Economics of Education Review 48 (2015) 30–40

Contents lists available at ScienceDirect

Economics of Education Review

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

The impact of time between cognitive tasks on performance: Evidence from advanced placement exams

Devin G. Pope[a,*], Ian Fillmore[b]

[a] Booth School of Business, University of Chicago, United States
[b] Economics Department of University of Chicago, United States

ARTICLE INFO

Article history:
Received 5 February 2013
Revised 10 April 2015
Accepted 20 April 2015
Available online 28 May 2015

JEL:
I20
D03

Keywords:
AP tests
Cognitive fatigue

ABSTRACT

Students are often required to perform several mental tasks in a short period of time, and their performance is likely to depend on how closely the tasks are scheduled. We examine this phenomenon in a particular context: Advanced Placement (AP) exams in the United States. We exploit variation in the AP exam schedule from year to year which results in a student who takes two exams in one year having more or less time between the exams than a student who takes the same two exams in a different year. We find evidence that more time between exams results in higher scores, particularly on the second exam, and that this effect varies across different types of students. Our estimates suggest that a student taking two exams ten days apart is 6–8% more likely to pass them both than a student taking the same exams only one day apart.

In educational settings, students are often faced with many projects and tasks that demand their attention. These competing demands require them to make careful trade-offs as to where they devote their time and energy, especially when they have two tasks scheduled close together. For physically demanding tasks, it is clear that the amount of time between them can significantly affect performance—running two consecutive miles is much harder than running two miles with a rest period in between. In fact after some physical events, such as ultra-marathons, athletes need several weeks of recuperation before they can return to peak performance (Chambers, Noakes, Lambert, & Lambert, 1998). However, it is less clear how time between cognitive tasks will affect performance. For example, imagine a student who must take two difficult exams a few days apart. Will her performance decrease because the exams are scheduled close together? Or is a one or two day separation enough to allow the student to prepare properly and return to peak mental acuity?

These questions are difficult to answer with observational data because the scheduling of tasks is endogenous.[1] A person who receives an assignment or volunteers to complete tasks that are scheduled close together may be very different from a person who does not. People may also organize their schedules to avoid having difficult tasks scheduled close together. Selection bias in both the types of tasks and the people who complete them can result in misleading conclusions about the importance of time between tasks on performance.

We identify the causal effect of time between cognitive tasks on performance by exploiting a novel natural experiment made possible by the timing of Advanced Placement (AP) exams. In May of each year, hundreds of thousands of high-school students in the United States take AP exams administered by the College Board. For most students, these exams are the culmination of a year of study in an AP course

* Corresponding author at 5807 S. Woodlawn Ave, Chicago, IL 60637, United States. Tel.: +1 773 702 2297.
E-mail address: devin.pope@chicagobooth.edu (D.G. Pope).

[1] These questions are difficult to answer in laboratory settings as well since the experiment would necessarily have to run multiple days and require large incentives to motivate survey participants.

http://dx.doi.org/10.1016/j.econedurev.2015.04.002
0272-7757/

intended to be comparable to college-level work.[2] Each year the College Board fixes an exam schedule which applies to all students in the country, and we exploit the fact that this schedule changes from year to year. We analyze administrative data for a 10% sample of all AP exam takers in the United States between 1996 and 2001 who took exactly two exams in the same year. Our sample consists of thousands of students who took the same two AP exams but who differed in the amount of time between those exams. We use this exogenous variation in the time between exams to identify its causal impact on exam performance.

Our results indicate that performance significantly improves with more days between exams. Increasing the number of days between exams from 1 to 10 improves the combined point total on the two exams, which ranges from 2 to 10, by approximately 0.11–0.14 points (0.05–0.07 standard deviations) and the probability of passing both exams by 6–8%. Rather remarkably, within the range of our data this relationship is essentially linear, which means that increasing the time between exams from 1 to 3 days has a similar impact on performance as going from 8 to 10 days. The effects that we find are stronger for some subgroups (e.g. females and Asians) than for others. We also find that the estimates are driven almost entirely by an increase in performance on the second exam.

There are several potential underlying mechanism for the effects that we find including cognitive fatigue or differences in the ability of students to make use of last-minute preparation time. Our data are unable to fully distinguish between these underlying mechanisms, although they do provide some clues. In the final section, we discuss these mechanisms in more detail.

Our findings contribute to large bodies of work in psychology exploring cognitive fatigue, cognitive load, and memory recall.[3] Cognitive or mental fatigue has a rich tradition in psychology (e.g. Ebbinghaus, 1896–1897; Offner, 1911). Studies have focused on the impact of fatigue on the ability to process information (Sanders, 1998), on future effort (Meijman, 2000), and on mood fluctuations (Broadbent, 1979; Holding, 1983). Much of this work has focused on the impact of task length (e.g. total exam time) on average performance. For example, Ackerman and Kanfer (2009) provide a nice review. They argue that the evidence is inconclusive regarding the impact of exam length on performance and produce empirical results that actually find that performance can increase with exam length. Overall, the evidence suggests that while cognitive fatigue may not immediately hurt automated tasks, it can have a sharp impact on more complex tasks (Holding, 1983; Kuhl and Goschke, 1994).

Related to mental and cognitive fatigue is the literature on "cognitive load" and memory (see Paas, Renkl, & Sweller (2004) and Cepeda, Pashler, Vul, Wixted, and Rohrer (2006) for related reviews). Cognitive load theory is based on the idea that working memory is limited and that performance, reasoning, and learning degrades as the working memory fills up. How short and long-term memory works has been the study of hundreds of cognitive psychologists and a thorough review of this literature is beyond the scope of this paper. As we discuss in the conclusions section, we do not attempt to test a particular underlying mechanism or cognitive theory for our findings, but rather we focus on the overall impact of time between tasks on performance in the particular domain of AP test taking.

The paper is organized in the following way. In Section 1, we provide background information about the Advanced Placement exam program and discuss the data that we use in our study. In Section 2, we lay out our empirical strategy. We report our results in Section 3, and we conclude with a discussion of our findings and their broader implications in Section 4.

1. Advanced placement exams and data

In May of each year, Advanced Placement (AP) exams are administered to high-school students by the College Board (the same company that administers the SAT college admissions exam). For most students, these exams are the culmination of a year's worth of study in an AP course intended to be comparable to college-level work. In 2013, more than 2.2 million students took at least one AP exam, resulting in over 3.9 million total exams taken.[4] Exams are currently offered on 34 different subjects and include both multiple-choice and free-response sections. They are graded by college professors and other individuals with expertise in the subject who are employed and trained by the College Board. Each exam is given an integer score from 1 (lowest) to 5 (highest), with the cutoffs for each number determined freshly every year for each subject exam. Students are highly motivated to perform well on these exams for at least two reasons. First, high scores on AP exams are thought to impress college admissions committees. Of equal importance, many colleges and universities offer college credit for passing marks (a score of 3 or higher) on AP exams.

We obtained administrative data for a 10% random sample of all AP exam takers from 1996 through 2001.[5] We restrict the sample to students who took exactly two exams in the same year, which results in 238,138 AP exams taken by 119,069 students. Table 1 lists the AP exams taken by the students in our dataset, ordered by subject popularity. United States History, English Language, English Literature, and Calculus were the most popular exams. Very few students took Physics C: Electricity and Magnetism, French Language and Culture, or Latin.

Table 2 provides basic summary statistics for the students in our sample. More than 80% are high-school seniors. The average AP exam score for these seniors is lower than the average score for juniors and sophomores, suggesting that

[2] There are currently 33 exams, each covering a different subject area such as Calculus, Chemistry or European History.

[3] In economics, our paper relates to work by Coviello, Ichino, and Persico (2010) on multitasking. They show that Italian judges who were randomly assigned to work on several trials in parallel spent more time than if they did the trials one after the other. There is also work in behavioral economics that explores the impact that time-inconsistent preferences can have on performance when there are varying amounts of task separation (Ariely and Wertenbroch (2002) and see DellaVigna (2009) for a review of this literature)

[4] This information was obtained from the College Board's website on Oct. 25, 2013. http://media.collegeboard.com/digitalServices/pdf/research/2013/Number-of-Exams-per-Student-2013.pdf

[5] We thank the College Board for making these data available to us.

Table 1
AP exams: Ordered by popularity.

Exam	# of students	% of students	Avg score
US history	45,684	38.37	2.98
English literature	32,941	27.67	3.04
English language	28,761	24.15	3.02
Calculus AB	25,914	21.76	3.01
Biology	19,685	16.53	3.19
Chemistry	13,207	11.09	2.88
Gov't & politics: United States	10,443	8.77	2.72
Spanish language	9176	7.71	3.41
European History	8340	7.00	3.21
Physics B	6474	5.44	2.83
Psychology	5479	4.60	3.24
Calculus BC	4913	4.13	3.52
Statistics	3990	3.35	2.78
Economics: macroeconomics	2831	2.38	2.72
Frenchl	2657	2.23	2.77
Computer science A	2226	1.87	2.79
Economics: microeconomics	2176	1.83	2.71
Art history	1849	1.55	3.14
Environmental science	1575	1.32	2.78
Spanish literature	1557	1.31	3.20
Physics C: mechanics	1417	1.19	2.80
Gov't & politics: comparative	1246	1.05	2.63
Computer science B	1173	0.99	3.34
Art studio: general	979	0.82	3.18
Latin: vergil	877	0.74	3.01
Music theory	760	0.64	3.38
German language	589	0.49	3.12
Art studio: drawing	454	0.38	3.37
Latin literature	445	0.37	2.68
French literature	286	0.24	3.22
Physics C: electricity & magnetism	34	0.03	3.35
Total	238,138	200.00	3.02

Notes: All averages and frequencies are for our sub-sample of students who took exactly two exams. This subsample came from a 10% random sample of all AP exam takers for the years 1996–2001. Not all exams were offered in every year.

there is positive selection on juniors and sophomores who take two AP exams in a given year. 55% of the students in our sample are female and 66% are white. Black and Hispanic students are underrepresented and receive lower scores on average. Approximately 10% of students had more than a week between exams, most had 2–7 days, and 18% had fewer than two days.

No matter the subject, each exam is offered in either a morning or afternoon session sometime during a two-week (10 exam days) period every May. All students around the entire country taking a particular exam are required to take it on a commonly specified day and time.[6] The appendix contains the AP exam schedules for the years 1996 to 2001. The College Board takes multiple factors into consideration when scheduling the exams; for instance, it tends to spread out popular exams over the two-week period. While the schedule for a given year is fixed for all students, it does vary somewhat from year to year, and this variation is central to our empirical strategy.

2. Empirical strategy

Estimating the causal impact of time between tasks on performance from observational data can be difficult. In most situations, time between tasks is likely to be endogenous. Selection bias in both the types of tasks and the types of people who complete them can provide misleading evidence about how the time between tasks affects performance. The AP exam program provides an ideal context in which to test the impact of time between two tasks on performance. By exploiting the year-to-year variation in the schedule of AP exams, we are able to isolate plausibly exogenous variation in the time between exams.

Our empirical strategy involves comparing students who took the same two AP exams in different years. For example, in 1998 the Calculus AP exam was offered on Friday of Week 1 and the United States History exam was offered on Monday of Week 2. In 1999, the Calculus exam was offered on Thursday of Week 1 and the United States History exam was offered on Friday of Week 1. Thus, students who were tested on these subjects in 1998 had three days between exams but in 1999 had only one day between exams. By comparing students' scores in 1998 with those in 1999, we can estimate the causal effect of going from 3 days to 1 day between exams.

In Appendix Table 2a and b, we list the exam schedules for all AP tests between 1996 and 2001. While the exam

[6] Students must take the exam during its scheduled session. If a student is taking two exams that are scheduled at the same time, she may take one of the exams during a make-up period several weeks later. For this reason we exclude any students who took exams that were scheduled at the same time.

Table 2
Summary statistics.

	Number	Percent	Avg score
Class			
Senior	95,637	80.3	2.97
Junior	22,514	18.9	3.23
Sophomore	918	0.8	3.51
Gender			
Female	65,539	55.0	2.95
Male	53,530	45.0	3.11
Race			
White	79,055	66.4	3.09
Asian	15,739	13.2	3.04
Hispanic	9046	7.6	2.75
Black	4,652	3.9	2.24
Other	10,577	8.9	3.12
Days between exams			
0	4239	3.6	2.88
1	16,660	14.0	3.03
2	21,800	18.3	2.97
3	16,557	13.9	3.05
4	16,192	13.6	3.06
5	12,385	10.4	3.12
6	11,975	10.1	3.04
7	7816	6.6	2.98
8	6199	5.2	2.94
9	3703	3.1	2.99
10	1308	1.1	3.16
11	235	0.2	3.16

Notes: All statistics are for our sub-sample of students taking exactly two exams, which is based on a 10% random sample of all AP exam takers. Days between exams indicates the number of full days between the two exams taken (e.g. 0 indicates that the two exams were taken on the same day).

schedules are fairly stable from year to year, changes are made frequently enough to provide useful variation. To get a sense for how many exam pairs are used in our identification strategy, consider the 632 total exam pairs. Of these 632 pairs, 226 of them had the same number of days between the two exams in all six years of our data. Therefore, students who took the two exams associated with one of these 226 exam pairs do not provide variation that is useful for our analysis. However, 406 of the 632 exam pairs had at least one change in the number of days between the two exams over our data period and thus provide useful variation for identification. In addition. 217 of the 632 exam pairs had more than one change in the number of days between the two exams over our data period and thus provide even more variation.

The baseline model that we estimate in this paper is

$$Y_{ijt} = \beta D_{jt} + \gamma X_i + \theta_j + \alpha_t + \varepsilon_{ijt} \tag{1}$$

where Y_{ijt} is an outcome variable (e.g. exam score) for student i taking exam-pair j in year t. D_{jt} is the number of days separating the two exams taken by student i in year t, X_i is a set of student-level controls, θ_j are exam-pair fixed effects,[7] and α_t are year fixed effects.

Our key identifying assumption is that year-to-year variation in the time between exams *within an exam-pair* is exogenous. We argue that this variation, which is a result of the College Board changing the exam schedule from year to year, creates a credible natural experiment. It is not entirely clear why they tinker with the exam schedule, but discussions with the College Board suggest that the changes are unlikely to be related to student characteristics. For example, a representative of the Board in a private correspondence indicated that a major reason that exam dates change from year to year is the introduction of new exams and the elimination of exams that are no longer offered. These additions and eliminations can lead to a reshuffling of the exam schedule—which is the source of the identifying variation used in our analysis.

Even if the changes made by the College Board are unrelated to student characteristics, one potential concern is that students react to these changes by selecting into or out of taking an exam-pair based on the exam schedule. Our findings could be biased if the better (or worse) students in any given year decide not to take an exam-pair because they are scheduled very close together. We do not expect this to be a problem for several reasons. Many other factors will likely play a more dominant role in affecting which AP exams a student takes such as the student's interests and talents as well as the (often limited) selection of AP courses offered at her high school. Moreover, taking an AP course is valuable in its own right, so that students are still incentivized to take AP courses independent of the exam schedule. And given that a student has already taken the course, the exam schedule is unlikely to affect her decision to take the exam.

We also provide two pieces of empirical evidence that suggest that this type of selection is not driving our findings. First, we demonstrate in the Results section that time between exams has a large and significant effect on the second exam a student takes but not on the first. Although simple self-selection (better or worse students choose not to take two exams when in close proximity) could explain why students score better or worse on both exams, selection effects alone are unable to explain a systematic difference in performance on the second exam relative to the first.

Second, we can directly test whether more students sign up to take exams that are scheduled further apart. To do this, we aggregate the data to the (exam-pair $\times$ year) level and regress the log of the number of students taking a given exam-pair on the number of days between the two exams while controlling for exam-pair fixed effects. We find little evidence that days between exams impacts the number of students taking the two exams.[8]

It is also worth considering whether the manner in which AP exams are graded could bias the results. Specifically, one might worry that these exams are graded on a "curve" and that if everyone does very poorly one year, the strictness of the grading is simply changed to ensure that a relatively stable percentage of people pass the exam. The College Board claims not to base its scores off of a strict curve. Rather, their stated goal is to match exam performance with how a college

[7] We define an exam-pair to be two exams that were given in a particular order—that is, an ordered exam-pair. For example, suppose that in 1996 U.S. History was offered earlier than Calculus, but in 1997 their order was reversed. We define these to be two distinct exam-pairs.

[8] Regressing the log number of exam takers on days between exams while including year and exam-group dummies gives a coefficient of −0.0018 and a robust standard error of 0.0062.

Table 3
The effect of days between exams on exam outcomes–OLS.

	First exam score + Second exam score			Number of exams passed		
Days between exams	0.016		0.009	0.008		0.006
	(0.004)***		(0.005)*	(0.002)***		(0.002)***
2–3 days between		0.040			0.02	
		(0.024)*			(0.009)**	
4–5 days Between		0.032			0.0192	
		−0.03			(0.011)*	
6–7 days between		0.077			0.0427	
		(0.033)**			(0.013)***	
8–11 days between		0.120			0.0575	
		(0.038)***			(0.015)***	
Female			−0.236			−0.075
			(0.012)***			(0.005)***
Sophomore			0.757			0.211
			(0.070)***			(0.023)***
Junior			0.459			0.152
			(0.016)***			(0.006)***
Hispanic			−0.862			−0.377
			(0.024)***			(0.009)***
Black			−1.582			−0.600
			(0.029)***			(0.012)***
Asian			−0.204			−0.090
			(0.018)***			(0.007)***
Other race			0.018			−0.008
			−0.021			−0.008
Exam group fixed effects	X	X	X	X	X	X
Year fixed effects	X	X	X	X	X	X
Day of week fixed effects			X			X
Average scores of other students		X			X	
R-squared	0.054	0.054	0.097	0.037	0.037	0.078
Observations	1,19,069	1,19,069	1,19,069	1,19,069	1,19,069	1,19,069

Notes: This table presents coefficients and robust standard errors from the OLS regression of total exam score on both exams (first exam score + second exam score) in Columns 1–3 and the number of exams passed (a score of 3 or higher) in Columns 4–6 on the days between exams and in some specifications additional controls. Each regression includes fixed effects for the two exams being taken by each student and year fixed effects.
* $p < .10$;
** $p < .05$;
*** $p < .01$.

freshman would have done in the typical corresponding college course. Thus, an AP score of 5 corresponds to an A, a 4 to a B, and so forth. Indeed, score distributions differ considerably across different subjects and across years in ways that are inconsistent with the College Board adhering to a simple “curve.” Furthermore, a curve would merely attenuate our results. If two exams are scheduled closer together this year than last, causing everyone this year to do worse than last year, the curve would automatically eliminate those differences. However, we do not expect attenuation to be a major concern because the curve for any given exam will largely be set by many students taking only one exam (the vast majority of AP exam takers do not take multiple exams).

3. Results

Table 3 contains the first set of results based on the model specified in Eq. (1). The first three columns of Table 3 use the combined total of the first and second exam scores as the dependent variable. Because scores on any individual exam range from 1 to 5, the dependent variable for these three columns ranges from 2 to 10. Column 1 contains the estimates from our baseline specification. We find a positive and statistically significant relationship between exam scores and days between exams. The coefficient suggests that having 1 more day between exams leads to a higher combined exam score of 0.016. Thus, increasing the number of days between exams from 1 to 10 increases the combined point total on the two exams by approximately 0.144 points (0.07 standard deviations).

The specification in Column 1 assumes a linear relationship between days between exams and total exam scores. It is possible that students benefit greatly from having a day or two between exams but benefit little beyond that. We test for this kind of nonlinear relationship in Column 2 by including dummy variables for the number of days between exams. The omitted category for this regression is that the exams were taken on the same day (very rare) or 1 day apart. The results suggest that students who take exams that are 2–5 days apart score 0.03–0.04 points higher than students with 0–1 days between exams. Exams taken 6–7 days apart yield a 0.08 increase relative to the omitted category and exams taken 8–11 days apart yield a 0.12 increase relative to the omitted category. Thus, these results provide support for an approximately linear relationship between days between exams and exam score outcomes. We further explore any nonlinear

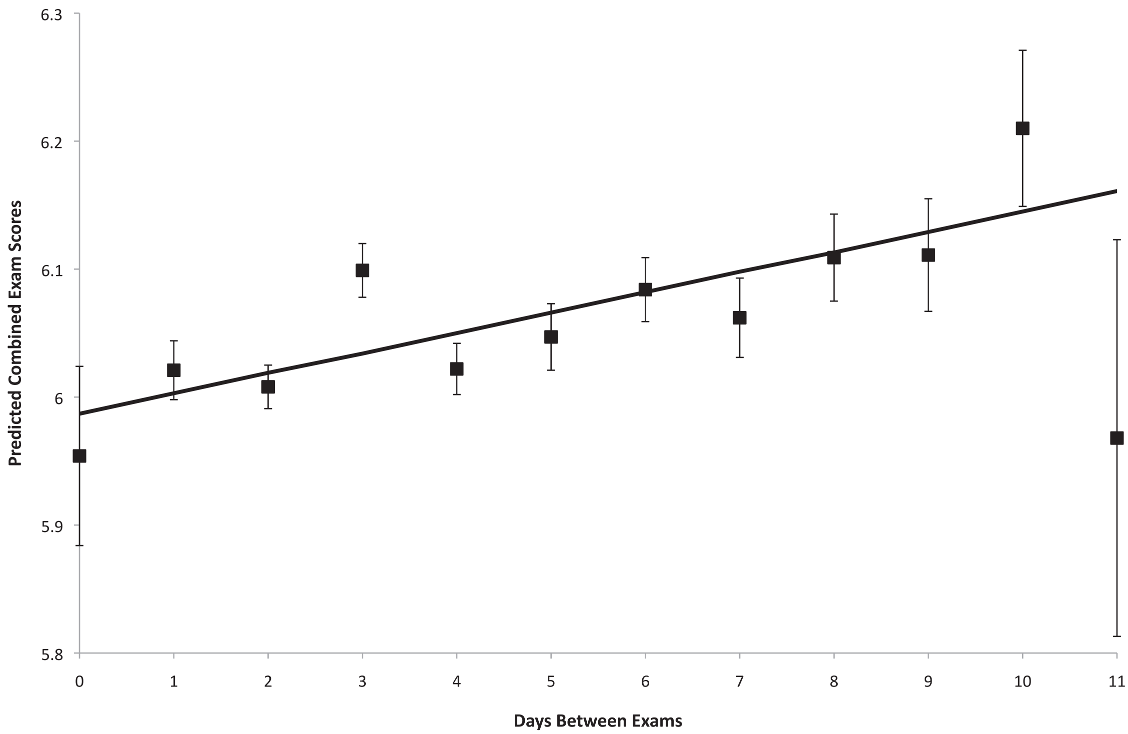

Fig. 1. Predicted values by days between exams. Each dot in this figure is the predicted combined exam score for the two AP exams when the number of days between the two exams varies from 0 to 11. These predicted values were obtained from estimating a regression similar to Column 1 of Table 3, but with a separate dummy variable for each possible day between exams. Standard error bars are also included. The linear line plots the predicted values obtained from the regression in Column 1 of Table 3. In both cases, the constant comes from normalizing the exam-pair fixed effects to have mean zero (default in Stata).

relationship that might exist between days between exams and combined exam score by estimating a model with a dummy variable for every possible number of days between exams (0–11). In Fig. 1, we plot the fitted values and standard error bars for the combined exam score by the number of days between exams. We also plot the linear relationship estimated in Column 1 of Table 3. Once again, the relationship between the number of days between exams and exam scores appears to be remarkably linear.

In Column 3 of Table 3 we again estimate the impact of days between exams on the combined exam score but also control for demographic variables (gender, race, and grade), dummies for the days of the week the exams were administered, and the average scores received by non-two-exam-takers (i.e. all students who took either one exam or three or more exams). Including this last covariate is a way of controlling for the “difficulty” of an exam in a particular year. The estimated effect of days between exams on performance shrinks somewhat when these controls are added and is only marginally significant.[9]

Although the combined exam score is one indication of performance on the two exams, perhaps a more important outcome to students themselves is whether they pass the exams. Typically, a student must achieve a score of 3 on an exam in order to pass and receive college credit. In Columns 4–6 of Table 3, we use as our dependent variable the number of exams passed, which can be equal to 0, 1, or 2. Going from one to ten days between exams increases the number of exams passed by 0.05–0.07, and this effect is statistically significant in all three specifications.

In Table 4, we estimate the same specifications as those found in Table 3, but do so using an ordered probit model, which is perhaps more appropriate than OLS given the discrete and ordered nature of the dependent variables.[10] Qualitatively, we find very similar results to those found in Table 3 using OLS. The coefficients of an ordered probit model lack a natural interpretation, and one common suggestion (e.g. Woolridge, 2002 , p. 506)) is to interpret the results by focusing on the predicted probabilities that the coefficients generate. For example, using the specifications in Column 4 and 6 of Table 4, we can calculate the probability of passing zero, one, or both exams for a given number of days between exams. We report these probabilities in Table 5. The probability of passing both exams when there is one day separating the two exams is 49.0–49.3% depending on the

[9] The drop in coefficient size from Column 1 to Column 3 of Table 3 due to the inclusion of control variables is not statistically significant. Thus, this drop could be due to statistical noise. However, the reduction in magnitude could also be a result of selection that is being accounted for by the control variables. For example, the average scores of other students could account for selection if students who took two exams happened to be better, relative to other exam takers, when the two exams had more days between them.

[10] In Appendix Table A1, we also present results from a Bivariate Probit Model. Once again, we find results that are consistent with those found using OLS.

Table 4
The effect of days between exams on exam outcomes–ordered probit.

	First exam score + Second exam score			Number of exams passed		
Days between exams	0.008		0.004	0.011		0.008
	(0.002)***		−0.003	(0.003)***		(0.003)***
2–3 days between		0.018			0.032	
		−0.012			(0.014)**	
4–5 days between		0.013			0.030	
		−0.015			(0.017)*	
6–7 days between		0.035			0.064	
		(0.017)**			(0.019)***	
8–11 days between		0.056			0.087	
		(0.019)***			(0.022)***	
Female			−0.122			−0.116
			(0.006)***			(0.007)***
Sophomore			0.398			0.369
			(0.037)***			(0.044)***
Junior			0.238			0.237
			(0.008)***			(0.010)***
Hispanic			−0.461			−0.548
			(0.013)***			(0.014)***
Black			−0.853			−0.867
			(0.016)***			(0.018)***
Asian			−0.107			−0.136
			(0.010)***			(0.011)***
Other race			0.006			−0.010
			−0.011			−0.012
Exam group fixed effects	X	X	X	X	X	X
Year fixed effects	X	X	X	X	X	X
Day of week fixed effects			X			X
Average scores of other students			X			X
Pseudo-R-squared	0.013	0.013	0.025	0.019	0.019	0.039
Observations	1,19,004	1,19,004	1,19,004	1,19,004	1,19,004	1,19,004

Notes: This table presents coefficients and robust standard errors from the ordered probit regression of total exam score on both exams (first exam score + second exam score) in Columns 1–3 and the number of exams passed (a score of 3 or higher) in Columns 4–6 on the days between exams and in some specifications demographic characteristics. Each regression includes fixed effects for the two exams being taken by each student and year fixed effects.

* $p < .10$;

** $p < .05$;

*** $p < .01$.

Table 5
Predicted probability of passing exams using ordered probit estimates.

		Number of days between exams			
		Baseline specification		Full controls	
		1 (%)	10 (%)	1 (%)	10 (%)
Number of exams passed					
	Zero	21.8	19.0	21.5	19.6
	One	29.2	28.1	29.1	28.3
	Two	49.0	52.9	49.3	52.2

Notes: This table provides the predicted probabilities of passing a given number of exams (0, 1, 2) with a given number of days between exams (1, 10) using the ordered probit specifications from Columns 4 and 6 of Table 4.

specification. With 10 days separating the two exams, the probability increases to 52.2–52.9%. Thus, the probability of passing both exams increases by approximately 6–8% (2.8–3.9% points) when we go from 1 to 10 days between exams.

A natural question is whether having more time between exams differentially affects the first exam relative to the second exam taken. We explore this question by running our baseline and full controls specifications (Columns 1 and 3 of Table 3) separately for the first and second exams. The results in Table 6 show that having more time between exams has no significant effect on the first exam, but it has a large and significant effect on the second exam. In fact, these results suggest that almost the entire combined effect in Table 3 is driven by the second exam.

Table 6
The effect of days between exams on the scores of the first and second exam.

	Exam 1 score	Exam 1 score	Exam 2 score	Exam 2 score
Days between exams	0.003	0.000	0.013	0.009
	(0.002)	(0.003)	(0.003)***	(0.003)***
Exam group fixed effects	X	X	X	X
Year fixed effects	X	X	X	X
Gender, race and grade controls		X		X
Day of week fixed effects		X		X
Average scores of other students		X		X
R-squared	0.060	0.087	0.064	0.104
Observations	1,19,069	1,19,069	1,19,069	1,19,069

Notes: This table presents coefficients and robust standard errors from the OLS regression of the score received on the first exam taken (Column 1), the second exam taken (Column 2), and the score ratio of exam 1 and exam 2 on the days between exams Each regression includes fixed effects for the two exams being taken by each student and year fixed effects. $^{*}p < .10$; $^{**}p < .05$; $^{***}p < .01$.

Table 7
OLS estimates of the effect of days between exams on total score: By subgroup.

	Gender	Gender	Race	Race	Grade	Grade	Gender * race	Gender * race
Days between exams	0.008	0.002	0.017	0.011	0.011	0.009	0.009	0.003
	(0.005)	(0.006)	(0.005)**	(0.006)	(0.005)*	(0.005)	(0.006)	(0.007)
Female * days	0.014	0.013					0.014	0.013
	(0.005)***	(0.005)***					(0.006)**	(0.006)**
Asian * days			0.023	0.024			0.015	0.015
			(0.008)***	(0.008)***			(0.012)	(0.012)
Black * days			−0.014	−0.014			0.016	0.016
			(0.012)	(0.012)			(0.021)	(0.021)
Hispanic * days			−0.038	−0.040			−0.041	−0.042
			(0.009)***	(0.009)***			(0.013)***	(0.013)***
Other * days			−0.009	−0.010			−0.001	−0.003
			(0.009)	(0.009)			(0.013)	(0.013)
Soph * days					0.017	0.021		
					(0.032)	(0.032)		
Junior * days					0.002	−0.001		
					(0.008)	(0.007)		
Female * asian * days							0.014	0.014
							(0.016)	(0.016)
Female * black * days							−0.047	−0.046
							(0.025)*	(0.025)*
Female * hispanic * days							0.004	0.004
							(0.015)	(0.015)
Female * other * days							−0.015	−0.012
							(0.017)	(0.017)
Exam pair fixed effects	X	X	X	X	X	X	X	X
Year dummies	X	X	X	X	X	X	X	X
Gender, race and grade controls	X		X		X		X	
Day of week fixed effects		X		X		X		X
Average scores of other students	X		X		X		X	
Observations	119069	119069	119069	119069	119069	119069	119069	119069
R-squared	0.058	0.097	0.086	0.097	0.061	0.097	0.089	0.097
F-stat (for interactions with Days)	7.94***	7.00***	7.96***	8.59***	0.16	0.23	5.05***	5.11***

Notes: This table presents coefficients and robust standard errors from the OLS regression of total exam score on both exams (first exam score + second exam score) on the days between exams and days between exams interacted with gender, race and grade dummies. The omitted categories are male, senior, and white. Each regression includes exam-pair and year fixed effects. Although not shown, each regression also includes main effects for the subgroup of interest.
* $p < .10$;
** $p < .05$;
*** $p < .01$.

In Table 7 we explore whether our estimates vary by demographic group. We interact days between exams with gender (Column 1–2), race (Column 3–4), grade (Column 5–6), and gender×race (Columns 7–8). Females benefit significantly more than males when the exams are farther apart; Asian and white students benefit most from having more time between exams; and we find little evidence for heterogeneity across grade levels. While the coefficient on the interaction between sophomores and days between exams appears to be large, it is imprecisely estimated due to the small number of sophomores in our sample. The last two columns of Table 7 allow for different effects by gender and race, and Fig. 2 plots the coefficients from these specifications. Again, Asian and white females (and perhaps Asian males) benefit the most from having more time between exams. Surprisingly, Hispanic males appear to do worse with more time between exams.

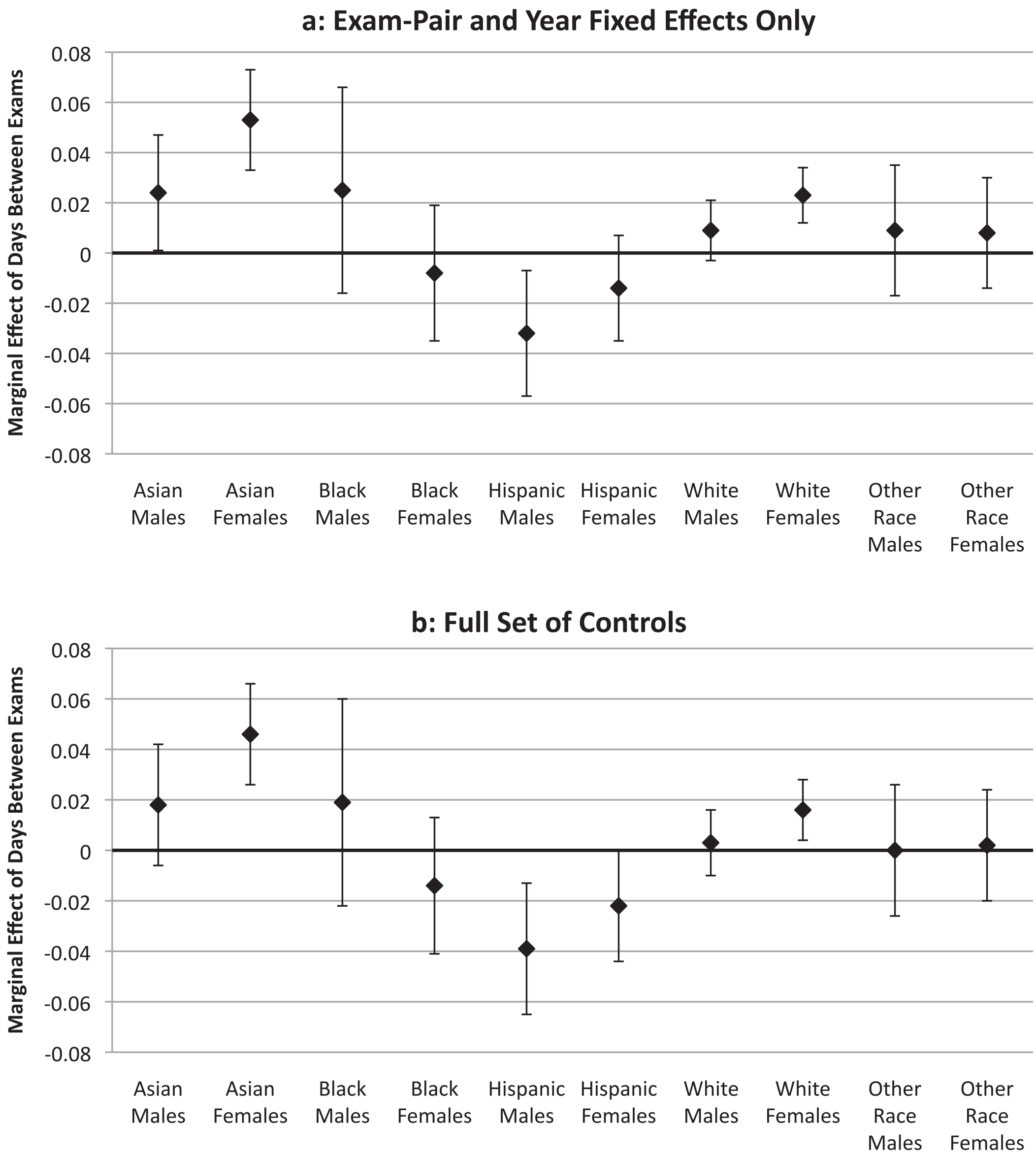

Fig. 2. Marginal effect of days between exams on total score – By race and gender. Each dot in this figure is the estimated marginal effect of Days between Exams on the combined exam score for each race-gender combination in our sample. The estimates come from the regressions in columns 7 and 8 in Table 7. The bars indicate 95% confidence intervals.

4. Conclusions

We use data on AP exam scores to estimate how the number of days between exams affects student performance. We find that a student who takes two exams does significantly better when they are further apart and show that this relationship is approximately linear within the range of our data (0–11 days). Some subgroups (e.g. females, Asians) are more sensitive to time between exams than others. Finally, most of the effect is concentrated on the second exam.

One could imagine various mechanisms for why more time between exams leads to better outcomes. For example, one possible explanation for our results is simple fatigue. Taking an AP exam is mentally and physically exhausting and it may be difficult to perform at peak ability when taking two exams in close succession. Another possible explanation is that last-minute preparation for exams ("cramming") is important and more difficult when exams are close together. A third and related mechanism is that when exams are close together, students foresee their possible fatigue or lack of cramming time, and preemptively allocate their energies to just one exam. Our data do not allow us to identify a specific mechanism behind our findings, but may provide some clues. For example, we find that the detrimental effect of temporally proximal exams is primarily associated with the second exam taken. Our fatigue mechanism predicts this effect. The cramming mechanism may also predict this effect, but not so directly. For example, if two exams are close together and a student has to do last-minute cramming for both exams at the same time, this could arguably affect both the first and the second test score. Thus, this evidence is suggestive of fatigue, but cannot rule out a cramming effect. On the other hand, the heterogeneity that we find across test takers is more suggestive in our minds of a cramming effect. Unless one thinks that fatigue operates differently across gender and race, the heterogeneities that we find seem to be more easily explained by which groups are more likely to engage in last minute preparation.

Independent of the exact mechanism for our effect, the findings we present have several direct implications. Testing agencies, students, teachers and parents can all benefit from understanding the impact that time between exams can have on performance and can use this knowledge to better plan test taking and other activities. While our results are clearly relevant to the population of AP test takers, it is possible that our findings generalize to test taking in general (although caution should be taken given that AP test takers are generally much stronger students than average). The results also speak to the sensitivity of test scores to ancillary factors. Even in a population of highly motivated students, the timing of exams has a non-trivial effect on performance. Finally, our results lead us to hypothesize that the timing of cognitive tasks may significantly affect how well individuals perform those tasks, even in contexts outside of education.

Appendix

Table A1, Table A2a, Table A2b.

Table A1
The effect of days between exams on passing – bivariate probit.

		Fail second exam (%)	Pass second exam (%)
1 day between exams:	Fail first exam	21.67	10.78
	Pass first exam	18.69	48.86
		Fail second exam (%)	Pass second exam (%)
10 days between exams:	Fail first exam	19.14	12.47
	Pass first exam	15.33	53.06

Notes: This table presents predicted probabilities from a bivariate probit model. The numbers in the boxes are the predicted probabilities of passing some combination of the first and second AP exams when there is one day between exams (upper box) or 10 days between exams (lower box).

Table A2a
AP examination schedules.

	1996 (May 6–10, 13–17)		1997 (May 5–9, 12–16)		1998 (May 11–15, 18–22)	
Week 1	Morning-8 a.m.	Afternoon-1 p.m.	Morning-8 a.m.	Afternoon-1 p.m.	Morning-8 a.m.	Afternoon—1 p.m.
Monday	French Lang.	Physics B Physics C	English Lit.	Computer Sci. A Computer Sci. AB	Economics: Microeconomics	Economics: Macroeconomics
Tuesday	Spanish Lang. Latin: Vergil Latin Lit.	English Lang.	Calculus AB Calculus BC	Statistics	French Lang. Computer Sci. A Computer Sci. AB	English Lang.
Wednesday	English Lit.	Music Theory	Spanish Lang.	English Lang.	Music theory	English Lit.
Thursday	Calculus AB Calculus BC	Psychology	Economics: Microeconomics	Economics: Macroeconomics	Spanish Lang.	Physics B Physics C
Friday	U.S. history	European history Studio art: portfolios due	U.S. history	European history Studio art: portfolios due	Calculus AB Calculus BC	Statistics Studio art: portfolios due
Week 2						
Monday	German Lang. Spanish Lit.	Chemistry	French Lang. Music theory	Physics B Physics C	German Lang. Spanish Lit.	U.S. history
Tuesday	Biology	Computer Sci. A Computer Sci. AB	Gov't & Politics: United States	Gov't & Politics: Comparative Chemistry	Gov't & Politics: United States	Gov't & Politics: Comparative European history
Wednesday	Gov't & Politics: United States	Gov't & Politics: Comparative	Biology	Latin: Vergil Latin Lit.	Biology	Environment Sci.
Thursday	Economics: Microeconomics	Economics: Macroeconomics	German Lang.	Psychology Spanish Lit.	Chemistry	Psychology
Friday	History of art	French Lit.	History of art	French Lit.	History of art Latin: Vergil Latin Lit.	French Lit. Spanish Lit.

Table A2b
AP examination schedules.

	1999 (May 10–14, 17–21)		2000 (May 8–12, 15–19)		2001 (May 7–11, 14–18)	
Week 1	Morning—8 a.m.	Afternoon—1 p.m.	Morning—8 a.m.	Afternoon—1 p.m.	Morning—8 a.m.	Afternoon—1 p.m.
Monday	German Lang.	English Lit.	French Lang.	Computer Sci. A Computer Sci. AB Statistics	English Lit.	German Lang. Physics C
Tuesday	French Lang.	Computer Sci. A Computer Sci. AB	Spanish Lang.	English Lang.	Spanish Lang.	Computer Sc. A Computer Sci. B
Wednesday	Spanish Lang.	English Lang.	English Lit.	German Lang.	English Lang.	French Lang. Int'l English Lang.
Thursday	History of art	Calculus AB Calculus BC	Calculus AB Calculus BC	History of art	Calculus AB Calculus BC	Art history
Friday	U.S. history	European history Studio art: portfolios due	U.S. history	European history Studio art: portfolios due	U.S. history Music theory	European history Studio art: portfolios due
Week 2						
Monday	Psychology Music theory	Chemistry	Gov't & Politics: United States	Gov't & Politics: Comparative Music Theory	Psychology Human geography	Chemistry
Tuesday	Gov't & Politics: United States	Gov't & Politics: Comparative Statistics	Chemistry	Psychology	Gov't & Politics: United States	Gov't & Politics: Comparative Environmental Sci.
Wednesday	Biology	Physics B Physics C	Biology	Physics B Physics C	Biology	Physics B Physics C
Thursday	Economics: Macroeconomics Environmental Sci.	Economics: Microeconomics	Economics: Macroeconomics Environmental Sci.	Economics: Microeconomics	Economics: Macroeconomics	Economics: Microeconomics Statistics
Friday	Latin: Vergil Latin Lit.	French Lit. Spanish Lit.	Latin: Vergil Latin: Lit.	French Lit. Spanish Lit.	Latin: Vergil Latin Lit.	French Lit. Spanish Lit.

References

Ackerman, P., & Kanfer, R. (2009). Test length and cognitive fatigue: an empirical examination of effects on performance and test-taker reactions. *Journal of Experimental Psychology: Applied, 15*(2), 163–181.

Ariely, D., & Wertenbroch, K. (2002). Procrastination, deadlines, and performance: self-control by precommitment. *Psychological Science, 13*(3), 219–224.

Broadbent, D. (1979). Is a fatigue test now possible? *Erognomics, 22*, 1277–1290.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: a review and quantitative synthesis. *Psychological Bulletin, 132*, 354–380.

Chambers, C., Noakes, T. D., Lambert, E. V., & Lambert, M. I. (1998). Time course of recovery of vertical jump height and heart rate versus running speed after a 90-km foot race. *Journal of Sports Sciences, 16*(7), 645–651.

Coviello, D., Ichino, A., & Persico, N. 2010. Don't spread yourself too thin: the impact of task juggling on workers' speed of job completion. NBER Working Paper #16502.

DellaVigna, S. (2009). Psychology and Economics: evidence from the field. *Journal of Economic Literature, 47*, 315–372.

Ebbinghaus, H. (1896-1897). (O. Wilhem, trans.) On a new method for testing mental abilities and its use with school children. *Zeitschrift fur Psychologie und Psysiologie der Sinnesorgane, 13*, 401–459.

Holding, D. (1983). Fatigue. In R. Hockey (Ed.), *Stress and fatigue in human performance* (pp. 145–164). Durham: John Wiley & Sons.

Kuhl, J., & Goschke, T. (1994). A theory of action control: mental subsystems, modes of control, and volitional conflict-resolution strategies. In J. Kuhl, & J. Beckmann (Eds.), *Volitional and personality: action versus state orientation* (pp. 93–124). Seattle, WA: Hogrefe and Huber.

Meijman, T. (2000). The theory of the stop-emotion: on the functionality of fatigue. In D. Pogorski, & W. Karwowski (Eds.), *Ergonomics and safety for global business quality and production* (pp. 45–50). Warschaw: CIOP.

Offner, M. (1911). *Mental Fatigue*. Baltimore: Warwick & York.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32*, 1–8.

Sanders, A. (1998). *Elements of human performance*. London: Lawrence Erlbaum Associates.

Woolridge, J. M. (2002). *Econometric analysis of cross section and panel data*. MIT Press.