Economics of Education Review 90 (2022) 102305

Contents lists available at ScienceDirect

Economics of Education Review

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

The deleterious effects of fatigue on final exam performance

Darrell J. Glaser, Michael A. Insler *

Department of Economics, United States Naval Academy, United States of America

ARTICLE INFO

JEL classification:
I20
I21
J24

Keywords:
Grades
Scheduling
Academic performance
Human capital

ABSTRACT

We utilize an interrupted time series approach to identify the causal effects of a Sunday break on final exam performance. Five different class-year cohorts at the U.S. Naval Academy (USNA) comprise our panel dataset with over 19,000 exam grades from more than 5,000 students. Exam schedules vary exogenously across years, and first semester freshmen have no say in their course schedules nor can they reschedule their exams. We find consistent empirical evidence that a day-long break situated in the middle of a final exam period increases the average exam grade after the break. This occurs even after we control for idiosyncratic scheduling characteristics during the exam cycle, other timing-related variables, and both course and student-specific variables. In particular, the average final exam grade taken one day after a break is 20%–30% of a letter grade higher than exams taken one day prior to the break. We also estimate that each passing day of a final exam cycle decreases performance by approximately 10%–15% of a letter grade, and that morning exams exhibit a 20%–25% of a letter grade decline in performance, compared to the afternoon.

1. Introduction

A challenging but crucial goal of education policymakers is to design effective assessment of student learning. Assessment can of course take many forms, but in the higher education setting, the individual-based, timed examination is perhaps the most prominent instrument. While test design is typically up to individual instructors or course coordinators, university administrators generally schedule testing periods broadly across an entire school or college. This is particularly prevalent for final examinations, which are often scheduled months in advance. If exam grades respond to periodic breaks from exams, the time of testing during the day, or to the amount of time between individual exams, university administrators may wish to consider such effects as they develop testing schedules.[1] Changes to exam schedules may be relatively low cost interventions for improvement, and may provide a useful signal before implementing related but larger scale policy changes such as shifts in school start times or nudges of students towards selection of more optimal course schedules.

This paper uses quasi-experimental data and Interrupted Time Series (ITS) methods to estimate causal effects of a study day dropped into the middle of an exam cycle on performance. It also identifies the effect of three other dimensions of final exam timing on performance: the number of days between exams (given one's idiosyncratic testing schedule); the day number at which a given exam occurs within the testing period (which often spans more than one week in our setting); and whether a given exam is administered in the morning, afternoon, or evening.

To examine these issues, we exploit a quasi-experimental environment provided at the United States Naval Academy (USNA), specifically for freshmen. At USNA, freshmen are required to pass a set of 11 "core" courses in various humanities, social science, math, and natural science disciplines. Students have limited opportunity to test out of (or "validate") these requirements, thus most freshmen possess similar course schedules. Core courses are large (but split up into classroom sections of approximately 20 students), closely coordinated across instructors and sections, employ the same final exam in a given year, and maintain matched rubrics for exam grading. Importantly, all students in a course sit for final exams concurrently, and students in these core courses cannot request alternate exam times (except under very rare circumstances described later in the paper). Exam schedules change from year to year based on exogenous variation in the academic calendar, which allows us to identify causal effect of an exam cycle break. Our data come from academic years 2013–2017 and include various student level characteristics from high school (e.g., SAT scores) as well as midterm course grades.

We estimate various specifications for an interrupted time series approach, with final exam grades regressed on the exam timing dimensions, observable student characteristics, course fixed effects, and a time trend. Initial estimates provide empirical evidence that indicates

* Corresponding author.
E-mail addresses: dglaser@usna.edu (D.J. Glaser), insler@usna.edu (M.A. Insler).

[1] Or at least, in this case, the timing of an exam should be considered as an important "control variable" when using course or final exam grades as a means of learning assessment.

https://doi.org/10.1016/j.econedurev.2022.102305
Received 25 August 2021; Received in revised form 6 August 2022; Accepted 11 August 2022
Available online 30 August 2022
0272-7757/Published by Elsevier Ltd.

that an entire day situated in the middle of final exam cycle increases average exam grades taken after the Sunday break. This occurs even after model specifications control for typical break days during the exam cycle, other timing-related variables and both course and student-related variables. In particular, our results indicate that the average final exam grade taken one day after the common break is 20%–30% of a letter grade higher than exams taken one day prior to the break. We also find that students perform substantially better on afternoon exams compared to morning exams. The data used for this is quasi-experimental, since USNA freshmen during their first (fall) semester have little to no say in course schedules or (non-academic) day-to-day activities. It is also nearly impossible for students to reschedule the timing of a final exam, and final exam schedules are randomized across the years of our sample as well. Each of these factors mitigate or essentially eliminate potential sources of selection bias that can interfere with the identification of this causal effect of exam-cycle breaks.

We then examine course-level fixed effects regressions, which indicate a *ceteris paribus* decrease in final exam grades from each passing day of a final exam cycle. Each day decreases grades by approximately 10%–15% of a letter grade, a result that remains robust to numerous specifications and sub-sample sensitivity checks. Furthermore, exams taken relatively early in the morning exhibit approximately a 20%–25% of a letter grade decline compared to afternoon exams. Finally, we find that each day of rest between exams (due to idiosyncratic differences across students' exam schedules *and thus in addition to* the Sunday break given to all students) improves grades by approximately 5% of a letter grade.

This paper is laid out as follows. Section 2 describes the institutional setting of USNA, the data, and potential external validity concerns. Section 3 develops the empirical methods, and Section 4 presents results and robustness checks. The final section offers additional discussion and concluding remarks.

2. Related literature

To our knowledge, there are no other systematic studies of the specific topic of a final exam timing in the higher education venue, however economists have investigated related questions and alternate settings. For example, researchers have studied how course or school start time affect academic achievement. Dills and Hernández-Julián (2008) estimate the effect of class start times at a large public institution, with a student body of somewhat average academic quality. They include only two semesters of data, and have no specific student characteristics, however, they include student fixed effects. Performance is measured over the entire semester course (course grades), and they find positive effects of classes that meet two days a week in the late afternoon relative to three. Their results also indicate a positive effect for classes that meet 3 days a week in the morning. They conclude that "spacing learning out over time may foster greater long-term memory of items".

Cotti, Gordanier, and Ozturk (2018) follow-up on the work of Dills and Hernández-Julián (2008), also including student fixed effects and measuring performance over the entirety of a course semester (course grades). In contrast, they find that frequency of class meetings (more classes, shorter periods in a week) matters, with positive effects observed for classes that meet two or four days relative to three. Once including instructor fixed effects, these results disappear. It is argued that positive sorting of instructors into classes that meet less frequently may drive the positive results of other papers where number of days a week that classes meet appears significant. That being said, they still find a positive effect of the start hour for classes (students perform better with a later start hour of the course).

Carrell, Maghakian, and West (2011) explore these effects at the U.S. Air Force Academy (USAFA), which has a similar academic to other liberal arts settings with above average student quality. Like USNA, their students have mandatory attendance of class and a similarly restricted scheduled lives. Their work finds that earlier school start times negatively affect student performance: Random assignments of students to first-period courses earned lower grades in that class and all other class during the day. Similarly, Lusher, Yasenov, and Luong (2019) finds that delayed start times tends to help student achievement.

A general implication of this body of research indicates that students perform better when learning takes place later in the day. If final exam grade and overall academic achievement in a course are driven by similar mechanisms, we would expect exam scores to improve in the afternoon and evening sessions, versus the morning.[2] Moreover, Goulas and Megalokonomou (2020) examine the high school venue and find that a "warm-up period" before an exam can help performance but fatigue—as the period drags on—can hurt it. Pope and Fillmore (2015) demonstrate that extra days between two AP exams bolsters high school students' performance, and the effect stems from the second exam. Our work seeks to further investigate these timing channels at the collegiate level, which may be of great interest to university administrators and policy makers, via various econometric specifications.

Researchers in higher education have studied other aspects of timing, such as how class meeting frequency affects student achievement. For example, Cotti et al. (2018) determine that grades are higher in classes that meet two times per week (versus three). This effect appears to be driven by instructor preferences for twice-per-week classes, which could manifest as either higher quality of instruction or easier grading in those classes. This finding is bolstered by Diette and Raghav (2018), who estimate no difference in student performance in twice-per-week versus thrice-per-week classes after controlling for instructor fixed effects. Based on the current body of literature, it is unclear how one's final exam performance might respond to the number of breaks embedded in her schedule, while also considering or how her grades might fare deeper into the examination period along with the time of day of exams. Our paper is the first investigate this range of timing effects on final exam outcomes in a higher education setting.

3. Data

We use administratively collected data of all USNA freshmen enrolled during fall semesters of 2013 through 2017.[3] We observe every grade assigned to these freshmen, along with course title and credit hours. For every freshman, we also observe both pre-USNA characteristics (SAT scores, recruited athlete status, gender, minority status, whether previously enlisted in the armed forces, and whether attended a preparatory school) and contemporaneous variables (midterm course grades—at the 6-week point, 12-week point, and "end-of-classes" point—as well as exam scheduling data, explained more below).[4]

USNA provides an ideal setting with quasi-randomized data such as this. For instance USNA possesses the ingredients to investigate peer effects, due to effectively random assignment of students to residential groups (called "companies") and core fall-of-freshman-year course-sections (Brady, Insler, & Rahman, 2017). In this paper, to estimate the effect of exam timing on student outcomes, relevant mechanisms that facilitate identification include students' lack of choice over their first year fall semester courses, along with their inability to request changes to their final exam schedule. Scheduling considerations are

[2] Some of the aforementioned researchers argue that sleep is the primary mechanism driving this effect, and they cite numerous studies in the sleep literature as evidence. Another mechanism may be students' tendency to target later-in-the-day periods (which may be preferred due to either sleeping habits or other reasons) for their preferred classes (in which student achievement tends to be better in general).

[3] We cannot use data prior to 2012 because we possess historical final exam records for academic years 2012–2017 only.

[4] The end of classes grade is effectively a student's cumulative course grade prior to sitting for the final exam.

exogenous and not subject to student ability or course preferences. We elaborate on these mechanisms below and demonstrate that final exam timing also varies randomly across time. This is a necessary condition for identification of the effect exam timing on performance.

3.1. First year course assignment

Freshmen have very little choice over their courses. All freshmen must pass or validate a set of 11 core courses in a range of subject areas. Excluding physical education courses, approximately half of the freshman class will take five classes in the fall, and the other half will take six (then six and five, respectively, in the spring). Students cannot choose whether they take five courses or six in their first semester, or if they take U.S. Government or American Naval History.[5]

It is possible for students to validate freshman year courses through USNA-administered placement exams or advanced placement scores (e.g., a student that validates Calculus 1 via AP scores will take Calculus 2 during the fall semester). The validation exams are the only form of student input into the course selection process. For this reason, we include robustness checks that focus only on the core courses most commonly taken by the majority of freshmen. These core courses taken by the majority of freshmen include: Calculus 1, Calculus 2, Chemistry 1, Intro to Cyber Security, US Government (a political science course), American Naval History, English 1, English 2 as well as two Navy-specific courses in Leadership and Seamanship. English courses do not administer final exams. The only course from this list that may be accessed by "advanced" freshmen is Calculus 2, but only would be considered advanced if taken during the fall semester. This is common, as 23% of the sample enroll in Calculus 2 in the fall, rather than Calculus 1. A sample restriction that excludes more advanced or specialized courses from our sample would circumvent potential endogeneity issues that might be related to the exam timing variables for such courses, which we elaborate upon in the next subsection. Our robustness checks later in the paper indicate that this is not a problem.

3.2. Final exam scheduling and timing

Core courses are relatively homogeneous across sections, and students cannot swap their sections to access different instructors (see discussion and evidence in Brady et al., 2017). For all courses used in our analysis, grading schemes are standardized, particularly rigidly for the final exam component, and a given course's final exam is taken concurrently by all students enrolled in the course, with very few exceptions. Statements provided by the Registrar and Academic Dean's Office inform us that the only excuses that might permit rescheduling of an individual freshman's core course final exams are varsity athletic events that directly conflict with a final exam (NCAA rules minimize these conflicts), leave for illness, family emergencies, or other military-related conflicts. Such conflicts are rare; we cannot observe exam scheduling changes, so these exceptions would not be reflected in our exam timing variables. Thus, a most conservative interpretation of estimates developed in the next section are as "intent to treat" effects of final exam schedules on student performance.

USNA final exams may be scheduled on any day of the week, except Sunday. The core discussion of this paper revolves around the causal effect of this mandatory break on average performance. Each day has three 3 hour-long exam blocks: The morning block begins at 7:55 a.m., the afternoon block begins at 1:30 p.m., and the evening block begins at 7:30 p.m. There are also timing considerations that are specific to the calendar year. Depending on this, the first and last days of exams do not always fall on the same day of the week, which creates variance in where students' off-days may lie within their individual exam schedules. The scheduling of the large, core, freshman year courses is prioritized by the Registrar to avoid conflicts and permit virtually all students to attend the chosen exam time, and not have two core exams on the same day. The calendar-related peculiarities mentioned above force the Registrar to make year-to-year changes in which days/times these courses' final exams occur.

Fig. 1 demonstrates substantial variation in the timing of final exams across academic years. Furthermore, each course appears sometimes before, and sometimes after, the Sunday break (which can be seen in the figure via the placement of the red stars in each academic year). Academic year 2014 contained additional break days due to timing of the Army/Navy football game; we elaborate on this unique year in Section 3.4 below. The key takeaways are that individual courses do not systematically appear in the early part of a final examination period or a latter part and there exists substantial variation around the break days.[6]

An additional method for assessing the impact of exam day assignments appears as Fig. 2. Here, histograms depict the density of exams taken relative to the key assignment variable, the day of the exam cycle. The Sunday break is defined where the assignment variable equals zero. Naval History (HH104) and Political Science (FP130) exams tend to occur prior to the Sunday break (with some exceptions), but other courses tend to occur on both sides of the break.

Finally, as an additional test that exam timing is effectively randomly assigned, we run a series of regressions to check whether exam timing of courses in our sample might somehow be related to observable student characteristics or outcomes. The results of these baseline tests appear below after we discuss the data in more detail.

3.3. External validity

USNA is a service academy with a liberal arts-style academic setting. Graduates earn a Bachelor of Science degree in one of approximately 25 majors. In this respect USNA is similar to USMA and USAFA (see Carrell, Fullerton, & West, 2009; Lyle, 2007, 2009, respectively). However, USNA's academic setting is distinct from USMA and USAFA in that the Naval Academy's faculty is at least fifty percent tenure-track civilian, career academics (this statistic tends to be as high as 60 percent in practice due to unfilled "billets" on the military side; see Keller, Lim, Harrington, O'Neill, & Haddad, 2013). In contrast, USMA's faculty model targets 25 percent civilian, while USAFA targets 29 percent (Keller et al., 2013). The civilian tenure-track faculty are required to have earned a Ph.D and are evaluated for tenure according to guidelines that are similar to "regular" colleges and universities. The military faculty predominantly hold Masters Degrees (a small percentage have Ph.Ds). For these reasons, USNA is academically comparable to other schools studied in the peer effects literature.

Empirical specifications with the cleanest identification rely on the subsample that excludes advanced/specialized courses which a small share of freshmen may take, if they were able to test out of the standard core courses. Thus the analyses that focus on the core may underrepresent stronger students taking especially advanced math, as well as a handful of weaker students enrolled in remedial math. When we exclude all non-core courses, robustness checks reveal very little change in the estimates. Student ability is not correlated with exam timing of

[5] These outcomes are determined by the Registrar and are simply based on one's company assignment (which is shown in Brady et al., 2017, to be effectively random). For example, the Registrar may assign 1st through 15th Companies to take U.S. Government in the fall, while 16th through 30th Companies take American Naval History (and vice versa in the spring).

[6] The exceptions to this are the Navy-specific courses of Seamanship and Leadership (not shown in the figures), where final exams tend to be scheduled earlier in the marking period. Hence and along with other reasons highlighted below, we restrict our main analysis to courses that are not Navy-specific. We include the Leadership and Seamanship courses in a single robustness check later in the paper.

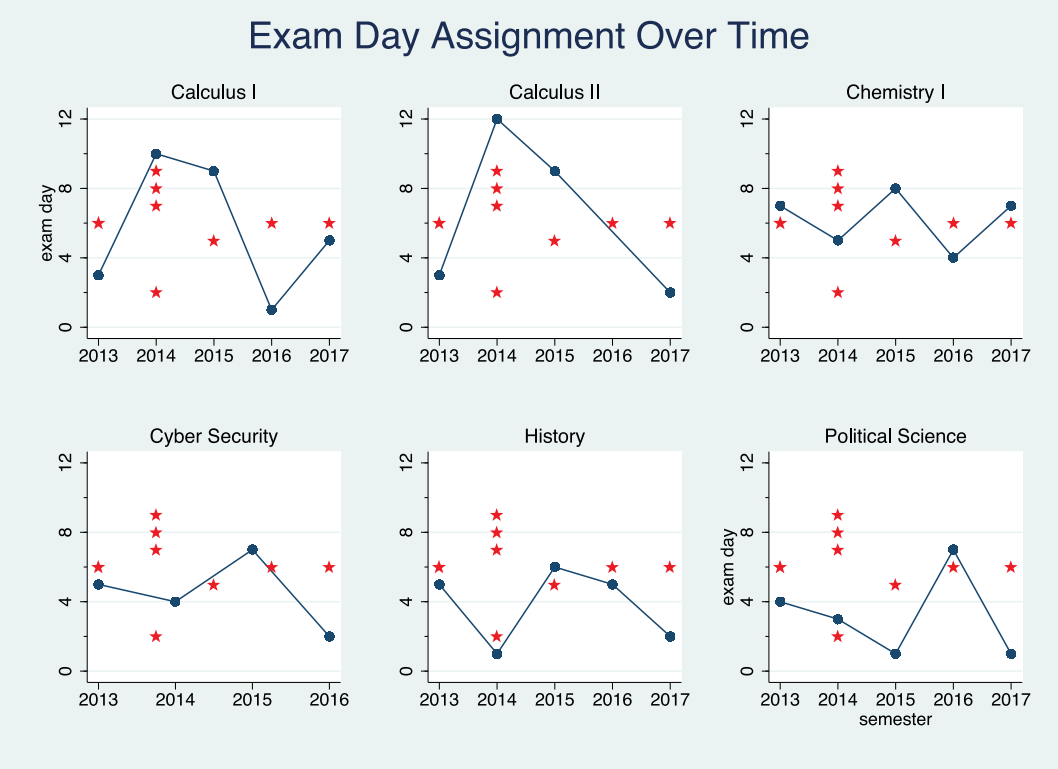

Fig. 1. Day of exam period when course finals scheduled.
Note: Each panel corresponds to a different fall semester core course. The vertical axis refers to the exam day (there can be up to twelve days in the final exam period); the horizontal axis refers to the particular academic year. Red stars in each panel display, for example, that the break-days took place on Day 6 in 2013, Day 2 and then Days 7–9 (due to Army-Navy Football game) in 2014, Day 5 in 2015, and so on. Similarly, the first panel, for example, shows that the Calculus I exam was scheduled before the break (on Day 3) in 2013, after the long break (Day 10) in 2014, and after the break (Day 9) in 2015, and so on.

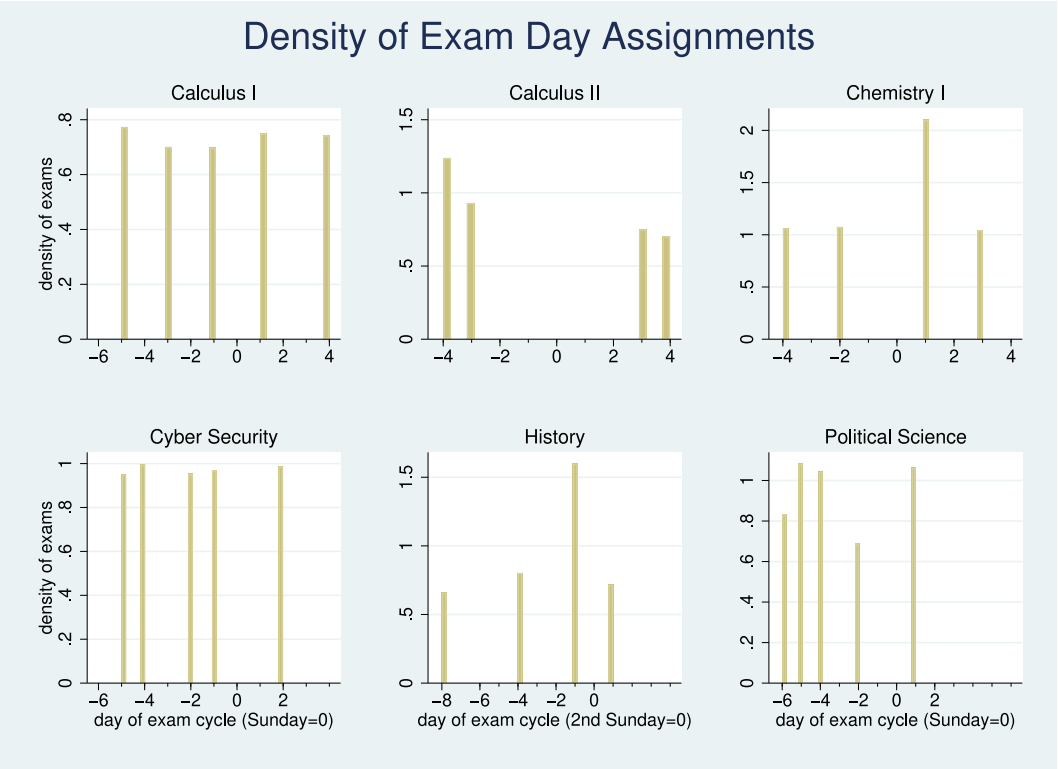

Fig. 2. Density of core exam day assignments.
Note: Each panel contains a histogram corresponding to a different fall semester core course. Histograms show the density of exams scheduled on a given day within the final exam period. While Naval History and Political Science tend to occur before the break, most courses' exams tend to fall on both sides of the break across the sampled years.

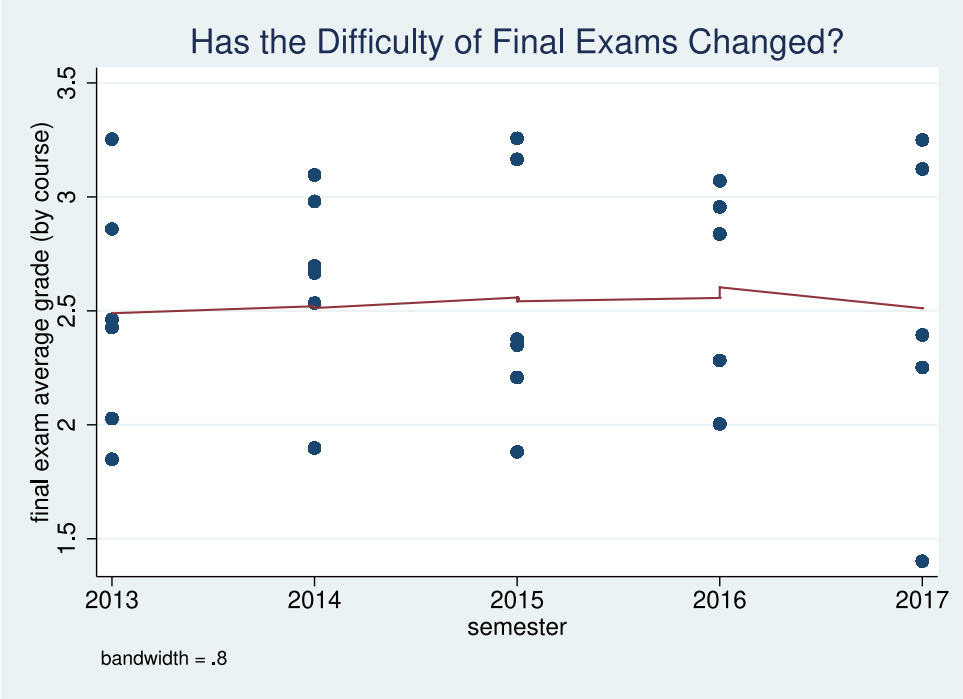

Fig. 3. Average final exam grades over time in the core.
Note: Figure shows average grades for final exams for each course in each academic year (blue dots) and the trend in the overall average (red line).

Table 1
Final exam grade frequencies.

(Core courses, fall semester)					
Final exam grade	Academic year				
	2013	2014	2015	2016	2017
A	21.7	22.2	24.7	25.3	26.9
B	31.1	33.6	30.9	32.5	32.3
C	27.7	25.4	24.0	25.0	21.1
D	13.3	12.7	13.5	11.8	11.9
F	6.2	6.2	6.9	5.4	7.9
Observations	3,247	3,205	3,342	3,281	3,529

Note: Table displays the distribution of final exam grades in our sample of fall semester core courses for each academic year, 2013–2017. Numbers in the table are in percentages.

core courses that are in the sample (and of course all models control for observable student ability as well). Therefore the identification strategy is not likely compromised, but we cannot completely rule out that the estimated exam timing effects may not apply to the small group of more advanced students if there is heterogeneity in response to treatment across student ability.

3.4. Descriptive statistics

The distribution of final exam grades tilts slightly towards better grades, however, a substantial percentage (17%–20%) of students receive grades of **D** and **F** on the final exam in any given semester. Table 1 shows the frequency of final exam grades during the fall semester for first-year students in core courses and indicates little change over time.

The information depicted in Fig. 3 supports the conjecture that grade inflation (or deflation) does not exist. The estimated coefficient from a simple regression of core course grades on a time trend is not significantly different from zero (with a p-value of 0.6). The non-parametric smoothing of the data also supports the assumption that no trend in average exam grades exists. That being said, estimates reported in the paper still include a time trend.[7]

Descriptive statistics for variables in the full sample and sub-samples used for robustness checks appear in Table 2.[8] The baseline samples used in our main models include 19,856 final exams taken over five fall semesters between the academic years of 2013 and 2017. The average course grade for students prior to taking the final exam (i.e., the end-of-classes grade) is slightly less than a **B**. For these courses, students pick-up an extra study day for approximately 40% of their exams (i.e., the variable "after Sunday break"), while students typically have roughly one full day to study between exams. Approximately 47% of exams occur in the morning time period, while 9% occur in the evening (but only for the Naval History course). The remaining 44% of final exams occur immediately after lunch at 1:30 in the afternoon. Other variables appear stable across the various samples. Roughly 25% of exams are taken by women, 26% by recruited athletes, 35% by minorities, 6% by students who were previously enlisted in the armed forces, and 19% by students who attended a preparatory school administered by the Navy. These fractions mirror the demographic make-up of typical USNA freshmen cohorts. The last two columns include descriptive statistics for samples used in robustness checks. Importantly, we note that sub-samples that exclude exams from non-core courses have students with on average lower observed academic ability. This is shown in the slightly lower SAT scores of the remaining students in these sub-samples. Ultimately our main findings are robust across the various samples, but as discussed above, we cannot completely rule out the existence of some unobserved heterogeneity in response to exam timing treatment across student types.

For robustness, we estimate some specifications that exclude data from 2014 (seen in the second column of Table 2), when an additional

[7] In most of these specifications, a small and negative time trend in final exam grades appears over the course of the sample.

[8] The Navy-specific core courses of Seamanship and Leadership have more **A** and **B** grades than other courses. Our analysis focuses on samples that exclude these courses.

Table 2
Descriptive statistics.

Variable	All courses	All courses (exclude 2014)	Core courses (exclude 2014)	Core courses (exclude "Honors" Calc 2, Naval History, and 2014)
percentage **A** grades	0.256	0.257	0.244	0.197
	(0.437)	(0.437)	(0.430)	(0.398)
percentage **B** grades	0.317	0.315	0.316	0.307
	(0.465)	(0.464)	(0.465)	(0.462)
percentage **C** grades	0.239	0.240	0.246	0.267
	(0.427)	(0.427)	(0.431)	(0.442)
percentage **D** grades	0.123	0.123	0.126	0.147
	(0.329)	(0.329)	(0.332)	(0.354)
percentage **F** grades	0.064	0.065	0.067	0.080
	(0.244)	(0.247)	(0.250)	(0.271)
after Sunday break[a]	0.380	0.400	0.398	0.423
	(0.485)	(0.490)	(0.490)	(0.495)
day of exam cycle	5.051	4.853	4.858	4.918
	(2.804)	(2.546)	(2.429)	(2.563)
days between exams	0.980	0.903	0.895	0.997
	(0.747)	(0.690)	(0.680)	(0.660)
morning exam[a]	0.470	0.454	0.459	0.559
	(0.499)	(0.498)	(0.498)	(0.496)
evening exam[a]	0.087	0.098	0.093	0.000
	(0.282)	(0.297)	(0.290)	–
week 6 grade	3.004	3.011	2.969	2.941
	(0.935)	(0.935)	(0.936)	(0.965)
week 12 grade	2.939	2.963	2.922	2.872
	(0.906)	(0.898)	(0.901)	(0.921)
end-of-classes grade	2.971	3.000	2.955	2.905
	(0.902)	(0.896)	(0.900)	(0.916)
SAT verbal (stdzd)	0.002	0.012	−0.044	−0.088
	(1.002)	(1.004)	(1.002)	(1.003)
SAT math (stdzd)	−0.019	−0.023	−0.098	−0.152
	(0.996)	(0.999)	(0.979)	(0.964)
recruited athlete[a]	0.261	0.260	0.272	0.279
	(0.439)	(0.439)	(0.445)	(0.447)
female[a]	0.251	0.259	0.262	0.264
	(0.434)	(0.438)	(0.440)	(0.441)
minority[a]	0.346	0.350	0.355	0.362
	(0.476)	(0.477)	(0.479)	(0.481)
prior enlisted service[a]	0.055	0.056	0.059	0.060
	(0.228)	(0.229)	(0.236)	(0.238)
attended Naval prep school[a]	0.188	0.191	0.207	0.216
	(0.391)	(0.393)	(0.405)	(0.412)
Observations	19,856	15,808	13,984	11,269

Note: Table displays summary statistics (sample means or proportions and their corresponding standard errors in parentheses) across the four samples used in the analysis.
[a]Indicates binary variables so the corresponding statistics are sample proportions rather than means.

and institutionally important extracurricular activity within student-life occurred that could potentially affect results: the annual Army-Navy football game. Because of this event, *two* Sunday breaks occurred in 2014. The first of these followed the day after the annual Army-Navy football game. Students in that fall semester had their first exam on a Friday, the football game was played on Saturday (with required attendance and no exams), and then students had an additional day off on Sunday. Exams were then scheduled Monday through Saturday of the following week, with an additional Sunday break, and then two more exam days the next week (thus exams took place across three distinct weeks).

We conduct balancing tests on the samples of student characteristics varied by Sunday breaks. An identification threat to this quasi-experiment would be if observed student attributes correlate with an unobserved variable that is also unexpectedly correlated with exam timing. For example, if standardized SAT scores (and an unobserved measure of ability) are for some unexpected reason correlated with exam timing, and if high ability students validate-out of Calculus 1, then the estimates of exam timing could exhibit a subtle bias attributable to the unobserved measure of ability. Additionally, the only course in our sample that has well defined variation across sections is Calculus 2: The vast majority of Calc 2 sections are considered "standard", while a small number each year (four or five in total) are considered "honors" or "supported".[9] To be extra cautious, we break out our balancing exercises (below) by section type of Calculus 2.

In these tests—shown in Table 3—the dependent variable is a dummy variable indicating whether exams occur after the Sunday break. In other words, we regress the treatment variable on the set of other covariates for each course. The last row of each column displays the p-value from and F-test for the joint significance of all student characteristic variables.

The treatment appears independent of student characteristics for most core courses. Only two coefficients are individually significant across the various models, which is to be expected given sampling variation. The overall F-tests' p-values in particular indicate that the advanced sections of Calculus 2 exhibit some non-random variation linked to exam timing. This implies that these sections had cohorts weighted with stronger students in semesters that took the exam before

[9] Recall that all sections in our sample are from the *fall semester*. Thus students taking Calculus 2 have passed a validation exam (testing out of Calculus 1). Beyond the validation exam, students have no input into their section assignment. Those who pass "with flying colors" are placed into an honors section, while those who marginally pass are placed into a section with additional support. All remaining validators land in standard sections. All section types take the *same final exam concurrently*.

Table 3
Balance tests of student characteristics.

	Core course							
Variable	Calc 1	Calc 2 standard	Calc 2 supported	Calc 2 honors	Chem 1	Cyber	Poli-Sci	Nav. Hist.
SAT verbal	−0.005	0.011	−0.010	0.047	−0.008	−0.008	−0.000	−0.008
	(0.013)	(0.020)	(0.031)	(0.050)	(0.009)	(0.011)	(0.014)	(0.012)
SAT Math	0.022	0.015	0.021	−0.168	−0.008	−0.016	0.011	0.002
	(0.016)	(0.022)	(0.033)	(0.053)	(0.009)	(0.012)	(0.013)	(0.011)
recruited	−0.007	0.069	0.010	0.072	−0.000	0.025	−0.030	−0.009
	(0.023)	(0.046)	(0.064)	(0.091)	(0.017)	(0.023)	(0.024)	(0.021)
female	0.024	0.015	0.066	−0.079	−0.012	−0.001	0.018	−0.020
	(0.023)	(0.039)	(0.054)	(0.077)	(0.016)	(0.021)	(0.024)	(0.019)
minority	−0.036	0.019	−0.085	0.165	0.002	−0.002	−0.018	−0.015
	(0.022)	(0.036)	(0.051)	(0.078)	(0.015)	(0.020)	(0.022)	(0.019)
prior enlisted	0.059	−0.081	−0.004	0.070	−0.013	0.067*	−0.028	0.030
	(0.043)	(0.067)	(0.092)	0.117	(0.034)	(0.039)	(0.045)	(0.040)
Naval prep school	0.056*	−0.029	−0.023	0.115	−0.019	0.009	0.036	0.017
	(0.029)	(0.058)	(0.068)	(0.107)	(0.021)	(0.028)	(0.030)	(0.028)
Observations	1841	747	374	211	4024	2332	1956	2506
$F-$stat.$p-$value	0.198	0.237	0.445	< 0.001	0.832	0.040	0.743	0.735

Note: Each column represents a linear probability model for specific core courses. Dependent variable is the treatment variable (i.e., exam after Sunday break); covariates include all relevant observable characteristics. Robust standard errors in parentheses.

the Sunday break. The Cyber security class cannot definitively reject the null-hypothesis for an F-statistic, but it is somewhat unclear as to why, since students cannot validate this class. This may be simply result from random variation in the sample. To be more confident that these trends are not driving our main results, we present robustness checks in the next section using sub-samples that exclude certain potentially problematic courses (see columns 3 and 4 of Table 2 for the corresponding summary statistics), as well as the academic year 2014 which experienced different timing effects described above.

4. ITS analysis

To determine the effect of a Sunday break on exam performance, we employ two approaches: graphical and non-parametric diagnostics and actual regressions. Fig. 4 depicts trends in exam grades based on binned local averages (by exam day), with smoothing defined across the period of the break as well. These non-parametric depictions set bandwidth sizes at 0.6 and 0.9 for comparison. The binned averages and trendlines illustrate what is likely some noise within the sample in the early days of the exam period because the only instances of exams scheduled on Day −3 (that is, three days before the universal break day) come from Calculus I and Calculus II. Thus a small sample of likely harder-than-average exams clustered on Day −3 causes the bin scatter for Fig. 4 to drop fairly substantially on that day. The figure is nevertheless suggestive of a trend-break following Day 0, and then a continuation of the trend into the final days of the exam cycle. Of course, this figure does not control for course difficulty or any other factors, but provide an important benchmark for discussion and motivation before we turn to better specified regression models.

Actual scores on final exams, y^*, are latent variables, but we observe the final exam letter grade for y_{ict} for each student i in course c during year t. Observed y increases in size from zero to four as the exam grades increase from an **F** to an **A**.

Regressions include controls for the exam time of day specified by the dummy variables $morning_{ct}$ and $evening_{ct}$. Morning exams begin at 7:55 a.m., and evening exams begin at 7:30 p.m. The reference-group exams begin at 1:30 p.m. The variable $rest_{ict}$ measures the amount of time students have to rest between exams (units are in days but can take fractional values). Other controls for student-specific variables appear in the vector **x** which includes: mid-semester course grades (6, 12, and end-of-classes grades), standardized SAT Math and Verbal scores and background characteristic controls for minority, gender, prior enlisted, prior attendance at Naval prep school and whether the student is a recruited athlete. The time $trend_t$ variable controls for possible grade inflation over the sampled years, while δ_c controls for any course-specific fixed effects.

Sunday breaks split the exam cycle assignments and serve as an exogenous interruption. The treatment is defined with the dummy variable $break_{ct}$. The variable d_{ct} represents the number of days that have passed since the exam cycle began for an exam in course c during year t. Let the Sunday break day within a particular exam cycle be defined by d_{bt}. The treatment variable, $break_{ct} = 1$, when $(d_{ct} > d_{bt})$, while $break_{ct} = 0$ when $d_{ct} < d_{bt}$. For example, if the exam period begin on a Wednesday, and a student takes an exam on Friday, then $d_{ct} = 3$ while $d_{bt} = 5$ (because Sunday—which has no exams—is the fifth day of the exam period). The running variable centers around the Sunday break and is thus $d_{ct} - d_{bt}$. For example an exam taken on a Saturday before the break would have $d_{ct} - d_{bt} = -1$, and an exam taken on the following Monday would have $d_{ct} - d_{bt} = 1$. In some additional specifications we employ an interaction variable, $break_{ct} * (day_{ct} - d_{bt})$, to model the possibility of alternative slopes after the break. As discussed in the data description above, students in 2014 received an additional Sunday break due to the Army-Navy football game. In models estimated on the sample including 2014 data, we include an additional dummy variable control for an exam that occurred on the first Friday of the exam period prior to the game. The only core course exam that occurred that year prior to the Army-Navy game was for the Naval History course.

The basic ITS model with constant slopes is specified as:

$$\begin{aligned} y_{ict} &= \alpha_0 + \gamma_1 break_{ct} + \gamma_2 \left(d_{ct} - d_{bt}\right) + \gamma_3 \left(d_{ct} - d_{bt}\right)^2 + \cdots \\ &\cdots + \alpha_1 morning_{ct} + \alpha_2 evening_{ct} + \alpha_3 rest_{ict} + \mathbf{x}'_{ict}\beta \\ &+ \delta_t trend_t + \delta_c + u_{ict} \end{aligned} \tag{1}$$

The key parameter γ_1 captures the treatment effect on exam performance from giving all students a common break. The parameters γ_2 and γ_3 determine the nonlinear and continuous effect of fatigue as time passes through the exam cycle. That being said, simply specifying the regression break does not necessarily identify a causal estimate. For instance, a non-linear relationship between the running variable and y_{ict} may underpin the model, and a break may not cause the change in y_{ict}. If this non-linearity exists and $\gamma_1 = 0$, then one expects to see a non-zero effect for γ_3.

4.1. Baseline ITS results

OLS estimates from baseline specifications that use two different samples (including and not including 2014) appear in Table 4. Each observation represents a student/course grade on the final exam. Hence,

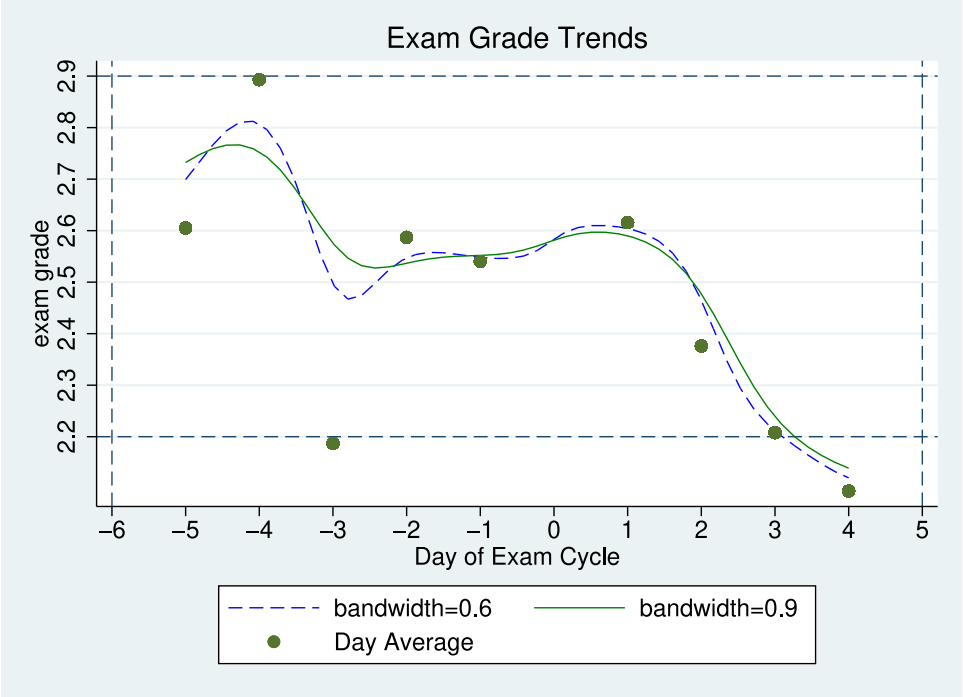

Fig. 4. Local linear regression of actual grades over exam cycles.
Note: Figure displays exam grade binned averages by day of the exam cycle (green dots), along with computed trends at two different bandwidths.

Table 4
ITS baseline estimates.

Variable	(I)	(II)	(III)	(IV)	(V)	(VI)
after Sunday break	0.230***	0.204***	0.294***	0.274**	0.222*	0.285***
	(0.079)	(0.043)	(0.094)	(0.106)	(0.118)	(0.093)
before Army-Navy game	−0.349***	–	−0.222*	–	−0.224	–
	(0.079)		(0.129)		(0.165)	
$d_{ct} - d_{bt}$	−0.169***	−0.120**	−0.147***	−0.091***	−0.149***	−0.107**
	(0.048)	(0.057)	(0.021)	(0.028)	(0.046)	(0.045)
$(d_{ct} - d_{bt})^2$	0.012*	0.006	0.008**	0.003	0.011**	0.004
	(0.005)	(0.005)	(0.003)	(0.003)	(0.004)	(0.005)
morning exam			−0.224***	−0.226**	−0.192**	−0.155*
			(0.129)	(0.092)	(0.085)	(0.087)
evening exam			−0.219***	−0.225***	−0.172***	−0.240***
			(0.042)	(0.063)	(0.058)	(0.073)
rest			0.056	0.086*	0.025	0.045
			(0.050)	(0.045)	(0.038)	(0.029)
Week 6 grade					0.076***	0.072**
					(0.023)	(0.034)
Week 12 grade					0.211***	0.206***
					(0.030)	(0.033)
Week 16 grade					0.454***	0.471***
					(0.020)	(0.021)
SAT verbal					0.055***	0.055***
					(0.015)	(0.013)
SAT math					0.132***	0.131***
					(0.028)	(0.028)
socioeconomic controls	no	no	no	no	yes	yes
2014 included	yes	no	yes	no	yes	no
overall R^2	0.003	0.009	0.010	0.013	0.441	0.435
graded observations	19,856	15,808	19,856	15,808	19,856	15,808

Note: Dependent variable is the ordering of exam grades where 0 represents an **F**, and 4 represents an **A**. Standard errors are clustered by course and shown in parentheses. Omitted/reference time slots are afternoon exams. All specifications include course fixed effects and an annual time trend.

the specifications including 2014 use data from 5828 students with 19,856 final exam grades. If we drop 2014 from the analysis, then we have 4665 students and 15,808 final exam grades.

Columns I and II of Table 4 report results for the specification with course fixed effects but without additional control variables. These results indicate that a Sunday break increases average exam performance by approximately one quarter of a letter grade relative to exams taken before the break. The initial break due to the Army-Navy football game actually appears to reduce performance prior to the break. This is not entirely surprising, since student time before the “big game” is

completely dominated by non-academic preparations for the game, and little studying time is available. Of course the power of this result is weak, since it only affects one iteration of one course, Naval History.

Upon including controls for exam time of day and idiosyncratic time off between exams (columns III and IV), results for the effect of the Sunday break remain strong and positive. The full set of controls are included in columns V and VI, which control for course performance heading into the final exam. These fully specified regressions indicate that the Sunday break improves performance by 0.22 to 0.29 of a letter grade. The Army-Navy game had a negative effect on exam performance prior to the game, but recall this occurred only one year and for one class.

The effect of time passing through the exam cycle appears through the quadratic functional form for the running variable, $d_{ct} - d_{bt}$. When including 2014 data, each additional day changes performance by $-0.15 + 0.022 * \left(d_{ct} - d_{bt}\right)$ (using column V results). The nonlinearity indicates that the negative impact of fatigue slowly diminishes over time. Perhaps this indicates how students get increasingly focused as the exam cycle passes. When 2014 data is excluded, the statistical significance of the non-linearity disappears and students simply perform worse by 0.11 of a letter grade for each passing day of the cycle.

Results for other time-related variables, such as the time of day for the exam, appear less consistent. Morning exams appear to have a negative impact on performance, reducing performance by about 0.16 to 0.2 of a letter grade. The impact of evening exams is more difficult to identify, since this also only occurred for Naval History and no other courses. Often students have more than one exam in a day (thereby giving them limited time to refresh and prepare), or sometimes they might get an extra day or two between exams. The *rest* variable tests whether more time between exams mitigates the effect of fatigue. In baseline specifications, *rest* between exams appears to have no effect on performance. Likely, the Sunday break variable conveys more prominent effects because, while most students enjoy an idiosyncratic rest day at some point, the Sunday break exists for all students in every final exam cycle and is typically near the middle of the period. The idiosyncratic rest days, on the other hand, are more variable and therefore more likely to occur at less impactful times, such as near the beginning or end of the exam period. This pattern, combined with likely multicollinearity between *rest* and other timing controls, likely supports our superior estimates of the Sunday break variable. Nevertheless, in the current analysis and the analyses below, if we assume based on the quasi-experimental setup that these additional exam timing variables are effectively randomly assigned to students, they should be uncorrelated with the error term and thus estimates of the α's, as well as γ_2 and γ_3, are plausibly causal in scope.

Other control variables generate consistently significant results across various specifications. Higher mid-term grades in courses not surprisingly have a positive effect on exam performance, as do both SAT math and verbal scores. Other control variables include dummy variables for recruited athletes, minorities, women, students who attended a remedial military prep school (NAPS) and students who served as enlisted sailors or marines (prior to the Academy). Overall estimates for the γ_j coefficients that relate to time and rest remain robust to the inclusion of this full set of socioeconomic controls and variables related to student ability (e.g. SAT scores, course mid-semester grades, etc.).

4.2. ITS results with different slopes

We generalize the specification in Eq. (1) to allow for the estimation of different rates of fatigue before and after the Sunday break. The ITS specification with different slopes before and after the break appears as:

$$\begin{aligned} y_{ict} \;=\; & \alpha_0 + \gamma_1 break_{ct} + \gamma_2\left(d_{ct} - d_{bt}\right) + \gamma_3\left(d_{ct} - d_{bt}\right)^2 + \cdots \\ & \cdots + \gamma_4\left(d_{ct} - d_{bt}\right) * break_{ct} + \gamma_5\left(d_{ct} - d_{bt}\right)^2 * break_{ct} + \cdots \\ & \cdots + \alpha_1 morning_{ct} + \alpha_2 evening_{ct} + \alpha_3 rest_{ict} \\ & + \mathbf{x}'_{ict}\beta + \delta_t trend_t + \delta_c + u_{ict} \end{aligned} \tag{2}$$

Table 5
ITS — Different slopes.

Variable	(I)	(II)	(III)	(IV)
break	0.667	0.734***	0.490***	0.328***
	(0.398)	(0.100)	(0.182)	(0.086)
before Army-Navy game	−0.164	–	−0.269	–
	(0.195)		(0.2444)	
$d_{ct} - d_{bt}$	−0.100	−0.301***	−0.060*	−0.100***
	(0.127)	(0.073)	(0.033)	(0.024)
$\left(d_{ct} - d_{bt}\right)^2$	−0.006	−0.032**	–	–
	(0.016)	(0.014)		
$\left(d_{ct} - d_{bt}\right) * break$	−0.140	0.087	−0.038	0.067*
	(0.210)	(0.167)	(0.032)	(0.039)
$\left(d_{ct} - d_{bt}\right)^2 * break$	0.035	0.069**	–	–
	(0.043)	(0.030)		
morning exam	−0.239***	−0.133	−0.241***	−0.146*
	(0.089)	(0.074)	(0.073)	(0.082)
evening exam	−0.264**	−0.302***	−0.252***	−0.247***
	(0.109)	(0.043)	(0.089)	(0.078)
rest	0.053*	0.051*	0.049*	0.050*
	(0.029)	(0.026)	(0.028)	(0.028)
student ability controls	yes	yes	yes	yes
socioeconomic controls	yes	yes	yes	yes
2014 included	yes	no	yes	no
overall R^2	0.450	0.422	0.452	0.430
graded observations	19,856	15,808	19,856	15,808

Note: Dependent variable is the ordering of exam grades where 0 represents an **F**, and 4 represents an **A**. Standard errors are clustered by course and shown in parentheses. Omitted/reference time slots are afternoon exams. All specifications include course fixed effects and an annual time trend.

The estimate for γ_1 represents the change in average exam grades from the break, while γ_4 and γ_5 capture the changes to these slopes after the break. These estimates are reported in Table 5, with columns III and IV assuming that $\gamma_3 = \gamma_5 = 0$. The effect of taking an exam z_b days prior to the break relative to z_a days after, in the purely linear model, is therefore given as $-(\gamma_1 + \gamma_2(z_b + z_a) + \gamma_4 z_a)$.

Column II allows for non-linear slopes on the running variable, while column IV assumes linearity. Note for interpretation that $\left(d - d_{bt}\right)$, which is equivalent to $(exam\ day - Sunday)$, has negative values for all days prior to the break and gets larger in absolute value as the exams get nearer to the initial day of exam cycle. Across the board, results support the hypothesis that a Sunday break has a positive impact on performance, while each day of the exam cycle slowly chips away at it. The non-linear specification exhibits larger declines over the cycle (larger slopes) and a larger initial jump in performance from the break.

The following thought exercise demonstrate these effects. For three-fourths of the sample, Sunday breaks occur six days into the exam cycle (while the others occur five days into the cycle). Using this upper limit (Sunday on sixth day) as a base, the average exam grade on the first day of an exam cycle (five days before the break) would be 0.2 higher than an exam taken one day after the break, but would be 0.4 higher than exams taken one day before the break.[10] The Sunday break clearly helps offset the fatigue of an ongoing exam cycle, essentially halving the grade loss from exam fatigue.

Following the same exercise and using nonlinear results in column II, the average exam taken on the first day of an exam cycle would be 0.234 higher than the average exam taken one day after the break. The average first exam would score 0.56 higher than an average exam taken one day before a Sunday break. By similar calculations, the average exam taken one day after the break would score 0.326 higher than the average exam taken before the break. The net gain at the margin of one day before relative to after the break is approximately one-third of a letter grade.

[10] The first difference is found from the coefficient results and subtracting $(-5 * -0.1) - (0.328 + (1 * 0.067) + (1 * 0.1))$. The second result is then more simply $(-5 * -0.1) - (-1 * -0.1)$.

Table 6
Sensitivity checks on core courses.

Variable	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)
break	0.896***	0.316**	0.895***	0.316**	0.873***	0.274**	0.873***	0.273*
	(0.113)	(0.110)	(0.115)	(0.112)	(0.085)	(0.111)	(0.094)	(0.113)
$d_{ct} - d_{bt}$	−0.272**	−0.093**	−0.272**	−0.092**	−0.307**	−0.096**	−0.307**	−0.094**
	(0.095)	(0.028)	(0.096)	(0.029)	(0.090)	(0.028)	(0.090)	(0.029)
$(d_{ct} - d_{bt})^2$	−0.027*	–	−0.028	–	−0.032**	–	−0.032**	–
	(0.015)		(0.096)		(0.012)		(0.012)	
$(d_{ct} - d_{bt}) * break$	−0.195	0.049	−0.195	0.045	−0.123	0.062	−0.124	0.057
	(0.247)	(0.043)	(0.256)	(0.046)	(0.222)	(0.041)	(0.226)	(0.043)
$(d_{ct} - d_{bt})^2 * break$	0.120***	–	0.120***	–	0.119***	–	0.119***	–
	(0.030)		(0.033)		(0.022)		(0.024)	
morning exam	−0.182**	−0.218**	−0.184*	−0.225*	−0.155	−0.217*	−0.156	−0.225*
	(0.094)	(0.091)	(0.096)	(0.094)	(0.107)	(0.093)	(0.110)	(0.097)
evening exam	−0.293***	−0.203**	−0.292***	−0.201*	–	–	–	–
	(0.069)	(0.085)	(0.070)	(0.088)				
rest	0.036*	0.042*	0.035	0.042	0.048	0.055	0.047	0.054
	(0.025)	(0.028)	(0.025)	(0.028)	(0.028)	(0.031)	(0.028)	(0.031)
drop Calc 2 (honors)	no	no	yes	yes	no	no	yes	yes
drop Naval History	no	no	no	no	yes	yes	yes	yes
clusters	9	9	8	8	8	8	7	7
observations	13984	13984	13773	13773	11480	11480	11269	11269

Note: Dependent variable is the ordering of exam grades where 0 represents an **F**, and 4 represents an **A**. Standard errors are clustered by course and shown in parentheses. Omitted/reference time slots are afternoon exams. All specifications include course fixed effects and an annual time trend. All specifications include student ability and socioeconomic controls.

Results for morning exams indicate that performance suffers in the morning by about 0.2 of a letter grade (depending on the specification). While statistically significant and robust, each additional day of rest between exams only increases performance by 0.05 of a letter grade, holding the rest of the exam timing variables constant. These results, while now more precisely estimated, remain small likely due to the reasons discussed above.

4.3. Sensitivity

Table 6 reports the results from sensitivity checks using alternative sub-samples that exclude academic year 2014 and all non-core classes (the majority of which are taken by advanced students).[11] Some samples additionally exclude Naval History or honors sections of Calculus 2 (see Table 2 for summary statistics of these sub-samples). However models using these sub-samples only exclude 2014 data when Naval History is included in the regression; when Naval History is excluded, the samples include all years of data.[12] All estimates in Table 6 include the same control variables and course fixed effects as in column (II) of Table 5.

Estimates do not change in any meaningful manner. Most importantly, the running variable, Sunday break and the associated interaction terms, yield very similar results to those reported in Table 5. The main differences are larger estimates of γ_1 in the nonlinear specifications. For the variable measuring days between exams (*rest*), the coefficients also do not show any meaningful change. Most fail statistical significance tests, but the coefficient magnitudes are robust. The same can be said for the variable measuring morning exams, where coefficients are not statistically significant for most two-tailed tests. Naval History is the only course with evening exams, indicating that the results for the evening exam coefficient in earlier tables lack statistical power from extremely limited variation (yet some are still statistically significant and negative).

[11] The percentage of non-core finals taken by first-year students ranges from 9 to 16.

[12] Recall that Naval History is the only course that was affected by the short exam week prior to the Army-Navy game and subsequent Sunday.

5. Conclusion

We estimate the causal effects of the timing of final exams and find consistent evidence that fatigue results in lower grades as the exam period proceeds. The deterioration due to fatigue, however, can be offset by anticipated breaks between exams. Specifically, we identify the causal effect of a Sunday break—common to all students—-from exams to be at least 20% of a letter grade, holding other timing considerations constant. These results are robust; in models that allow for non-linear fatigue effects the effect of the break on final exam performance may be much larger. Furthermore, exams taken in the morning, rather than the afternoon, exhibit a large (approximately 20%) drop in the odds of a better grade. In addition to the effects of a break in the final exam period that is common to all students, we also find that each day-long break within a student's idiosyncratic exam schedule leads to superior exam performance by about 0.05 of a letter grade.

The implication of these results is two-fold. First, college faculty may wish to consider the timing of their final exams when they assess learning outcomes in their courses primarily from final exams, especially when timing changes semester to semester. Second, administrators who schedule exams for an entire institution may consider adding an extra study day in the middle of final exam periods, clustering more exams in the afternoon when possible, or moving the start time of the first morning exam. Of course this could mean that final exam blocks of time extend a few more days (which could affect students and faculty in other ways), but overall performance on final exams may increase by these relatively low cost scheduling changes. While policy implementation may be complicated in venues where students' schedules are more diverse than students at the U.S. Naval Academy, in general our results support simple but important guidelines such as, "more breaks will help students perform better", and "students' exam scores are lower in the morning and late into the exam period". At the very least, if such reforms are challenging to implement on a university-wide scale, administrators could target particular disciplines or courses deemed especially important.

Acknowledgment

All authors approved version of the manuscript to be published.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Brady, R. R., Insler, M. A., & Rahman, A. S. (2017). Bad company: Understanding negative peer effects in college achievement. *European Economic Review*, *98*, 144–168.

Carrell, S. E., Fullerton, R. L., & West, J. E. (2009). Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor Economics*, *27*, 439–464.

Carrell, S. E., Maghakian, T., & West, J. E. (2011). A's from zzzz's? The causal effect of school start time on the academic achievement of adolescents. *American Economic Journal: Economic Policy*, *3*(3), 62–81.

Cotti, C., Gordanier, J., & Ozturk, O. (2018). Class meeting frequency, start times, and academic performance. *Economics of Education Review*, *62*, 12–15.

Diette, T. M., & Raghav, M. (2018). Do GPAs differ between longer classes and more frequent classes at liberal arts colleges? *Research in Higher Education*, *59*(4), 519–527.

Dills, A. K., & Hernández-Julián, R. (2008). Course scheduling and academic performance. *Economics of Education Review*, *27*, 646–654.

Goulas, S., & Megalokonomou, R. (2020). Marathon, hurdling, or sprint? The effects of exam scheduling on academic performance. *The B.E. Journal of Economic Analysis & Policy*, *20*(2).

Keller, K. M., Lim, N., Harrington, L. M., O'Neill, K., & Haddad, A. (2013). *The mix of military and civilian faculty at the united states air force academy: Finding a sustainable balance for enduring success*. Santa Monica, CA: RAND Project Air Force.

Lusher, L., Yasenov, V., & Luong, P. (2019). Does schedule irregularity affect productivity? Evidence from random assignment into college classes. *Labour Economics*, *60*, 115–128.

Lyle, D. (2007). Estimating and interpreting peer and role model effects from randomly assigned social groups at West point. *The Review of Economics and Statistics*, *89*(2), 289–299.

Lyle, D. (2009). The effects of peer group heterogeneity on the production of human capital at West point. *American Economic Journal: Applied Economics*, *1*(4), 69–84.

Pope, D., & Fillmore, I. (2015). The impact of time between cognitive tasks on performance: Evidence from advanced placement exams. *Economics of Education Review*, *48*, 30–40.