DISCUSSION PAPER SERIES

Time of Day, Cognitive Tasks and Efficiency Gains

Alessio Gaggero
Denni Tommasi

DISCUSSION PAPER SERIES

IZA DP No. 13657

Time of Day, Cognitive Tasks and Efficiency Gains

Alessio Gaggero
University of Granada and Campus Universitario de Cartuja

Denni Tommasi
Monash University and IZA

SEPTEMBER 2020

ABSTRACT

Time of Day, Cognitive Tasks and Efficiency Gains*

The link between time-of-day and productivity on cognitive tasks is crucial to understand workplace efficiency and welfare. We study the performance of University students taking at most one exam per day in the final two weeks of the semester. Exams are scheduled at different time-of-day in a quasirandom fashion. We find that peak performance occurs around lunchtime (1.30pm), as compared to morning (9am) or late afternoon (4.30pm). This inverse-U shape relationship between time-of-day and performance (i) is not driven by stress or fatigue, (ii) is consistent with the idea that cognitive functioning is an important determinant of productivity and (iii) implies that efficiency gains of up to 0.14 standard deviations can be achieved through simple re-arrangements of the time of exams. While researchers have shown that biological factors influence changes in productivity between day and night shifts, we establish that such relationship is also important within a standard day-light shift. A simple back of the envelope calculation applied to an external context that is likely to benefit from our results, elective surgeries, suggests that a different sorting of the cognitive tasks performed by surgeons may lead to an increase in the number of patients saved.

Corresponding author:
Denni Tommasi
Department of Econometrics and Business Statistics
Monash University
Building H, Caulfield Campus
900 Dandenong Road
Melbourne VIC 3145
Australia

E-mail: denni.tommasi@monash.edu

1 Introduction

The productivity of working hours is an important determinant of workplace efficiency and welfare (Pencavel, 2018). Many authors have documented that individual productivity decreases during the day due to stress and fatigue (see e.g. Brachet, David, and Drechsler (2012) and Pope (2016) among others). This explanation does not account for the possibility that productivity may also vary depending on the type of task and the specific time-of-day in which the task is performed. This intuition comes from the fact that alertness and mental focus, which are important determinants of cognitive functioning, vary substantially during the day (Higuchi et al., 2000) and, in turn, could also affect the productivity on cognitive tasks specifically. This implication is not obvious, however, because individuals may adopt a variety of behavioral strategies that could help them improve their performance on cognitive tasks that are not scheduled at optimal time-of-day. Establishing the relevance of time-of-day effects may reveal insights that could change the way we think about task sorting for skilled individuals inside an organization.

In this paper, we examine the relationship between time-of-day and cognitive tasks performance during daytime. The key challenge to establish a causal link is that, in most settings, individuals (i) perform several type of tasks during the day, which makes it hard to separate the effect of stress and fatigue from other possible determinants, and (ii) choose the time-of-day at which they perform certain tasks. To overcome both problem, we exploit an ideal setting, provided by the higher education system in the UK, that allows us to identify the causal effect of interest. In our framework, students can take at most one exam per day in one of three time-slots: Morning, 9am; Lunch, 1.30pm; and Afternoon, 4.30pm.[1] Our identification strategy relies on quasi-random assignment of the time-of-day of the exams. Specifically, the University exam office is in charge of the schedules and, by following a set of clearly specified rules, exams are simultaneously scheduled into different timeslots by means of a software. Then, schedules are published on-line for both students and Professors a few weeks prior to the exam session.

We use administrative data on the universe of students from one of the largest public Universities in the UK, spanning from 2014-15 to 2018/19. The data contain the complete records on half million student-exam level observations. Taking advantage of these detailed information, we are able

---

[1]Our analysis uses University exams as proxy for cognitive tasks because they "[...] require a person to mentally process new information (i.e., acquire and organize knowledge/learn) and allow them to recall, retrieve that information from memory and to use that information at a later time in the same or similar situation (i.e., transfer)." (Kester and Kirschner, 2012).

to compare the performance of students taking exams at different time-of-day in the final two weeks of the semester. Since much of a student's future is determined by their academic performances, test scores are often used as proxies for cognitive skills that are valued in the labor market (Bradley and Green, 2020). Indeed, not only test scores are a significant determinant of wage determination (Murnane et al., 1995) and labor-force quality (Hanushek and Kimko, 2000), but they are also a significant predictor of economic growth (Hanushek and Woessmann, 2008). Furthermore, from a biological standpoint, individuals enter adulthood at around 20 years old (Roenneberg et al., 2004). Since our sample is made of University students who largely satisfy this definition, they have, on average, similar biological characteristics as other adults performing cognitive tasks in the labor market. This makes our findings interesting to a broader population of skilled workers we wish to speak to.

The main results are threefold. First, we find that, on average, the productivity profile of students quasi-randomly assigned to different time-of-day has an inverse U-shape relationship, where peak performance occurs around lunchtime (1.30pm). At this time-of-day, their exam marks increase by 0.069 percentage of a standard deviation (SD) with respect to a morning exam (9am), which we use as control group, and decrease again in late afternoon (4.30pm). Second, these time-of-day effects are largely heterogeneous between the Fall and Spring exam sessions. For example, students taking an exam during lunchtime in January, at the end of the first semester, increase their performance by 0.094 SD, whereas students taking an exam during lunchtime in May-June, at the end of the second semester, increase their performance by only 0.043 SD. Third, time-of-day effects are stronger for cognitive tasks involving fluid intelligence (e.g. problem solving, logic thinking and abstract reasoning) (see e.g. Ghisletta and Lecerf (2018) for a formal definition). Specifically, we find that students taking a STEM exam during lunchtime in the Fall increase their performance by 0.139 SD, whereas we find no similar effects for students taking a non-STEM exam.[2] Balancing, robustness checks and falsification tests support a causal interpretation of the findings. Finally, there are no clear systematic differences in time-of-day effects between men and women, or low- and high-performing students.

The estimated gains in productivity at specific time-of-day may be caused by different mechanisms. On one hand, time-of-day may affect the behavior of individuals around the time of the

---

[2] "STEM" education is a term used to refer collectively to the teaching of the disciplines within its umbrella: Science, Technology, Engineering and Mathematics.

exam, such as their study strategy, food intake or sleep patterns. On the other hand, since time-of-day affects cognitive functioning, this could in turn affect their performances at the moment of the task. Our dataset does not contain information that allow us to identify, separately, the magnitude of each mechanism driving the results. However, as we will discuss in detail, the inverse-U shape relationship found in our paper is consistent with recent findings on the circadian rhythm, which is a biological process that governs our sleep-wake cycles (Cardinali, 2008). This suggests that the latter mechanism is likely to be the main explanation. The consistency between our causal estimates and the theory of circadian rhythm is important because it provides a rationale for the formulation of new policies that can improve the productivity of workers performing cognitive tasks in many sectors of the economy. The main implication is that tasks involving fluid intelligence are more affected by time-of-day and should be moved to the early afternoon, particularly at times of year when sunlight exposure is limited. Depending on the type of job, an optimal sorting of the tasks may lead to incredibly relevant efficiency gains. Indeed, in the final discussion of the paper, we apply our insights to one important context that is likely to benefit from our research, namely the sorting of elective surgeries, which we use as a case study. We show to what extent a simple change in schedules can lead to a marked decrease in mortality rate.

The paper proceeds as follows. Section 2 reviews the related literature and our contributions. Section 3 describes the institutional setting, presents the data and discusses the conceptual framework and identification strategy. Section 4 outlines the results, including several robustness and falsification tests, an explanation for the observed results as well as a discussion of the implications of our findings. Finally, Section 5 concludes.

2 Related Literature

Our paper is primarily motivated by the extensive literature on workplace productivity. The question of how productivity varies within a given work-shift goes back to Florence (1924), on the underpinnings of Taylorism, and Hamermesh (1990, 1996). Focusing on productivity changes during daylight is interesting because, as documented by Hamermesh (1999), the fraction of work performed during the traditional regular working day has grown over time.[3] Recently, studies in

---

[3]A related literature focuses on how productivity changes between day and night work-shifts. For example, Folkard and Tucker (2003) document a decline in productivity during the night shifts as a result of a decline in work-safety. Many studies in the medical literature also document an increase in sleep deprivation of medical residents which, in turn, lowers their clinical performance at night (e.g. Weinger and Ancoli-Israel (2002), Veasey et al. (2002) and Philibert (2005) among others).

this area provide causal evidence of decreasing returns to the number of hours worked (see e.g. Pencavel (2015) and Collewet and Sauermann (2017) for recent findings). Specifically related to individuals performing also cognitive tasks, Brachet et al. (2012) compare the performance of paramedics in short and long shifts and find that performance deteriorates significantly toward the end of a long shift. Whereas Pope (2016) studies differences in productivity among students and draws a parallelism with the cognitive skills used in a work setting. This author assesses the effect of class scheduling on test scores of high-schoolers (aged 12-17) and finds that productivity declines for classes scheduled later in the day. All these papers suggest that fatigue accumulated throughout the day may drive the results. However, the type of task and the time-of-day in which a cognitive task is performed may also have an impact. To the best of our knowledge, this is the first paper to isolate time-of-day as mediating factor for improvements in performance of individuals carrying out cognitive tasks. By doing so, we reach different conclusion for optimal task sorting of skilled workers inside an organization.

The second literature we relate to focuses on scheduling and students performance. Much of the research in this area has focused on class scheduling and school start time. Carrell et al. (2011) randomize the start time of 1st year college students (aged<20) and find that starting the school day later in the morning (7:50am), as opposed to early in the morning (7am), has a significant positive effect on their achievement. Dills and Hernández-Julián (2008), Edwards (2012), Heissel and Norris (2017), Cotti et al. (2018), Shin (2018) and Williams and Shapiro (2018) reach the same conclusion. The explanation they provide is in line with the standard theories of the circadian rhythm, namely that peak performance for teenagers occurs later in the morning. Pope (2016) investigates at what time students learn more and finds that, likely due to increase fatigue, students learn more in classes scheduled in the morning (8am - 10am) rather than in the afternoon (12:50pm - 2:45pm). We differentiate from this literature because we study the effects of time-of-day of an exam (the task itself), rather than the scheduling of classes, school start time or learning, and we focus on older students.[4]

Finally, we relate to the circadian rhythm literature. Circadian rhythms (or circadian clocks) are internal cycles that repeat *circa* every 24 hours, and regulate many biological processes, including peak and through wakefulness and cognitive functioning (see e.g. Cardinali (2008) for a review).

---

[4]Somewhat related, Pope and Fillmore (2015), Bensnes and Strom (2019) and Goulas and Megalokonomou (2020), study the effects of exam scheduling on student performance. However, these articles investigate different measures of across-days scheduling, whereas we are interested in the effects of time-of-day on performances.

In population, there are significant inter-individual differences in timing of behavior (sleep-wake cycles), denoted chronotypes (Roenneberg et al., 2007a), and timing of optimal (cognitive) performance (Schmidt et al., 2007).[5] In controlled laboratory experiments, medical psychologist have shown that cognitive abilities are susceptible to time-of-day and chronotype effects (Goldstein et al., 2007). However, this is not necessarily informative about whether, and to what extent, time-of-day actually matters for cognitive task performances in real-world settings.[6] Indeed, outside of the laboratory, individuals may adopt a variety of behavioral strategies that could compensate for their performances on cognitive tasks that are not scheduled at optimal time-of-day. Moreover, all these studies are based on small samples of adolescents, hence they provide little information about the population of workers we wish to speak to. We contribute to this literature by providing an ideal setting to study the link between time-of-day and cognitive task performances.

3 Background, Data, and Identification

This Section is organized as follows. First, we provide information about the institutional setting, the allocation rules of exam scheduling and test the conditional independence assumption (CIA) underlying our approach. Next, we describe the sample selection and the data we use. Finally, we explain the conceptual framework.

3.1 Institutional Setting and Conditional Random Assignment

We use administrative data covering the population of students enrolled in one of the largest public Universities in the UK. Data span five academic years, from 2014/15 to 2018/19. In line with the UK higher education system, the University offers undergraduate (UG), postgraduate taught (PGT), and postgraduate research (PGR) programs. Students enrolled in taught programs (either UG or PGT) must choose a combination of (previously approved) compulsory and elective modules in order to qualify for an academic degree. While the overall assessment of certain modules solely depends upon the final exam, the assessment of other modules may also depend upon the mark(s) students achieved in different assignments preceding the final exam (such as coursework, problem

---

[5]For example, some individuals may reach peak cognitive function in the morning (early type), whereas others are more alert in the evening (late type).

[6]A few papers in biology look at chronotype effects in a school setting. Generally, they show that students with a late type obtain significant lower academic achievements (see e.g. van der Vinne et al. (2015) and Itzek-Greulich et al. (2016); see also Zerbini and Merrow (2017) for a review).

sets, presentations, group activities, etc).[7]

In our context, exams are split in two main exam sessions, Fall and Spring, and an additional resit session in the Summer. The Fall session is held in January and it is approximately 12 days long. The Spring session begins towards the end of May and continues in June, lasting approximately 15 days. Students who fail one or more exams can resit the exam(s) at the end of August.[8] Within each main session, exams take place between Monday to Saturday, and are scheduled in one of three different time slots, which we label: "Morning", at 9am; "Lunch", at 1.30pm; and "Afternoon", at 4.30pm. Importantly, the University Student Services office is in charge of the schedules of the exams. By means of a software, called "Exam Scheduler", exams are simultaneously scheduled into different timeslots and available locations. The office takes students availability into account (to prevent exam clashes) and follows a set of clearly specified rules to ensure the integrity of the process. Three of these rules are crucial for our identification strategy: (i) Student timetables depend on the number of exams that a student has to take in a given exam period, the size of the cohort taking the exam, whether a School has requested a particular date/time for the exam,[9] and the duration of the exam;[10] otherwise the timeslot is randomized; (ii) Chief examiners cannot modify the schedule of their exams, unless they obtain a formal approval after submitting a motivated request to the office;[11] (iii) Exam schedules are published on-line for both students and Professors a few weeks prior to the exam session.[12]

Hence, our CIA entails that, conditional on the allocation rules used by the University's exam office, time-of-day of the exam is as good as randomly assigned and, consequently, orthogonal to a student's characteristics, as well as to a Professor's characteristics. To test this assumption, we regress each treatment on five pre-treatment characteristics of the students, together with the set of conditions defining the allocation rule. Specifically, the variables used as condition are: number of exams per session, class size, School fixed effects (FE), and exam weight (in credits). We use the

---

[7]Note that, as it is common in the UK system, for first year UG exams, final marks do not count towards the degree. In Section 4.2 we test the sensitivity of our findings after accounting for first year UG students.

[8]The resit session is also intended for students who, for a variety of reasons (e.g. personal, medical, religious, etc), either could not take the exams in the scheduled date or believed their exam performance was negatively affected by "extenuating circumstances".

[9]The office informed us that, from time to time, Schools do request specific times for their exams, but "it is not often that we can accommodate them, so this is not always a factor when scheduling examinations." Hence, we consider it as a minor condition.

[10]The specific condition is that, if an exam is longer than 2.5 hours, it is scheduled in the morning, otherwise the timeslot is randomized. In our sample, only 8 percent of the exams are longer than 2.5 hours. Hence, we consider also this condition as minor.

[11]Unfortunately, we do not have information about which exams have been re-scheduled following these formal requests. However, we know from the administrators at the exam office that this number is negligible.

[12]The other rules are: (iv) Exams longer than 2.5 hours can be scheduled in the afternoon slot only if the exam is scheduled on a Saturday; (v) The examination of modules taught both in the main and in satellites University campuses must be scheduled in conjunction to ensure the integrity of the exam; (vi) Similarly, the schedules must be in conjunction with students with alternative examination arrangements, who are required to take the exam in a different location (e.g. smaller room, computer room, etc); (vii) Students should have at least 24 hours between consecutive exams.

latter because our data on duration are limited to the last 2 academic years of the sample and exam weight is highly correlated with exam duration (over 90 percent), which allows us to run the test on the entire sample. The treatments are defined as a dummy for whether the exam was taken at 9am (Morning), at 1.30pm (Lunch) and at 4.30pm (Afternoon). The pre-treatment variables are the main personal characteristics of the students available in our dataset, including also their previous year's Grade point average (GPA).[13] All regressions include also the conditions. We perform these balancing tests for each of the 9 final samples we use for the analysis below.

We report the results in Table 1. There are 135 parameters estimated and only 23 are significant. Concerning Student GPA prior to the exams, only 5 out of 27 parameters are significant and none of them is economically meaningful. Concerning the other characteristics, only the dummy Asian exhibits a pattern, as it seems that Asian students are slightly less likely to take exams in the morning. We investigate this fact further. The cross-correlation Table A4 in Appendix A.1 shows that there is some correlation between time-of-day of exams from one year to another. If Asian students tend to enrol in courses with exams that happen to be more systematically scheduled later in the day, this observed heterogeneity might explain the slight imbalance.[14] In Table A2 in Appendix A.1, we perform the same balancing tests as above controlling also for exam FE. As one can see, virtually all coefficients on the dummy Asian become zero.[15] This indicates that the mild selection that we observe in the Table above is likely not driven by unobserved characteristics. Notice that, in order to further consolidate the validity of our causal analysis, in section 4 we test the robustness of our findings by estimating a specification including also exam FE. In another robustness check, we also consider the possibility that the allocation process is not completely (conditionally) random and assess the degree of omitted variable bias by implementing the coefficient stability test developed by Altonji et al. (2005) and Oster (2019). All main results will hold.[16] Thus, based on these tests, and the way the timetable is implemented, we believe that the allocation of students to time-of-day of exams is quasi-random, conditional on the characteristics used by the exam office.

---

[13]By computing their previous year's GPA, we perform the balancing tests on smaller samples than the ones used in estimation. In Table A1 in Appendix A.1, we include also the balancing tests without this variable, which allows us to validate the quasi-randomness of time-of-day on the actual samples that we use in estimation.

[14]This could happen because, systematically, specific exams exhibit the same characteristics every year (duration, class size, etc), which are taken into account by the software when the exam office is scheduling the timetable of the exams.

[15]As before, by computing student GPA prior to the exam session, we perform the balancing tests on a smaller samples than the ones used in estimation. In Table A3 in Appendix A.1, we include also the balancing tests without this variable and exam FE, which allows us to validate the quasi-randomness of time-of-day on the actual samples that we use in estimation.

[16]In Appendix A.2, we also estimate the same models using exam duration in place of exam weight (in credits). Our data on duration are limited to the last 2 academic years of the sample, hence this test is performed using only this subset of the data. Tables A5, A6, A7 and A8 report similar findings.

Table 1: Balancing Tests of Student Characteristics by Treatment Status

	Morning Exams			Lunch Exams			Afternoon Exams		
	(1) Pooled Sessions	(2) Fall Session	(3) Spring Session	(4) Pooled Sessions	(5) Fall Session	(6) Spring Session	(7) Pooled Sessions	(8) Fall Session	(9) Pooled Session
Panel A: Full-Sample									
Student GPA (t-1)	0.005	0.005	0.004	-0.005	-0.007	-0.003	0.001	0.002	-0.001
	(0.003)	(0.005)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)
Female	-0.007*	-0.007	-0.007	0.003	0.005	0.002	0.004	0.002	0.006*
	(0.004)	(0.005)	(0.005)	(0.003)	(0.005)	(0.005)	(0.003)	(0.004)	(0.004)
Black	0.002	0.010	-0.007	0.001	0.000	0.002	-0.003	-0.010	0.005
	(0.006)	(0.008)	(0.007)	(0.005)	(0.007)	(0.007)	(0.005)	(0.007)	(0.005)
Asian	-0.025**	-0.045***	-0.004	0.023**	0.021	0.023*	0.001	0.023*	-0.019*
	(0.011)	(0.016)	(0.014)	(0.010)	(0.014)	(0.013)	(0.009)	(0.014)	(0.010)
White	0.003	0.005	-0.001	-0.004	0.001	-0.007	0.001	-0.005	0.008*
	(0.004)	(0.007)	(0.006)	(0.004)	(0.006)	(0.005)	(0.004)	(0.006)	(0.004)
Observations:	205,494	94,340	111,154	205,494	94,340	111,154	205,494	94,340	111,154
Adjusted R^2:	0.125	0.093	0.214	0.069	0.055	0.129	0.061	0.086	0.087
Panel B: STEM									
Student GPA (t-1)	0.001	0.004	-0.001	-0.003	-0.002	-0.003	0.002	-0.003	0.003
	(0.004)	(0.006)	(0.005)	(0.004)	(0.006)	(0.005)	(0.003)	(0.004)	(0.004)
Female	-0.010*	-0.016**	-0.005	0.005	0.009	0.002	0.005	0.007	0.004
	(0.005)	(0.008)	(0.007)	(0.005)	(0.007)	(0.006)	(0.004)	(0.006)	(0.005)
Black	0.006	0.018	-0.013	-0.006	-0.017	0.007	0.000	-0.001	0.006
	(0.008)	(0.011)	(0.010)	(0.008)	(0.011)	(0.009)	(0.006)	(0.009)	(0.008)
Asian	-0.015	-0.032	0.003	0.025*	0.018	0.028	-0.010	0.014	-0.031**
	(0.016)	(0.022)	(0.020)	(0.015)	(0.021)	(0.020)	(0.013)	(0.019)	(0.016)
White	0.002	0.004	-0.003	-0.008	-0.006	-0.008	0.006	0.002	0.011*
	(0.007)	(0.010)	(0.008)	(0.006)	(0.009)	(0.008)	(0.005)	(0.008)	(0.006)
Observations:	107,214	48,569	58,645	107,214	48,569	58,645	107,214	48,569	58,645
Adjusted R^2:	0.127	0.090	0.229	0.080	0.049	0.152	0.082	0.085	0.109
Panel C: NON-STEM									
Student GPA (t-1)	0.014**	0.008	0.012*	-0.010**	-0.014**	-0.004	-0.003	0.006	-0.008*
	(0.005)	(0.008)	(0.007)	(0.005)	(0.006)	(0.006)	(0.004)	(0.007)	(0.005)
Female	-0.004	0.001	-0.011*	0.002	0.001	0.004	0.002	-0.002	0.006
	(0.005)	(0.007)	(0.006)	(0.005)	(0.006)	(0.007)	(0.004)	(0.006)	(0.005)
Black	-0.004	-0.000	-0.005	0.008	0.014	-0.001	-0.003	-0.014	0.006
	(0.007)	(0.010)	(0.009)	(0.007)	(0.009)	(0.009)	(0.006)	(0.009)	(0.006)
Asian	-0.032**	-0.052***	-0.015	0.015	0.013	0.017	0.018	0.039**	-0.002
	(0.014)	(0.020)	(0.017)	(0.012)	(0.018)	(0.016)	(0.012)	(0.020)	(0.012)
White	0.006	0.005	0.004	-0.002	0.007	-0.007	-0.004	-0.012	0.002
	(0.005)	(0.008)	(0.007)	(0.005)	(0.007)	(0.007)	(0.005)	(0.008)	(0.006)
Observations:	98,280	45,771	52,509	98,280	45,771	52,509	98,280	45,771	52,509
Adjusted R^2:	0.145	0.105	0.236	0.067	0.068	0.121	0.058	0.087	0.093

Notes: Dependent variables: dummy for whether the exam was taken at 9am (Morning Exam), dummy if it was taken at 1.30pm (Lunch Exam), dummy if it was taken at 4.30pm (Afternoon Exam). Standard errors are clustered by exam-year. Observations are at the student-exam-year level. In each specification, from Column (1)-(9), we control only for the conditions and year FE: number of exams per session fixed effects (FE), class size, School FE and exam weight (in credits) FE. We report only the coefficients on the main variables of interest for the balancing tests. In Panel A, we report estimates of the full sample. In Panel B, we report estimates for STEM schools only. In Panel C, we report estimates for Non-STEM schools only. Furthermore, in Column (1), (4) and (7), we report the results on the pooled sessions. In Column (2), (5) and (8), we report the results for the Fall exam session only. In Column (3), (6) and (9), we report the results for the Spring exam session only. *p < 0.10, **p < 0.05, ***p < 0.01.

3.2 Sample Selection and Data Description

The sample selection proceeds in three steps. First, we include only exams of taught programs (UG and PGT), and particularly exams including also a final assessment. Second, we focus on the two main sessions, Fall and Spring, because the August resit session may be systematically different (also, it accounts for only 0.06 percent of the sample). Third, we focus on exam scheduled within officials sessions and time-slots (Morning, Lunch, Afternoon) (only 2.31 percent of the sample is scheduled off session or at different time slots).[17] The final sample covers 43,913 UG and 7,642 PGT full-time students, over five academic years. The are 503,358 student-exam observations and 7,669 exam-year clusters.

Our data contains the full profile of a student's academic performance, including exam marks separated from the specific marks obtained in the various assignments (if any) contributing to the final mark. Additionally, the dataset contains specific information about scheduling, namely: the date, time, and room of each exam that took place in the University. Finally, the data includes several student-level characteristics (such as gender, age, ethnicity, etc) to use as control variables. By linking student-level exam marks with exam schedules, our dataset is uniquely suited to investigate time-of-day effects on student performance.

In Table 2, we report summary statistics of the main variables of interest. The outcome variable is "Exam mark," which captures a student's cognitive performance on the day of the exam.[18] In line with the UK higher education system, exam marks range between 0 and 100.[19] The average exam mark is around 60 (median value 62),[20] the average mark students achieve in their assignment(s) is around 69, and the average final mark (which is a weighted average of exam mark and assignment mark) is roughly 62 (median value 63).[21] Furthermore, the table reports that 45 percent of the exams are scheduled in the morning, around 34 percent are scheduled at lunch time, and around 21 percent are scheduled in the afternoon.[22] The vast majority of exams, 48 percent, is 2-hours

---

[17]Results do not change if these observations are included and are available upon request.

[18]We use a student's "Exam mark," rather than her "Final mark," as outcome variable because it allows us to observe the performance of young adults engaged in a complex cognitive task, with is a precise and objective measure of performance. Additionally, from a labor economics perspective, exams marks are of particular relevance as they involve a series of skills which are directly transferable to a workplace, including the ability to work under pressure, time-management, and the ability (either written or verbal) to efficiently communicate an idea. As mentioned in Section 2, several papers use exam mark as a measure of cognitive task performance.

[19]For quality control, exam marks are always double-marked internally and a sample is sent to an external examiner. Moreover, the scores obtained are typically translated into distinct degree classifications according to whether the student is UG or PGT.

[20]Throughout the analysis, however, we follow the literature and use the standardized exam mark achieved by students. This is obtained by normalizing the exam mark by its standard deviation.

[21]The summary statistics of assignment(s) marks refer to the sample of exams which do include assignments as part of their coursework.

[22]Note that an exam is more likely to be allocated to a morning slot because it allows the conjugate examination of modules taught in different satellites University campuses.

Table 2: Descriptive Statistics

	Mean	S.D.	Min	Max
Key Variables:				
Exam mark	60.32	16.66	0	100
Assignments mark	68.96	15.24	0.5	100
Final mark	62.02	14.10	0	100
Morning exam [0,1]	0.45	0.50	0	1
Lunch exam [0,1]	0.34	0.47	0	1
Afternoon exam [0,1]	0.21	0.41	0	1
Exam Characteristics:				
1-hour Exam [0,1]	0.14	0.35	0	1
2-hours Exam [0,1]	0.48	0.50	0	1
2.5-hours+ Exam [0,1]	0.08	0.26	0	1
10 Credits Exam [0,1]	0.47	0.50	0	1
20 Credits Exam [0,1]	0.41	0.49	0	1
30 Credits Exam [0,1]	0.03	0.16	0	1
30+ Credits Exam [0,1]	0.02	0.14	0	1
Number of Assignments per Exam	0.99	1.05	0	11
Class size	151.79	103.34	1	602
Student Characteristics:				
Age	20.78	2.35	17	64
Female [0,1]	0.47	0.50	0	1
Black [0,1]	0.07	0.25	0	1
Asian [0,1]	0.11	0.31	0	1
White [0,1]	0.63	0.48	0	1
Undergraduates [0,1]	0.93	0.25	0	1
First Year UG [0,1]	0.37	0.48	0	1
STEM [0,1]	0.50	0.50	0	1
Number of exams per year	6.32	2.12	1	12
Spring exam session [0,1]	0.55	0.50	0	1
Number of exams per session	3.70	1.40	1	9
Total credits per year	91.25	24.49	10	200
Total credits per session	56.39	28.02	10	400
Academic Year:				
2014/15	0.19	0.40	0	1
2015/16	0.19	0.40	0	1
2016/17	0.20	0.40	0	1
2017/18	0.19	0.39	0	1
2018/19	0.23	0.42	0	1
Clusters	7,667			
Observations	503,358			

Notes: The Table reports the mean, standard deviation, minimum and maximum values of the main variables used in the paper. There is a total of 503,358 student-exam level observations, covering 51,555 students. Of these, 224,933 (278,425) are in the Fall (Spring), and 251,691 (251,667) are from STEM (Non-STEM) Schools. Statistics on assignments mark are for the sample of exams which do include assignments as part of their coursework. Statistics for exam duration are limited to the last 2 academic years of the sample.

long. In terms of exam weight, around 47 and 41 percent of the exams are worth 10 and 20 credits, respectively. The average age of the students is 21. Around 47 percent of the students are females. Whereas 64, 11, and 7 percent of the students are White, Asian, and Black, respectively, and the remainder consists of other ethnicities. The vast majority of the sample consists of UG students (93 percent), and around 50 percent of the students are enrolled in a STEM school. Additionally, students take an average of 6 (3) exams per year (session), which amounts to 91 (56) credits in total.

3.3   Conceptual framework

Traditionally, economists investigating the productivity of working hours study the relationship between output and efficient labor input (Hamermesh, 1996; Pencavel, 2015). Within this literature, efficiency is a function of total hours of work. In our setting, however, individuals perform only one cognitive task per day, which is scheduled at a specific time-of-day. Hence, we model each worker's production function as follows:

$$Y = f(W(T), B(T), \boldsymbol{\theta}) + \epsilon \tag{1}$$

where Y is a measure of cognitive task performance, T is the time of the task, $W(T)$ is a set of worker's behaviors which are relevant for the performance (e.g. effort, work strategy, sleep patterns, food intake) and are affected by time-of-day,[23] $B(T)$ is a measure of cognitive functioning, which follows the circadian rhythm, $\boldsymbol{\theta}$ is a vector of task and worker's characteristics, and ϵ is the error term.

In practice, since $W(T)$ and $B(T)$ are unobserved, our research may be interpreted as an investigation of the overall effect of time-of-day on cognitive performance, which is the policy relevant variable. That is, we estimate:

$$\frac{\partial Y}{\partial T} = \frac{\partial Y}{\partial B}\frac{\partial B}{\partial T} + \frac{\partial Y}{\partial W}\frac{\partial W}{\partial T} \tag{2}$$

which may be different from the partial effect of each component. Equation (2) allows to better understand the underlying hypothesis behind our investigation. Depending on the time of the task, a worker will try to do her best to maximize the output. Hence, in principle, it is not obvious that

---

[23]For example, a worker may choose a specific work strategy depending on the time-of-day a task is scheduled. Similarly, a worker may employ a particular food intake strategy for a morning taks, while she may choose a different strategy had the task being schedule at lunchtime or in the afternoon.

the effect of T on Y is significant because, even if a task is not scheduled at an optimal time-of-day, a worker may adopt a variety of behavioral strategies that could compensate for her performances. One could think of this behavior as "adaptation". If instead we observe a positive and significant effect, we interpret this as a sign of efficiency gain (with respect to the counterfactual). That is, even if a task is anticipated, a worker finds a certain time-of-day more optimal than others to perform a cognitive task.

Further about the mechanisms, a natural hypothesis behind Equation (2) is that the two components, $W(T)$ and $B(T)$, are substitute to each other. When $B(T)$ is high, because at time T the brain is more efficient with respect to the baseline, then $W(T)$ is low, because the worker does not have to exert so much effort. Whereas, if $B(T)$ is low, the worker needs to compensate the lowering of brain capacity, hence $W(T)$ has to be high in order to boost the performance. Hence, under mild Assumptions (outlined in Appendix A.4), we cannot identify separately the effect of each component. Nevertheless, since they both have a positive effect on productivity, but they are negatively correlated with each other, the time-of-day effect that we estimate is a lower bound of $B(T)$. We can think of it as a standard attenuation bias.

Given Equation (1), the linear specification we estimate is simply:

$$Y_{i,e,t} = \alpha + T'_{e,t}\beta + X'_{i,e,t}\gamma + \alpha_i + \upsilon_{i,e,t}, \tag{3}$$

where $Y_{i,e,t}$ is the standardized mark achieved by student i, in exam e, in academic year t. The term $T_{e,t}$ is a vector of exam-specific time-of-day variables, namely: A dummy equal to 1 if an exam is at lunch, and 0 otherwise, and a dummy equal to 1 if an exam is in the afternoon, and 0 otherwise. Hence, the main parameters of interest, denoted by the vector β, capture the average treatment effect (ATE) of time-of-day for students taking a lunch or afternoon exam relatively to a morning exam, which is used as control group. The vector of covariates $X_{i,e,t}$ contains the set of conditions for the validity of the conditional random assignment as well as a large set of student and exams' characteristics that might affect exam marks.[24] Finally, the error term $\upsilon_{i,e,t}$ captures common unobservable shocks. In our research design, the quasi-randomization of students to time-of-day ensures that $T'_{e,t}$ is not correlated with $\upsilon_{i,e,t}$ in Equation (3), conditional on the vector $X_{i,e,t}$ variables that

---

[24]This includes: exam weight (in credits), class size, students' age, gender, ethnicity, nationality, a dummy for whether the student is under-graduate, a dummy for whether the exam belongs to STEM schools, academic year FE, year of enrollment FE, stage of study FE, school FE, the number of assignments and the average assignments mark, the number of exam credits, and the total number of credits the students achieved in a particular examination session and in the year.

were used by the University exam office to schedule the exams.[25] To enhance efficiency, we estimate the model for the pooled sample of students but allow the level of the outcome to differ between students by including individual FE, α_i. This captures students' time-invariant characteristics, both observable and unobservable, that might affect their performance, including family background, motivation and ability.[26]

4 Empirical Analysis

This section is organized as follows. First, we estimate the time-of-day effects on student performances using Equation (3). Next, we look at whether, and to what extent, time-of-day effects vary differentially by season, school, gender and student ability. Furthermore, we use a series of robustness checks and falsification tests to provide further support to the main results. Standard errors are always clustered by exam-year.[27] Finally, we provide a simple back of the envelope calculation for what the estimates imply for an important context that is likely to benefit from our research: sorting of elective surgeries.

4.1 Main Results

Table 3 reports the first set of estimates. In Column (1), we regress the standardized exam mark only on the main variables of interest, $T_{e,t}$, using the Pooled sample (Fall and Spring sessions together). The average effect of a lunch (afternoon) exam is 0.046 (0.050) percentage of standard deviation (SD) higher than a morning exam. In Column (2), we include also the conditions (number of exams per session fixed effects (FE), class size, School FE and exam weight (in credits) FE), which allow us to satisfy the CIA. The estimated coefficients remain relatively stable at 0.051 (0.042) SD for a lunch (afternoon) exam. In Column (3), we add the remaining control variables, $X_{i,e,t}$, as well as student fixed effects (FE), α_i, which might affect a student performance. The average effect remains positive, 0.068 (0.036) SD, for taking a lunch (afternoon) exam. The latter is our preferred specification.

We also perform a test of equality of Lunch and Afternoon exams, which is rejected at 5% level.

---

[25]In Appendix A.4, we point out the exact identifying assumptions underlying our strategy.

[26]Individual FE are not needed for identification but we include them in order to increase the precision of the estimates. In Table A12 in Appendix A.3 we report all the results also without these FE.

[27]Results do not change when using a different clustering strategy and are available upon request.

We use (i) the latter result and (ii) the fact that the parameter attached to Lunch is consistently larger (in magnitude) than the parameter attached to Afternoon in all specifications estimated in this paper, to argue that we detect an inverse-U shape relationship between time-of-day and cognitive performance.

Next, in Columns (4), (5) and (6), we report the same set of estimates for the Fall session only, and in Columns (7), (8) and (9), for the Spring session only, using our preferred specification. Time-of-day effects are significantly stronger in the Fall examination session. Indeed, looking at Column (6), having a lunch (afternoon) exam increases performance by 0.094 (0.076) SD with respect to a morning exam. Finally, when we look at the Spring session, the results confirm that students perform better at lunch time, although the point estimate is less than half the effect in the Fall. Whereas there is no significant effect for afternoon exams. Overall, the results in this table show positive effects of time-of-day on student performance, suggesting significant gains in scores for taking a lunch or an afternoon exam.

Table 3: Effects of Time-of-Day on Student Performance: Full Sample

	Pooled Sessions			Fall Session			Spring Session		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Lunch [0,1]	0.046***	0.051***	0.068***	0.042	0.075***	0.094***	0.049**	0.028	0.043**
	(0.017)	(0.016)	(0.015)	(0.026)	(0.025)	(0.024)	(0.022)	(0.020)	(0.019)
Afternoon [0,1]	0.050***	0.042**	0.036**	0.054**	0.072***	0.076***	0.049**	0.008	-0.010
	(0.018)	(0.017)	(0.015)	(0.026)	(0.024)	(0.023)	(0.024)	(0.022)	(0.021)
Test of equality of Lunch and Afternoon exams in linear regressions above:									
$Pr > F$	0.807	0.619	0.037	0.679	0.900	0.463	0.989	0.370	0.008
Conditions		✓	✓		✓	✓		✓	✓
Controls			✓			✓			✓
Student FE			✓			✓			✓
Clusters:	7,663	7,663	7,624	3,486	3,486	3,434	4,351	4,351	4,284
Observations:	503,358	503,358	500,958	224,933	224,933	219,236	278,414	278,414	274,677
Adjusted R^2:	0.000	0.022	0.462	0.000	0.025	0.457	0.001	0.029	0.499

Notes: Dependent variable: standardized final exam mark. Key variables: dummies for whether the exam was taken at 1.30pm (Lunch Exam) or at 4.30pm (Afternoon Exam), where the 9am (Morning Exam) is the excluded category. Standard errors are clustered by exam-year. Observations are at the student-exam-year level. In Column (1), (4) and (7), we regress the dependent variables on the key variables only. In Column (2), (5) and (8), we control also for the conditions. In Column (3), (6) and (9), we add also the rest of the covariates and student FE, which allow us to control for all time-invariant observable and unobservable students' characteristics. The latter represent our preferred specifications. Furthermore, in Column (1), (2) and (3), we report the results on the pooled sessions. In Column (4), (5) and (6), we report the results for the Fall exam session only. In Column (7), (8) and (9), we report the results for the Spring exam session only. Below the test of equality of Lunch and Afternoon exams is reported the p-value of the test statistic. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

In order to better understand the time-of-day effects on performance, in Table 4 we continue our analysis by separating the results between students in STEM versus non-STEM schools. In Columns (1), (2), and (3), we report the estimates for STEM exams, for the Pooled, Fall, and

Spring examination sessions, respectively. Similarly, in Columns (4), (5), and (6), we report the estimates for non-STEM exams, for the Pooled, Fall, and Spring examination sessions, respectively. The estimates in this Table indicate that time-of-day matters almost exclusively for STEM exams. The parameters in Column (1) imply an increase of 0.081 (0.046) SD in exam mark when having a lunch (afternoon) exam as compared to a morning exam. As before, also these estimates are significantly stronger in the Fall session. Looking at Column (2), once we focus on exams taken in the Fall session, the estimated coefficients almost double in magnitude, suggesting that, by moving a morning to a lunch (afternoon) exam, a student increases her performance by 0.139 (0.093) SD. As Columns (4), (5), and (6) highlight, time-of-day appears to matter only marginally, if at all, for non-STEM exams.

Table 4: Effects of Time-of-Day on Student Performance: By School

	STEM			NON-STEM		
	(1) Pooled Sessions	(2) Fall Session	(3) Spring Session	(4) Pooled Sessions	(5) Fall Session	(6) Spring Session
Lunch [0,1]	0.081***	0.139***	0.030	0.037**	0.016	0.038*
	(0.022)	(0.035)	(0.028)	(0.017)	(0.028)	(0.022)
Afternoon [0,1]	0.046*	0.093***	-0.034	0.010	0.028	-0.007
	(0.025)	(0.035)	(0.034)	(0.017)	(0.027)	(0.022)
Test of equality of Lunch and Afternoon exams in linear regressions above:						
$Pr > F$	0.165	0.210	0.047	0.140	0.670	0.048
Clusters:	3,465	1,639	1,886	4,143	1,781	2,384
Observations:	250,361	109,498	138,422	248,198	107,729	134,316
Adjusted R^2:	0.482	0.483	0.515	0.450	0.444	0.495

Notes: Dependent variable: standardized final exam mark. Key variables: dummies for whether the exam was taken at 1.30pm (Lunch Exam) or at 4.30pm (Afternoon Exam), where the 9am (Morning Exam) is the excluded category. Standard errors are clustered by exam-year. Observations are at the student-exam-year level. In each Column (1)-(6) we use the preferred specification which controls for student FE and all covariates. In Column (1) and (4) we report the results for the pooled sessions, split between STEM vs Non-STEM exams. In Column (2) and (5) we report the results for the Fall exam session only, split between STEM vs Non-STEM exams. In Column (3) and (6) we report the results for the Spring exam session only, split between STEM vs Non-STEM exams. Below the test of equality of Lunch and Afternoon exams is reported the p-value of the test statistic. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Three final remarks are in order. First, we investigate whether, and to what extent, the observed time-of-day effects differ by gender. We do so by including in model (3) an interaction term between gender and time-of-day dummies. The results are reported in Table A9 in Appendix A.3 and suggest no systematic differences between the two subgroups.[28] Second, we also investigate, extensively,

---

[28]This result is consistent with the circadian rhythm literature. Indeed, according to this body of work, differences in circadian rhythms between men and women are starkest during teenage-hood, but start to decrease around the age of 20 (Roenneberg et al., 2004), and completely disappear around the age of 50, which coincides with the average age at menopause (Hollandera et al., 2001).

whether, and to what extent, the observed time-of-day effects differ by students ability. Table A10 in Appendix A.3 reports the results of one such specification where we include in model (3) an interaction term between ability and time-of-day dummies, where High Ability $= \mathbb{1}(\text{Assigment mark} \geq 70)$. As one can see, the coefficients are not significant.[29] Finally, by reporting the main results in Figure 1, one can grasp the form of this inverse-U shape relationship between time-of-day and performances.

Figure 1: Time-of-Day and Performances
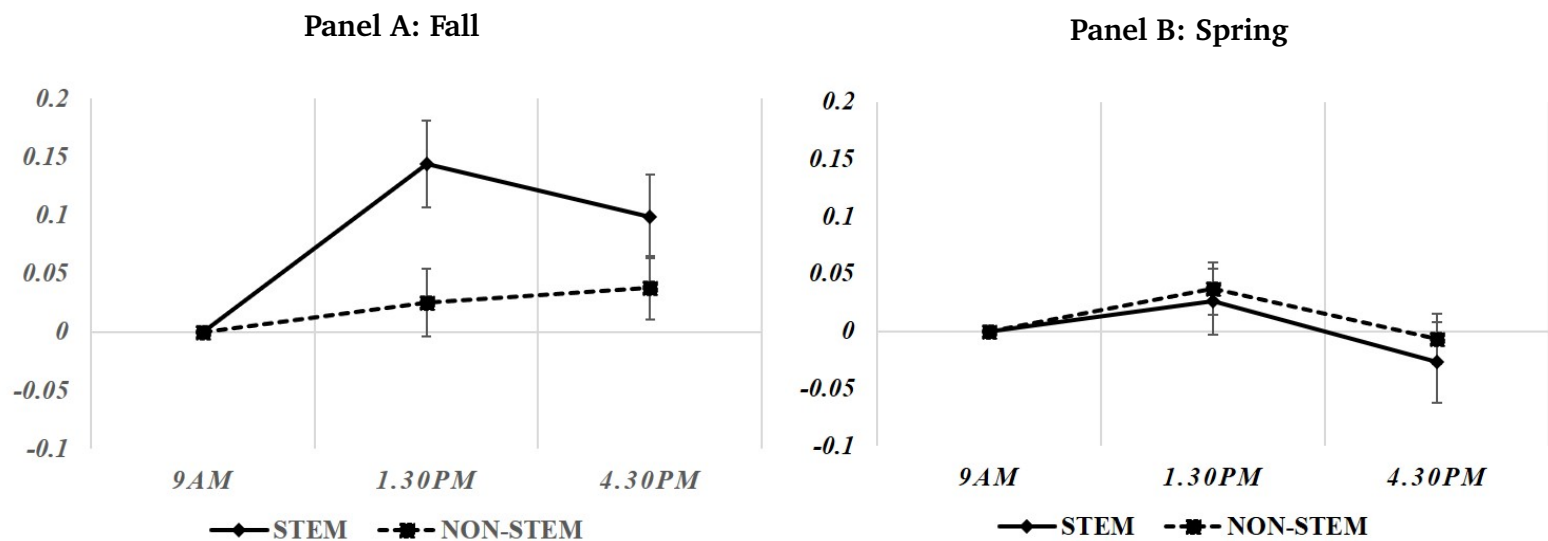
Notes: The Figures provide a graphical representation of the main findings of the paper. In Panel A, we report the estimated coefficients, with associated standard error, for the sample of students enrolled in the Fall exam session (January), by school. Exams taken at 9am represents the control group. Similarly, in Panel B, we report the estimated coefficients, and associated standard errors, for the sample of students in the Spring session (May-June), by school. One can grasp this inverse-U shape relationship between time-of-day and performances.

4.2 Robustness Checks and Falsification Tests

In order to explore the sensitivity of our results, we exploit at best the richness of our data and perform several robustness checks. We discuss them in turn and report the results in Table 5. Panel A provides the benchmark comparison.

First of all, the regression model (3) satisfies the CIA if we can control for exam duration. As we mentioned before, since we have information about exam duration only for the last 2 years of the sample, we estimated the model using a proxy for it, namely exam weight FE. This is potentially a limitation and could cause concerns. As a robustness check, we take exam duration from the last 2 years and impute duration data for the same exams that are also present in the first 3 years of the

---

[29]We explored several variations of this specification, including (i) a quantile regression framework and (ii) a version of Equation (3) in which the outcome variable is replaced by a dummy variable for each degree classification, namely distinction (exam mark ≥ 70), merit/credit (exam mark [60-70]), pass (exam mark [50-60] for PGs, and [40-60] for UGs) and fail (exam mark <50 for PGs, and exam mark <40 for UGs). The qualitative conclusion does not change and the results are available upon request.

sample.[30] We then estimate the same regression model including a dummy variable taking value 1 if the exam is longer than 2.5 hours, and 0 otherwise. Panel B shows that our results are robust to this inclusion. Along the same line, and in light of the mild imbalance detected in Section 3.1, we re-estimate model (3) using exam FE (without student FE). This allows us to control for time-invariant observable and unobservable attributes of the exams, such as the systematic characteristics that exams have, which are taken into account by the exam office in scheduling the timetable of the exams. Panel C shows that, also in this case, results do not change. Results using student FE and exam FE are also very similar (and available upon request).

We wish to stress that, albeit the results are consistent throughout, from an identification viewpoint, the specification with exam FE is not our preferred one. This is because exam FE control for much more than simply the average time-invariant attributes of the exams that allow us to satisfy the CIA. They also control for the average difficulty in the content of an exam. As we argue at length in Section 4.3, time-of-day effects are important for tasks involving fluid intelligence (working memory, logic thinking, problem solving, and abstract reasoning), rather than crystallized intelligence (knowledge and vocabulary). Hence, by introducing exam FE, we absorb part of the underlying time-of-day effects.[31]

Continuing the investigation, our ideal experiment entails the comparison of students taking different exams at different time-of-day. Hence, one may be concerned that, by comparing a student taking different exams, at different point in time of the exam session, we might be picking up other confounding effects besides time-of-day, including accumulated cognitive fatigue effects, warm-up effects, day-specific and seasonality effects. Accordingly, in Panel D, we include in model (3) a set of days FE in order to take into account these potential confounders. Also in this case, results do not change. Finally, one may be concerned that the environment in which the students take their exams might play an important role in the observed results. Once again, we exploit the richness of our data and, in Panel E, we include exam room FE in order to control for potential room-specific effects that might affect a student performance during the exams, including cold/heat, light/dark, size, air quality, seat disposition, number of invigilators, etc. Results are again robust.

Furthermore, one may be concern about the differential impacts time-of-day might have on first

---

[30]It is worth highlighting again that only 8% of the exams in the sample are longer than 2.5 hours. The large majority of these are first year UG students. This means that, very likely, these are compulsory exams which are regularly offered every year. Hence, the underlying (implicit) assumption of this robustness check is that, if an exam is longer than 2.5 hours in a given year, very likely it was longer than 2.5 hours also a few years earlier.

[31]Notice that exam FE may account also for Professor teaching style, and their leniency or strictness in exam marking. Given the similarity of the results when accounting for these FE, we conclude that such unobserved characteristics are not relevant in explaining our results.

	Full-Sample			STEM			NON-STEM		
	(1) Pooled Sessions	(2) Fall Session	(3) Spring Session	(4) Pooled Sessions	(5) Fall Session	(6) Spring Session	(7) Pooled Sessions	(8) Fall Session	(9) Spring Session
Panel A: Benchmark									
Lunch [0,1]	0.068***	0.094***	0.043**	0.081***	0.139***	0.030	0.037**	0.016	0.038*
	(0.015)	(0.024)	(0.019)	(0.022)	(0.035)	(0.028)	(0.017)	(0.028)	(0.022)
Afternoon [0,1]	0.036**	0.076***	-0.010	0.046*	0.093***	-0.034	0.010	0.028	-0.007
	(0.015)	(0.023)	(0.021)	(0.025)	(0.035)	(0.034)	(0.017)	(0.027)	(0.022)
Test of equality of Lunch and Afternoon exams in linear regressions above:									
$Pr > F$	0.037	0.463	0.008	0.165	0.210	0.047	0.140	0.670	0.048
Clusters:	7,624	3,434	4,284	3,465	1,639	1,886	4,143	1,781	2,384
Observations:	500,958	219,236	274,677	250,361	109,498	138,422	248,198	107,729	134,316
Adjusted R^2:	0.462	0.457	0.499	0.482	0.483	0.515	0.450	0.444	0.495
Panel B: Duration									
Lunch [0,1]	0.061***	0.085***	0.048**	0.068***	0.126***	0.047	0.040**	0.027	0.032
	(0.017)	(0.025)	(0.022)	(0.025)	(0.036)	(0.033)	(0.019)	(0.029)	(0.024)
Afternoon [0,1]	0.026	0.058**	-0.013	0.020	0.041	-0.034	0.015	0.046	-0.021
	(0.017)	(0.023)	(0.024)	(0.028)	(0.035)	(0.039)	(0.019)	(0.029)	(0.025)
Test of equality of Lunch and Afternoon exams in linear regressions above:									
$Pr > F$	0.035	0.253	0.006	0.089	0.020	0.033	0.196	0.503	0.031
Clusters:	6,469	2,886	3,668	2,900	1,372	1,578	3,550	1,498	2,082
Observations:	432,138	184,387	239,616	207,796	88,609	115,723	222,076	93,864	121,983
Adjusted R^2:	0.471	0.473	0.506	0.493	0.505	0.524	0.458	0.455	0.499
Panel C: Exam FE									
Lunch [0,1]	0.027**	0.057***	-0.009	0.042**	0.098***	-0.017	0.002	0.016	-0.006
	(0.013)	(0.019)	(0.016)	(0.020)	(0.028)	(0.026)	(0.014)	(0.024)	(0.017)
Afternoon [0,1]	0.013	0.051**	-0.027	0.014	0.076**	-0.045	0.015	0.035	-0.008
	(0.015)	(0.021)	(0.019)	(0.023)	(0.032)	(0.031)	(0.016)	(0.025)	(0.021)
Test of equality of Lunch and Afternoon exams in linear regressions above:									
$Pr > F$	0.302	0.782	0.320	0.189	0.464	0.320	0.403	0.450	0.912
Clusters:	7610	3458	4317	3462	1645	1891	4148	1813	2426
Observations:	503305	224905	278380	251682	111802	139864	251623	113103	138516
Adjusted R^2:	0.261	0.257	0.275	0.266	0.270	0.276	0.268	0.246	0.293
Panel D: Day FE									
Lunch [0,1]	0.070***	0.092***	0.046**	0.087***	0.141***	0.047*	0.041**	0.014	0.045**
	(0.014)	(0.023)	(0.018)	(0.022)	(0.033)	(0.028)	(0.016)	(0.027)	(0.019)
Afternoon [0,1]	0.034**	0.080***	-0.019	0.047*	0.100***	-0.037	0.008	0.028	-0.015
	(0.015)	(0.022)	(0.020)	(0.025)	(0.035)	(0.033)	(0.017)	(0.026)	(0.021)
Test of equality of Lunch and Afternoon exams in linear regressions above:									
$Pr > F$	0.022	0.624	0.001	0.120	0.269	0.011	0.062	0.621	0.006
Clusters:	7,624	3,434	4,284	3,465	1,639	1,886	4,143	1,781	2,384
Observations:	500,958	219,236	274,677	250,361	109,498	138,422	248,198	107,729	134,316
Adjusted R^2:	0.466	0.461	0.504	0.490	0.488	0.526	0.456	0.450	0.502
Panel E: Room FE									
Lunch [0,1]	0.068***	0.083***	0.041**	0.098***	0.126***	0.044	0.025	0.025	0.025
	(0.015)	(0.023)	(0.019)	(0.022)	(0.033)	(0.028)	(0.018)	(0.028)	(0.023)
Afternoon [0,1]	0.034**	0.074***	-0.003	0.052**	0.078**	-0.025	0.007	0.028	-0.005
	(0.015)	(0.021)	(0.021)	(0.025)	(0.033)	(0.034)	(0.017)	(0.025)	(0.023)
Test of equality of Lunch and Afternoon exams in linear regressions above:									
$Pr > F$	0.034	0.679	0.033	0.070	0.194	0.036	0.323	0.927	0.216
Clusters:	7,611	3,432	4,274	3,462	1,639	1,884	4,130	1,779	2,371
Observations:	500,920	219,215	274,648	250,345	109,486	138,406	248,161	107,713	134,282
Adjusted R^2:	0.473	0.469	0.513	0.500	0.503	0.536	0.460	0.456	0.512

Notes: Dependent variable: standardized final exam mark. Key variables: dummies for whether the exam was taken at 1.30pm (Lunch Exam) or at 4.30pm (Afternoon Exam), where the 9am (Morning Exam) is the excluded category. Standard errors are clustered by exam-year. Observations are at the student-exam-year level. In each Column (1)-(9) we use the preferred specification which controls for student FE and all covariates. In Columns (1)-(3) we report estimates of the full sample, while in Columns (4)-(6) and Columns (7)-(9) we report results for STEM and NON-STEM schools, separately. Specifically, in Columns (1), (4), and (7) are reported the results for the pooled sample. In Columns (2), (5), and (8) we report the results for the Fall exam session only. In Columns (3), (6), and (9) we report the results for the Spring exam session only. Panel (A) presents the benchmark results. Panel (B) report the results when including in the regression a dummy for exams longer than 2 hours. Finally, In Panel (C), (D) and (E) we include, respectively, exam, day, and room FE. Below the test of equality of Lunch and Afternoon exams is reported the *p*-value of the test statistic. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

year UG students. This is because, in line with the UK higher educational system, the exam marks for these students do not count towards their final degree. Hence, this sub-group of individuals, which accounts for 20% of the observations, might be intrinsically different from the rest of the sample because they do not respond to the same set of incentives. Moreover, first year UG students may not be accustomed with University exam procedures, or with the rhythm of study and preparation, and hence they may lack experience. Ex ante, it is not obvious whether their (potential) heterogeneous effects would push upward or downward the ATE. In order to check that this sub-group is not affecting the results in a different manner, we include in model (3) a set of interaction terms with the time-of-day dummies. Table A11 in Appendix A.3 presents the results which suggest that, indeed, first year UG students might behave (slight) differently from other students, although the interaction terms are hardly ever significant.[32] Even if we cannot exclude that alternative mechanisms play a role for first year UG students, the main findings of our paper, which are summarized in Panel A of Table 5, remain consistent throughout the specifications.

Another concern is that the different results between Fall and Spring may be driven by other confounders occurring in May-June but not in January. For example, Bensnes (2016) finds a connection between pollen exposure and reduction in exam performance. If a student taking an exam around lunchtime or afternoon in the Spring is more affected by day-time pollen exposure, then pollen levels should be controlled for. We conduct a robustness analysis including daily city-level pollen data obtained from the British Meteorological Office. The results are identical to the benchmark in Panel A of Table 5 and are omitted.

We also consider the possibility that the allocation process is not completely (conditionally) random. This could happen, for example, in case students' choose their exams according to time-of-day preferences and can lead to correlation between students' unobserved characteristics and the number of exams taken at specific time-of-day. Hence, we assess the degree of omitted variable bias in our estimated effects of time-of-day on cognitive performances by implementing the coefficient stability test developed by Altonji et al. (2005) and Oster (2019). The intuition of this method is that, in order to evaluate the degree of omitted variable bias, one should test both the movements of the main coefficients after the inclusion of the observed controls and also the movements in R-square. Table A13 in Appendix A.3 presents the validation results for the analysis of the estimated effects

---

[32]Specifically, after including these interaction terms, the ATE of scheduling change (slightly) downward in the Fall and (slightly) upward in the Spring. For example, in Column (2) of Table A11, the effect of taking a lunch exam in the Fall is 0.081 SD, whereas in Column (6) of Table 3, it was 0.094 SD. Vice versa, in Column (3) of Table A11, the effect of taking a lunch exam in the Spring is 0.060 SD, whereas in Column (9) of Table 3, it was 0.043 SD. The same dynamics retain for STEM and Non-STEM exams.

of time-of-day on student performances. As one can see, the bounding sets, under the most restrictive scenario of the test, are all very close to the estimated coefficients. The results of these tests suggest that the omitted variable bias from the possible correlation between students' unobserved preferences and time-of-day is not a concern for us.

To conclude this Section, we propose a set of falsification tests to further strengthen the validity of our findings. The main intuition behind this analysis is to show that our estimated effects do not exist when, in fact, they should not. Recall that exam time is unknown to students (and Professors) until a few weeks prior the examination period. Therefore, in principle, students cannot adjust their (pre-final exam) assignment performance on the basis of their final exam scheduling. Hence, we exploit the richness of our data and regress the average assignment marks on final exam time. We exclude from the analysis students without any assignment mark and perform the estimation on each of the 9 sub-samples used in the previous Section. The regression model is:

$$\bar{a}_{i,e,t} = \mu_i + T'_{e,t}\rho + X'_{i,e,t}\pi + v_{i,e,t}. \tag{4}$$

where $\bar{a}_{i,e,t}$ is the average assignment mark for student i, in exam e, in academic year t. The rest of the covariates are exactly the same as in our preferred specification. Table 6 outlines the results. As one can see, almost all the estimated coefficients are close to zero or statistically insignificant. This is an important result because it shows that students do not attempt to manipulate or substitute the exam time with an adjustment of their within-semester assignment performance.

4.3   Explanations

While we have found a causal link between time-of-day of an exam and performances, it is also important to understand why such relationship would exist. Looking at the economics literature, three potential mechanisms may explain our results. First of all, fatigue, unrest, stress, warm-up and the length of a work shift, are common factors used to explain patterns of productivity during the day (e.g. Hamermesh, 1990, 1996). However, in our research design, individuals perform only one cognitive task per day. Hence, these mechanisms are likely to play a minor role in explaining differences in performance between students taking exams scheduled at different time-of-day. Second, one could argue that exams in the morning are systematically more difficult than those later in the day, which could induce different behaviors. However, in the previous sections we showed that

Table 6: Falsification Test

	Full Sample			STEM			NON-STEM		
	(1) Pooled Sessions	(2) Fall Session	(3) Spring Session	(4) Pooled Sessions	(5) Fall Session	(6) Spring Session	(7) Pooled Sessions	(8) Fall Session	(9) Spring Session
Lunch [0,1]	-0.011	0.002	-0.017*	-0.008	0.014	-0.011	-0.004	0.001	-0.009
	(0.008)	(0.013)	(0.010)	(0.013)	(0.018)	(0.016)	(0.008)	(0.016)	(0.008)
Afternoon [0,1]	0.001	-0.001	0.001	-0.001	0.001	-0.011	0.006	-0.004	0.022**
	(0.009)	(0.014)	(0.013)	(0.016)	(0.022)	(0.022)	(0.008)	(0.013)	(0.010)
Clusters:	7624	3434	4284	3465	1639	1886	4143	1781	2384
Observations:	500958	219236	274677	250361	109498	138422	248198	107729	134316
Adjusted R^2:	0.890	0.896	0.887	0.859	0.876	0.852	0.928	0.930	0.929

Notes: Dependent variable: standardized assignment mark. Key variables: dummies for whether the exam was taken at 1.30pm (Lunch Exam) or at 4.30pm (Afternoon Exam), where the 9am (Morning Exam) is the excluded category. Standard errors are clustered by exam-year. Observations are at the student-exam-year level. In each Column (1)-(9) we use the preferred specification which controls for student FE and all covariates. In Columns (1)-(3) we report estimates of the full sample, while in Columns (4)-(6) and Columns (7)-(9) we report results for STEM and NON-STEM schools, separately. Specifically, in Columns (1), (4) and (7), we report the results for the pooled sample, split between Full vs STEM vs Non-STEM exams. In Columns (2), (5) and (8), we report the results for the Fall exam session only, split between Full vs STEM vs Non-STEM exams. In Columns (3), (6) and (9), we report the results for the Spring exam session only, split between Full vs STEM vs Non-STEM exams. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. .

this is not likely to be a valid explanation.[33] Third, one could argue that the observed effects may be driven by how long students have been awake. That is, if you take an exam in the morning, you might be affected by "sleep inertia",[34] as compared to taking an exam later in the day. Although we do not have information about sleep patterns, this explanation would be potentially relevant if one could make the case that students taking an exam at 9am are tested within a very small window after waking up (the first half an hour), which is unlikely in our case. Together with the rest of the robustness checks that we performed in the previous section, we conclude that this literature does not seem to be equipped to explain our findings.

We turn instead to the circadian rhythm literature. As mentioned before, our internal biological clock regulates almost each aspect of human life, including cognition. In a given population, peak sleep and wake times show a near-Gaussian distribution, with extreme early types waking up when extreme late types fall asleep (Cardinali, 2008). Three main findings of this literature help us rationalize all our results at once.

Firstly, age is one of the main determinants of a chronotype (Roenneberg et al., 2004).[35] Children are generally earlier chronotypes, progressively delaying their clock during development, reaching a maximum of their "lateness" at around the age of 20, and then advancing their clock with

---

[33]Furthermore, if this was true, we should also be able to explain why morning exams are more difficult only in the Fall, and not in the Spring, and only for STEM students.

[34]Sleep inertia refers to the transitional state between sleep and wake, which is marked by impaired performance if we wake up too early and we are still in our biological night.

[35]Toh et al. (2001), Vink et al. (2001) and Archer et al. (2003) show that chronotypes depend on genetic factors and Roenneberg et al. (2003) on environmental factors.

increasing age. Therefore, from a biological point of view, it is expected that, if an exam is scheduled early in the morning, the performance of the average (biologically speaking) student should be worse compared to a performance later in the day. Drawing from this literature, two potential mechanisms can explain the observed findings of our paper: circadian process and sleep deprivation (Schmidt et al., 2007). The former implies that the average performance is lower in morning exams because students are not tested at the peak of their cognitive abilities. The latter, instead, implies that the observed time-of-day effects may occur as a consequence of the potential sleep deprivation resulting from the asynchrony between being a late chronotype and the early wake-up required by a morning exam.[36]

Secondly, there are a host of external signals, often called *Zeitgebers*, which act as cues capable of synchronising this inner clock. Light exposure is among the most important factors (Duffy and Wright, 2005; Roenneberg et al., 2007b; Wright et al., 2013). Indeed, there is now compelling evidence that chronotypes are significantly affected by seasonal changes (Shawa et al., 2018). Hubert et al. (1998) first noted that chronotype was earlier in Summer than in Winter. This result was later confirmed in a thorough study by Kantermann et al. (2007) who show how timing of sleep follows the seasonal progression of dawn. Therefore, also in this case, from a biological point of view it is expected to observe differences in time-of-day effects between the Fall and Spring exam sessions: in the Fall session, the average student appears to be severely disrupted by an early morning exam due to lack of light exposure, while in the Spring session this effect may be mitigated by the earlier sunrise.

Thirdly, Goldstein et al. (2007) and Zerbini et al. (2017) provide evidence that time-of-day and chronotype effects vary systematically depending on the type of tasks individuals perform. Specifically, significant time-of-day effects were found for tasks involving fluid intelligence (working memory, logic thinking, problem solving, and abstract reasoning), but were not found on crystallized intelligence (knowledge and vocabulary). Our results echo these findings and suggest that tasks involving fluid intelligence benefit from later exam timing. Altogether, it seems that $B(T)$ is more likely to be the main driver, and the effect that we estimate is, as we acknowledged, likely to be a lower bound of the true value.

---

[36]Specifically, since the secretion of the sleep-inducing hormone melatonin will only occur late at night for young adults, an early wake-up to attend a morning exam will likely result in sleep loss and, consequently, a reduction in cognitive function (Krause et al., 2017).

4.4 Implications

The main implication of our study for economics is that individuals, workers and organizations around the World may obtain efficiency gains by sorting their tasks according to the pattern of the circadian rhythm. In general, tasks involving problem solving skills should be prioritized and moved to early afternoon at times of year when sunlight exposure is limited. Whereas the rest of the tasks should be moved to the beginning or the end of the work day.

Concerning our specific setting, sorting of tasks can be primarily implemented by schools to increase student outcomes. Several papers in the education literature have investigated the effectiveness of various interventions aimed at improving student outcomes. For example, by reducing class sizes (Bandiera et al., 2010), delaying school start time (Carrell et al., 2011), or introducing teacher performance evaluations (Taylor and Tyler, 2012). Many of these policies, although effective, can be quite expensive to implement. Our results suggest that Universities have an additional and simple tool they can use to help improve student performances, namely: to optimize the daily exam schedules.[37] This new policy may be implemented in a variety of ways, depending on the context. For example, since the scheduling of exams is centrally managed, moving STEM exams to the early afternoon and other exams to morning or late afternoon, could increase test scores. There are, however, constraints on how much Universities can implement this policy. The most important constraint is on the supply of classrooms at each given hour. However, this issue may be solved by a more careful optimization of the capacity constraints.

Concerning other workplace environments, sorting of tasks can be implemented to improve the productivity of workers and organizations. There is little doubt that innovations achieving this goal with little or no increase in costs are a goldmine for organizations and policymakers. This is because the adoption of superior ideas may lead to improvements in performance, higher markups for firms, or, as our opening example suggests, better health status of patients. The literature in personnel, labor economics and management science have long studied possible solutions to achieve efficiency gains. For example, via better compensation schemes (Lazear, 2000), better workplace practices, IT, human capital investments and innovations (Black and Lynch, 2001, 2004), team

---

[37]Note that we interpret the results as indicating higher productivity and argue that students could do better with a different exam scheduling. We do not interpret our results as higher learning or knowledge. Indeed, in our context, students have, on average, the same knowledge at the moment of the exam, because we have (conditional) random assignment. This point suggests one further implication. (Conditional) Random assignment exists in our specific framework, which makes is suitable to identify causal effects. But it does not exist in many other settings. In each of these settings, the non-random scheduling of the exams may produce welfare losses and unfairness: that is, some students may be penalized and have their knowledge measured more incorrectly than others. This, in turn, might have negative consequences on the labor market.

incentives (Hamilton et al., 2003), higher wages (Charness and Kuhn, 2007), limiting task juggling (Coviello et al., 2014), or promoting on-the-job trainings (Konings and Vanormelingen, 2015). Also in this case, our results suggest that scheduling of workers performing cognitive tasks should be re-arranged. The actual implementation of this new policy may take different forms, depending on the context. For example, Hospital boards should schedule risky surgical interventions early in the afternoon. Obviously, some interventions are urgent and require immediate action, regardless of the time-of-day. However, our suggestion can still be applied for those operations that can be scheduled in advance, which, according to Prin et al. (2018), represent around 90 percent of the total surgeries. The same recommendation would apply to CEOs, managers or consultants, concerning their (most mentally demanding) business meetings. Finally, also academic researchers can benefit from our recommendation.

4.5   Sorting of cognitive tasks: A numerical example

We conclude this section with a simple exercise. In order to understand the relevance of our find-ings, we apply our insights to one important context that could greatly benefit from our research: elective surgeries. These medical procedures are particularly relevant because they are also com-plex cognitive tasks that are scheduled in advance and performed by highly skilled workers. To do this, we draw from Ingraham et al. (2010), who collected data on elective surgeries between 2005-2008 for a sample of 168 hospitals across the United States. As reported in their study, a total of 406,174 elective surgeries were performed, and 3,243 people (0.8%) died due to complications following the surgery. Accordingly, we implement a back of the envelope calculation using our cen-tral estimate of 14 percent SD increase in productivity associated with a cognitive task performed at 1.30 pm relative to 9 am. The formula of the effect size is:

$$ES = (M_m - M_l)/SD_M$$

where M_m (M_l) is the mortality rate for an operation scheduled in the morning (lunch) and it is given by the number of people who died in the morning (lunch), D_m (D_l), divided by the number of operations occurred in the morning (lunch), N_m (N_l). Whereas SD_M is the standard deviation of the overall mortality rate. We do not know M_m and M_l, but we can calculate SD_M from the knowledge of the overall mortality rate, which is obtained pooling all elective surgeries occurred during this

period. Given our raw data, $SD_M \approx 0.089$.[38] We maintain three simplifying assumptions: (i) The effect size found in our paper applies to this context as well ($ES \approx 0.14$); (ii) A person dies from an elective surgery due to the misperformance of the surgeon; (iii) There is an equal split of the number of operations at each point in time (hence $N_m = N_l$). All these assumptions are arbitrary and can be relaxed depending on the data at hand.

Extrapolating from this relationship implies that, if we could have moved all the operations from morning to lunch, around 40 less people in the sample would have died.[39] Our findings therefore suggest that a simple innovation in task sorting of skilled workers may lead to potentially large efficiency gains. Of course, issues of external validity and maintained assumptions suggest that the exact result from this back of the envelope calculation should be viewed with considerable caution. Besides the assumption on the effect size, one issue with external validity is that the number of operations may not be equally distributed throughout the day. However, even if we assumed that only a smaller number of operations were scheduled in the morning, our proposed innovation would still lead to considerable welfare gains.

5   Conclusion

Although the time-of-day in which a cognitive task is performed may affect productivity, economists have paid little attention to this relationship. In our paper, we exploit a unique setting to provide causal evidence about this link. Specifically, in the context of University education, students (our workers) take different exams (their cognitive tasks) at the end of each semester, and their scores (productivity) generally make up a large portion of their final grade. The identification strategy relies on a clear assignment rule and on quasi-random assignment of the time-of-day of the exams. We find that the productivity profile of students during the day has an inverse U-shape relationship, where peak performance occurs around lunchtime (1.30pm), relative to morning (9am), and decreases again in late afternoon (4.30pm). Moreover, these time-of-day effects are much larger in January, with respect to May-June, and they are much larger for students taking STEM exams, with respect to Non-STEM. We support this causal relationship with a battery of robustness and falsification tests.

These results suggest that the time-of-day is a crucial determinant for the performance of stu-

---

[38]Let $E[M] = 0.08$, then $VAR[M] = E[M] * (1 - E[M]) = 0.008 * 0.992 = 0.009$. Hence $SD[M] = \sqrt{0.009} \approx 0.089$.

[39]$M_m - M_l = 0.14 * 0.089 \approx 0.012$. Hence, $D_m - D_l = (0.012 * 3243) \approx 40$.

dents. These gains in productivity may be caused by different mechanisms. Although we do not have information that allow us to identify, separately, the magnitude of each specific mechanism driving this relationship, our results can be rationalized by recent findings in the circadian rhythm literature. Since much of a student's future is determined by their academic performances, exams can be associated with cognitive tasks a worker has to complete in a work setting. In light of this, we conclude that efficiency gains can be obtained for workers and organizations in many sectors of the economy. The main policy implication is that tasks involving problem solving skills are more affected by time-of-day and should be moved to the early afternoon at times of year when sunlight exposure is more limited. Sorting of tasks can be primarily implemented by schools to increase student outcomes. Similar implications would apply also to surgeons, doctors, consultants, CEOs as well as academic researchers.

References

ALTONJI, J. G., T. E. ELDER, AND C. R. TABER (2005): "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 113, 151–184. [8], [20]

ARCHER, S. N., D. L. ROBILLIARD, D. J. SKENE, M. SMITS, A. WILLIAMS, J. ARENDT, AND M. VON SCHANTZ (2003): "A length polymorphism in the circadian clock gene Per3 is linked to delayed sleep phase syndrome and extreme diurnal preference," *Sleep*, 26, 413–415. [22]

BANDIERA, O., V. LARCINESE, AND I. RASUL (2010): "Heterogenous Class size effects: New evidence from a panel of university students," *The Economic Journal*, 120, 1365–1398. [24]

BENSNES, S. S. (2016): "You sneeze, you lose:: The impact of pollen exposure on cognitive performance during high-stakes high school exams," *Journal of Health Economics*, 49, 1 – 13. [20]

BENSNES, S. S. AND B. STROM (2019): "Earning or Learning? How Extending Closing Time in the Retail Sector Affects Youth Employment and Education," *Oxford Bulletin of Economics and Statistics*, 81, 299–327. [5]

BLACK, S. E. AND L. M. LYNCH (2001): "How to Compete: The Impact of Workplace Practices and Information Technology on Productivity," *The Review of Economics and Statistics*, 83, 434–445. [24]

——— (2004): "What's Driving the New Economy?: The Benefits of Workplace Innovation," *The Economic Journal*, 114, F97–F116. [24]

BRACHET, T., G. DAVID, AND A. M. DRECHSLER (2012): "The effect of shift structure on performance," *American Economic Journal: Applied Economics*, 4, 219–246. [2], [5]

BRADLEY, S. AND C. GREEN (2020): *The Economics of Education (Second Edition)*, Academic Press, second edition ed. [3]

CARDINALI, D. P. (2008): *Chronoeducation: How the biological clock influences the learning process*, Cambridge University Press, 110–126. [4], [5], [22]

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. [5], [24]

CHARNESS, G. AND P. KUHN (2007): "Does Pay Inequality Affect Worker Effort? Experimental Evidence," *Journal of Labor Economics*, 25, 693–723. [25]

COLLEWET, M. AND J. SAUERMANN (2017): "Working hours and productivity," *Labour Economics*, 47, 96–106. [5]

COTTI, C., J. GORDANIER, AND O. OZTURK (2018): "Class meeting frequency, start times, and academic performance," *Economics of Education Review*, 62, 12 – 15. [5]

COVIELLO, D., A. ICHINO, AND N. PERSICO (2014): "Time Allocation and Task Juggling," *American Economic Review*, 104, 609–23. [25]

DILLS, A. K. AND R. HERNÁNDEZ-JULIÁN (2008): "Course scheduling and academic performance," *Economics of Education Review*, 27, 646 – 654. [5]

DUFFY, J. AND K. WRIGHT (2005): "Entrainment of the Human Circadian System by Light," *Journal of biological rhythms*, 20, 326–38. [23]

EDWARDS, F. (2012): "Early to rise? The effect of daily start times on academic performance," *Economics of Education Review*, 31, 970 – 983. [5]

FLORENCE, P. S. (1924): *The Economics of Fatigue and Unrest*, (London: George Allen Unwin. [4]

FOLKARD, S. AND P. TUCKER (2003): "Shift work, safety and productivity," *Occupational Medicine*, 53, 95–101. [4]

GHISLETTA, P. AND T. LECERF (2018): *Crystallized and Fluid Intelligence*, Oxford Bibliographies. [3]

GOLDSTEIN, D., C. S. HAHN, L. HASHER, U. WIPRZYCKA, AND P. D. ZELAZO (2007): "Time of day, intellectual performance, and behavioural problems in morning versus evening type adolescents: is there a synchrony effect?" *Personality and Individual Differences*, 42, 431–440. [6], [23]

GOULAS, S. AND R. MEGALOKONOMOU (2020): "Marathon, Hurdling or Sprint? The Effects of Exam Scheduling on Academic Performance," *The B.E. Journal of Economic Analysis & Policy*. [5]

HAMERMESH, D. (1996): *Workdays, Workhours and Work Schedules: Evidence for the United States and Germany*, Kalamazoo, MI: W.E. Upjohn Institute for Employment Research. [4], [12], [21]

HAMERMESH, D. S. (1990): "Shirking or Productive Schmoozing: Wages and the Allocation of Time at Work," *ILR Review*, 43, 121–S–133–S. [4], [21]

——— (1999): "The Timing of Work Over Time," *The Economic Journal*, 109, 37–66. [4]

HAMILTON, B. H., J. A. NICKERSON, AND H. OWAN (2003): "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation," *Journal of Political Economy*, 111, 465–497. [25]

HANUSHEK, E. A. AND D. D. KIMKO (2000): "Schooling, Labor-Force Quality, and the Growth of Nations," *The American Economic Review*, 90, 1184–1208. [3]

HANUSHEK, E. A. AND L. WOESSMANN (2008): "The Role of Cognitive Skills in Economic Development," *Journal of Economic Growth*, 46, 607–668. [3]

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*. [5]

HIGUCHI, S., Y. LIU, T. YUASA, A. MAEDA, AND Y. MOTOHASHI (2000): "Diurnal Variation in the P300 Component of Human Cognitive Event Related Potential," *Chronobiology International*, 17, 669–678. [2]

HOLLANDERA, L. E., E. W. FREEMAN, M. D. SAMMELA, J. A. BERLINA, J. A. GRISSO, AND M. BATTISTINI (2001): "Sleep quality, estradiol levels, and behavioral factors in late reproductive age women," *Obstetrics and Gynecology*, 98, 391–397. [16]

HUBERT, M., M. DUMONT, AND J. PAQUET (1998): "Seasonal and Diurnal Patterns of Human Illumination Under Natural Conditions," *Chronobiology International*, 15, 59–70. [23]

INGRAHAM, A., M. COHEN, M. RAVAL, C. KO, AND A. NATHENS (2010): "Comparison of hospital performance in emergency versus elective general surgery operations at 198 hospitals," Tech. Rep. 6. [25]

ITZEK-GREULICH, H., C. RANDLER, AND C. VOLLMER (2016): "The interaction of chronotype and time of day in a science course: Adolescent evening types learn more and are more motivated in the afternoon," *Learning and Individual Differences*, 51, 189 – 198. [6]

KANTERMANN, T., M. JUDA, M. MERROW, AND T. ROENNEBERG (2007): "The Human Circadian Clock's Seasonal Adjustment Is Disrupted by Daylight Saving Time," *Current Biology*, 17, 1996 – 2000. [23]

KESTER, L. AND P. A. KIRSCHNER (2012): *Cognitive Tasks and Learning*, Boston, MA: Springer US, 619–622. [2]

KONINGS, J. AND S. VANORMELINGEN (2015): "The Impact of Training on Productivity and Wages: Firm-Level Evidence," *The Review of Economics and Statistics*, 97, 485–497. [25]

KRAUSE, A. J., E. B. SIMON, B. A. MANDER, S. M. GREER, J. M. SALETIN, A. N. GOLDSTEIN-PIEKARSKI, AND M. P. WALKER (2017): "The sleep-deprived human brain," *Nature Reviews Neuroscience*, 18, 404–418. [23]

LAZEAR, E. P. (2000): "Performance Pay and Productivity," *American Economic Review*, 90, 1346–1361. [24]

MURNANE, R. J., J. B. WILLETT, AND F. LEVY (1995): "The Growing Importance of Cognitive Skills in Wage Determination," *The Review of Economics and Statistics*, 77, 251–266. [3]

OSTER, E. (2019): "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, 37, 187–204. [8], [20]

PENCAVEL, J. (2015): "The Productivity of Working Hours," *Economic Journal*, 125, 2052–2076. [5], [12]

——— (2018): *Diminishing Returns at Work: The Consequences of Long Working Hours*, Oxford University Press. [2]

PHILIBERT, I. (2005): "Sleep Loss and Performance in Residents and Nonphysicians: A Meta-Analytic Examination," *Sleep*, 28, 1392–402. [4]

POPE, D. G. AND I. FILLMORE (2015): "The impact of time between cognitive tasks on performance: Evidence from advanced placement exams," *Economics of Education Review*, 48, 30 – 40. [5]

POPE, N. G. (2016): "How the Time of Day Effects Productivity: Evidence from School Schedules," *Review of Economics and Statistics*. [2], [5]

PRIN, M., J. GUGLIELMINOTTI, AND O. MTALIMANJA (2018): "Emergency-to-Elective Surgery Ratio: A Global Indicator of Access to Surgical Care," *World J Surg*, 42, 1971–1980. [25]

ROENNEBERG, T., T. KUEHNLE, M. JUDA, T. KANTERMANN, K. ALLEBRANDT, M. GORDIJN, AND M. MERROW (2007a): "Epidemiology of the human circadian clock," *Sleep Medicine Reviews*, 11, 429–438. [6]

ROENNEBERG, T., T. KUEHNLE, P. P. PRAMSTALLER, J. RICKEN, M. HAVEL, A. GUTH, AND M. MERROW (2004): "A marker for the end of adolescence," *Current Biology*, 14, 1038–1039. [3], [16], [22]

ROENNEBERG, T., C. KUMAR, AND M. MERROW (2007b): "The Human Circadian Clock Entrains to Sun Time," *Current biology : CB*, 17, R44–5. [23]

ROENNEBERG, T., A. WIRZ-JUSTICE, AND M. MERROW (2003): "Life between clocks: daily temporal patterns of human chronotypes." *Journal of biological rhythms*, 18 1, 80–90. [22]

SCHMIDT, C., F. COLETTE, C. CAJOCHEN, AND P. PEIGNEUX (2007): "A time to think: circadian rhythms in human cognition," *Cognitive Neuropsychology*, 24, 755–789. [6], [23]

SHAWA, N., D. E. RAE, AND L. C. RODEN (2018): "Impact of seasons on an individual's chronotype: Current perspectives," *Nature and Science of Sleep*, 10, 345–354. [23]

SHIN, J. (2018): "Sleep More, Study Less? The Impact of Delayed School Start Time on Sleep and Academic Performance," working paper. [5]

TAYLOR, E. S. AND J. H. TYLER (2012): "The Effect of Evaluation on Teacher Performance," *The American Economic Review*, 102, 3628–3651. [24]

TOH, K. L., C. R. JONES, Y. HE, E. J. EIDE, W. A. HINZ, D. M. VIRSHUP, L. J. PTÁČEK, AND Y. H. FU (2001): "An hPer2 phosphorylation site mutation in familial advanced sleep phase syndrome," *Science*, 291, 1040–1043. [22]

VAN DER VINNE, V., G. ZERBINI, A. SIERSEMA, A. PIEPER, M. MERROW, R. A. HUT, T. ROENNEBERG, AND T. KANTERMANN (2015): "Timing of Examinations Affects School Performance Differently in Early and Late Chronotypes," *Journal of Biological Rhythms*, 30, 53–60. [6]

VEASEY, S., R. ROSEN, B. BARZANSKY, I. ROSEN, AND J. OWENS (2002): "Sleep Loss and Fatigue in Residency TrainingA Reappraisal," *JAMA*, 288, 1116–1124. [4]

VINK, J. M., J. M. VINK, A. S. GROOT, G. A. KERKHOF, AND D. I. BOOMSMA (2001): "Genetic analysis of morningness and eveningness," *hronobiology International*, 18, 809–822. [22]

WEINGER, M. B. AND S. ANCOLI-ISRAEL (2002): "Sleep Deprivation and Clinical Performance," *JAMA*, 287, 955–957. [4]

WILLIAMS, K. M. AND T. M. SHAPIRO (2018): "Academic achievement across the day: Evidence from randomized class schedules," *Economics of Education Review*, 67, 158 – 170. [5]

WRIGHT, K., A. MCHILL, B. BIRKS, B. GRIFFIN, T. RUSTERHOLZ, AND E. CHINOY (2013): "Entrainment of the Human Circadian Clock to the Natural Light-Dark Cycle," *Current biology : CB*, 23. [23]

ZERBINI, G. AND M. MERROW (2017): "Time to learn: How chronotype impacts education," *PsyCh Journal*. [6]

ZERBINI, G., V. VAN DER VINNE, L. K. M. OTTO, T. KANTERMANN, P. W. KRIJNEN, T. ROENNEBERG, AND M. MERROW (2017): "Lower school performance in late chronotypes: underlying factors and mechanisms," *Scientific Reports*. [23]

A Appendix

A.1 Balancing Tests Using Exam Weight

Table A1: Balancing Tests, no Student GPA (t-1), no exam FE

	Morning Exams			Lunch Exams			Afternoon Exams		
	(1) Pooled Sessions	(2) Fall Session	(3) Spring Session	(4) Pooled Sessions	(5) Fall Session	(6) Spring Session	(7) Pooled Sessions	(8) Fall Session	(9) Pooled Session
Panel A: Full Sample									
Female	-0.002	-0.003	-0.002	0.005*	0.003	0.005	-0.002	-0.000	-0.004
	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.002)	(0.003)	(0.003)
Black	0.000	0.006	-0.005	-0.002	-0.005	0.001	0.002	-0.001	0.004
	(0.004)	(0.006)	(0.005)	(0.004)	(0.005)	(0.005)	(0.003)	(0.005)	(0.004)
Asian	-0.005	-0.007	0.005	0.001	0.001	-0.003	0.004	0.006	-0.001
	(0.010)	(0.014)	(0.014)	(0.010)	(0.014)	(0.014)	(0.007)	(0.011)	(0.008)
White	-0.000	0.006	-0.006	0.000	-0.000	0.001	0.000	-0.006	0.005
	(0.003)	(0.005)	(0.004)	(0.003)	(0.005)	(0.004)	(0.003)	(0.005)	(0.003)
Observations:	503358	224933	278414	503358	224933	278414	503358	224933	278414
Adjusted R^2:	0.089	0.088	0.147	0.054	0.060	0.086	0.064	0.059	0.089
Panel B: STEM									
Female	-0.007*	-0.013**	-0.003	0.006	0.007	0.005	0.001	0.007	-0.002
	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.003)	(0.004)	(0.003)
Black	0.005	0.016*	-0.005	-0.004	-0.012	0.002	-0.000	-0.003	0.002
	(0.006)	(0.009)	(0.007)	(0.005)	(0.008)	(0.007)	(0.005)	(0.007)	(0.005)
Asian	0.018*	0.010	0.026**	-0.021**	-0.017	-0.023**	0.003	0.007	-0.003
	(0.009)	(0.013)	(0.012)	(0.008)	(0.012)	(0.012)	(0.008)	(0.011)	(0.010)
White	-0.003	0.004	-0.009	0.002	0.006	-0.001	0.001	-0.010	0.010**
	(0.005)	(0.008)	(0.006)	(0.005)	(0.007)	(0.006)	(0.004)	(0.007)	(0.004)
Observations:	251691	111807	139873	251691	111807	139873	251691	111807	139873
Adjusted R^2:	0.087	0.068	0.150	0.054	0.056	0.079	0.068	0.065	0.091
Panel C: NON-STEM									
Female	0.003	0.006	-0.000	0.003	-0.002	0.005	-0.006*	-0.004	-0.005
	(0.004)	(0.006)	(0.005)	(0.004)	(0.006)	(0.005)	(0.003)	(0.005)	(0.004)
Black	-0.007	-0.004	-0.009	0.003	0.001	0.004	0.004	0.003	0.005
	(0.005)	(0.007)	(0.006)	(0.005)	(0.007)	(0.006)	(0.004)	(0.006)	(0.005)
Asian	-0.031	-0.026	-0.017	0.023	0.013	0.017	0.007	0.013	-0.000
	(0.019)	(0.026)	(0.026)	(0.019)	(0.025)	(0.026)	(0.012)	(0.020)	(0.014)
White	0.001	0.007	-0.002	-0.000	-0.007	0.003	-0.001	0.000	-0.001
	(0.004)	(0.006)	(0.005)	(0.004)	(0.005)	(0.005)	(0.003)	(0.005)	(0.004)
Observations:	251667	113126	138541	251667	113126	138541	251667	113126	138541
Adjusted R^2:	0.101	0.132	0.153	0.063	0.087	0.101	0.061	0.061	0.090

Notes: Dependent variables: dummy for whether the exam was taken at 9am (Morning Exam), dummy if it was taken at 1.30pm (Lunch Exam), dummy if it was taken at 4.30pm (Afternoon Exam). Standard errors are clustered by exam-year. Observations are at the student-exam-year level. In each specification, from Column (1)-(9), we control only for the conditions and year FE: number of exams per session fixed effects (FE), class size, School FE and exam duration. We report only the coefficients on the main variables of interest for the balancing tests. In Panel A, we report estimates of the full sample. In Panel B, we report estimates for STEM schools only. In Panel C, we report estimates for Non-STEM schools only. Furthermore, in Column (1), (4) and (7), we report the results on the pooled sample. In Column (2), (5) and (8), we report the results for the Fall exam session only. In Column (3), (6) and (9), we report the results for the Spring exam session only. *p < 0.10, **p < 0.05, ***p < 0.01.

Table A2: Balancing Tests, with Student GPA (t-1), with exam FE

	Morning Exams			Lunch Exams			Afternoon Exams		
	(1) Pooled Sessions	(2) Fall Session	(3) Spring Session	(4) Pooled Sessions	(5) Fall Session	(6) Spring Session	(7) Pooled Sessions	(8) Fall Session	(9) Pooled Session
Panel A: Full-Sample									
Student GPA (t-1)	-0.000	-0.000	-0.000	0.000	0.000	-0.000	0.000	0.000	0.001
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Female	0.000	-0.000	0.000	0.001	0.003	-0.001	-0.001	-0.003	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Black	-0.002	0.006	-0.008*	0.004	0.001	0.006	-0.002	-0.007	0.002
	(0.004)	(0.006)	(0.005)	(0.003)	(0.005)	(0.004)	(0.003)	(0.005)	(0.004)
Asian	-0.004	-0.008	0.001	0.005	0.009*	0.001	-0.001	-0.002	-0.001
	(0.004)	(0.006)	(0.005)	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)
White	0.002	0.001	0.002	-0.001	0.002	-0.004	-0.000	-0.003	0.001
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)
Observations:	205379	94289	111070	205379	94289	111070	205379	94289	111070
Adjusted R^2:	0.492	0.475	0.580	0.417	0.436	0.486	0.415	0.467	0.435
Panel B: STEM									
Student GPA (t-1)	-0.003	-0.004	-0.002	0.003	0.004	0.002	0.000	-0.000	0.000
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Female	0.001	-0.000	0.003	0.001	0.005	-0.002	-0.003	-0.005*	-0.001
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)
Black	0.001	0.007	-0.007	0.001	-0.008	0.010*	-0.001	0.000	-0.004
	(0.005)	(0.007)	(0.006)	(0.005)	(0.007)	(0.006)	(0.004)	(0.006)	(0.005)
Asian	-0.003	-0.006	0.001	0.002	0.004	-0.001	0.001	0.002	0.000
	(0.005)	(0.008)	(0.007)	(0.005)	(0.007)	(0.007)	(0.004)	(0.007)	(0.006)
White	0.004	-0.002	0.008*	-0.004	0.001	-0.007	-0.000	0.000	-0.002
	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)
Observations:	107157	48546	58604	107157	48546	58604	107157	48546	58604
Adjusted R^2:	0.483	0.456	0.556	0.405	0.397	0.467	0.411	0.442	0.422
Panel C: NON-STEM									
Student GPA (t-1)	0.007***	0.008***	0.006*	-0.005**	-0.006**	-0.005*	-0.001	-0.002	-0.000
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)
Female	-0.001	-0.001	-0.002	0.002	0.002	0.001	-0.001	-0.000	0.001
	(0.002)	(0.004)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
Black	-0.006	0.001	-0.009	0.007	0.008	0.002	-0.001	-0.009	0.007*
	(0.005)	(0.007)	(0.005)	(0.004)	(0.006)	(0.005)	(0.004)	(0.007)	(0.004)
Asian	-0.006	-0.014	0.001	0.005	0.010	-0.000	0.001	0.003	-0.001
	(0.006)	(0.010)	(0.007)	(0.005)	(0.007)	(0.007)	(0.005)	(0.008)	(0.007)
White	-0.002	-0.000	-0.004	0.001	0.002	-0.000	0.001	-0.002	0.005*
	(0.003)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)
Observations:	98222	45743	52466	98222	45743	52466	98222	45743	52466
Adjusted R^2:	0.507	0.506	0.615	0.435	0.485	0.514	0.434	0.506	0.467

Notes: Dependent variables: dummy for whether the exam was taken at 9am (Morning Exam), dummy if it was taken at 1.30pm (Lunch Exam), dummy if it was taken at 4.30pm (Afternoon Exam). Standard errors are clustered by exam-year. Observations are at the student-exam-year level. In each specification, from Column (1)-(9), we control only for the conditions, year FE and exam FE: number of exams per session fixed effects (FE), class size, School FE and exam weight (in credits) FE. We report only the coefficients on the main variables of interest for the balancing tests. In Panel A, we report estimates of the full sample. In Panel B, we report estimates for STEM schools only. In Panel C, we report estimates for Non-STEM schools only. Furthermore, in Column (1), (4) and (7), we report the results on the pooled sample. In Column (2), (5) and (8), we report the results for the Fall exam session only. In Column (3), (6) and (9), we report the results for the Spring exam session only. *p < 0.10, **p < 0.05, ***p < 0.01.

Table A3: Balancing Tests, no Student GPA (t-1), with exam FE

	Morning Exams			Lunch Exams			Afternoon Exams		
	(1) Pooled Sessions	(2) Fall Session	(3) Spring Session	(4) Pooled Sessions	(5) Fall Session	(6) Spring Session	(7) Pooled Sessions	(8) Fall Session	(9) Pooled Session
Panel A: Full-Sample									
Female	0.001	0.000	0.001	0.000	-0.001	0.001	-0.001	0.000	-0.002
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Black	-0.003	-0.000	-0.005*	0.003	0.001	0.005*	-0.000	-0.001	-0.000
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
Asian	-0.004	-0.004	-0.002	0.001	0.001	-0.000	0.003	0.003	0.003
	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)	(0.002)	(0.003)	(0.003)
White	0.001	0.004*	-0.002	-0.001	-0.002	0.001	-0.001	-0.002	0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Observations:	503305	224905	278380	503305	224905	278380	503305	224905	278380
Adjusted R^2:	0.469	0.451	0.542	0.391	0.396	0.448	0.369	0.400	0.403
Panel B: STEM									
Female	0.000	-0.001	0.001	0.002	0.002	0.001	-0.002	-0.000	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Black	-0.001	0.004	-0.006	0.001	-0.007	0.009**	0.000	0.002	-0.003
	(0.003)	(0.005)	(0.004)	(0.003)	(0.005)	(0.004)	(0.003)	(0.004)	(0.004)
Asian	-0.002	-0.002	-0.000	-0.003	-0.001	-0.002	0.004	0.004	0.003
	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.003)	(0.005)	(0.004)
White	0.003	0.002	0.003	-0.002	0.000	-0.002	-0.002	-0.003	-0.001
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)
Observations:	251682	111802	139864	251682	111802	139864	251682	111802	139864
Adjusted R^2:	0.457	0.427	0.532	0.373	0.372	0.422	0.367	0.387	0.393
Panel C: NON-STEM									
Female	0.001	0.001	0.000	-0.001	-0.002	0.000	0.000	0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Black	-0.005*	-0.003	-0.005	0.006**	0.006	0.003	-0.001	-0.003	0.002
	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)
Asian	-0.008**	-0.009*	-0.005	0.004	0.002	0.002	0.003	0.007	0.003
	(0.004)	(0.006)	(0.005)	(0.004)	(0.006)	(0.004)	(0.004)	(0.005)	(0.004)
White	-0.001	0.006**	-0.006***	0.001	-0.005*	0.005**	0.000	-0.002	0.002
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)
Observations:	251623	113103	138516	251623	113103	138516	251623	113103	138516
Adjusted R^2:	0.487	0.497	0.555	0.415	0.432	0.482	0.378	0.425	0.420

<u>Notes</u>: Dependent variables: dummy for whether the exam was taken at 9am (Morning Exam), dummy if it was taken at 1.30pm (Lunch Exam), dummy if it was taken at 4.30pm (Afternoon Exam). Standard errors are clustered by exam-year. Observations are at the student-exam-year level. In each specification, from Column (1)-(9), we control only for the conditions, year FE and exam FE: number of exams per session fixed effects (FE), class size, School FE and exam duration. We report only the coefficients on the main variables of interest for the balancing tests. In Panel A, we report estimates of the full sample. In Panel B, we report estimates for STEM schools only. In Panel C, we report estimates for Non-STEM schools only. Furthermore, in Column (1), (4) and (7), we report the results on the pooled sample. In Column (2), (5) and (8), we report the results for the Fall exam session only. In Column (3), (6) and (9), we report the results for the Spring exam session only. *p < 0.10, **p < 0.05, ***p < 0.01.

Table A4: Cross-correlation Between Exam Characteristics

	Morning exam	Lunch exam	Afternoon exam	Morning exam [t-1]	Lunch exam [t-1]	Afternoon exam [t-1]	STEM Exam	Fall Session	Student GPA [t-1]
Morning exam	1.000								
Lunch exam	-0.634	1.000							
	(0.000)								
Afternoon exam	-0.494	-0.359	1.000						
	(0.000)	(0.000)							
Morning exam [t-1]	0.239	-0.130	-0.143	1.000					
	(0.000)	(0.000)	(0.000)						
Lunch exam [t-1]	-0.148	0.108	0.057	-0.611	1.000				
	(0.000)	(0.000)	(0.000)	(0.000)					
Afternoon exam [t-1]	-0.118	0.034	0.105	-0.502	-0.377	1.000			
	(0.000)	(0.013)	(0.000)	(0.000)	(0.000)				
STEM	0.063	0.010	-0.087	0.065	-0.003	-0.072	1.000		
	(0.000)	(0.387)	(0.000)	(0.000)	(0.804)	(0.000)			
Fall exam	0.037	-0.049	0.011	0.015	-0.028	0.013	0.037	1.000	
	(0.001)	(0.000)	(0.351)	(0.283)	(0.042)	(0.332)	(0.001)		
Student GPA [t-1]	0.071	-0.031	-0.052	0.045	-0.021	-0.030	0.228	-0.042	1.000
	(0.000)	(0.026)	(0.000)	(0.001)	(0.135)	(0.034)	(0.000)	(0.002)	

Notes: Correlation between exam characteristics calculated at the exam-level.

A.2 Balancing Tests Using Duration

Table A5: Balancing Tests, with Student GPA (t-1), no exam FE

	Morning Exams			Lunch Exams			Afternoon Exams		
	(1) Pooled Sessions	(2) Fall Session	(3) Spring Session	(4) Pooled Sessions	(5) Fall Session	(6) Spring Session	(7) Pooled Sessions	(8) Fall Session	(9) Pooled Session
Panel A: Full Sample									
Student GPA (t-1)	0.002	0.006	0.000	-0.002	-0.006	0.001	0.000	-0.000	-0.001
	(0.003)	(0.005)	(0.004)	(0.003)	(0.005)	(0.004)	(0.003)	(0.004)	(0.003)
Female	-0.005	-0.005	-0.005	0.002	0.004	0.000	0.003	0.002	0.005
	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.003)	(0.004)	(0.004)
Black	0.002	0.008	-0.006	0.000	0.000	0.001	-0.002	-0.009	0.005
	(0.006)	(0.008)	(0.007)	(0.005)	(0.007)	(0.007)	(0.005)	(0.008)	(0.005)
Asian	-0.021**	-0.037**	-0.006	0.018*	0.009	0.024*	0.003	0.028*	-0.018*
	(0.011)	(0.016)	(0.013)	(0.010)	(0.014)	(0.013)	(0.009)	(0.015)	(0.010)
White	0.004	0.004	-0.002	-0.005	-0.001	-0.006	0.001	-0.003	0.007*
	(0.005)	(0.007)	(0.006)	(0.004)	(0.006)	(0.006)	(0.004)	(0.006)	(0.004)
Observations:	188990	86229	102761	188990	86229	102761	188990	86229	102761
Adjusted R^2:	0.171	0.109	0.290	0.090	0.061	0.172	0.070	0.094	0.102
Panel B: STEM									
Student GPA (t-1)	-0.002	0.004	-0.004	0.001	0.002	0.001	0.001	-0.006	0.003
	(0.004)	(0.007)	(0.005)	(0.004)	(0.006)	(0.005)	(0.003)	(0.004)	(0.004)
Female	-0.007	-0.014*	-0.001	0.001	0.006	-0.002	0.005	0.008	0.003
	(0.005)	(0.008)	(0.007)	(0.005)	(0.008)	(0.007)	(0.004)	(0.006)	(0.005)
Black	0.007	0.017	-0.012	-0.008	-0.019*	0.006	0.001	0.001	0.005
	(0.008)	(0.011)	(0.010)	(0.008)	(0.011)	(0.010)	(0.007)	(0.010)	(0.008)
Asian	-0.010	-0.016	-0.006	0.019	-0.003	0.036*	-0.009	0.020	-0.030*
	(0.016)	(0.023)	(0.021)	(0.015)	(0.021)	(0.021)	(0.014)	(0.020)	(0.016)
White	0.003	0.003	-0.004	-0.010	-0.011	-0.006	0.007	0.008	0.010
	(0.007)	(0.010)	(0.009)	(0.006)	(0.009)	(0.008)	(0.006)	(0.009)	(0.006)
Observations:	96192	43115	53077	96192	43115	53077	96192	43115	53077
Adjusted R^2:	0.164	0.124	0.272	0.100	0.058	0.186	0.094	0.105	0.128
Panel C: NON-STEM									
Student GPA (t-1)	0.013**	0.009	0.010	-0.010**	-0.015**	-0.002	-0.003	0.006	-0.008*
	(0.005)	(0.008)	(0.007)	(0.005)	(0.006)	(0.006)	(0.005)	(0.007)	(0.005)
Female	-0.005	0.002	-0.013**	0.004	0.002	0.007	0.001	-0.004	0.006
	(0.005)	(0.007)	(0.006)	(0.005)	(0.006)	(0.007)	(0.004)	(0.006)	(0.005)
Black	-0.003	-0.002	-0.001	0.006	0.015	-0.004	-0.003	-0.012	0.006
	(0.007)	(0.010)	(0.008)	(0.007)	(0.009)	(0.009)	(0.006)	(0.009)	(0.006)
Asian	-0.032**	-0.056***	-0.009	0.012	0.012	0.011	0.020*	0.045**	-0.002
	(0.013)	(0.020)	(0.015)	(0.012)	(0.018)	(0.015)	(0.012)	(0.020)	(0.012)
White	0.005	0.005	0.003	-0.002	0.006	-0.006	-0.004	-0.011	0.003
	(0.005)	(0.008)	(0.006)	(0.005)	(0.007)	(0.007)	(0.005)	(0.008)	(0.006)
Observations:	92798	43114	49684	92798	43114	49684	92798	43114	49684
Adjusted R^2:	0.191	0.101	0.339	0.088	0.068	0.177	0.065	0.090	0.105

Notes: Dependent variables: dummy for whether the exam was taken at 9am (Morning Exam), dummy if it was taken at 1.30pm (Lunch Exam), dummy if it was taken at 4.30pm (Afternoon Exam). Standard errors are clustered by exam-year. Observations are at the student-exam-year level. In each specification, from Column (1)-(9), we control only for the conditions and year FE: number of exams per session fixed effects (FE), class size, School FE and exam duration. We report only the coefficients on the main variables of interest for the balancing tests. In Panel A, we report estimates of the full sample. In Panel B, we report estimates for STEM schools only. In Panel C, we report estimates for Non-STEM schools only. Furthermore, in Column (1), (4) and (7), we report the results on the pooled sample. In Column (2), (5) and (8), we report the results for the Fall exam session only. In Column (3), (6) and (9), we report the results for the Spring exam session only. *p < 0.10, **p < 0.05, ***p < 0.01.

Table A6: Balancing Tests, no Student GPA (t-1), no exam FE

	Morning Exams			Lunch Exams			Afternoon Exams		
	(1) Pooled Sessions	(2) Fall Session	(3) Spring Session	(4) Pooled Sessions	(5) Fall Session	(6) Spring Session	(7) Pooled Sessions	(8) Fall Session	(9) Pooled Session
Panel A: Full Sample									
Female	-0.001	-0.004	0.001	0.004	0.004	0.003	-0.003	0.000	-0.004
	(0.003)	(0.005)	(0.004)	(0.003)	(0.005)	(0.004)	(0.002)	(0.004)	(0.003)
Black	0.002	0.008	-0.004	-0.003	-0.006	0.000	0.001	-0.002	0.004
	(0.004)	(0.006)	(0.005)	(0.004)	(0.006)	(0.005)	(0.003)	(0.005)	(0.004)
Asian	-0.004	-0.005	0.001	-0.001	-0.005	0.001	0.006	0.011	-0.002
	(0.011)	(0.016)	(0.015)	(0.011)	(0.015)	(0.015)	(0.008)	(0.012)	(0.009)
White	0.002	0.005	-0.003	-0.001	0.001	-0.001	-0.001	-0.006	0.004
	(0.004)	(0.005)	(0.004)	(0.003)	(0.005)	(0.004)	(0.003)	(0.005)	(0.003)
Observations:	434956	190567	244378	434956	190567	244378	434956	190567	244378
Adjusted R^2:	0.148	0.117	0.222	0.082	0.071	0.128	0.080	0.081	0.108
Panel B: STEM									
Female	-0.003	-0.013**	0.002	0.002	0.007	-0.001	0.001	0.006	-0.001
	(0.004)	(0.006)	(0.005)	(0.004)	(0.006)	(0.005)	(0.003)	(0.005)	(0.003)
Black	0.005	0.014	-0.005	-0.004	-0.010	0.001	-0.001	-0.004	0.003
	(0.006)	(0.009)	(0.008)	(0.006)	(0.008)	(0.008)	(0.005)	(0.008)	(0.006)
Asian	0.020**	0.023*	0.020	-0.027***	-0.035***	-0.017	0.006	0.011	-0.003
	(0.010)	(0.014)	(0.013)	(0.009)	(0.012)	(0.013)	(0.008)	(0.012)	(0.011)
White	-0.001	0.000	-0.006	-0.001	0.008	-0.004	0.001	-0.008	0.010**
	(0.006)	(0.008)	(0.006)	(0.005)	(0.008)	(0.006)	(0.005)	(0.007)	(0.005)
Observations:	209286	91315	117960	209286	91315	117960	209286	91315	117960
Adjusted R^2:	0.126	0.107	0.198	0.080	0.071	0.122	0.088	0.104	0.110
Panel C: NON-STEM									
Female	0.001	0.004	-0.002	0.004	-0.001	0.007	-0.005	-0.003	-0.006
	(0.004)	(0.006)	(0.006)	(0.004)	(0.006)	(0.006)	(0.003)	(0.005)	(0.004)
Black	-0.002	0.002	-0.005	-0.001	-0.002	-0.000	0.003	0.000	0.005
	(0.005)	(0.007)	(0.006)	(0.005)	(0.007)	(0.006)	(0.004)	(0.006)	(0.005)
Asian	-0.030	-0.037	-0.015	0.021	0.016	0.015	0.009	0.022	-0.001
	(0.020)	(0.028)	(0.026)	(0.020)	(0.027)	(0.026)	(0.013)	(0.021)	(0.015)
White	0.006	0.009	0.001	-0.003	-0.008	0.001	-0.003	-0.002	-0.002
	(0.004)	(0.006)	(0.005)	(0.004)	(0.006)	(0.005)	(0.004)	(0.005)	(0.004)
Observations:	225670	99252	126418	225670	99252	126418	225670	99252	126418
Adjusted R^2:	0.178	0.161	0.251	0.093	0.104	0.141	0.078	0.073	0.110

Notes: Dependent variables: dummy for whether the exam was taken at 9am (Morning Exam), dummy if it was taken at 1.30pm (Lunch Exam), dummy if it was taken at 4.30pm (Afternoon Exam). Standard errors are clustered by exam-year. Observations are at the student-exam-year level. In each specification, from Column (1)-(9), we control only for the conditions and year FE: number of exams per session fixed effects (FE), class size, School FE and exam duration. We report only the coefficients on the main variables of interest for the balancing tests. In Panel A, we report estimates of the full sample. In Panel B, we report estimates for STEM schools only. In Panel C, we report estimates for Non-STEM schools only. Furthermore, in Column (1), (4) and (7), we report the results on the pooled sample. In Column (2), (5) and (8), we report the results for the Fall exam session only. In Column (3), (6) and (9), we report the results for the Spring exam session only. *p < 0.10, **p < 0.05, ***p < 0.01.

Table A7: Balancing Tests, with Student GPA (t-1), with exam FE

	Morning Exams			Lunch Exams			Afternoon Exams		
	(1) Pooled Sessions	(2) Fall Session	(3) Spring Session	(4) Pooled Sessions	(5) Fall Session	(6) Spring Session	(7) Pooled Sessions	(8) Fall Session	(9) Pooled Session
Panel A: Full-Sample									
Student GPA (t-1)	0.000	0.000	0.000	0.000	0.000	-0.001	-0.000	-0.001	0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Female	0.000	-0.000	-0.000	0.002	0.003	0.000	-0.002	-0.002	0.000
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Black	-0.003	0.004	-0.008	0.004	0.002	0.006	-0.001	-0.006	0.002
	(0.004)	(0.006)	(0.005)	(0.004)	(0.005)	(0.005)	(0.003)	(0.005)	(0.004)
Asian	-0.005	-0.010	0.001	0.006	0.011**	-0.000	-0.001	-0.001	-0.001
	(0.004)	(0.007)	(0.005)	(0.004)	(0.005)	(0.006)	(0.004)	(0.006)	(0.005)
White	0.001	-0.001	0.003	-0.002	0.002	-0.004	0.000	-0.002	0.001
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)
Observations:	188929	86192	102717	188929	86192	102717	188929	86192	102717
Adjusted R^2:	0.468	0.441	0.568	0.384	0.389	0.471	0.381	0.433	0.404
Panel B: STEM									
Student GPA (t-1)	-0.003	-0.003	-0.002	0.003	0.005	0.002	-0.001	-0.001	-0.000
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Female	0.001	-0.001	0.002	0.002	0.005	-0.000	-0.003	-0.004	-0.002
	(0.002)	(0.003)	(0.003)	(0.002)	(0.004)	(0.003)	(0.002)	(0.003)	(0.003)
Black	-0.001	0.005	-0.006	0.001	-0.007	0.011*	-0.001	0.002	-0.005
	(0.005)	(0.008)	(0.007)	(0.005)	(0.007)	(0.006)	(0.004)	(0.007)	(0.005)
Asian	-0.003	-0.007	0.004	0.002	0.005	-0.003	0.000	0.002	-0.001
	(0.006)	(0.008)	(0.008)	(0.006)	(0.007)	(0.008)	(0.005)	(0.007)	(0.006)
White	0.004	-0.003	0.011**	-0.004	0.002	-0.008	0.000	0.002	-0.003
	(0.003)	(0.005)	(0.005)	(0.003)	(0.005)	(0.005)	(0.003)	(0.004)	(0.004)
Observations:	96173	43100	53066	96173	43100	53066	96173	43100	53066
Adjusted R^2:	0.449	0.405	0.539	0.359	0.331	0.444	0.375	0.400	0.397
Panel C: NON-STEM									
Student GPA (t-1)	0.007***	0.009***	0.006*	-0.005**	-0.007**	-0.005	-0.001	-0.002	-0.001
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
Female	-0.001	-0.001	-0.003	0.002	0.002	0.002	-0.002	-0.000	0.000
	(0.003)	(0.004)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
Black	-0.006	-0.000	-0.008	0.006	0.008	0.000	0.000	-0.008	0.008*
	(0.005)	(0.008)	(0.006)	(0.005)	(0.006)	(0.006)	(0.004)	(0.007)	(0.004)
Asian	-0.007	-0.017	0.000	0.006	0.013*	-0.000	0.002	0.004	0.000
	(0.006)	(0.010)	(0.007)	(0.005)	(0.007)	(0.008)	(0.006)	(0.009)	(0.007)
White	-0.003	-0.001	-0.005	0.001	0.003	-0.001	0.002	-0.002	0.005*
	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)
Observations:	92756	43092	49651	92756	43092	49651	92756	43092	49651
Adjusted R^2:	0.493	0.486	0.607	0.415	0.457	0.505	0.401	0.476	0.429

Notes: Dependent variables: dummy for whether the exam was taken at 9am (Morning Exam), dummy if it was taken at 1.30pm (Lunch Exam), dummy if it was taken at 4.30pm (Afternoon Exam). Standard errors are clustered by exam-year. Observations are at the student-exam-year level. In each specification, from Column (1)-(9), we control only for the conditions, year FE and exam FE: number of exams per session fixed effects (FE), class size, School FE and exam duration. We report only the coefficients on the main variables of interest for the balancing tests. In Panel A, we report estimates of the full sample. In Panel B, we report estimates for STEM schools only. In Panel C, we report estimates for Non-STEM schools only. Furthermore, in Column (1), (4) and (7), we report the results on the pooled sample. In Column (2), (5) and (8), we report the results for the Fall exam session only. In Column (3), (6) and (9), we report the results for the Spring exam session only. *p < 0.10, **p < 0.05, ***p < 0.01.

Table A8: Balancing Tests, no Student GPA (t-1), with exam FE

	Morning Exams			Lunch Exams			Afternoon Exams		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Pooled Sessions	Fall Session	Spring Session	Pooled Sessions	Fall Session	Spring Session	Pooled Sessions	Fall Session	Pooled Session
Panel A: Full-Sample									
Female	0.001	0.001	0.001	0.001	-0.001	0.002	-0.001	-0.000	-0.002
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)
Black	-0.002	0.001	-0.005	0.003	0.000	0.005*	-0.000	-0.001	-0.000
	(0.003)	(0.004)	(0.003)	(0.002)	(0.004)	(0.003)	(0.002)	(0.003)	(0.003)
Asian	-0.005	-0.004	-0.005	-0.000	-0.001	-0.000	0.005*	0.004	0.005
	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
White	0.000	0.002	-0.001	0.000	-0.000	0.001	-0.000	-0.002	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.001)	(0.002)	(0.002)
Observations:	434936	190556	244361	434936	190556	244361	434936	190556	244361
Adjusted R^2:	0.452	0.429	0.529	0.370	0.366	0.433	0.338	0.369	0.376
Panel B: STEM									
Female	-0.000	-0.001	0.001	0.002	0.002	0.002	-0.002	-0.000	-0.003
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.001)	(0.002)	(0.002)
Black	-0.003	0.004	-0.008*	0.003	-0.006	0.011**	0.000	0.002	-0.003
	(0.004)	(0.006)	(0.005)	(0.004)	(0.006)	(0.005)	(0.003)	(0.005)	(0.004)
Asian	-0.003	-0.003	-0.004	-0.004	-0.003	-0.003	0.008**	0.005	0.006
	(0.004)	(0.006)	(0.006)	(0.004)	(0.006)	(0.006)	(0.004)	(0.005)	(0.005)
White	0.003	-0.000	0.006*	-0.002	0.003	-0.005	-0.001	-0.003	-0.001
	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.002)	(0.003)	(0.003)
Observations:	209280	91310	117954	209280	91310	117954	209280	91310	117954
Adjusted R^2:	0.419	0.384	0.499	0.327	0.309	0.390	0.321	0.332	0.360
Panel C: NON-STEM									
Female	0.002	0.001	0.001	-0.001	-0.002	0.000	-0.001	0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Black	-0.004	-0.002	-0.004	0.004	0.004	0.002	-0.000	-0.003	0.001
	(0.003)	(0.005)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
Asian	-0.008**	-0.010	-0.005	0.004	0.002	0.001	0.004	0.008	0.004
	(0.004)	(0.006)	(0.005)	(0.004)	(0.006)	(0.005)	(0.004)	(0.006)	(0.005)
White	-0.002	0.004	-0.007***	0.002	-0.004	0.006**	0.001	-0.000	0.001
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)
Observations:	225656	99246	126407	225656	99246	126407	225656	99246	126407
Adjusted R^2:	0.486	0.500	0.556	0.415	0.439	0.480	0.358	0.411	0.394

Notes: Dependent variables: dummy for whether the exam was taken at 9am (Morning Exam), dummy if it was taken at 1.30pm (Lunch Exam), dummy if it was taken at 4.30pm (Afternoon Exam). Standard errors are clustered by exam-year. Observations are at the student-exam-year level. In each specification, from Column (1)-(9), we control only for the conditions, year FE and exam FE: number of exams per session fixed effects (FE), class size, School FE and exam duration. We report only the coefficients on the main variables of interest for the balancing tests. In Panel A, we report estimates of the full sample. In Panel B, we report estimates for STEM schools only. In Panel C, we report estimates for Non-STEM schools only. Furthermore, in Column (1), (4) and (7), we report the results on the pooled sample. In Column (2), (5) and (8), we report the results for the Fall exam session only. In Column (3), (6) and (9), we report the results for the Spring exam session only. *p < 0.10, **p < 0.05, ***p < 0.01.

A.3 Additional results

Table A9: Effects of Scheduling on Student Performance: By Gender

	Full-Sample			STEM			NON-STEM		
	(1) Pooled Sessions	(2) Fall Session	(3) Spring Session	(4) Pooled Sessions	(5) Fall Session	(6) Spring Session	(7) Pooled Sessions	(8) Fall Session	(9) Spring Session
Lunch [0,1]	0.066***	0.072***	0.050**	0.076***	0.105***	0.047	0.046**	0.030	0.039
	(0.015)	(0.024)	(0.020)	(0.022)	(0.035)	(0.029)	(0.019)	(0.031)	(0.024)
Afternoon [0,1]	0.038**	0.083***	0.002	0.049*	0.125***	-0.040	0.022	0.039	0.010
	(0.016)	(0.023)	(0.021)	(0.026)	(0.036)	(0.036)	(0.019)	(0.029)	(0.024)
Interaction Terms:									
Lunch * Male	0.005	0.041	-0.012	0.008	0.058	-0.027	-0.020	-0.033	-0.003
	(0.016)	(0.027)	(0.019)	(0.022)	(0.039)	(0.026)	(0.015)	(0.024)	(0.018)
Afternoon * Male	-0.004	-0.012	-0.021	-0.005	-0.050	0.008	-0.027*	-0.024	-0.040*
	(0.016)	(0.023)	(0.021)	(0.024)	(0.033)	(0.030)	(0.016)	(0.023)	(0.022)
Clusters:	7624	3434	4284	3465	1639	1886	4143	1781	2384
Observations:	500958	219236	274677	250361	109498	138422	248198	107729	134316
Adjusted R^2:	0.462	0.457	0.499	0.482	0.483	0.516	0.450	0.444	0.495

Notes: Dependent variable: standardized final exam mark. Key variables: dummies for whether the exam was taken at 1.30pm (Lunch Exam) or at 4.30pm (Afternoon Exam), where the 9am (Morning Exam) is the excluded category. Standard errors are clustered by exam-year. Observations are at the student-exam-year level. In each Column (1)-(9) we use the preferred specification which controls for student FE and all covariates. In Columns (1)-(3) we report estimates of the full sample, while in Columns (4)-(6) and Columns (7)-(9) we report results for STEM and NON-STEM schools, separately. Specifically, in Columns (1), (4), and (7) are reported the results for the pooled sample. In Columns (2), (5), and (8) we report the results for the Fall exam session only. In Columns (3), (6), and (9) we report the results for the Spring exam session only. Below the test of equality of Lunch and Afternoon exams is reported the p-value of the test statistic. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. .

Table A10: Effects of Scheduling on Student Performance: By Ability Level

	Full-Sample			STEM			NON-STEM		
	(1) Pooled Sessions	(2) Fall Session	(3) Spring Session	(4) Pooled Sessions	(5) Fall Session	(6) Spring Session	(7) Pooled Sessions	(8) Fall Session	(9) Spring Session
Lunch [0,1]	0.097***	0.113***	0.062**	0.118***	0.150***	0.083*	0.058**	0.051	0.038
	(0.023)	(0.036)	(0.031)	(0.038)	(0.057)	(0.050)	(0.024)	(0.040)	(0.030)
Afternoon [0,1]	0.070***	0.083**	0.033	0.075*	0.078	0.010	0.044*	0.048	0.034
	(0.023)	(0.035)	(0.032)	(0.045)	(0.068)	(0.059)	(0.023)	(0.039)	(0.030)
Interaction Terms:									
Lunch * High Ability	-0.010	0.033	-0.029	-0.018	0.058	-0.065	-0.019	-0.041	0.003
	(0.024)	(0.044)	(0.028)	(0.035)	(0.060)	(0.042)	(0.022)	(0.045)	(0.022)
Afternoon * High Ability	-0.006	0.015	-0.013	-0.013	0.018	0.004	-0.005	0.034	-0.028
	(0.028)	(0.037)	(0.036)	(0.046)	(0.064)	(0.058)	(0.026)	(0.038)	(0.032)
Clusters:	5171	2221	2995	2357	1041	1352	2799	1166	1632
Observations:	313328	130967	170633	170919	69759	96531	140767	59687	72955
Adjusted R^2:	0.445	0.436	0.489	0.477	0.466	0.523	0.411	0.420	0.439

Notes: Dependent variable: standardized final exam mark. Key variables: dummies for whether the exam was taken at 1.30pm (Lunch Exam) or at 4.30pm (Afternoon Exam), where the 9am (Morning Exam) is the excluded category. Standard errors are clustered by exam-year. Observations are at the student-exam-year level. In each Column (1)-(9) we use the preferred specification which controls for student FE and all covariates. In Columns (1)-(3) we report estimates of the full sample, while in Columns (4)-(6) and Columns (7)-(9) we report results for STEM and NON-STEM schools, separately. Specifically, in Columns (1), (4), and (7) are reported the results for the pooled sample. In Columns (2), (5), and (8) we report the results for the Fall exam session only. In Columns (3), (6), and (9) we report the results for the Spring exam session only. Below the test of equality of Lunch and Afternoon exams is reported the p-value of the test statistic. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. .

Table A11: Effects of Time-of-Day on Student Performance: First Year UG Students

	Full-Sample			STEM			NON-STEM		
	(1) Pooled Sessions	(2) Fall Session	(3) Spring Session	(4) Pooled Sessions	(5) Fall Session	(6) Spring Session	(7) Pooled Sessions	(8) Fall Session	(9) Spring Session
Lunch [0,1]	0.080***	0.081***	0.060***	0.083***	0.097***	0.053*	0.057***	0.037	0.044*
	(0.015)	(0.022)	(0.021)	(0.023)	(0.031)	(0.032)	(0.019)	(0.032)	(0.025)
Afternoon [0,1]	0.039**	0.058***	0.009	0.022	0.029	-0.011	0.031*	0.061**	0.005
	(0.017)	(0.022)	(0.024)	(0.027)	(0.034)	(0.038)	(0.018)	(0.026)	(0.026)
Interaction Terms:									
Lunch * 1st Year UG	-0.029	0.049	-0.045	-0.008	0.142*	-0.063	-0.043	-0.043	-0.014
	(0.030)	(0.055)	(0.036)	(0.046)	(0.086)	(0.053)	(0.035)	(0.055)	(0.046)
Afternoon * 1st Year UG	-0.002	0.069	-0.042	0.078	0.230***	-0.040	-0.045	-0.078	-0.028
	(0.031)	(0.055)	(0.039)	(0.051)	(0.085)	(0.063)	(0.037)	(0.066)	(0.046)
Clusters:	7624	3434	4284	3465	1639	1886	4143	1781	2384
Observations:	500958	219236	274677	250361	109498	138422	248198	107729	134316
Adjusted R^2:	0.459	0.455	0.495	0.478	0.480	0.510	0.448	0.442	0.493

Notes: Dependent variable: standardized final exam mark. Key variables: dummies for whether the exam was taken at 1.30pm (Lunch Exam) or at 4.30pm (Afternoon Exam), where the 9am (Morning Exam) is the excluded category. Standard errors are clustered by exam-year. Observations are at the student-exam-year level. In each Column (1)-(9) we use the preferred specification which controls for student FE and all covariates. In Columns (1)-(3) we report estimates of the full sample, while in Columns (4)-(6) and Columns (7)-(9) we report results for STEM and NON-STEM schools, separately. Specifically, in Columns (1), (4), and (7) are reported the results for the pooled sample. In Columns (2), (5), and (8) we report the results for the Fall exam session only. In Columns (3), (6), and (9) we report the results for the Spring exam session only. Below the test of equality of Lunch and Afternoon exams is reported the p-value of the test statistic. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. .

Table A12: Effects of Time-of-Day on Student Performance: Main Specifications

	Full-Sample			STEM			NON-STEM		
	(1) Pooled Sessions	(2) Fall Session	(3) Spring Session	(4) Pooled Sessions	(5) Fall Session	(6) Spring Session	(7) Pooled Sessions	(8) Fall Session	(9) Spring Session
Panel A: Conditions									
Lunch [0,1]	0.051***	0.075***	0.028	0.057**	0.127***	-0.007	0.038*	0.010	0.054**
	(0.016)	(0.025)	(0.020)	(0.025)	(0.039)	(0.029)	(0.019)	(0.027)	(0.025)
Afternoon [0,1]	0.042**	0.072***	0.008	0.053*	0.097**	-0.012	0.021	0.033	0.012
	(0.017)	(0.024)	(0.022)	(0.027)	(0.040)	(0.037)	(0.020)	(0.030)	(0.026)
Test of equality of Lunch and Afternoon exams in linear regressions above:									
$Pr > F$	0.619	0.900	0.370	0.891	0.475	0.878	0.436	0.481	0.125
Clusters:	7663	3486	4351	3471	1650	1900	4192	1836	2451
Observations:	503358	224933	278414	251691	111807	139873	251667	113126	138541
Adjusted R^2:	0.022	0.025	0.029	0.023	0.023	0.031	0.028	0.033	0.034
Panel B: All Covariates									
Lunch [0,1]	0.066***	0.086***	0.048**	0.078***	0.136***	0.034	0.035*	0.017	0.042*
	(0.016)	(0.024)	(0.019)	(0.024)	(0.034)	(0.029)	(0.018)	(0.028)	(0.022)
Afternoon [0,1]	0.032*	0.072***	0.007	0.048*	0.108***	-0.003	0.005	0.024	-0.012
	(0.016)	(0.024)	(0.021)	(0.026)	(0.038)	(0.033)	(0.019)	(0.028)	(0.025)
Test of equality of Lunch and Afternoon exams in linear regressions above:									
$Pr > F$	0.041	0.569	0.047	0.261	0.472	0.250	0.133	0.814	0.030
Clusters:	7663	3486	4351	3471	1650	1900	4192	1836	2451
Observations:	503358	224933	278414	251691	111807	139873	251667	113126	138541
Adjusted R^2:	0.173	0.163	0.197	0.176	0.175	0.196	0.187	0.167	0.219
Panel C: Benchmark									
Lunch [0,1]	0.068***	0.094***	0.043**	0.081***	0.139***	0.030	0.037**	0.016	0.038*
	(0.015)	(0.024)	(0.019)	(0.022)	(0.035)	(0.028)	(0.017)	(0.028)	(0.022)
Afternoon [0,1]	0.036**	0.076***	-0.010	0.046*	0.093***	-0.034	0.010	0.028	-0.007
	(0.015)	(0.023)	(0.021)	(0.025)	(0.035)	(0.034)	(0.017)	(0.027)	(0.022)
Test of equality of Lunch and Afternoon exams in linear regressions above:									
$Pr > F$	0.037	0.463	0.008	0.165	0.210	0.047	0.140	0.670	0.048
Clusters:	7624	3434	4284	3465	1639	1886	4143	1781	2384
Observations:	500958	219236	274677	250361	109498	138422	248198	107729	134316
Adjusted R^2:	0.462	0.457	0.499	0.482	0.483	0.515	0.450	0.444	0.495
Panel D: All Covariates & Exam FE									
Lunch [0,1]	0.027**	0.057***	-0.009	0.042**	0.098***	-0.017	0.002	0.016	-0.006
	(0.013)	(0.019)	(0.016)	(0.020)	(0.028)	(0.026)	(0.014)	(0.024)	(0.017)
Afternoon [0,1]	0.013	0.051**	-0.027	0.014	0.076**	-0.045	0.015	0.035	-0.008
	(0.015)	(0.021)	(0.019)	(0.023)	(0.032)	(0.031)	(0.016)	(0.025)	(0.021)
Test of equality of Lunch and Afternoon exams in linear regressions above:									
$Pr > F$	0.302	0.782	0.320	0.189	0.464	0.320	0.403	0.450	0.912
Clusters:	7610	3458	4317	3462	1645	1891	4148	1813	2426
Observations:	503305	224905	278380	251682	111802	139864	251623	113103	138516
Adjusted R^2:	0.261	0.257	0.275	0.266	0.270	0.276	0.268	0.246	0.293

Notes: Dependent variable: standardized final exam mark. Key variables: dummies for whether the exam was taken at 1.30pm (Lunch Exam) or at 4.30pm (Afternoon Exam), where the 9am (Morning Exam) is the excluded category. Standard errors are clustered by exam-year. Observations are at the student-exam-year level. n Columns (1)-(3) we report estimates of the full sample, while in Columns (4)-(6) and Columns (7)-(9) we report results for STEM and NON-STEM schools, separately. Specifically, in Columns (1), (4), and (7) are reported the results for the pooled sample. In Columns (2), (5), and (8) we report the results for the Fall exam session only. In Columns (3), (6), and (9) we report the results for the Spring exam session only. Panel (A) presents the results controlling only for the conditions. Panel (B) report the results controlling for all covariates except student FE. Panel (C) report the results controlling also for student FE. Panel (D) report the results controlling for all covariates and exam FE, instead of student FE. Below the test of equality of Lunch and Afternoon exams is reported the p-value of the test statistic. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table A13: Coefficient Stability Tests

	Full-Sample			STEM			NON-STEM		
	(1) Pooled Sessions	(2) Fall Session	(3) Spring Session	(4) Pooled Sessions	(5) Fall Session	(6) Spring Session	(7) Pooled Sessions	(8) Fall Session	(9) Spring Session
Conditions:									
Lunch [0,1]	0.051***	0.075***	0.028	0.057**	0.127***	-0.007	0.038*	0.010	0.054**
	(0.016)	(0.025)	(0.020)	(0.025)	(0.039)	(0.029)	(0.019)	(0.027)	(0.025)
Afternoon [0,1]	0.042**	0.072***	0.008	0.053*	0.097**	-0.012	0.021	0.033	0.012
	(0.017)	(0.024)	(0.022)	(0.027)	(0.040)	(0.037)	(0.020)	(0.030)	(0.026)
Clusters:	7663	3486	4351	3471	1650	1900	4192	1836	2451
Observations:	503358	224933	278414	251691	111807	139873	251667	113126	138541
Adjusted R^2:	0.022	0.025	0.029	0.023	0.023	0.031	0.028	0.033	0.034
Panel B: Student FE, Covariates									
Lunch [0,1]	0.068***	0.094***	0.043**	0.081***	0.139***	0.030	0.037**	0.016	0.038*
	(0.015)	(0.024)	(0.019)	(0.022)	(0.035)	(0.028)	(0.017)	(0.028)	(0.022)
Afternoon [0,1]	0.036**	0.076***	-0.010	0.046*	0.093***	-0.034	0.010	0.028	-0.007
	(0.015)	(0.023)	(0.021)	(0.025)	(0.035)	(0.034)	(0.017)	(0.027)	(0.022)
Clusters:	7624	3434	4284	3465	1639	1886	4143	1781	2384
Observations:	500958	219236	274677	250361	109498	138422	248198	107729	134316
Adjusted R^2:	0.462	0.457	0.499	0.482	0.483	0.515	0.450	0.444	0.495
Panel C: Bounding Set									
Lunch [0,1]	[0.036, 0.068]	[0.049, 0.093]	[0.025, 0.043]	[0.041, 0.080]	[0.097, 0.139]	[-0.003, 0.030]	[0.029, 0.037]	[-0.003, 0.015]	[0.049, 0.037]
Afternoon [0,1]	[0.019, 0.035]	[0.038, 0,076]	[-0.005, -0.009]	[0.027, 0.045]	[0.043, 0.093]	[-0.009,- 0.034]	[0.003, 0.010]	[0.027, 0.027]	[-0.014,- 0.007]

Notes: The table presents the validation results for the analysis of the effect of time-of-day on performance. Dependent variable: standardized final exam mark. Key variables: dummies for whether the exam was taken at 1.30pm (Lunch Exam) or at 4.30pm (Afternoon Exam), where the 9am (Morning Exam) is the excluded category. Panel A, presents estimates when controlling for the conditional independence assumption variables, while in Panel B we include in addition other covariates as well as student FE. Finally, Panel C reports estimates of the bounding set assuming a Rmax equals to one. Standard errors are clustered by exam-year. Observations are at the student-exam-year level. In Columns (1)-(3) we report estimates of the full sample, while in Columns (4)-(6) and Columns (7)-(9) we report results for STEM and NON-STEM schools, separately. Specifically, in Column (1), (4) and (7), we report the results on the pooled sample. In Column (2), (5) and (8), we report the results for the Fall exam session only. In Column (3), (6) and (9), we report the results for the Spring exam session only. $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

A.4 Identification: Details

For simplicity, and slight abuse of notation, we re-write Equation (3) as follows:

$$Y_i(h, \boldsymbol{\theta}) = T_i(h, \boldsymbol{\theta}) + \epsilon_i(h, \boldsymbol{\theta}) \tag{A1}$$

where $Y_i(h, \boldsymbol{\theta})$ is a worker's productivity at performing a task, which is a function of hour of the day h and a vector of task and worker's characteristics $\boldsymbol{\theta}$, including also the cognitive level required by the task. Furthermore, $T_i(h, \boldsymbol{\theta})$ is the individual time-of-day effect and $\epsilon_i(h, \boldsymbol{\theta})$ is a standard random unobserved error with mean zero.

In our dataset, we observe each worker performing multiple and different tasks. For the moment, suppose we observe two performances. Ideally, we wish to identify:

$$E[\Delta_Y(h, \boldsymbol{\theta})] = E[Y_i(h, \boldsymbol{\theta}) - Y_i(0, \boldsymbol{\theta})] \tag{A2}$$

which is the average treatment effect (ATE) of performing a task $\boldsymbol{\theta}$ at hour h on productivity with respect to a baseline hour (9am in our case). Since each worker is observed performing tasks with different characteristics $\boldsymbol{\theta}$, we need to make some identifying assumptions.

Assumption 1. *Random assignment:*

i. $E[Y_i(h, \boldsymbol{\theta})] = E[Y_i(h, \boldsymbol{\theta}')] = E[Y_i(h)],$

ii. $E[Y_i(0, \boldsymbol{\theta})] = E[Y_i(0, \boldsymbol{\theta}')] = E[Y_i(0)].$

Assumption 1 says that the potential outcomes of the workers are independent of the vector of characteristics $\boldsymbol{\theta}$. This means that, on average, workers need to perform similar tasks at different time-of-day. Under assumption 1, the population model simplifies to:

$$Y_i(h) = T_i(h) + \epsilon_i(h), \tag{A3}$$

where $T_i(h) \perp \epsilon_i(h)$ by design.

Assumption 2. *Time-of-day:*

$$T_i(h) = W_i(h) + B_i(h).$$

Assumption 2 says that time-of-day is made of two components: a "worker" and "brain" compo-
nents. $W_i(h)$ captures the worker's individual behavior to increase her performance or minimize her
fatigue (e.g. effort, work strategy, food intake) that affects productivity. $B_i(h)$ captures the effect of
the circadian rhythm on worker's productivity. That is, a brain is more productive at performing a
cognitive task at specific hours of the day. Under assumption 2, the population model becomes:

$$Y_i(h) = W_i(h) + B_i(h) + \epsilon_i(h). \tag{A4}$$

We take expectations over the difference in performance between two tasks using the following
mild assumption:

Assumption 3. *Random shocks:*

$$E[\epsilon_i(h)] = E[\epsilon_i(0)] = 0.$$

Hence we have:

$$
\begin{aligned}
E[\Delta_Y(h)] &= E[Y_i(h) - Y_i(0)] \\
&= E[W_i(h) - W_i(0)] + E[B_i(h) - B_i(0)] \\
&= E[\Delta_W(h)] + E[\Delta_B(h)]
\end{aligned}
\tag{A5}
$$

The third line of Equation (A5) shows that the effect that we wish to identify is affected by two
components that are functions of the same factor: we call $E[\Delta_W(h)]$ the expected worker reaction
effect to the hour of the day and $E[\Delta_B(h)]$ the expected brain reaction effect to the hour of the day.
These two components are substitute of each other. Hence, under Assumptions 1, 2 and 3, we cannot
identify separately the effect of each component $E[\Delta_W(h)]$ and $E[\Delta_B(h)]$. However, whatever their
values are, if the worker was perfectly efficient, the total effect of time-of-day on productivity should
be around zero. Contrary, if the effect of time-of-day is significantly different from zero, then this
tells us that workers performing cognitive tasks suffer from efficiency (or welfare) losses driven
by the time-of-day of their tasks. Importantly, since both components have a positive effect on
productivity, but they move in opposite direction with each other, the time-of-day effect that we
estimate is a lower bound of $E[\Delta_B(h)]$. We can think of it as a standard attenuation bias.