The Effect of Final Exam Spacing on Student Performance

Shahar Sansani[a] **(corresponding author)**

Afik Rahamim[b]

a: School of Economics
College of Management, Academic Studies
7 Yitzhak Rabin Blvd.
Rishon Lezion, 75190
Israel
Email: sansanis@colman.ac.il, ssansani@gmail.com
Telephone: +972 50-797-5872

b: School of Economics
College of Management, Academic Studies
7 Yitzhak Rabin Blvd.
Rishon Lezion, 75190
Israel
Email: afik.ra@gmail.com
Telephone: +972 54-255-4211

Abstract:
In this paper, we examine the relationship between exam spacing and exam performance. Our approach exploits scheduling differences between two groups of undergraduate Economics students. The treatment group and the control group have similar exam spacing for one 'early exam', but the treatment group has four additional days between exams for another 'later exam'. We find that four more days of available study time is associated with an increase of 4.81 points (out of 100) on the final exam for females, while having no effect on the scores of males.

Keywords: Higher Education; Study Time; Gender; Exam Performance

JEL Codes: I23; J16

Electronic copy available at: https://ssrn.com/abstract=3012522

Introduction

Numerous studies in the economics of education literature focus on the predictors of student grades and exam performance. In this paper, we examine the effect of available study time before an exam on undergraduate economics students' exam scores. We measure 'exam spacing', or available study time, as the number of class-free and exam-free days students had before a final exam. The effect of available study time on student grades is difficult to estimate because it is hard to isolate it from other factors determining student success because students often set their own schedule. For instance, students with higher ability may be students who choose a more evenly spaced exam schedule or may take more classes, leading to less spacing between exams. To overcome this selection issue, we rely on a difference-in-differences (DiD) approach. We compare the difference in exam scores for two groups of students in one 'early exam' to the difference in scores on a 'later exam'. The two groups of students had equal amounts of time to study for the early exam, but one group had four more days to study for the later exam.

Our work is related to two strands of literature. The first is research that examines the effect of study time on student grades. Because we examine available study time immediately prior to an exam, we posit that this correlates highly with actual study time.[1] There are several papers that have used quasi-experimental approaches in order to bypass the endogeneity of study time and student grades. Stinebrickner and Stinebrickner (2008) find a significant positive relationship between study time and grade point average using the presence of video games in

[1] In their review of the 'cramming' (significant studying immediately before an exam) literature, McIntyre and Munson (2008) note that 'cramming generally describes the study strategy of more than 25% of the students by almost any general definition of the phenomenon.' (p. 3). Even if students do not significantly increase the amount of studying before exams, it is hard to argue that study time decreases before exams. In the case of the college examined, library hours are extended and extra office hours are offered in order to accommodate the noticeable increase in student studying.

Electronic copy available at: https://ssrn.com/abstract=3012522

one's dorm room as an instrument for study time. Bonesrønning and Opstand (2012) find a positive relationship between study time and test scores. They use student fixed effects and rely on within-student variation in study time, thus controlling for unobserved heterogeneity between students. Although the consensus in these two papers is that extra study time leads to higher grades, the results of other papers that do not control for unobserved student heterogeneity have been mixed. Some papers have found a positive relationship between study time and student outcomes (e.g. Stinebrickner and Stinebrickner 2004), but there are others that find no relationship (see, e.g. Schuman et al. 1985; Guillaume and Khackikian 2011), and others even a negative, albeit small, relationship between study time and student outcomes (e.g. Krohn and O'Connor 2005). Combined, these studies highlight both that the effect of study time on grades is ambiguous to a degree, and that the reason for the differential results found may be that unobserved factors across students need to be controlled for, which is what we do in our paper.

One of the differences between our control and treatment groups is that the treatment group includes a higher percentage of students who work.[2] Therefore, our study also addresses the literature on the effect on student grades of working while in school, where the evidence is also mixed. Darolia (2014) finds little impact of working either part-time or full-time on student performance, while Kalenkoski and Pabilonia (2010) find that more hours worked negatively affects students' grades. Because an increasing number of students work while in college (Scott-Clayton 2012), the impact of these additional time constraints need to be considered when

[2] For a different study, a survey of economics students at the College included a question on whether a student did not work, worked part-time, or worked full-time. In the evening group (treatment) 81 percent worked full-time, 17 percent part-time, and 2 percent did not work. In the morning group (control), 29 percent worked full-time, 51 percent worked part-time, and 20 percent did not work.

devising optimal education policy, so we take this trend of increased work among college students into account in the interpretation of our results.[3]

We control for the omitted variables that may bias the relationship between available study time and student grades by comparing the difference in scores on exams where two groups of students had the same number of lecture and exam free days before the exam, with the difference in scores on an exam in which one group had four more lecture and exam free days before the exam. Our key identifying assumption is that the only (significant) difference between the difference in scores on the early exam and the difference in scores on the late exam is that the treatment group had more available days to study for the late exam. Based on this assumption, the difference-in-differences (DiD) estimate represents the causal effect of extra available study time on student grades. We discuss possible violations of this assumption in a later section.

Our study adds to the existing literature in several ways. First, we examine a factor affecting student grades – available study time before an exam – that to our knowledge has not been studied directly. Exam spacing would be a very straightforward policy to implement. Second, our unique data set allows us to bypass, to a large extent, the omitted variables bias and sample selection issues that are evident in most studies examining the factors affecting student grades.

We do not find conclusive results for the effect of extra available study time on student exam scores as a whole. In our main estimation, we find that four extra days before the later exam are associated with a statistically insignificant 2.22 (out of 100) more points on the exam. However, when decomposing the results by gender, we find that the extra days of study

[3] Beyond the scope of this work is determining to what degree increased studying affects the long-term retention of course material and student performance in subsequent courses, as opposed to simply increasing the specific course grade.

increased female scores by 4.81 points while having no effect on male scores. In general, the point estimates for females are greater than the point estimates for males. Taken as a whole, our results suggest that there may be beneficial effects to relaxing the time constraint before exams (i.e., spacing out exams more). That said, some of the estimates are not statistically significant at conventional levels, and the results vary by academic year, so further research in different contexts is needed to elucidate this relationship.

Data and Methodology

Our data set is composed of first-year undergraduate students in the School of Economics from the 2012-2013 through 2015-2016 academic years at a large private college in Israel. We combine administrative data on students (gender, age and pre-college exams scores) with their final exam scores in three first-semester courses: Mathematics A (hereafter Mathematics), Principles of Microeconomics (hereafter Microeconomics), and Statistics A (hereafter Statistics).

Our estimation strategy relies on differences in final exam schedules. Students studying economics can choose between one of two schedules: morning or evening.[4] 'Morning' students study two semesters (fall and spring) per year over the course of three years, while 'evening' students study three semesters (fall, spring, and summer) per year over the course of three years. Evening students take less courses every semester, but the course requirements to obtain a degree are identical. For courses that occur during the same semester, like the three courses above, students take the same final exams at the same time.

[4] Within the morning and evening groups, students are also divided into different classes by the level of mathematics that they studied in high school. Students stay with the same peers for all classes, so peer effects do not play a role in our study.

Because ‘morning’ students take more courses per semester, their final exam schedule is denser. We exploit the fact that although the order of morning students’ final exam schedule during the years examined was: 1) Mathematics, 2) Microeconomics, 3) Principles of Marketing,[5] 4) Statistics, the start of evening students’ final exam schedule was: 1) Mathematics, 2) Microeconomics, 3) Statistics. Because the final exam for courses that are common to both schedules occur on the same day, both groups of students have the same time between the Mathematics and Microeconomics exam, but the evening students have more available time before the Statistics exam because they do not have the Principles of Marketing exam. Thus, we use the difference in scores on the Microeconomics exam as the baseline difference between the students of the two groups[6], and the difference in scores on the Statistics exam as the effect of the treatment. These courses are typical of the economics degree in the United States and Europe as most departments require Statistics as well as Calculus classes, which is similar to the Mathematics course (Siegfried and Walstad 2014; Monteiro and Lopes 2007).

Figure 1 depicts our identification strategy by showing the exact date of exams at the end of the fall semester of the 2014-2015 academic year (the time between exams in other years is identical) and the number of days between each exam.

In order to estimate the effect of available study time on student grades, we compare the grades of the two groups on the Microeconomics exam (where each group had 5 days before the exam) to the grades of the two groups on the Statistics exam, where the evening group had four more days before the exam. Formally, we estimate models of the following form:

[5] For the fall 2015 semester (exams occur at the start of 2016), the third exam for the morning students was Computer Applications for Economists instead of Principles of Marketing.

[6] In all of the years examined, the exam in Mathematics occurs the same number of days after the end of the semester. However, students may end the semester on different days, so using Mathematics to determine the pre-treatment difference between the two groups is not as sound as using the Microeconomics course. Therefore, we only use the Mathematics exam as a robustness check of the main result.

$$\text{Grade}_{ige} = \alpha + \beta_1\text{Evening}_g + \beta_2\text{Statistics}_e + \beta_3\text{EvStatistics}_{ge} + \varepsilon_{ige}$$

Where *i* represents each student, *g* represents whether the student is in the morning or evening group, and *e* represents the different exams. *Grade* represents the student's final exam grade in the three courses examined. *Evening* is a dummy variable representing whether the student is in the evening program (treatment), *Statistics* is a dummy variable representing whether the score is for an exam in statistics, and *EvStatistics* is a dummy variable representing an exam for an evening student in statistics. The coefficient of interest is β_3, which represents the DiD estimate of having extra class-free and exam-free days before the final exam. That is, we are interested in whether the difference in grades between the morning and evening groups on the Statistics exam (post-treatment) is different than the difference in grades on the Microeconomics exam (pre-treatment). The data we have on age and pre-college exam scores do not affect the results because these variables are differenced out in the regression – student covariates are identical pre-treatment and post-treatment.

We run the analysis for all students combined, and then by gender. Females perform better than males in all levels of education (Voyer and Voyer 2014) and are more motivated and disciplined than males (Meece and Painter 2008; Duckworth and Seligman 2006). As such, they may also be affected differently by having more days before an exam. Masui et al. (2014) find that once study time as well as other student characteristics are controlled for, gender does not have a consistent effect on student grades. With these studies in mind, the expected differential effect by gender of having more days before an exam is ambiguous. It may be that females take more advantage of the extra days of study, or perhaps they are affected less by the extra days of study before an exam because they study more consistently throughout the semester.

We limit our sample to the students who took all three exams in the evening group and all four exams in the morning group, in order to not have 'control' individuals in the treatment group. Most of the students who did not take all exams are students who transferred to the college from other institutions and therefore received exemptions from certain courses. That said, when including all exam takers, our results are qualitatively the same, both in terms of magnitude and statistical significance.

Results

Table 1 presents summary statistics depicting observable differences between our treatment (evening) and control (morning) groups. Several differences emerge between the two groups which necessitate the use of the DiD strategy. That is, we cannot simply compare their scores on the exam where the treatment group had more days before the exam. Rather, we need a baseline difference between the control and treatment groups. The treatment (evening) group relative to the control (morning) group is less male (48% to 57%), older (24.85 years versus 23.85 years), and has higher scores on pre-college exams. The evening group scored higher on the *Bagrut* exam, 90.13 to 88.93, and higher on the *Psychometry* exam, 544.4 to 526.5.[7] The evening group also studied Mathematics at a higher level in high school, with an average level of 3.66 versus 3.58 units (out of 5). The number of math units is slightly higher than the country-wide average of 3.49 (in 2014) (MOE 2016), but this is not surprising given that our sample

[7] The *Bagrut* and *Psychometry* are two country-wide pre-college exams. Students must score a minimum on at least one of these for college admission. The *Bagrut* exam allows students to pick the subjects they would like to be tested in and therefore is a less sound indicator of student ability. The *Psychometry* exam is comparable to the Scholastic Aptitude Test (SAT) in the United States and is therefore a better measure of student ability.

consists of economics students. The overall *Psychometry* score of 533 in our sample is equal to the country-wide average (in 2013) (NITE 2016).

Also in Table 1 are the average grades in the three courses used for the morning (control) and evening (treatment) group. The average grades are 72.88 and 69.34 for the morning (control) and evening (treatment) groups, respectively. Two possible explanations for the higher pre-college exam scores but lower grades in the evening group than the morning group are the evening group's work burden and familial responsibilities given their older age. The effect of giving the evening group more time before an exam (spacing out their exams more) is precisely what we are estimating in this paper.

In Table 2 we depict the differences in grades on the Microeconomics exam where both groups had the same number of days before the exam, and the Statistics exam where the treatment group had more days before the exam. Columns 1 through 3 depict all years combined. We find that extra days of study lead to a 2.22, statistically insignificant, increase in scores. Because there are four extra days of studying for the evening group versus the morning group, this comes out to 0.55 points per day. The difference between the two groups on the control exam is a statistically significant 4.12 points (p-value = 0.003), while the difference on the treatment exam is a statistically insignificant 1.91 points (p-value = 0.121). When examining the results independently by year (columns 4 through 15), we find that while the results are noisy, in all cases the DiD estimate is positive.

In Tables 3 and 4 we depict the results by gender. In Table 3 are the results for males. We find a DiD estimate that is very close to zero and statistically insignificant. This is not surprising given that the DiD estimate is positive in two years and negative in the other two years. Overall, males score 2.68 points higher on the pre-treatment exam and 2.51 points higher on the post-treatment exam. In Table 4, we examine the effect of extra available days before an exam on females. Here we find a statistically significant (p – value = 0.087) DiD estimate of 4.81 points. There is a difference of 4.71 points on the pre-treatment exam, and this difference is erased on the post-treatment exam. In all four years the DiD estimate is positive and all point estimates are at least 2 exam points.[8] The overall effect for females decreases to 3.34 and is not statistically significant when the 2014-2015 academic year is removed, although the result holds when other years are omitted. We examine the difference in exam scores between the morning and evening groups in the 2014-2015 academic year over their entire first two years of study and find that grade differences between the two groups persist. Therefore, these large differences in females' first semester scores in this academic year are not an anomaly.

Robustness Checks

Our identifying assumption is that the difference in scores between the two groups on the Microeconomics exam is the counterfactual for the difference in scores on the Statistics exam, absent any difference in available study time. Here, we address a number of issues in order to strengthen our identifying assumption.

The first issue is that the grades in the treatment and control groups are in two different courses. Perhaps the Statistics exam is an easier exam and therefore the difference in scores would have been smaller regardless of one group having more available time before the exam than the other group. Although we cannot rule this out completely, we do not believe that this explains the whole story. We test this by using the Mathematics exam as an alternative measure of the pre-treatment difference in scores, instead of Microeconomics, because mean Mathematics

[8] The overall higher scores for males is in line with the findings of Krohn and O'Conner (2005).

scores, 74.3, are closer to the mean scores in Statistics, 79.2. As can be seen in Figure 1, the Mathematics exam is the first exam students take at the end of the semester. It does not represent as sound a control group as the Microeconomics exam because it is hard to determine precisely how much time students had to study for the Mathematics exam. That is, there are not going to be distinct differences in the number of days before these two exams as there is between the Mathematics and Microeconomics exams. We find a nearly identical point estimate as in Table 2. In Table 2, using Microeconomics as a pre-treatment difference in scores, the point estimate is 2.22, and using Mathematics (Table 5, column 3), the point estimate is 2.75. In the two years in which the grades in Mathematics were either on par with the Statistics scores (2014) or higher than the Statistics scores (2015), the DiD estimate, while not statistically significant, is a positive, 3.36 points (Table 5, column 6).

Additionally, we compare the average percentiles of scores for the morning and evening groups on the Microeconomics and Statistics exams in order to determine whether the student class ranks in the control and treatment groups changed. These results are in Table 5, columns 7 through 9. On the Microeconomics exam, where both groups had the same number of lecture- and exam-free days before the exam, the average percentile for the morning group was 50.94 and the average percentile for the evening group was 44.68. On the Statistics exam, where the evening group had more available days before the exam, the average percentile for the morning group was 50.28 and the average percentile for the evening group was 47.23. The evening group is increasing their average percentile of grades relative to the morning group. The difference in average percentile on the pre-treatment exam is 6.26, and 3.05 on the post-treatment exam. This is an indication that scores are not just increasing for both groups, but that the evening group is increasing their scores relative to the control group.

A second issue that is not controlled for is differences in instructors. Perhaps our results are explained by instructors for the evening group being of *higher* quality than the instructors for the morning group in Statistics (post-treatment exam), while the instructors in the evening group for the Microeconomics exam (pre-treatment exam) are of *lower* quality than the morning group. Ideally, we would run a model with instructor fixed effects. However, including instructor fixed effects reduces the identification to only being off instructors who taught both morning and evening groups. If an instructor has a stronger class in the morning than in the evening, or vice versa, then the DiD estimate is not based on the same students. What we can do is note the significant, although incomplete, overlap many instructors have in terms of teaching both morning and evening groups. For instance, in the fall semester of 2015, there were three lecturers for the post-treatment exam (Statistics). Two lecturers taught both a morning group and an evening group. The third lecturer taught a morning group in Statistics, and both morning and evening groups in Microeconomics. In addition, in order to have the morning and evening programs be as qualitatively similar as possible, department policy is to have generally the same rank of faculty, for instance with regard to academic degree, in both morning and evening groups.

A third issue is that students in the control group, planning ahead for the fact that they only have four days before the Statistics exam (post-treatment), may start studying for that exam throughout the final exam period and not only when they finish the last exam before the Statistics exam. Although this may be the case for some students, if anything, this would bias our results downwards. That is, their scores would have been lower, and the DiD estimate larger, if they had not started to study ahead of time. Unfortunately, we do not have data on when students started studying for each exam.

Discussion and Conclusion

Overall, our results show that four extra days of available study time do not lead to a statistically significant increase in student grades. However, when decomposing the results by gender, we find a statistically significant (p-value = 0.087) increase of 4.81 points for females. Females score almost five points less on the exam they had the same number of exam- and lecture-free days before (Microeconomics), and equalize their scores on the exam they had the same number of free days before (Statistics). For males, on the other hand, there was no effect – there is a difference of about 2.5 points on both exams.

We posit the following explanation for our results overall and the differential impact on males and females. The morning (control) group is comprised of students who have lower pre-college exam scores and work less than the evening (treatment) group. The morning group is also younger, on average, so presumably they may have less familial obligations as well. The extra available time the control group has overall may be the explanation for their overall higher exam scores, even though their college-entrance scores are inferior. The increase in available time for the treatment group relative to the control group ahead of the post-treatment exam may explain the decrease in the gap in scores for females. Previous studies have noted that males make less use of time for academic purposes than females (Brint and Cantwell 2010) and that females spend more time studying than males (Krohn and O'Conner 2005), which may explain why we find significant results only for females; they make better use of the extra study time than males.

Although our findings are suggestive of a policy that may impact students' final exam grades, overall they are not very precise, and not always robust to taking out one particular year. Further research is needed to address the robustness of the results to different populations and different exams. In the United States, where all final exams are typically condensed to between 5 and 10 days, extra spacing between final exams may be especially pertinent.

References

Bonesrønning, H., and L. Opstad. 2012. "How much is Students' College Performance Affected by Quantity of Study?" *International Review of Economics Education* 11 (2): 46-63.

Brint, S., and A. M. Cantwell. 2010. "Undergraduate Time Use and Academic Outcomes: Results from the University of California Undergraduate Experience Survey 2006." *Teachers College Record* 112 (9): 2441-2470.

Darolia, R. 2014. "Working (and Studying) Day and Night: Heterogeneous Effects of Working on the Academic Performance of Full-Time and Part-Time Students." *Economics of Education Review* 38: 38-50.

Duckworth, A. L., and M. E. Seligman. 2006. "Self-Discipline Gives Girls the Edge: Gender in Self-Discipline, Grades, and Achievement Test Scores." *Journal of Educational Psychology* 98 (1): 198-208.

Guillaume, D. W., and C. S. Khachikian. 2011. "The Effect of Time-on-Task on Student Grades and Grade Expectations." *Assessment & Evaluation in Higher Education* 36 (3): 251-261.

Kalenkoski, C. M., and S. W. Pabilonia. 2010. "Parental Transfers, Student Achievement, and the Labor Supply of College Students." *Journal of Population Economics* 23 (2): 469-496.

Krohn, G. A., and C. M. O'Connor. 2005. "Student Effort and Performance over the Semester." *The Journal of Economic Education* 36 (1): 3-28.

Masui, C., J. Broeckmans, S. Doumen, A. Groenen, and G. Molenberghs. 2014. "Do Diligent Students Perform Better? Complex Relations between Student and Course Characteristics, Study Time, and Academic Performance in Higher Education." *Studies in Higher Education* 39 (4): 621-643.

McIntyre, S. H., and J. M. Munson. 2008. "Exploring Cramming: Student Behaviors, Beliefs, and Learning Retention in the Principles of Marketing Course." *Journal of Marketing Education* 30 (3): 226-243.

Meece, J. L., and J. Painter. 2008. "Gender, Self-Regulation, and Motivation." In D. H. Schunk and B. J. Zimmerman (Eds.), *Motivation and Self-Regulated Learning: Theory, Research, and Applications* (pp. 339-367). New York: Taylor and Francis Group.

MOE (Ministry of Education) (2016). National Average of Bagrut Test Scores, 2013-2014. http://ecatmate.education.gov.il

Monteiro, H., and A. F. Lopes. 2007. "A Benchmarking of the Undergraduate Economics Major in Europe and in the United States." *International Review of Economics Education* 6 (2): 9-26.

NITE (National Institute for Testing and Evaluation), 2015. https://nite.org.il/index.php/he/research-conferences/statistics/graphs-1999-2013.html

Schuman, H., E. Walsh, C. Olson, and B. Etheridge. 1985. "Effort and Reward: The Assumption that College Grades are Affected by Quantity of Study." *Social Forces* 63 (4): 945-966.

Scott-Clayton, J. 2012. “What Explains Trends in Labor Supply among US Undergraduates?” *National Tax Journal* 65 (1): 181-210.

Siegfried, J. J., and W. B. Walstad. 2014. “Undergraduate Coursework in Economics: A Survey Perspective.” *The Journal of Economic Education* 45 (2): 147-158.

Stinebrickner, R., and T. R. Stinebrickner. 2004. “Time-Use and College Outcomes.” *Journal of Econometrics* 121 (1): 243-269.

Stinebrickner, R., and T. R. Stinebrickner. 2008. “The Causal Effect of Studying on Academic Performance.” *The BE Journal of Economic Analysis & Policy* 8 (1): Article 14.

Voyer, D., and S. D. Voyer. 2014. “Gender Differences in Scholastic Achievement: A Meta-Analysis.” *Psychological Bulletin* 140 (4): 1174-1204.

Table 1: Summary Statistics

	Number of Observations	
Year	Morn. (Cont.)	Even. (Treat.)
2013	218	88
2014	158	81
2015	135	69
2016	158	74
Total	669	312

	Comparison of Groups		
	Morn. (Cont.)	Even. (Treat.)	Diff.
Percent Male	0.57	0.48	0.09***
			(0.02)
Age	23.85	24.85	-1.00***
			(0.11)
Bagrut Score	88.93	90.13	-1.20***
			(0.36)
Psychometry Score	526.5	544.4	-17.90***
			(5.69)
Math Units	3.58	3.66	-0.08***
			(0.03)
Average Grade	72.88	69.34	3.53***
			(1.17)

Notes: The morning (Mor.) group is the control (Cont.) group and the evening (Even.) group is the treatment (Treat.) group. The control group includes students that study two semesters per year and the treatment group includes students that study three semesters per year. Therefore, the treatment group has more days between exams. *** significant at 0.01; ** at 0.05; * at 0.10. Standard errors in parentheses.

Table 2: Effect of Extra Exam Spacing on Final Exam Scores - All Students

	All Years Combined			2015-2016			2014-2015			2013-2014			2012-2013		
	Morn. (Cont.)	Even. (Treat.)	Diff.	Morn. (Cont.)	Even. (Treat.)	Diff.	Morn. (Cont.)	Even. (Treat.)	Diff.	Morn. (Cont.)	Even. (Treat.)	Diff.	Morn. (Cont.)	Even. (Treat.)	Diff.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Pre-Treatment Exam Score: Microecon.	63.01	58.89	4.12***	54.89	54.12	0.76	65.41	54.50	10.91***	66.08	65.73	0.34	65.20	60.01	5.19**
			(1.41)			(2.94)			(3.17)			(2.84)			(2.25)
Post-Treatment Exam Score: Statistics	79.81	77.91	1.91	80.28	81.66	-1.38	78.65	73.46	5.19*	78.98	79.50	-0.52	80.80	76.82	3.98**
			(1.23)			(2.74)			(2.66)			(2.66)			(1.91)
Diff-in-Diff			2.22			2.15			5.65			0.86			1.21
			(1.87)			(4.02)			(4.15)			(3.89)			(2.95)
Observations			1962			464			408			478			612

Notes: The morning (Mor.) group is the control (Cont.) group and the evening (Even.) group is the treatment (Treat.) group. The Difference-in-differences estimate is the coefficient β_3 in the following regression: $Grade_{ige} = \alpha + \beta_1 Evening_g + \beta_2 Statistics_e + \beta_3 EvStatistics_{ge} + \varepsilon_{ige}$. *** significant at 0.01; ** at 0.05; * at 0.10. Standard errors in parentheses.

Table 3: Effect of Extra Exam Spacing on Final Exam Scores - Male Students

	All Years Combined			2015-2016			2014-2015			2013-2014			2012-2013		
	Morn. (Cont.)	Even. (Treat.)	Diff.	Morn. (Cont.)	Even. (Treat.)	Diff.	Morn. (Cont.)	Even. (Treat.)	Diff.	Morn. (Cont.)	Even. (Treat.)	Diff.	Morn. (Cont.)	Even. (Treat.)	Diff.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Pre-Treatment															
Exam Score: Microecon.	64.56	61.88	2.68	58.36	58.68	-0.32	67.54	60.30	7.24	65.99	68.72	-2.73	65.74	58.80	6.94*
			(1.89)			(3.79)			(4.57)			(3.74)			(3.23)
Post-Treatment															
Exam Score: Statistics	81.96	79.45	2.51	81.86	84.51	-2.65	81.83	76.30	5.53	80.20	80.72	-0.52	83.26	76.00	7.26***
			(1.60)			(3.63)			(3.65)			(3.44)			(2.40)
Diff-in-Diff			0.17			2.34			1.71			-2.21			-0.32
			(2.48)			(5.24)			(5.85)			(5.09)			(4.02)
Observations			1046			230			196			268			352

Notes: The morning (Mor.) group is the control (Cont.) group and the evening (Even.) group is the treatment (Treat.) group. The Difference-in-differences estimate is the coefficient β_3 in the following regression: $Grade_{ige} = \alpha + \beta_1 Evening_g + \beta_2 Statistics_e + \beta_3 EvStatistics_{ge} + \varepsilon_{ige}$. *** significant at 0.01; ** at 0.05; * at 0.10. Standard errors in parentheses.

Table 4: Effect of Extra Exam Spacing on Final Exam Scores - Female Students

	All Years Combined			2015-2016			2014-2015			2013-2014			2012-2013		
	Morn. (Cont.)	Even. (Treat.)	Diff.	Morn. (Cont.)	Even. (Treat.)	Diff.	Morn. (Cont.)	Even. (Treat.)	Diff.	Morn. (Cont.)	Even. (Treat.)	Diff.	Morn. (Cont.)	Even. (Treat.)	Diff.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Pre-Treatment															
Exam Score: Microecon.	61.12	56.42	4.71**	51.50	49.57	1.93	63.43	49.31	14.12***	66.19	63.29	2.90	64.39	61.58	2.81
			(2.10)			(4.41)			(4.38)			(4.36)			(3.25)
Post-Treatment															
Exam Score: Statistics	77.15	77.25	-0.10	78.74	78.81	-0.07	75.70	71.22	4.48	77.33	79.89	-2.57	76.70	78.48	-1.78
			(1.85)			(4.10)			(3.84)			(4.01)			(3.00)
Diff-in-Diff			4.81*			2.01			9.65*			5.47			4.59
			(2.80)			(6.01)			(5.82)			(5.92)			(4.43)
Observations			916			234			212			210			260

Notes: The morning (Mor.) group is the control (Cont.) group and the evening (Even.) group is the treatment (Treat.) group. The Difference-in-differences estimate is the coefficient β_3 in the following regression: $Grade_{ige} = \alpha + \beta_1 Evening_g + \beta_2 Statistics_e + \beta_3 EvStatistics_{ge} + \varepsilon_{ige}$. *** significant at 0.01; ** at 0.05; * at 0.10. Standard errors in parentheses.

Table 5: Robustness Checks

	Using Mathematics Exam for Baseline Difference							Mean Percentile of Scores		
	All Years Combined			Condensed Sample						
	Morn. (Cont.)	Even. (Treat.)	Diff.	Morn. (Cont.)	Even. (Treat.)	Diff.		Morn. (Cont.)	Even. (Treat.)	Diff.
	(1)	(2)	(3)	(4)	(5)	(6)		(7)	(8)	(9)
Pre-Treatment Exam Score: Mathematics	75.80	71.21	4.60***	81.37	76.43	4.93***	Pre-Treatment Percentile: Microecon.	50.94	44.68	6.26
			(1.42)			(1.76)				
Post-Treatment Exam Score: Statistics	79.81	77.96	1.85	78.83	77.26	1.57	Post-Treatment Percentile: Statistics	50.28	47.23	3.05
			(1.23)			(1.88)				
							Observations	669	312	
Diff-in-Diff			2.75			3.36				
			(1.88)			(2.57)				
Observations			1962			886				

Notes: Table reports estimations using Mathematics exam for baseline difference in scores and the change in mean percentile of scores on the pre-treatment and post-treatment exams. Condensed sample includes only those years where the mean Mathematics A score was on par or higher than the mean Statistics A score. The morning (Mor.) group is the control (Cont.) group and the evening (Even.) group is the treatment (Treat.) group. The Difference-in-differences estimate is the coefficient β_3 in the following regression: $Grade_{ige} = \alpha + \beta_1 Evening_g + \beta_2 Statistics_e + \beta_3 EvStatistics_{ge} + \varepsilon_{ige}$. *** significant at 0.01; ** at 0.05; * at 0.10. Standard errors in parentheses.