school day progresses. Perhaps because teacher effectiveness improves with repetition, students in the second or third section of an instructor's course perform, on average 0.04 standard deviations better than their peers in the first section.

Subgroup analysis reveals that the predicted negative effects of fatigue are more extreme for students in the bottom tercile of predicted aptitude. These students experience the highest penalty from schedules featuring multiple consecutive classes without a break. Students in the top tercile are less affected by their schedules. This suggests that schools can raise mean performance by assigning struggling student's schedules that space breaks optimally and best match their biological rhythms.

In counterfactual simulations, we show that if the worst students were given the most favorable schedules, bottom-tercile students would increase mean achievement by 0.034 standard deviations, equivalent to increasing their teacher quality by 0.33 standard deviations in all courses.[1] We conclude with a discussion of policies, obstacles, costs, and benefits facing the implementation of a rescheduled school day. In comparison with many of the inputs commonly studied in the education production function, such as teacher quality and class size, rescheduling classes to better align with students' optimal learning times is a potentially cost-effective intervention that may be easier for schools to implement than a later start time.

2. Background

To fully understand how the organization of the school day schedule can influence academic achievement, it is important to have a basic understanding of the theory of cognitive fatigue and the biology of sleep, wakefulness, and daily fluctuations in cognitive function. It is no surprise that repetitive or low-stimulation environments can bring on a feeling of boredom or mental tiredness, collectively called cognitive fatigue (Persson et al., 2007). Cognitive fatigue has been identified in laboratory settings as well as education. Fatigue and boredom are a persistent issue in education, though typically very difficult to measure (Vogel-Walcutt, 2012). Self-reported student measures of fatigue do not necessarily correspond with decreased aptitude in repetitive settings. The time it takes for cognitive fatigue to impact a subject's performance is context-specific, but in situations similar to a classroom setting it has been shown to be anywhere from 20 min (Jackson, Sabina & Eugene, 2014) to over 2 h (Ackerman & Kanfer, 2009).

The biological rhythm that governs our sleep-wake cycle is called the circadian rhythm, a hard-wired "clock" in the brain. During adolescence, young people experience major changes in their circadian rhythms. They develop more adult-like sleep patterns, experience more daytime sleepiness, and begin to prefer later bed and wake-up times (Carskadon, Vieira, & Acebo, 1993; Crowley et al., 2007; Wolfson & Carskadon, 1998). Circadian timing also affects the times of day when a person is more alert, independent of sleep (Blake, 1967). For adolescents, alertness rises in the late morning, drops off in mid-afternoon, and peaks again in the early evening (Cardinali, 2008). An individual's ability to learn fluctuates throughout the day based on their biological rhythm (Goldstein, Hahn, Hasher, Wiprzycka, & Zelazo, 2007; Schmidt et al., 2007). Standard academic schedules are "out of sync" with teens' circadian rhythms; they require students to wake up earlier than is optimal and to attend classes at times that are asynchronous with their optimal cognitive functioning.[2]

Our understanding of cognitive function, sleep and wakefulness, suggests two factors in a student's daily schedule can affect his or her grades. The first is the cognitive load a student has experienced before the start of a class. We refer to this as the *student fatigue effect* or cognitive fatigue. The second is the timing of a class: students may perform less well if classes are scheduled when they're naturally less alert. We refer to this as the *time-of-day effect*. We expect student fatigue to unambiguously hinder academic performance. The time-of-day effect may vary throughout the day.

Because academic achievement reflects the interaction between learning and teaching, we also estimate the *instructor schedule effect*. The expected effect of instructor schedule is ambiguous. Unlike students, instructors are frequently engaged in the same class multiple times per day. Tiredness, wakefulness and mental fatigue could mean instructors are less effective as the day goes on, but they may also become more effective as they repeat material.

A few of strands of the literature have investigated the role of time and scheduling on academic and workplace outcomes. Because of difficulties in identification, these studies do not separate out student

[1] This is scaled relative to earlier work (Carrell & West, 2010; Chetty, Friedman, & Rockoff, 2014; Kane & Staiger, 2008) that finds a 1.0 standard-deviation increase in teacher quality benefits students by between 0.1 and 0.2 standard deviations. Throughout the paper, we assume a 0.1 standard deviation improvement in student achievement from a 1 standard deviation improvement in instructor quality when relating the effects of schedules to an increase in instructor quality.

[2] This is not to understate the importance of sleep, which itself is important to cognitive functioning. Several studies find an inverse relationship between sleep and academic performance at both the secondary and the post-secondary level (Curcio, Ferrara, & Gennaro, 2006; Trocket, Barnes, & Egget, 2000; Wolfson & Carskadon, 1998).

fatigue and time-of-day effects. The impact of school start times on student achievement has been studied using natural variation across schools and cohorts for identification. The overall findings of the literature support that delaying school start times benefits students (Diette & Raghav, 2017; Edwards, 2012; Groen & Pabilonia, 2017; Heissel & Norris, 2018), though not every study has found significant impacts (Hinrichs, 2011). Delayed start times may also have health-related benefits (Wahlstrom, 2002).

Carrell et al. (2011), who use the same data that we use in this study, find that students' grades throughout the day benefit from later start times. They make across-cohort comparisons to show that cohorts facing the earliest start times reduced grades throughout the whole schoolday. Our work is in part motivated by this finding. Shifting an entire school's start and end time may be impractical.[3] Our aim is to use variation in students schedules within a semester to understand how features of a schedule beyond its start time impact achievement and explore how schedules may be restructured to improve achievement.

Other studies have looked at differential achievement across morning and afternoon classes, separate from start times. Most find that classes meeting in the afternoon are associated with higher achievement than morning classes (Cortes, Bricker, & Rohlfs, 2012; Cotti, Gordanier, & Ozturk, 2018; Dills & Hernandez-Julian, 2008; Lusher & Yasenov, 2018). Others have found the opposite (Pope, 2015), showing that learning decreases throughout the school day. We believe our work generalizes these results. If the student fatigue is effect negative, while the time-of-day effect is positive in the afternoon, the intrinsic link between the two (morning classes mechanically must occur before afternoon ones) creates an ambiguous result for any model that estimates the combination of these two effects.

3. Data

Data for this study come from the United States Air Force Academy (USAFA). USAFA is a fully accredited post-secondary institution with annual enrollment of approximately 4500 students, offering 32 majors in the humanities, social sciences, basic sciences, and engineering. Despite its military setting, USAFA is comparable on many dimensions to other selective colleges and universities in the United States. Like other selective post-secondary schools, USAFA faculty hold graduate degrees from high quality programs in their fields. Approximately 40% of classroom instructors have terminal degrees, similar to large universities where graduate students teach introductory courses. Class size at USAFA is rarely larger than 25 students and students are encouraged to engage with faculty members in and outside the classroom. Students at USAFA are high achievers with average math and verbal SAT scores at the 88th and 85th percentiles of the nationwide SAT distribution, respectively. Only 14% of applicants were admitted to USAFA in 2007. Students are drawn from each congressional district in the United States through a highly competitive admission process that ensures geographic diversity.[4]

A number of USAFA's institutional characteristics aid in the causal identification of our research question. First, the school day at USAFA is very structured. There are four 53-min class periods each morning and three each afternoon after an 85-min lunch break. This structure is similar to many high school settings. Class attendance is mandatory and all students are required to attend breakfast 25 min before the first period of the day. Hence, wake up time is constant among students.

Students are randomly assigned to their instructors and schedules, conditional on course placement (e.g., remedial math, Calculus I, Calculus II, etc.), athlete status, and gender. The USAFA registrar assigns students to required course sections. Freshmen and sophomores take exclusively required core courses and thus have little input into their own schedules. Foreign language courses are an exception. Foreign language is required, but students may choose which foreign language to study. They are randomly assigned to a section in their foreign language of choice.

As they advance, students begin selecting into majors and taking major-specific elective courses, but still do not have control over the time of day of their classes. Our identification comes from comparing students taking the same course in the same term, but with different schedules. Although self-selection of students into majors and electives is a potential contaminant of our estimates, we do not believe this aspect of upperclassmen's schedules will significantly bias our results.

USAFA's grading structure for core courses allows for a consistent measure of student achievement; faculty members teaching the same course in each semester use an identical syllabus, give the same exams during a common testing period, and assign course grades jointly with other instructors, ensuring standardized grades within a course-semester.

Common testing periods provide two key benefits for identification. First, they ensure that there is no information leakage, where students who take a test in a morning class pass information on to friends in the afternoon sections. Second, it means that our estimates of time-of-day effects are about the classroom learning that occurs during those time periods, not about variation test-taking aptitude throughout the day.

USAFA runs on an M/T schedule. On M days, students have one set of classes and on T days they have a different set of classes. The M/T schedules alternate days of the week.[5] These institutional characteristics provide us with random variation in class schedules for all students, which, along with extensive background data on students, allow us to examine how course scheduling affects student achievement.

Our dataset covers all students, courses, and grades from academic years 2004 through 2008 at USAFA. We observe 6981 students, 4788 of whom are observed as freshmen during their first term at the academy. Student characteristics are summarized in Table 1. For each student we have pre-treatment demographic data and measures of their academic, fitness, and leadership aptitude. Academic aptitude is measured by SAT verbal and math scores and an academic composite computed by the USAFA admissions office, which is a weighted average of an individual's high school GPA, class rank, and the quality of the high school they attended. The measure of pre-treatment fitness aptitude is scored on a test required by all applicants before entrance. The measure of pre-treatment leadership aptitude is a composite also computed by the USAFA admissions office as a weighted average of high school and community activities. Other individual-level controls include indicators for whether a student is black, Hispanic, Asian, female, or a recruited athlete, whether they attended a military preparatory school, and the number of class credits students have on a schedule-day.

Table 2 provides a summary of our data at the student-course level, the level of observation used in analyses. We observe 232,862 total courses taken (Column 1), of which 187,525 are considered core academic courses (Column 2). Excluded courses are physical education, military science and independent study courses. We do not include grades from these courses as observations in our primary analysis, but they are considered a part of a student's schedules when constructing

[3] See Jacob and Rockoff (2011) for a full discussion. They find that shifting the school-day schedule to a later start time could cost anywhere from $150 to $1900 per student.

[4] The make-up of USAFA students is comparable to that of other selective institutions. The SAT scores of USAFA students are comparable to those of students in flagship schools such as UCLA and UNC Chapel Hill, while the heavily male USAFA student body is more similar to that of other technical institutions such as Georgia Tech and Rensselaer Polytechnic Institute.

[5] Thus each student has two different class schedules within one semester. Language courses are an exception and meet every day during a period. Students are coded as in class for both M and T days of their language courses. One week, M-days will occur on Mon–Wed–Fri, the next they will occur on Tues–Thur.

Table 1
Summary statistics–students.

	All students	Freshmen
Black	0.0392	0.0374
	(0.194)	(0.190)
Hispanic	0.0776	0.0819
	(0.268)	(0.274)
Asian	0.0852	0.0942
	(0.279)	(0.292)
Female	0.185	0.191
	(0.389)	(0.393)
Prep school	0.179	0.170
	(0.384)	(0.376)
SAT verbal/100	6.526	6.451
	(5.942)	(0.655)
SAT math/100	6.797	6.690
	(6.413)	(0.638)
Academic classes	5.093	5.068
	(0.637)	(0.769)
Credits per day	9.515	8.854
	(2.708)	(2.203)
Observations	6,981	4,788

Note: Summary statistics at the student level. Prep School refers to a fifth year of high school at a military preparatory school. Academic classes summarizes the average number of non-physical education courses taken by a student in a single semester. Credits per Day shows the average course load for a single M/T schedule day. Standard deviations in parentheses.

Table 2
Summary statisticscourses.

	All courses	Core courses	Core first year
STEM course	0.391	0.485	0.544
	(0.488)	(0.500)	(0.498)
Consecutive classes	0.686	0.749	0.631
	(0.929)	(0.956)	(0.902)
Cumulative classes	1.630	1.716	1.663
	(1.410)	(1.422)	(1.338)
Grade	0.0249	0.0406	0.0515
	(0.996)	(0.997)	(0.996)
Ace	0.252	0.257	0.228
	(0.434)	(0.437)	(0.419)
Fail	0.0378	0.0441	0.0745
	(0.191)	(0.205)	(0.263)
Class size	19.64	19.74	22.26
	(6.208)	(6.216)	(6.344)
Average sections	15.59	17.77	28.61
	(12.63)	(13.10)	(13.65)
7 am hour	0.192	0.158	0.150
	(0.393)	(0.364)	(0.357)
8 am hour	0.131	0.158	0.128
	(0.337)	(0.364)	(0.334)
9 am hour	0.244	0.223	0.187
	(0.430)	(0.416)	(0.390)
10 am hour	0.121	0.145	0.135
	(0.326)	(0.352)	(0.341)
1pm Hhour	0.190	0.167	0.190
	(0.392)	(0.373)	(0.392)
2 pm hour	0.0808	0.0989	0.139
	(0.272)	(0.298)	(0.346)
3 pm hour	0.0411	0.0505	0.0716
	(0.198)	(0.219)	(0.258)
Observations	232,862	187,525	24,264

Note: Summary statistics for all courses taken from academic year 2004–2005 through 2008–2009. Unit of observation is a student-class. Ace refers to receiving a letter grade of "A" in a course, Fail corresponds to a letter grade of D+ or lower. Class size refers to the number of students in a single section of a course. Time-of-day references the hour in which instruction for a class begins (See Table 3). Standard deviations in parentheses.

measures of fatigue.

Column 3 summarizes core courses taken by first-year students in their first term. Students at USAFA are required to take a set of approximately 30 core courses, mainly in their first two years, in mathematics, basic sciences, social sciences, humanities, and engineering. For first-year students, we focus on the mandatory introductory courses in mathematics, chemistry, engineering, computer sciences, English, and history.

For all analysis, we consider both the full sample of student-courses (Column 2) as well as the subsample of fall semester freshmen (Column 3). This subsample is free of the course-selection issue faced by upperclassmen who have chosen majors. Further, these students are mostly 18 and 19-year-olds only months removed from high school. To the extent that our results will generalize to high school settings, this subsample will be the best comparison group.[6]

48% of all core courses taken are STEM (science, technology, engineering, and math) courses, and 54% of freshman courses are in STEM fields.[7] Students take approximately 10 credits per day; given the M/T schedule, this means the average course load is around 20 credits–roughly seven full-credit courses.

Table 2 summarizes our measures of student fatigue. *Consecutive Classes* captures how many prior consecutive classes a student has had without a break. The first class of a student's day, or a class following a break, is assigned a value of 0, while a second, third, or fourth class in a row is given a value of 1, 2, and 3, respectively. Lunch is considered a break. Roughly 40% of all classes have a positive value for *consecutive*; the average value is 0.75. *Cumulative Classes* ignores breaks. It counts (starting at 0) how many total prior classes a student has had up to that point in her daily schedule.

The M/T schedule-days at USAFA create an additional layer of variation in student fatigue. For example, Student A may have classes during 8 am, 10 am, and 2 pm hours on her M schedule-day, while Student B has classes during the 7am, 8 am, 1 pm, and 2 pm hours. For their 2 pm course, Student A has had two cumulative classes, but zero consecutive classes (since she had no 1pm course), while Student B has had three cumulative classes and one consecutive class. If academic achievement is affected by having had to focus and learn earlier in the day, the performance of Students A and B in their 2 pm courses will be affected by the time of day the class is held *and* by the number of classes each has had that day, both consecutive and cumulative. On their T schedule day, both students may have different (or null, if the hour is free) values of consecutive classes and cumulative classes for the 2 pm hour.

The class schedule changed twice during the time period studied. Table 3 shows the class schedules for our sample period. We typically discuss a course by the hour that instruction begins, for example the 2pm hour refers to a course that begins between 2:00 and 2:59 pm. This groups together classes by their period in USAFA's schedule. Alternate definitions will be considered in our results. Each beginning hour can alternatively be thought of as a course period (e.g. the 7 am hour always corresponds to the first period of the school day, while a course beginning in the 10 am hour is always the fourth period of the day).

We measure academic performance using the final percentage score each student earned in a course. To account for differences in course

[6] Concerns about external validity of USAFA students are valid. However, we note that USAFA has been used to study a number of education-related questions and, in each instance, findings using USAFA data (and the causality it affords) have subsequently been confirmed in other, more general settings. School start times is one example already discussed. The effects of teacher quality is another (Carrell & West, 2010; Chetty et al., 2014). Peer effect experiments have also shown similar results inside and outside of USAFA (Booij, Leuven, & Oosterbeek, 2017; Carrell, Sacerdote, & West, 2013).

[7] When included, Military Science is considered non-STEM in our analysis, but robustness checks have shown that it's designation does not meaningfully impact results.

Table 3
Daily class schedule at the U.S. Air Force Academy.

Name	Period	AY2004 - AY2005	AY2006	AY2007 - AY2009
7 am hour	1	7:30	7:00	7:50
8 am hour	2	8:30	8:05	8:50
9 am hour	3	9:30	9:10	9:50
10 am hour	4	10:30	10:15	10:50
1 pm hour	5	13:00	13:00	13:30
2 pm hour	6	14:00	14:05	14:30
3 pm hour	7	15:00	15:10	15:30

Note: List of USAFA schedules over time period of analysis. A course period always corresponds to 53 min of instruction. In a given academic year, the schedule timing is the same for both M and T days.

difficulty or grading across years, we normalize all scores to a mean of zero and a variance of one within a course-semester. We refer to this measure as the student's normalized grade.[8]

Varsity athletes are excluded from the main sample to avoid the impacts that practice timing has on course schedules. We include athletes when normalizing course grades and computing classroom peer measures. Because athletes tend to have below-median grades, the average standardized grade for our sample is slightly positive.

The average class size is 19.7 students. The variable *Average Sections* measures how many sections of a course meet in a given term. Because of the large number of required courses and relatively small class settings, the average class in our data is one of 19 different sections being offered. For first-year courses, the average class size is 22.3 students per section with an average of 28 sections per course. Given that there are fourteen periods in USAFA's schedule (seven on each M/T day), there are often multiple sections of a course being taught simultaneously by different professors.

In the bottom rows of Table 2 we summarize the proportion of classes that meet during each time slot. Courses are relatively evenly distributed with 9 am being the most common hours for core courses to meet (22%). 3 pm courses, the latest of the day, are the exception. Only 5% of core courses begin during this hour.

Our analysis relies on conditional random assignment of students to their schedules. Before modeling achievement, we test this assumption by regressing student background characteristics on time-of-day variables. Similar to prior work on these data (Carrell & West, 2010), we find no reason to reject the assumption of random assignment of students to schedules.[9]

4. Methodology

Our primary analysis leverages the distinct characteristics of USAFA to simultaneously estimate effects of student fatigue, the time-of-day a class is held and instructor schedules. We estimate the following model:

$$Y_{icjspt} = \alpha + \beta Fatigue_{icjspt} + \mu InstructorSchedule_{cjspt} + \psi Time_p + \delta_1 X_{icspt} + \delta_2 Peers_{cjspt} + \phi_{cts} + \gamma_{jt} + \rho_i + \epsilon_{icjtspt} \tag{1}$$

Identification in this model comes from comparing students who take the same course in the same term on the same schedule day, but with different class timing and with variation in their daily schedules and the schedules of their instructors. Our primary outcome Y_{icjspt} is the normalized grade for student i in course c with instructor j on schedule-day s in hour (period) p in year t.

Fatigue is a vector of student's daily schedule characteristics, namely the number of prior consecutive and cumulative classes. *Consecutive Classes*, which ranges from 0 to 3, is captured with dummy variables, while *Cumulative Classes* (ranges 0 to 6) is measured continuously. We include a quadratic term for *Cumulative Classes* which is de-meaned before squaring.[10]

InstructorSchedule is a vector of instructor schedule characteristics, analogous to student fatigue and captured with dummy variables. Like students, professors schedules create variation both in the number of consecutive classes taught and the number of cumulative classes taught before a given class.

$Time_p$ is a vector of dummy variables, corresponding to each hour that classes begin instruction. The 7 am hour dummy is omitted.[11] The variables measure the time-of-day effect. They capture how students perform in the same class with the same grading standards when learning takes place at different times of the day. By grouping together all classes beginning in the same hour (see Table 3), each value also corresponds to a class period. This means that, in addition to capturing the true time-of-day effect, these variables will also absorb structural schedule effects, that may be separate from the specific time they occur. For example, the 1pm hour (always the fifth period of the day) is also the after-lunch class. With our data, we cannot separately identify the time-of-day from these structural characteristics.[12]

The vector X_{icjspt} includes the following student and section characteristics: SAT math and SAT verbal test scores, academic and leadership composites, fitness score, race, gender, whether s/he attended a military preparatory school, how many credit hours the student had on that schedule-day and section size. To control for classroom peer effects, we include $Peers_{icjtsp}$, the average pre-treatment characteristics of all students in the class except for individual i.[13]

ϕ_{cst} are course-by-year-by-schedule-day fixed effects, which control for unobserved mean differences in academic achievement or grading standards across courses, years, and schedule-days. Professor by year fixed effects, γ_{jt}, control for fixed differences in instructors within a given year. This ensures we are accounting for fixed instructor differences while leveraging the variation in their schedules.

We also show specifications that include individual student fixed effects, ρ_i. These estimates exploit within-student variation in schedules across the M/T schedule-days and make comparisons based on how students perform in a class relative to their own average performance in all other classes. These models do not include individual characteristics. Standard errors are clustered by student in all models.

4.1. Classmate comparison

Our primary analysis simultaneously identifies the effects of student fatigue, time-of-day and instructor schedules. Our second approach narrows in on student fatigue by utilizing only within-section comparisons. Rather than comparing two students taking the same course at different times of the day, we are comparing students in the same

[8] We also consider whether a student "Aces" (letter grade of "A") or "Fails" (D + or worse) a course to see the impacts on the extremes of the grade distribution. These estimates are available in the online appendix.

[9] Results are available in the online appendix.

[10] We define the quadratic term as $(Cumulative_{icjspt} - \bar{Cumulative}_{pt})^2$ which allows the estimated slope on *Cumulative* to be interpreted as the marginal effect of a one-unit increase at the mean.

[11] We also consider models with more flexible controls for time-of-day including a continuous polynomial and separate dummy variables for each separate start time.

[12] It should be noted that instructors may be affected by the time-of-day effects as well. Hence, our time-of-day estimates are a weighted combination of timing's impact on students and instructors. We believe the majority of the time-of-day estimates are driven by students' natural wakefulness cycles, since the school day is particularly out of sync with the circadian rhythm of adolescents, but we cannot rule out that instructor quality may vary with time-of-day as well.

[13] Formally, the *Peers* variables are defined as follows: $\frac{\sum_{k \in cjsptk \neq i} X_k}{n_{cjtsp} - 1}$, where X_k represents student background characteristics.

section of a class, but who had different schedules earlier in the day. We include section, rather than course, fixed effects. In essence, a student's schedule immediately before a given class can be thought of as a "treatment" on their ability to learn at that time. By comparing students in the same section, we are holding peers, instructor quality, instructor schedules and time-of-day constant.

We estimate this model both with our *Fatigue* vector of schedule traits and additionally by measuring student's *LeadUp* class.

$$Y_{icjtsp} = \alpha + \beta LeadUp_{icjtsp} * \psi Time_p + \delta_1 X_{ict} + \delta_2 Peers + \phi_{ctspj} + \gamma_{jt} + \rho_i + \epsilon_{icjtsp} \tag{2}$$

The primary difference compared with Eq. (1) is section fixed effects, ϕ_{ctspj}, which replace course-schedule day fixed effects. The $LeadUp_{icjtsp}$ variable is comprised of four mutually exclusive possibilities representing schedule the prior period: Free Period, P.E., STEM Class, Non-STEM Class. Interacting the *LeadUp* variables with time-of-day dummies, $\psi Time_p$ allows for the effect of a student's prior classes to vary over the day.[14]

5. Results

The results from Eq. (1) are shown in Table 4. Columns 1–5 show estimates for the full study body while columns 6 and 7 show estimates for the subsample of fall-semester freshmen. All specifications include individual and peer characteristic controls as well as instructor and course-by-term fixed effects. The outcome shown in all columns is the normalized grade earned in the class.

Column 1 focuses on measures of student fatigue, Column 2 examines time-of-day effects without fatigue, Column 3 looks at both sets of effects together, Column 4 adds instructor schedule variables, and Column 5 adds individual fixed effects and is our preferred specification as it allows us to make inferences on performance relative to a student's own average performance.

Consecutive classes have a consistently negative impact on performance.[15] The top two rows correspond to a student who has had either one or two or more classes immediately prior to the current class, respectively. A student sitting in their second consecutive class is expected to perform 0.031 standard deviations worse than if she took the same course after a break. We take this as solid evidence of cognitive fatigue. When student's schedules require them to sit in multiple classes in a row, they perform significantly worse in the latter classes, likely because of a decreased ability to absorb material.

The effect of cumulative classes, the total number of prior courses a student has taken on a given day, varies more across our models, but is significantly negative in our preferred specification. The effect is estimated to be positive in Column 1, which is due to the omission of time-of-day controls and so large numbers of prior cumulative classes are conflated with the timing of those courses. Once time-of-day is included, the estimate of cumulative classes becomes insignificant in Columns 3 and 4. With individual fixed effects included in Column 5, the effects of cumulative classes are negative and significant, with a weakly positive quadratic term. This suggests that students suffer both from the immediate effect of consecutive classes and the cumulative effect of heavy course loads in a single day.

We examine the impact of class timing beginning in Column 2. Classes that begin in the 7 am hour are omitted. The penalty for students taking a 7 am hour is consistent and robust. All time-of-day coefficients are positive and significant, which suggests holding class every hour of the day after the first benefits students. These effects are large in magnitude. Students taking a 9 am or later are expected to perform 0.16 standard deviations better than students taking the same class at 7am. Relative to the difference among other hours of the day, the penalty of early start times is the first-order effect. This is consistent with prior work (Carrell et al., 2011) who find students in the 2006–2007 academic year who faced the earliest start times (see Table 3) had lower grades throughout the entire school day relative to students who had their first class later in the day. Time-of-day estimates from Column 5 are plotted in Fig. 1 Panel A against the raw, unadjusted mean grades by hour of the day.

8 am hour classes are likewise bad for students, consistent with the literature suggesting that early start times are out of sync with students' circadian rhythms. In our preferred specification, 8 am is significantly worse than all later hours of the school day. In fact, we can reject the null hypothesis that the morning and afternoon are equivalent for student learning, regardless of if we define "the morning" to include the 7 am through 10 am hours, 8 am through 10 am hours or just 9 am and 10 am hours.[16]

Instructor schedule variables are included starting in Column 4. Teachers improve the more they teach that day. Students in a class with an instructor who has taught the material once earlier in the day perform 0.03 standard deviations better than students in the instructor's first class, all else equal. If the instructor has taught the material two or more times the benefit increases to 0.05 standard deviations. The effect is similar in magnitude to the penalty a student faces from consecutive classes. Instructors teaching consecutive classes do not appear to significantly impact student performance in any of our specifications.[17]

To our knowledge, this is the first set of results to causally identify these three types of effects in the same setting.[18] The negative impacts of student fatigue are particularly striking. Student assigned to two or three classes in a row are put at a disadvantage relative to peers who have breaks in their schedule.

What can be done to offset the effects of cognitive fatigue? The answer depends on the setting and schedule at hand. College students who have agency over their own schedules can aim to avoid scheduling back-to-back classes, when possible. Many higher education settings offer a wide range of course times with classes meeting on alternating days of the week so a student could reasonably take a full course load without ever needing to have back to back classes. A high school with a structured schedule like USAFA's may be able to increase mean achievement by a general reallocation of their breaks or of free periods. For example, the 3 pm hour at USAFA is the least utilized time of day in terms of courses taught. This means many students end the day an hour early, cramming their schedules earlier and necessarily facing more consecutive classes.

To show the interactions of time-of-day, student fatigue, and instructor schedules, we aggregate their expected effects in Table 5. Each column considers a different daily schedule a student might have. Each

[14] 7 am hour observations are dropped owing to lack of variation in *LeadUp*.

[15] The consecutive class measures are mutually exclusive. *1 Consecutive Class* refers to having had exactly one class immediately prior to the current class. *2+ Consecutive Classes* refers to having had two or more classes immediately prior.

[16] For each of these statements we perform a linear hypothesis test on the coefficients from Column 5. The tests are, respectively: H_0: $(\psi_{8am} + \psi_{9am} + \psi_{10am})/4 = (\psi_{1pm} + \psi_{2pm} + \psi_{3pm})/3$; p-value $= 0.001$. H_0: $(\psi_{8am}\psi_{9am} + \psi_{10am})/3 = (\psi_{1pm} + \psi_{2pm} + \psi_{3pm})/3$; p-value $= 0.009$. H_0: $(\psi_{9am} + \psi_{10am})/2 = (\psi_{1pm} + \psi_{2pm} + \psi_{3pm})/3$ the p-value $= 0.034$

[17] *1+ Consecutive Taught* is included in all subsequent models containing instructor schedule controls, but is not significant in any of them and is unreported.

[18] Also, as an alternate to our main specification, we interact the variables *1 Consecutive* and *2+ Consecutive* with dummy variables indicating a morning class (8am–10am) and an afternoon class (1pm–3pm). While not reported, we cannot reject the null hypothesis that the effect of student fatigue due to consecutive classes is the same in the morning and afternoon.

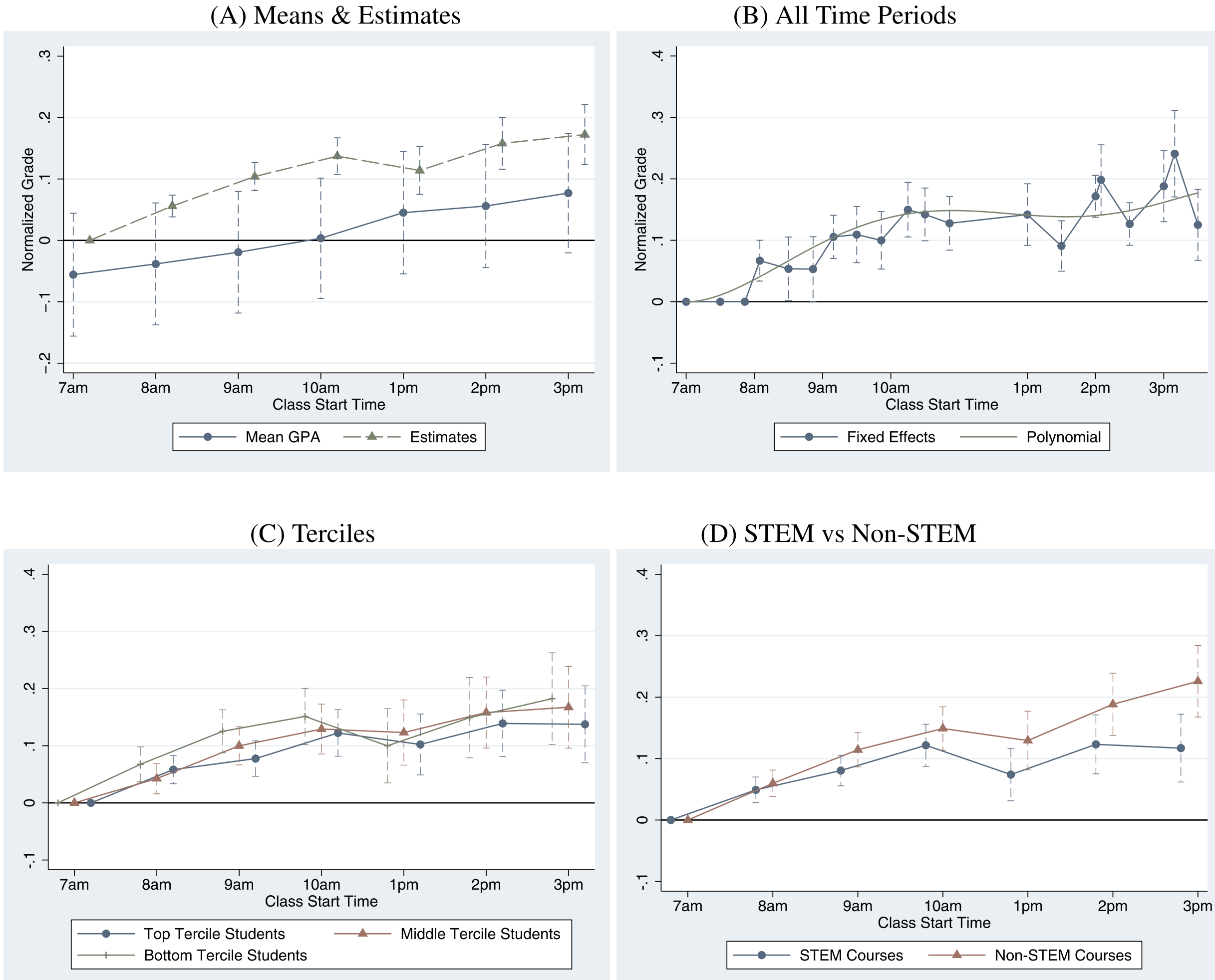


Fig. 1. Plotted Time-of-Day Regression Coefficients Outcome: Normalized Grade. *Note:The figures above show estimates of the time-of-day effect on standardized grades. Graph (A) shows raw, unadjusted mean grades along with time-of-day coefficients from Table 4 Column 5. Graph (B) plots time-of-day estimates from Columns 1 and 2 of Table 6. (C) shows esitmates from Columns 1–3 of Table 7. The STEM estimates from (D) correspond with stratified results available in the online appendix. 90% C.I.s are shown.*

cell provides the expected student performance relative to a 7 am class, factoring in all coefficients from Table 4, Column 5.[19] The row "Average" reports a student's expected standardized grade, relative to 7am, for the corresponding schedule.

The table reinforces that the benefits of a later start time are of first-order importance in our results. Columns 2 and 3 show that two students with the same course load, but where one effectively begins school 2 h later, have an expected 0.062 standard deviation difference in grades across all off their classes. This difference is equivalent to one student having all teachers two-thirds of a standard deviation better than the other.

Columns 3–5 consider students who all face a 7am class, but have their breaks spaced differently throughout the day. The differences here are smaller, but still significant. A student with a "Balanced" schedule earns expected grades 0.035 and 0.018 standard deviations higher than students with a "Morning" or "Long Lunch" schedule, respectively.

Columns 6 and 7 of Table 4 show estimates for the subsample of fall-semester freshmen. These students are taking almost exclusively required courses. While we have some concerns about major choice and course electives causing scheduling dependencies for our full sample, we have no such concerns for freshmen. Over twenty sections are taught each term for a majority of the courses taken by freshmen. Hence, if estimates for this subsample are similar to the overall sample, we feel confident asserting that it is unlikely our main effects are being driven by selection.

Column 6 corresponds to Column 4 for the full sample, and Column 7 includes student fixed effects, corresponding to Column 5. Because this sample observes a student over only one semester, the inclusion of student fixed effects is particularly demanding; thus Column 6 is our preferred specification for this subsample going forward. Estimates for these students are consistent with the full sample, but typically larger in magnitudes. The expected penalty of a consecutive class is nearly twice as large (−0.059). The impact of cumulative classes is negative, but insignificant. Time-of-day estimates similarly suggest that students assigned to 7am classes are put at a disadvantage.

For freshman, instructor effects are not significant. This may be due to the smaller sample or to the fact that introductory material is more standardized or routine for instructors. They may have little to improve on after their first lecture of the day. The row labeled "Freshman" of Table 5 is analogous to the "Average" row, but uses estimates from Column 6 of Table 4 (hourly estimates are not shown). Similar to the

[19] For example, the 10am estimate in the "Balanced" column suggests that a student who is taking a 10am hour class and has faced 2 cumulative classes and 1 consecutive class is expected to perform 0.093 standard deviations better than if they had taken the same class at 7am. Instructor effects are incorporated by multiplying the coefficients estimates by the probability of having an instructor who has taught the course 0, 1 or 2+ times already that day. The effect of instructor's consecutive classes is ignored.

Table 4
Effects of fatigue, timing and teacher load.

	(All students)					(Freshman fall)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1 Consecutive class	−0.024***		−0.027***	−0.028***	−0.031***	−0.059***	−0.043**
	(0.006)		(0.007)	(0.007)	(0.006)	(0.020)	(0.017)
2+ Consecutive classes	−0.020***		−0.018	−0.018	−0.019**	−0.059**	−0.060**
	(0.008)		(0.011)	(0.011)	(0.008)	(0.029)	(0.024)
Cumulative classes	0.025***		0.004	0.004	−0.018***	−0.014	−0.020**
	(0.003)		(0.006)	(0.006)	(0.004)	(0.012)	(0.010)
Cumulative-Sq	−0.003**		0.001	0.001	0.002	0.006	0.008**
	(0.001)		(0.002)	(0.002)	(0.001)	(0.005)	(0.004)
8 am hour		0.023**	0.035***	0.010	0.056***	0.093***	0.120***
		(0.010)	(0.011)	(0.012)	(0.009)	(0.033)	(0.027)
9 am hour		0.063***	0.075***	0.034**	0.104***	0.140***	0.164***
		(0.009)	(0.013)	(0.016)	(0.012)	(0.039)	(0.032)
10 am hour		0.112***	0.120***	0.056***	0.137***	0.145***	0.171***
		(0.011)	(0.016)	(0.021)	(0.015)	(0.050)	(0.042)
1 pm hour		0.109***	0.097***	0.034	0.114***	0.114*	0.104*
		(0.012)	(0.022)	(0.028)	(0.020)	(0.067)	(0.055)
2 pm hour		0.154***	0.154***	0.067**	0.158***	0.228***	0.204***
		(0.015)	(0.024)	(0.030)	(0.021)	(0.077)	(0.063)
3 pm hour		0.159***	0.157***	0.049	0.172***	0.160*	0.164**
		(0.018)	(0.026)	(0.034)	(0.025)	(0.089)	(0.073)
1+ Consecutive taught				0.011	0.005	0.030	−0.010
				(0.011)	(0.008)	(0.028)	(0.024)
1 Cumulative taught				0.031***	0.030***	−0.018	0.022
				(0.012)	(0.009)	(0.030)	(0.025)
2+ Cumulative taught				0.058***	0.050***	−0.005	0.055
				(0.018)	(0.013)	(0.047)	(0.039)
Indv. controls	Y	Y	Y	Y	N	Y	N
Peer controls	Y	Y	Y	Y	Y	Y	Y
Teacher FEs	Y	Y	Y	Y	Y	Y	Y
Individual FEs	N	N	N	N	Y	N	Y
N	187,525	187,525	187,525	187,525	187,525	24,264	24,264
Adj-R2	0.147	0.147	0.147	0.147	0.524	0.224	0.575

Standard errors in parentheses * ($p<0.10$), ** ($p<0.05$), *** ($p<.01$) Note: The table above shows the estimates from Eq. 1. Outcome is a student's standardized grade in a single course. Individual controls include gender, race, SAT scores as well as the admission office's academic composite, leadership and fitness scores. Peer controls include all the variables from individual controls, measured for a single section of a course. Teacher fixed effects are estimated separately for each semester of instruction. The quadratic term Cumulative-squared is de-meaned before squaring so the estimate for the variable Cumulative can be interpreted as the marginal effect at the mean. Standard errors are clustered at the individual student level.

full sample, schedules without a 7am hour class provide the highest expected grades. Conditional on an early morning class, the "Balanced" schedule with evenly spaced free periods performs better than the alternatives.

Overall, results from Columns 6 and 7 suggest that, if there is selection bias in the full sample's scheduling, it is not economically relevant.[20] The larger magnitude of a schedule's effects on freshman is consistent with school schedules being most out of sync with adolescents (Crowley et al., 2007). As students progress at USAFA they may be able to adjust to their schedules by explicitly changing their habits, but they may also benefit from natural changes in daily rhythms associated with aging into adulthood.

We now consider the robustness of our main result. Table 6 looks at variations of our primary specification. Column 1 measures class start time continuously with a flexible polynomial. Column 2 estimates a separate dummy variable for each start time listed in Table 3, rather than grouping them together by hour. The time-of-day estimates from these two columns are shown graphically in Panel B of Fig. 1. In both models, the effects of student fatigue are not statistically different from our preferred estimates. This suggests that, despite grouping together start times across different years, our primary model does a good job controlling for time-of-day effects.

Column 3 estimates a version of Eq. (2) by the inclusion of section fixed effects. The effects of fatigue are identified by comparing two students taking the same section of a class (same time-of-day, same instructor, same peers), but with variation in their prior schedules. Estimates of student fatigue are again consistent. The penalty associated with having to sit through back-to-back consecutive classes is near −0.03 standard deviations in all models. The last three columns estimate the equivalent models for freshman. More flexible controls for time increase the effect of consecutive classes for freshman to −0.07, but the difference is not significantly different from our preferred model.

In Table 7 we consider heterogeneity of these effects across the distribution of predicted student achievement. We group students in terciles of pre-enrollment academic aptitude.[21] Doing so helps us understand how to optimize class schedules so that the classes and/or students that benefit the most from being assigned "prime" times are the ones given those times. The first three columns show estimates for the full sample, the last three the fall freshman subsample. It is important to note that since USAFA is a highly selective institution, even the bottom-tercile students are among the top 15% of high school students nationwide. Fatigue has the largest impact on the bottom

[20] Further, we estimate variations of Columns 6 and 7 for every "age group" in our sample (spring-semester freshmen, fall-semester sophomores, etc.), and the results for each group are consistent in sign and magnitude with those for the overall population and for fall-semester freshmen. These results are available upon request.

[21] We regress average GPA on pre-enrollment characteristics (SAT scores, academic composite score) and generate a predicted-GPA for each student based on these characteristics. We use this predicted-GPA to rank students and group them in terciles.

Table 5
Expected grades by time of day.

	Full	Afternoon	Morning	Balanced	Long lunch
7 am hour	0.000		0.000	0.000	0.000
8 am hour	0.020		0.020		0.020
9 am hour	0.064	0.118	0.064	0.101	0.064
10 am hour	0.090	0.111	0.090	0.093	
1 pm hour	0.076	0.101	0.076	0.086	
2 pm hour	0.088	0.105			0.136
3 pm hour	0.117	0.126		0.145	0.114
Average	0.065	0.112	0.050	0.085	0.067
Other Specifications					
Freshman	0.068	0.096	0.039	0.076	0.065
Freshman top	0.077	0.107	0.048	0.051	0.058
Freshman mid	0.084	0.107	0.012	0.068	0.067
Freshman low	0.014	0.016	-0.007	0.038	0.021

Note: Each column shows estimates of a hypothetical one-day schedule. In the top panel, results show the predicted grade (relative to a 7 am hour class) based on coefficient estimates from Table 4, Column 5. For example, under the Balanced schedule, a student taking a 10 am hour class is expected to perform 0.093 standard deviations better than if they took the same course in the 7 am hour. The estimate captures the fact that, under this particular schedule, the class would be the third cumulative class of the day and second consecutive one. Teacher effects account for the probability that a teacher has taught 0, 1 or 2+ courses at any given hour of the day. Average shows the average grade for a given schedule, relative to a 7 am hour class. The bottom panel provides the resulting averages (without showing the expected hourly class values) for the subsamples: freshman fall semester, freshman top tercile, frehsman middle tercile, and freshman bottom tercile. Results correspond to Table 4, Column 6 and Table 7 Columns 1 and 3, respectively. USAFA's M/T scheduling feature means that each student will each have two separate daily schedules within a term.

tercile of USAFA students. A consecutive class reduces a bottom-tercile student's expected grade by 0.042 standard deviations, compared with 0.030 or 0.019 for top-tercile and middle-tercile students, respectively. Two or more consecutive classes only have a significantly negative impact on bottom tercile students.

Time-of-day dummy effects are included in all specifications, but not meaningfully different across the subgroups, which can be seen in Panel C of Fig. 1.

There are a few explanations for the differences across the ability distribution. Top students may be better able to maintain focus, regardless of their schedule. Or, they may face similar fatigue during class time, but are better able to learn material on their own outside of class. It could also be that top tercile students face the same penalty from student fatigue in their schedules but it is not picked up in our data because they are too far above the cutoff for an "A". In this case, top-tercile students do suffer losses in learning, but it cannot be measured by our model. We cannot separate out these various explanations as to why different student terciles respond to student fatigue differently.

Columns 4–6 look at effects by terciles for our freshman subsample. A similar pattern can be found. Students in the lowest tercile of predicted ability again suffer the largest average penalties from taking consecutive classes, with and a very large −0.13 standard deviation predicted effect when sitting in a third or fourth class in a row. Top tercile students show no significant fatigue due to consecutive classes, but are the only group to face significant fatigue from their cumulative class load. The point estimates for fatigue's impact on middle tercile students are in line with the full sample estimates from Table 4, but lower precision makes all estimates of fatigue insignificant for this group. Lower-ability students may be more dependent on absorbing knowledge during lectures and thus are more adversely affected when unable to focus in class. Given that USAFA students are high achievers to begin with, the cognitive penalties could be even larger for other student populations. Some ill-timed classes are inevitable, but a school could consider offering more favorable schedules to struggling students. We explore this idea with simulations below.

The bottom three rows of Table 5 show the effect of different schedules on each tercile's predicted grades. Top tercile freshmen students have very little predicted difference among the three final columns due to the negligible impact of consecutive classes on their predicted grades (the effect of cumulative classes, while significantly negative for this group, will average out across any 5-course schedule). Low tercile students, on the other hand, clearly benefit from the "Balanced" schedule which minimizes the number of consecutive classes they face.

Additional robustness checks include stratifying by STEM vs Non-

Table 6
Robustness: alternate measures of class timing and section fixed effects.

	(All students)			(Freshman fall)		
	(1) Cont. Time	(2) Time FEs	(3) Section FEs	(4) Cont. Time	(5) Time FEs	(6) Section FEs
1 Consecutive class	−0.028***	−0.031***	−0.028***	−0.062***	−0.070***	−0.078***
	(0.006)	(0.006)	(0.008)	(0.017)	(0.018)	(0.018)
2+ Consecutive classes	−0.016*	−0.021**	−0.021*	−0.071***	−0.069***	−0.073***
	(0.008)	(0.008)	(0.012)	(0.025)	(0.026)	(0.027)
Cumulative classes	−0.019***	−0.017***	0.005	−0.017	−0.017	−0.014
	(0.004)	(0.004)	(0.006)	(0.011)	(0.011)	(0.011)
Cumulative-Sq	0.002*	0.001	0.001	0.008*	0.007	0.007
	(0.001)	(0.001)	(0.002)	(0.004)	(0.004)	(0.005)
1 Cumulative taught	0.025***	0.029***		−0.009	0.002	
	(0.009)	(0.009)		(0.027)	(0.027)	
2+ Cumulative taught	0.047***	0.049***		−0.005	0.004	
	(0.013)	(0.013)		(0.042)	(0.042)	
Indv. controls	N	N	Y	Y	Y	Y
Peer controls	Y	Y	Y	Y	Y	Y
Teacher FEs	Y	Y	N	Y	Y	N
Individual FEs	Y	Y	N	N	N	N
Section FEs	N	N	Y	N	N	Y
N	187,525	187,525	187,525	24,264	24,264	24,264
Adj-R2	0.524	0.524	0.152	0.232	0.232	0.236

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < .01$. Note: The table above shows the estimates from Eq. (1) when including the variables listed above. All regressions include controls for student characteristics and classroom peer effects as well as course by year by schedule-day fixed effects. The quadratic term Cumulative-squared is de-meaned before squaring so the estimate for the variable Cumulative can be interpreted as the marginal effect at the mean. Standard errors are clustered by student.

Table 7
Subgroup analysis: terciles.

	(All students)			(Freshman fall)		
	(1) Top tercile	(2) Mid tercile	(3) Low tercile	(4) Top tercile	(5) Mid tercile	(6) Low tercile
1 Consecutive class	−0.030***	−0.019*	−0.042***	−0.015	−0.065	−0.097**
	(0.010)	(0.010)	(0.012)	(0.033)	(0.040)	(0.038)
2+ Consecutive classes	−0.022	0.000	−0.035**	0.029	−0.069	−0.129**
	(0.013)	(0.014)	(0.017)	(0.043)	(0.054)	(0.050)
Cumulative classes	−0.019***	−0.022***	−0.008	−0.038**	−0.007	0.008
	(0.006)	(0.007)	(0.008)	(0.017)	(0.021)	(0.020)
Cumulative-Sq	0.001	0.001	0.001	0.009	0.010	0.004
	(0.002)	(0.002)	(0.002)	(0.007)	(0.008)	(0.008)
1 Cumulative taught	0.030**	0.026*	0.038**	−0.026	−0.006	0.012
	(0.015)	(0.016)	(0.017)	(0.047)	(0.057)	(0.054)
2+ Cumulative taught	0.052**	0.069***	0.038	0.007	−0.046	0.067
	(0.022)	(0.024)	(0.026)	(0.076)	(0.090)	(0.085)
Time-of-day	Y	Y	Y	Y	Y	Y
Indv. controls	N	N	N	Y	Y	Y
Peer controls	Y	Y	Y	Y	Y	Y
Teacher FEs	Y	Y	Y	Y	Y	Y
Individual FEs	Y	Y	Y	N	N	N
N	66,145	62,834	58,546	8447	7758	8059
Adj-R2	0.547	0.499	0.444	0.240	0.141	0.157

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < .01$. Note: The table above shows the estimates from Eq. (1) when including the variables listed above. The outcomes for the regressions shown in columns (1)–(6) is the normalized grade in the course. All regressions include controls for student characteristics and classroom peer effects as well as course by year by schedule-day fixed effects, teacher fixed effects, quadratic values of student fatigue and instructor schedules. Terciles account for predicted academic performance and are based on admission composite scores. The quadratic term Cumulative-squared is de-meaned before squaring so the estimate for the variable Cumulative can be interpreted as the marginal effect at the mean. Standard errors are clustered by student.

STEM courses and considering the binary outcomes of failing or acing a course. The estimates are not substantively different than the main results and are available in an online appendix.

Lastly, we estimate Eq. (2) on the freshman sample and present the results in Fig. 2. This specification includes section fixed effects; variation comes from comparing students taking the same class at the same time with the same professor, but with different prior schedules on that day. Each bar in the figure represents a single coefficient, β, from Eq. (2), sorted by time and color-coded by *LeadUp* scenario. Having had a free period immediately prior is the omitted reference group. The second graph includes individual student fixed effects.[22] For two students in the same 9 am section, one who had a free period beforehand and one who had a non-STEM class, the student with the free period is expected to perform 0.15 standard deviations better, a stark difference. Physical education (P.E.) is similarly beneficial in the morning. A free period prior is a strong predictor of success in a course. The 3pm hour is an interesting exception. Here, both P.E. and a free period beforehand lead to an expected *decrease* in performance, although results are not significant. One explanation could be that these students are mentally "checked out." Lunch, combined with either a P.E. or no class, means that students have had a nearly three-and-a-half-hour break from the classroom. It may be difficult for students to re-focus for a single afternoon class after an extended break.

5.1. Simulations

Our main results focus on the impact that time of day, student fatigue, and instructor schedules have on grades in individual classes. Now we examine the overall impact course rescheduling could have on student achievement. We perform two simulations in which we assess how achievement would differ if students were assigned schedules based on their academic aptitude. The simulations aim to estimate the extent to which course schedules could be used to reduce inequality in student outcomes while raising mean achievement.

We focus on first semester freshman because there is a high amount of course overlap among these students and so it is more natural to simulate a reassignment of schedules than it would be for upperclassmen who are taking elective courses. For the first simulation, we use estimates from Table 4 Column 6. In the first simulation, we assume that the schedule impacts all students equally. In the second simulation, we use estimates from Table 7, which assumes that course schedules impacts students differentially.

Because of the M/T day schedule at USAFA, each student can be thought of as having a fourteen-period schedule which spans two seven-period days. Our simulations take the 4536 freshman schedules (which represent 1900 different combinations of class periods) observed in our sample and calculates the estimated impact of each, independent of the actual student. Then we rank those schedules from best to worst and simulate reassigning them to students in reverse order.

First, each student is assigned a predicted own-GPA, determined using their background characteristics. This is used to rank students by predicted academic ability.[23]

Second, we create estimates akin to those in Table 5, using our preferred freshman results. These estimates represent predicted impact of taking a generic academic course at a particular time of day, given the student's whole schedule of courses on that day.[24]

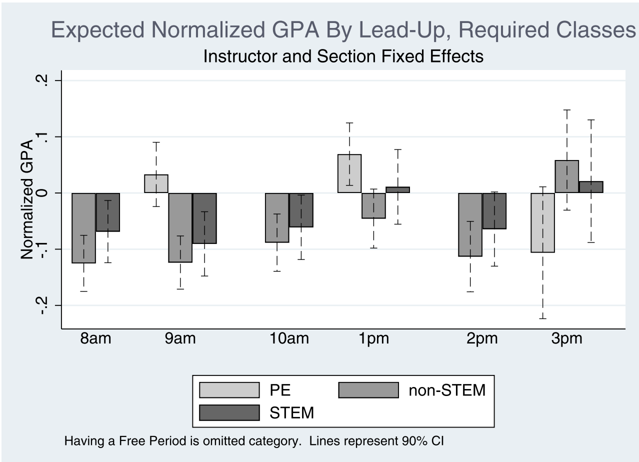
We then average together the individual course impacts for a student into one overall predicted schedule-GPA. For example, using the designations from Table 5, if a student had a "Balanced" schedule on their M-day (Average impact: 0.076) and a "Long Lunch" schedule on their T-day (0.065) their overall schedule would be assigned a predicted impact of 0.071.

Each predicted schedule-GPA value is independent of the student

[22] Physical education is a two-period class, but only meets starting starting in the 7 am, 9 am, and 1 pm hours.

[23] $\hat{Y}_i$ is the predicted own-GPA of a student obtained using only X_i coefficients from our freshman estimates of Eq. (1).

[24] $\hat{Y}_{cjsp}$ is the predicted schedule-GPA of one course on a student's schedule and uses the coefficient estimates of *Fatigue, Time*, average teaching counts, and year-by-schedule day fixed effects. For example, in fall semester 2009, all 9 am hour courses where the student also had an 8am hour course will have the same value of $\hat{Y}_{cjsp}$.



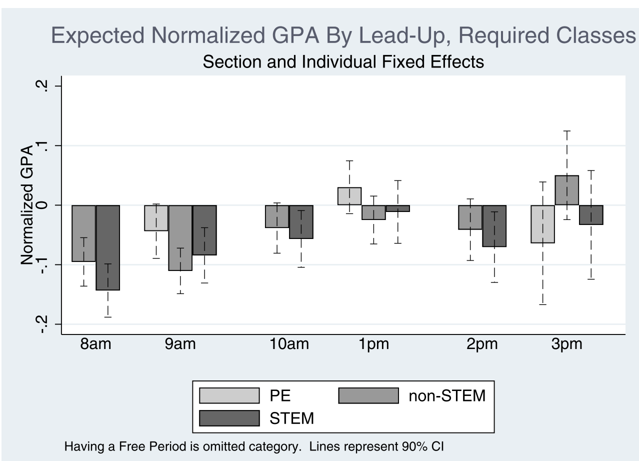


Fig. 2. Effects of Preceding, *Lead Up*, Class. *Note: Results correspond to estimates of Eq. (2) for freshman in their fall semester. Having a "Free" period is the reference category. Estimates shown both with instructor and section fixed effects (top) and with individual and section fixed effects (bottom). 90% confidence intervals shown in dotted lines, all regressions cluster standard errors at the individual level.*

who was actually assigned that schedule. Likewise, our predicted own-GPA is independent of the schedule or grades actually received by the student.

We then re-assign schedules to students in reverse order, conditional on cohort and class counts. Students with low predicted GPAs are given the most favorable schedules and the high-ability students are given the least favorable ones. If a student is enrolled in five academic courses, she will be assigned a different schedule also with five academic courses.[25] We limit ourselves to the set of existing schedules in our data to ensure that results would be feasible within USAFA's current scheduling constraints, such as faculty size and classroom availability. The simulations do not account for possible general equilibrium effects induced by schedule reassignment such as changes in peer effects, but the simulations do hold hypothetical class size constant.

The first simulation assumes that the time-of-day and fatigue effects are homogeneous across students by estimating schedule-GPA using the same estimates from Table 4, Column 6 for all students. The results are shown in Table 8. This is a zero-sum simulation as any gain a student gets from a better schedule is offset by another student receiving a worse schedule. The results show a narrowing of the overall grade distribution with no change in average performance across students. Specifically, the standard deviation of grades decreases from 0.448 to 0.420, a decrease in variance of around 9%. Bottom-tercile students experience a 0.023 standard deviation increase in overall performance, on average, but a similar loss is predicted for the top tercile students.

The second simulation allows for heterogeneous effects of schedules based on students' predicted ability. The simulation is done in a similar fashion, but we use coefficients from Columns 4–6 of Table 7 to estimate the schedule-GPA of each course.[26] Once again, students are assigned schedules in an inverse relationship to their predicted ability. The worst student is assigned the most-favorable schedule, while the best student is assigned the least-favorable schedule. Depending on the student's tercile, however, the coefficients used to estimate the schedule's impact will vary.[27]

By assuming that top-, middle-, and bottom-tercile students experience differing effects of fatigue, there is an opportunity to both narrow the overall GPA distribution and raise mean performance. Results from the second simulation are shown in the lower half of Table 8. They show that re-assigning schedules raises expected performance by 0.012 standard deviations for all students. Variance in student achievement is reduced by 11%. These gains are concentrated in the bottom tercile of student ability: this group experiences an average GPA increase of 0.034 standard deviations. The gain due to rescheduling for the bottom-tercile students is equivalent to increasing their teacher quality in all courses by 0.33 standard deviations. Middle-tercile students experience gains near the overall average, while top-tercile students lose in this scenario. Their grades decrease by an average of −0.006 standard deviations, one-fifth the gains of the bottom-tercile students, reflecting the fact that they are generally more robust to undesirable schedules.

[25] However, we do not condition on the exact set of classes. A student enrolled in Chemistry, Calculus, Spanish 1 and American History may be simulated with a schedule that comes from a student who took Biology, Calculus, French 1 and European History. If we forced the exact courses to match, the bins of possible reassignments would be too small. Further, estimates in Table 5 use expected instructor values. For example, on average, 25% of professors in 9 am hour classes are teaching for the second time that day and so the estimates reflect a 25% chance of "1 + Consecutive Taught" = 1. This way, the predicted impact of a schedule is really only due to course timing. All "Long Lunch" schedules are treated the same, even though in reality, two students may have schedules that meet during the same times, but one has all instructors teaching for the first time that day and the other does not.

[26] For each of the 4536 observed fourteen-period student schedule, three scores are calculated: one each for the schedule's predicted impact on top, middle, and bottom-tercile students. A schedule's predicted impact is independent of student characteristics and is only based on the time of day the course takes place and the timing of courses a student took earlier on that day.

[27] An example using the estimates and terminology from the bottom rows of Table 5. Assume there are three students, *T, M*, and *B* who are in the top, middle, and bottom terciles respectively. The actual M/T schedules assigned to these students are: student *T* has a "Long Lunch" schedule on both M/T days. Student *M* has a "Balanced" schedule one day and "Morning" schedule the other. Student *B* has a "Morning" and an "Afternoon Schedule". Using the tercile-specific estimates, these schedules have an average predicted impact of 0.034 for the three students. In the simulation, student *B* will get assigned a schedule first and be given the best of the three options for a bottom-tercile student. In this case, the average impact of *T*'s "Long Lunch" schedule (0.021) is higher than the predicted impact of either *M* (.038/2 + −.007/2)or *B*'s (0.016/2 + −.007/2) original schedule and so student *B* will be given *T*'s schedule. *M* then gets assigned the better of the two remaining schedules based on estimates for middle-terciles students. *B*'s schedule is best for a middle-tercile student (both options contain "Morning" on one day, but "Afternoon" is better for a middle-tercile student than "Balanced") . *T* will be given the last remaining schedule, which was *M*'s original schedule of "Morning" and "Balanced." Because "Morning" schedules do not impact top-tercile students the same way they do bottom-tercile ones, the simulated reassignment has increased the average schedule impact across the three students increases to 0.043. *B*'s individual expected schedule impact goes from 0.005 to 0.021.

Table 8
Schedule reassignment simulation.

	All			Bottom		Middle		Top	
	Mean	SD	N	Mean	SD	Mean	SD	Mean	SD
Homogeneous schedules									
ActualGPA	0.00	0.448	4536	−0.580	0.225	0.022	0.121	0.681	0.156
SimulatedGPA	0.00	0.428	4536	−0.557	0.215	0.025	0.113	0.655	0.150
Difference	**0.0**			**0.023**		**0.003**		**−0.026**	
Heterogenous schedules									
ActualGPA	0.0	0.428	4536	−0.452	0.240	−0.059	0.119	0.476	0.210
SimulatedGPA	0.012	0.408	4536	−0.418	0.220	−0.045	0.111	0.470	0.200
Difference	**0.012**			**0.034**		**0.014**		**−0.006**	

Note: The table above shows the estimates from simulations where students were inversely re-assigned schedules based on predicted own-GPA and predicted overall schedule impact. Schedule impacts were predicted separately by ability tercile using results from Columns 4–6 of Table 7.

6. Discussion and conclusions

We study how the organization of classes throughout the school day affects academic achievement. We consider the effects of three distinct components of course schedules: student fatigue, the time of day a course meets, and instructor schedules. Our results show that all three significantly impact performance.

Student fatigue is such that two similar students taking the same class with the same teacher, but with different schedules, could be expected to receive grades as different as 0.15 standard deviations. Cognitive fatigue is well established in experimental settings (Vogel-Walcutt, 2012) and we show that it exists within the school day environment. This research extends our understanding of what outside factors affect academic achievement and provides an opportunity to increase achievement, and, presumably human capital, by rescheduling the times that classes are held.

Our findings support the idea that the way in which school schedules are currently organized hinders student performance. Consistent with earlier work using this sample of students (Carrell et al., 2011) as well as many other studies on different populations (Diette & Raghav, 2017; Edwards, 2012; Lusher & Yasenov, 2018; Wahlstrom et al., 2014), we find consistently negative effects associated with early morning (especially 7am hour, but also 8am hour) classes. Students seem to learn better in the afternoon–times that are better aligned with their circadian rhythms. These results are consistent with work showing that for adolescents scores on intelligence tests are significantly lower during the early-morning hours (Goldstein et al., 2007) . Instructors, if anything, improve as the day goes on, exhibiting short-term learning by doing.

The institutional characteristics of USAFA provide us a unique chance to identify the effect of student fatigue uncontaminated by student selection in schedules or courses, lack of common grading standards or differing testing times and conditions. The course and grading structure at USAFA allow for a causal interpretation of our results. Assignment to classes and professors is conditionally random, attendance in all classes is mandatory, and all students enrolled in a course in a given semester take their exams during a common testing period and are graded on a collective curve. Hence, we can be certain that the effects we find are due to variation in schedules and reflect differences in learning/understanding of class material and not differences in grading standards.

Subgroup results show that bottom-tercile students are most susceptible to the effects of timing and fatigue. Lower-ability students at USAFA are still likely to be in the 85th percentile nationally, so while not nationally representative, we have little reason to think that USAFA students would be more adversely impacted by their schedules than a typical student. Unfortunately, our data do not allow us to determine *why* differences across ability groups exist. There are a number of possible explanations for this difference: high-achieving students may be better able to learn even when they are tired; they may be better able to teach themselves material they missed in class; or they may actually suffer from cognitive fatigue, but are still above a threshold where it impacts their grade. Each of these hypotheses would be an interesting area for further research within the social or biological sciences.

We perform simulations that show, in a hypothetical framework, assigning the worst students to more optimal schedule is a pathway to increasing overall mean achievement. However, implementing such a policy would also require consideration of changes in peer effects (Carrell, Fullerton, & West, 2009) or class size (Diette & Raghav, 2015; Monks & Schmidt, 2011) induced by reassignment. Our estimates suggest that the benefits of having an afternoon class relative to a early morning one would be equivalent to increasing an average class size by 7 students, or 35%.[28]

In our data, the shortest break for a student is around an hour long. We cannot identify the minimum amount time needed to reset a student's fatigue against the impact of consecutive classes. Based on research looking at the onset of fatigue (Jackson et al., 2014), we hypothesize that a 15- or 20-min break, perhaps at the expense of a shorter lunch, during the school day could be a way to offset student fatigue, without needing to reassign student schedules.

We recommended policies, or rules-of-thumb, based on our results. First, our findings consistently and strongly support the hypothesis that classes starting 9am or later are more favorable to student learning than earlier ones. Our aggregated schedule scores consistently show 7am and 8am classes to be the worst for students. However, shifting a school's entire schedule may be expensive or unpopular among administration, teachers, parents, and coaches. A goal of this paper is to think through what schools can do within their existing schedule structure to improve student outcomes. To offset early start times, schools might consider scheduling as much P.E. as possible in the morning. We also show a clear downside to consecutive classes, especially for the lowest-performing students. Thus scheduling free periods so they provide breaks throughout the day could improve performance. In our sample, the afternoon periods were both the best time for learning and the least utilized class periods. Stretching out the school day and providing more breaks for students has clear benefits.

We also show that teachers improve at teaching the same material as the day goes on. Placing struggling students in an instructor's second or third lecture of the day will likely be beneficial. While most of our discussion focuses on policies that schools could implement, many students in higher education have control over their own schedules. A student looking to proactively schedule classes to take advantage of their own alertness could focus on spacing throughout the day.

[28] This is a back-of-the-envelope calculation uses Monks and Schmidt (2011) estimates of negative effects of class size on self-reported learning in a similar higher-ed environment.

Acknowledgments

The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the USAF, DoD, or the U.S. government. This article was completed with data managed in collaboration with Lt. Col. Scott Carrell and Jim West. We thank Kathryn Bruchmann, Scott Carrell, Seth Gershenson, Andrew Goodman-Bacon, Hilary Hoynes, Matthew Larsen, Lisa Schulkind, and Adam Shapiro for their valuable input as well as seminar participants at Santa Clara University, UC Davis, the Southern Economic Association Annual Meeting, the APPAM Annual Conference, and the SREE Annual Conference. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.econedurev.2018.10.007.

References

Ackerman, P. L., & Kanfer, R. (2009). Test length and cognitive fatigue: An empirical examination of effects on performance and test-taker reactions. *Journal of Experimental Psychology. Applied, 15*(2), 163–181.
Blake, M. (1967). Time of day effects on performance in a range of tasks. *Psychonomic Science, 9*(6), 349–350.
Booij, A. S., Leuven, E., & Oosterbeek, H. (2017). Ability peer effects in university: Evidence from a randomized experiment. *The Review of Economic Studies, 84*(2), 547–578.
Cardinali, D. (2008). Chronoeducation: How the biological clock influences the learning process. In A. Battro, K. Fischer, & P. Lena (Eds.). *The educated brain* (pp. 110–126). Cambridge University Press.
Carrell, S. E., Fullerton, R. L., & West, J. E. (2009). Does your cohort matter? measuring peer effects in college achievement. *Journal of Labor Economics, 27*(3), 439–464.
Carrell, S. E., Maghakian, T., & West, J. E. (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*(3), 62–81.
Carrell, S. E., Sacerdote, B. I., & West, J. E. (2013). From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica, 81*(3), 855–882.
Carrell, S. E., & West, J. E. (2010). Does professor quality matter? evidence from random assignment of students to professors. *Journal of Political Economy, 118*(1).
Carskadon, M., Vieira, C., & Acebo, C. (1993). Association between puberty and delayed phase preference. *Sleep, 16*(3).
Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review, 104*(9), 2593–2632. https://doi.org/10.1257/aer.104.9.2593.
Cortes, K., Bricker, J., & Rohlfs, C. (2012). The role of specific subjects in education production functions: Evidence from morning classes in chicago public high schools. *The B.E. Journal of Economic Analysis & Policy, 12*.
Cotti, C., Gordanier, J., & Ozturk, O. (2018). Class meeting frequency, start times, and academic performance. *Economics of Education Review, 62*, 12–15.
Crowley, S., Acebo, C., & Carskadon, M. (2007). Sleep, circadian rhythms, and delayed phase in adolescents. *Sleep Medicine, 8*, 602–612.
Curcio, G., Ferrara, M., & Gennaro, L. D. (2006). Sleep loss, learning capacity, and academic performance. *Sleep Medicine Reviews, 10*, 323–337.
Diette, T. M., & Raghav, M. (2015). Class size matters: Heterogeneous effects of larger classes on college student learning. *Eastern Economic Journal, 41*(2), 273–283.
Diette, T. M., & Raghav, M. (2017). Does early bird catch the worm or a lower GPA? Evidence from a liberal arts college. *Applied Economics, 49*(33), 3341–3350.
Dills, A., & Hernandez-Julian, R. (2008). Course scheduling and academic performance. *Economics of Education Review, 27*, 646–654.
Edwards, F. (2012). Early to rise? The effect of daily start times on academic performance. *Economics of Education Review, 31*(6), 970–983.
Goldstein, D., Hahn, C., Hasher, L., Wiprzycka, U., & Zelazo, P. D. (2007). Time of day, intellectual performance, and behavioral problems in morning versus evening type adolescents: Is there a synchrony effect? *Personality and Individual Differences, 42*(3).
Groen, J., & Pabilonia, S. W. (2017). Snooze or lose: High school start times and academic achievement. IZA Discussion Paper No. 11166, IZA.
Heissel, J., & Norris, S. (2018). Rise and shine: The effect of school start times on academic performance from childhood through puberty. *Journal of Human Resources, 53*(4), 957–992.
Hinrichs, P. (2011). When the bell tolls: The effects of school starting times on academic achievement. *Education Finance and Policy, 6*(4), 486–507.
Jackson, S. A., Sabina, K., & Eugene, A. (2014). Low cognitive load and reduced arousal impede practice effects on executive functioning, metacognitive confidence and decision making. *PLoS One, 9*(12), e115689.
Jacob, B. A., & Rockoff, J. E. (2011). *Organizing schools to improve student achievement: start times, grade configurations, and teacher assignments.* Brookings Institution, Hamilton Project.
Kane, T. J., & Staiger, D. O. (2008). *Estimating teacher impacts on student achievement: An experimental evaluationWorking Paper 14607*. National Bureau of Economic Researchhttps://doi.org/10.3386/w14607.
Lusher, L., & Yasenov, V. (2018). Gender performance gaps: Quasi-experimental evidence on the role of gender differences in sleep cycles. *Economic Inquiry, 56*(1), 252–262.
Monks, J., & Schmidt, R. M. (2011). The impact of class size on outcomes in higher education. *The B.E. Journal of Economic Analysis & Policy, 11*(1), 1–19.
Persson, J., Welsh, K. M., Jonides, J., & Reuter-Lorenz, P. A. (2007). Cognitive fatigue of executive processes: Interaction between interference resolution tasks. *Neuropsychologia, 45*(7), 1571–1579.
Pope, N. (2015). How time of day affects productivity: Evidence from school schedules. *The Review of Economics and Statistics, 98*(1).
Schmidt, C., Collette, F., Cajochen, C., & Peigneux, P. (2007). A time to think: Circadian rhythms in human cognition. *Cognitive Neuropsychology, 24*(7), 755–789.
Trocket, M., Barnes, M., & Egget, D. (2000). Health-related varibales and academic performance among first-year college students: Implications for sleep and other behaviors. *Journal of American College Health, 49*.
Vogel-Walcutt, J., Fiorella, L., & Carper Teresa and Schatz Sae (2012). The definition, assessment, and mitigation of state boredom within educational settings: A comprehensive review. *Educational Psychology Review, 24*, 89–111.
Wahlstrom, K. (2002). Changing times: Findings from the first longitudinal study of later high school start times. *NASSP Bulletin, 86*(633).
Wahlstrom, K., Dretzke, B., Gordon, M., Peterson, K., Edwards, K., & Gdula, J. (2014). *Examining the impact of later high school start times on the health and academic performance of high school students: A multi-site study.* Center for Applied Research and Educational Improvement, St Paul, MN: University of Minnesota.https://conservancy.umn.edu/handle/11299/162769.
Wolfson, A., & Carskadon, M. (1998). Sleep schedules and daily functioning in adolescents. *Child Development, 69*(4), 875–887.