Available study time and undergraduate student exam performance

Article · January 2019

DOI: 10.1108/JARHE-12-2017-0158

Shahar Sansani
University of the Pacific

Available Study Time and Undergraduate Student Exam Performance

Shahar Sansani[a] **Afik Rahamim**[b]

July, 2018

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 having four more days before the later exam is not significantly correlated with the score on that exam.

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

JEL Codes: I23; J16

Final version published as:

Sansani, Shahar, & Rahamim, Afik. (2019). "Available Study Time and Undergraduate Student Exam Performance", *Journal of Applied Research in Higher Education*, Vol. 11, No. 1, pp.: 20-35.

Purpose: The purpose of this paper is to determine whether the number of lecture-free and exam-free days before a final exam affects students' scores overall and by gender.

Methodology: The paper exploits scheduling differences in final exams between two groups of students who take identical final exams. 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', allowing for a difference-in-differences (DiD) analysis. A survey of contemporary students is conducted to complement the empirical results.

Findings: Overall, there are no statistically significant differences in the grades on the exams between the control group and the group that had four more study days. When examined by gender, the point estimate on females is large in magnitude but statistically insignificant at conventional levels (p-value = 0.087).

Research limitations: The study uses data on undergraduate students studying Economics in Israel. More research in other contexts is needed to determine the robustness of the findings.

Originality: This is the first paper to study the effect of the number of days students have between final exams on student final exam scores. The results can aid in determining optimal final exam schedules.

Introduction

Numerous studies in the economics of education literature focus on the predictors of student grades and exam performance. This paper examines the effect of available study time before an exam on undergraduate economics students' exam scores. Available study time represents the number of lecture-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 student characteristics determining student success. For instance, students with higher ability may be those students who are more concerned with higher grades and choose classes where mean grades are higher. To overcome this selection issue, this paper uses a difference-in-differences (DiD) approach to compare the difference in exam scores for two groups of students in one 'early exam' to the difference in scores on a 'later exam', where all students take the same exams. 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.

This 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. Changing the number of days between exams is a straightforward policy for higher education administrators to implement. If there is a positive relationship between final exam spacing and student grades, increasing the number of days between exams may lead to greater student mastery of the material. On the other hand, if no relationship is found, decreasing the number of days may be an option. This would enable longer breaks between semesters and or school years. This study is also unique in that our 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. The DiD strategy allows us to control for factors like differences in motivation and ability between students.

This work is related to several strands of literature. The first is research that examines the effect of study time on student grades. Because this paper examines available study time immediately prior to an exam, plausibly this correlates highly with actual study time. 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 per cent of the students by almost any general definition of the phenomenon.’ 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. In addition, our survey of students reveals that 85% of those that work take time off from work during the exam period in order to free up time for studying. 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 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 this paper does.

This work is also related to the literature on optimal study strategies for both short-term success on exams and long-term retention of course material. This literature has mostly examined the effectiveness of distributed practice (i.e. studying throughout the semester) versus massed practice (i.e. cramming before an exam). This literature has examined both the effect on immediate recall of material and also retention of the material for longer periods of time. In their review of the literature, Cepeda et al. (2006) show that while the relationship between the time between study sessions and recall of the material depends on when the material needs to be recalled, in general longer time between study sessions leads to greater retention. Most related to our work are the papers that have examined undergraduate students. Kapler, Weston, and Wiseheart (2015) find that students exhibit greater recall of material when they study eight days after initially learning the material versus one day after initially learning the material. Zulkiply et al. (2012) also find that spacing generates better results on the task performed in their experiment, although students believed that massed study would produce superior results. Rohrer and Taylor (2006) examine undergraduate students and show that mathematics material is better retained four weeks after being studied if students studied across two sessions rather than one session. These papers are related to our study because those students with less days before an exam have to plan ahead and begin studying earlier. As a result, they must space out their study time more. On the other hand, those students with more days before the exam can plan on studying all of the material in the days before the exam and therefore can rely more on cramming.

One of the differences between the control and treatment groups analyzed in this paper is that the treatment group includes a higher percentage of students who work, so this paper is also related to the research on the effect on student grades of working while in school since those students that work have more available study days before some of the final exams. 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 per cent worked full-time, 17 per cent part-time, and 2 per cent did not work. In the morning group (control), 29 per cent worked full-time, 51 per cent worked part-time, and 20 per cent did not work. The evidence on the effect on student grades of working while in school is 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 devising optimal education policy, so this trend of increased work among college students is accounted for in the interpretation of the results.

The analysis is run 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). Brint and Cantwell (2010) find that males make less use of time for academic purposes than females, while Krohn and O'Conner (2005) find that females spend more time studying than males. As such, the extra days before the exam may have a different effect on females versus males. 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.

This paper controls for 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-free and exam-free days before the exam, with the difference in scores on an exam in which one group had four more lecture-free and exam-free days before the exam. The 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. For instance, factors that may affect exam scores like student motivation, student ability, family structure, and work commitments are the same throughout the exam period for a particular student. Based on this assumption, the difference-in-differences (DiD) estimate represents the causal effect of extra available study time on student grades. The discussion of possible violations of this assumption appears in a later section.

Overall, this paper finds no effect of extra available study time on student exam scores. In the main estimation, four extra days before the later exam are not significantly correlated with the score on that exam. When decomposing the results by gender, extra days of study are associated with an increase of female scores by 4.81 points while having no effect on male scores, but this result for females is not statistically significant at conventional levels (p-value =

0.087). These results run counter to our survey of students which revealed that students believe that if they were granted extra days before exams their scores would increase. While further research in different contexts is needed to elucidate whether extra days before a final exam have an impact on student scores, the results of this study back the notion that having a condensed exam schedule (i.e. less days between exams), much like many institutions in the United States, does not hurt student exam scores. Moreover, since condensed exam schedules plausibly lead students to study more throughout the semester, this spacing out of study time contributes to long-term retention of the material (Cepeda et al., 2006).

Data and methodology

The data set is composed of 981 first-year undergraduate students in the School of Economics from the 2012-2013 through 2015-2016 academic years at a large college in Israel. The data are a combination of 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).

The estimation strategy relies on differences in final exam schedules. Students studying economics can choose between one of two schedules: morning or evening. 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. ‘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. The final course grade in all of the years examined is based solely on the final exam grade. In order to avoid potential bias in the grading of exams, department policy is that all short answer questions are graded by the same teaching assistant, and students write their identification number, rather than their name, on the final exam.

Because 'morning' students take more courses per semester, their final exam schedule is denser. The estimation strategy exploits the fact that while the start of morning students' final exam schedule during the years examined was: 1) Mathematics, 2) Microeconomics, 3) Principles of Marketing, 4) Statistics, the start of evening students' final exam schedule was: 1) Mathematics, 2) Microeconomics, 3) Statistics. 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, but the rest of the exam order and the timing of exams are the same. 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, this paper uses the difference in scores on the Microeconomics exam as the baseline difference between the students of the two groups, 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). 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 (e.g. morning students on Wednesday and evening students on Friday), so using Mathematics to determine the pre-treatment difference between the two groups is not as sound as using the Microeconomics course. Therefore, the Mathematics exam is only used as a robustness check of the main result and is not part of the main analysis.

Figure 1 depicts the 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.

[Figure 1 about here]

In order to estimate the effect of available study time on student grades, this paper compares 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, models of the following form are estimated:

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

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 lecture-free and exam-free days before the final exam. That is, 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). Age may affect exam scores because it could be a proxy for time constraints in terms of familial obligations and pre-college exam scores may be a proxy for ability. However, both of these variables do not affect the results because these variables are differenced out in the regression – student covariates are identical pre-treatment and post-treatment.

The sample is limited 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, the results are qualitatively the same, both in terms of magnitude and statistical significance.

In order to enable more accurate implications of our findings, we conducted a survey of contemporary students to get a sense of when and how much students study, what their work obligations are, and whether they believe that having more days between exams would increase their scores (see Appendix). In addition, the survey allowed us to determine if the responses differed by whether the student studies in the morning group (control) or the evening group (treatment). While the survey does not include the students who make up our sample, the survey was conducted on students studying the same courses who had nearly identical schedules as the students in the sample in terms of number of courses and time between exams.

Results

Table 1 presents summary statistics depicting observable differences between the treatment (evening) and control (morning) groups. Several differences emerge between the two groups which necessitate the use of the DiD strategy. That is, the differences in student characteristics do not allow a simple comparison of scores on the exam where the treatment group had more days before the exam. Rather, a baseline difference between the control and treatment groups is needed. The treatment (evening) group relative to the control (morning) group is less male (48 per cent to 57 per cent), older (24.85 years versus 23.85 years), and has higher scores on pre-college exams. There are also 669 and 312 students in the morning (control) and evening (treatment) groups, respectively. 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. 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. 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 the sample consists of economics students. The overall *Psychometry* score of 533 in our sample is equal to the country-wide average (in 2013) (NITE, 2016).

[Table 1 about here]

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 this paper estimates.

In Table 2 are 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. The table shows that extra days are not significantly correlated with exam scores. 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). Examining the results independently by year (columns 4 through 15), there are no cases where there is a statistically significant difference in scores.

[Table 2 about here]

Tables 3 and 4 depict the results by gender. In Table 3 are the results for males. The DiD estimate 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. That is, having twice as many days (eight versus four) right before an exam does not lead to a decrease in the difference in scores between these two groups. Table 4 depicts the effect of extra available days before an exam on females. The DiD estimate is statistically insignificant at conventional levels (p – value = 0.087). That said, the point estimate is 4.81 points, about a fourth of a standard deviation in student grades. 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. This result for females is largely driven by the large difference of 14.51 points on the pre-treatment exam between the morning and evening group in the 2014-2015 academic year.

[Table 3 about here]

[Table 4 about here]

Our survey of students, which we detail in the Appendix, indicates that while there are differences in age and work obligations mentioned earlier between the control group and treatment group, there are no significant differences in study habits. Table A1 shows that over 80% of morning and evening students take time off during the final exam period in order to study for exams. In addition, around three-quarters of students in both groups report that they study throughout the semester and do not leave their studying for a final exam to the last minute. Finally, over 84% of students in both groups responded that they feel their grade would increase if there was more time between exams, and nearly all students report that they study every day or nearly every day during the final exam period. Taken as whole, students report utilizing nearly all days during the final exam period. As such, it is not surprising that they believe that their grade would increase if there were more days between exams. Our results, however, do not support this view. We elaborate on the implications of our findings, based on both the response to the survey and our estimation results, in the conclusion.

Robustness checks

The main identifying assumption in the estimations 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, a number of issues are addressed in order to strengthen the identifying assumption and address some of the study limitations.

The first issue is that the grades in the treatment and control exams are in two different courses. Perhaps the Statistics exam is an easier (more difficult) exam and therefore the difference in scores would have been smaller (larger) regardless of one group having more available time before the exam than the other group. While this cannot be ruled out completely, we address this issue by using the Mathematics exam as an alternative measure of the pre-treatment difference in scores, instead of Microeconomics, because mean Mathematics 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 because the morning and evening groups may end the semester on different days. 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. Table 5 shows that nearly identical point estimates are found when using Mathematics as a baseline difference in scores versus using Microeconomics. 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, 3.36 points, close to the 2.22 estimate in the baseline specification (Table 5, column 6).

[Table 5 about here]

Table 5 (columns 7 through 9) also depicts 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. 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 slightly 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 the statistically insignificant results overall are explained by instructors for the evening group being of *lower* 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 *higher* quality than the morning group. Ideally, a model with instructor fixed effects would be run. 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. While a model with instructor fixed effects cannot be estimated, there is still 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, relative to those in the treatment 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, or before the final exam period begins, and not only when they finish the last exam before the Statistics exam. While no data is available on when students started studying for each exam, the results of our survey indicate that there are no statistical differences in how much students in the two groups study during the semester and in the final exam period.

Discussion and conclusion

Overall, the results show that four extra days of available study time do not affect student grades. There is a difference of 4.12 points on the exam they had the same number of lecture-free and exam-free days before (Microeconomics), and 1.91 points on the exam where one group had four more days before (Statistics), but the DiD estimate is not statistically significant. For females there is an increase of 4.81 points (out of 100), that is not statistically significant at conventional levels (p-value = 0.087), while for males the DiD estimate is close to 0.

Our survey of contemporary students, combined with the literature on optimal study strategies, offers some explanation on why overall there is no positive effect on grades of having more free days before an exam. First, our survey shows that both morning (control) and evening (treatment) students study throughout the semester. As such, extra study days before the exam that one group has relative to another pale in comparison to the studying that students have done throughout the semester. Therefore, these extra days right before the exam do not influence their grades. That is, students in the evening group, who on average work much more than students in the morning group and presumably have less time to devote to their studies throughout the semester, do not decrease the gap in grades with the extra days they have to study. The finding of no statistically significant increase in grades contrasts with students' perception that their grade will increase if they have more time before a final exam. This is in line with Zulkiply et al. (2012) who find that students believe cramming will produce superior results.

Our findings can also be interpreted in the context of the literature on the effectiveness of studying throughout the semester (distributed practice) versus cramming (massed practice) that has been examined through several experiments with undergraduate students (e.g. Kapler, Weston, and Wiseheart (2015); Rohrer and Taylor (2006)). This previous research has found that while for short-term recall and tasks cramming may be just as effective as learning throughout the semester, for long-term retention of the material studying throughout the semester (spacing) is more effective (Cepeda et al. 2006). Therefore, our results show that having more space between exams may be hurting students. This is because having more days between exams encourages students to put off studying by giving them more days to cram right before the exam.

These students miss the long-term benefits to studying throughout the semester, which they will be forced to do if the exam schedule is condensed.

While our findings indicate that overall students do not benefit from having extra days before a final exam, further research is needed to test for the robustness of this result to different populations and different exams. Moreover, the large difference in magnitude between females and males of almost 5 points, while statistically insignificant, is large and stresses the need for further research. Gender differences have been found in other aspects of education such as motivation, discipline, and time use ((Meece and Painter, 2008; Duckworth and Seligman, 2006; Brint and Cantwell (2010)).

The results of this study have implications for college administrators because there is currently heterogeneity in the spacing between final exams across institutions, raising an inquiry about ideal spacing. For instance, at the college examined, students have approximately five days on average between each exam depending on their study track, resulting in a final exam period of over a month. Providing this magnitude of spacing between exams leads to scheduling inconveniences for students, including but not limited to final exams flowing into the subsequent semester and students not having a break between semesters. Moreover, making the campus hospitable for students studying during the final exam period leads to overhead costs associated with electricity, security, cafeteria hours, and extended library hours; these costs would be lower if students were on break. This study finds that these extra costs do not translate into higher exam scores.

If the relationship between days between exams and student scores operates in the other direction as well (i.e. there is no effect whether the number of days are increased or decreased), then perhaps decreasing the time between exams would not hurt student outcomes. If this is the case, perhaps the college examined, as well as other institutions with similar final exam schedules, should decrease the time between exams. For example, in many institutions in the US final exams are typically condensed to between five and 10 days. Decreasing the time between exams would benefit students in terms of longer breaks between semesters and ease the budget constraints of higher education institutions. Further research in other institutions, exam settings, and fields of study is needed to test for the robustness of these findings.

References

Bonesrønning, H. and Opstad, L. (2012), "How much is students' college performance affected by quantity of study?" *International Review of Economics Education,* Vol. 11 No. 2, pp. 46-63.

Brint, S. and Cantwell, A. M. (2010), "Undergraduate time use and academic outcomes: Results from the University of California Undergraduate Experience Survey 2006", *Teachers College Record*, Vol. 112 No. 9, pp. 2441-2470.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., and Rohrer, D. (2006), "Distributed practice in verbal recall tasks: A review and quantitative synthesis," *Psychological Bulletin*, No. 132 Vol. 3, pp. 354-380.

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,* Vol. 38, pp. 38-50.

Duckworth, A. L. and Seligman, M. E. (2006), "Self-discipline gives girls the edge: Gender in self- discipline, grades, and achievement test scores", *Journal of Educational Psychology,* Vol. 98 No. 1, pp. 198-208.

Guillaume, D. W. and Khachikian, C. S. (2011), "The effect of time-on-task on student grades and grade expectations", *Assessment & Evaluation in Higher Education,* Vol. 36 No. 3, pp. 251-261.

Kalenkoski, C. M. and Pabilonia, S. W. (2010), "Parental transfers, student achievement, and the labor supply of college students", *Journal of Population Economics*, Vol. 23 No. 2, pp. 469-496.

Kapler, I. V., Weston, T., and Wiseheart, M. (2015), "Spacing in a simulated undergraduate classroom: Long-term benefits for factual and higher-level learning," *Learning and Instruction*, Vol. 36, pp. 38-45.

Krohn, G. A. and O'Connor, C. M. (2005), "Student effort and performance over the semester", *The Journal of Economic Education,* Vol. 36 No. 1, pp. 3-28.

Masui, C., Broeckmans, J., Doumen, S., Groenen, A. and Molenberghs G. (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,* Vol. 39 No. 4, pp. 621-643.

McIntyre, S. H. and Munson, J. M. (2008), "Exploring cramming: student behaviors, beliefs, and learning retention in the principles of marketing course", *Journal of Marketing Education,* Vol. 30 No. 3, pp. 226-243.

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

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

Monteiro, H. and Lopes, A. F. (2007), "A benchmarking of the undergraduate economics major in Europe and in the United States", *International Review of Economics Education,* Vol. 6 No. 2, pp. 9-26.

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

Rohrer, D., and Taylor, K. (2006), "The effects of overlearning and distributed practise on the retention of mathematics knowledge", *Applied Cognitive Psychology*, Vol. 20 No. 9, pp. 1209-1224.

Schuman, H., Walsh, E., Olson, C. and Etheridge, B. (1985), "Effort and reward: The assumption that college grades are affected by quantity of study", *Social Forces,* Vol. 63 No. 4, pp. 945-966.

Scott-Clayton, J. (2012), "What explains trends in labor supply among us undergraduates?" *National Tax Journal,* Vol. 65 No. 1, pp. 181-210.

Siegfried, J. J. and Walstad, W. B. (2014), "Undergraduate coursework in economics: A survey perspective", *The Journal of Economic Education,* Vol. 45 No. 2, pp. 147-158.

Stinebrickner, R. and Stinebrickner, T. R. (2004), "Time-use and college outcomes", *Journal of Econometrics,* Vol. 121 No. 1, pp. 243-269.

Stinebrickner, R. and Stinebrickner, T. R. (2008), "The causal effect of studying on academic performance". *The BE Journal of Economic Analysis & Policy*, Vol. 8 No. 1, Article 14.

Voyer, D. and Voyer, S. D. (2014), "Gender differences in scholastic achievement: A meta-analysis", *Psychological Bulletin,* Vol. 140 No. 4, pp. 1174-1204.

Zulkiply, N., McLean, J., Burt, J. S., and Bath, D. (2012), "Spacing and induction: Application to exemplars presented as auditory and visual text", *Learning and Instruction*, Vol. 22 No. 3, pp. 215-221.

Appendix

To get a better sense of student study habits and their feelings towards the number of days they have between exams, a short survey of contemporary students who take the same subjects as the students in our sample was conducted. The survey was conducted at the beginning of the second semester, about a month after the students took their first semester exams, during the last ten minutes of class. Participation in the survey was voluntary and included no information specifically identifying the student. One student chose not to participate in the survey. A total of 184 students participated, 72 from the morning (control) group and 102 from the evening (treatment) group.

The original survey was in Hebrew. Below is the translation of the survey questions:

1. Age: _______

2. Male Female

3. In which track do you study?

Morning Evening

4. Are you employed?

Yes, full-time Yes, part-time No

If yes, do you take time off of work during the final exam period?

Yes No

5. Do you think that if there were eight days between exams instead of the usual four days, that your score would increase?

Yes No Maybe

6. Do you study throughout the semester (and not solely before final exams)?

Always Sometimes No Only in the tougher courses

7. During the final exam period you study:

Every day Every day (other than the Sabbath) Almost every day Every few days

[Table A1 about here]

Figure 1:

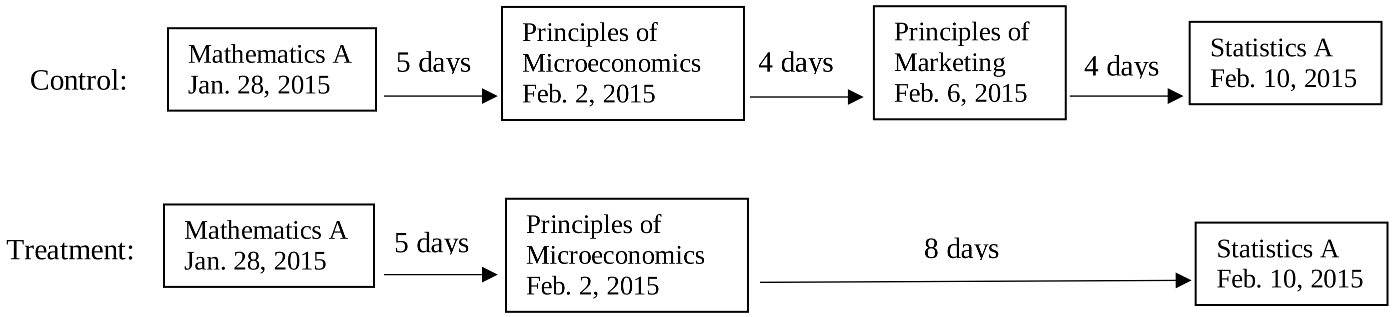
Control:
Mathematics A
Jan. 28, 2015
5 days
Principles of
Microeconomics
Feb. 2, 2015
4 days
Principles of
Marketing
Feb. 6, 2015
4 days
Statistics A
Feb. 10, 2015
Treatment:
Mathematics A
Jan. 28, 2015
5 days
Principles of
Microeconomics
Feb. 2, 2015
8 days
Statistics A
Feb. 10, 2015

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. *** difference significantly different from 0 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. 'Observations' represents the total number of exams in each specification, with each student taking two exams, one in each subject. The Difference-in-differences estimate is the coefficient β_3 in the following regression: $Grade_{ige} = \alpha + \beta_1 Evening_g + \beta_2 Statistics_e + \beta_3 EvStatistics_{ge} + \varepsilon_{ige}$. Difference is significantly different from 0 using a t-test at 0.01***; 0.05**; 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. 'Observations' represents the total number of exams in each specification, with each student taking two exams, one in each subject. The Difference-in-differences estimate is the coefficient β_3 in the following regression: $Grade_{ige} = \alpha + \beta_1 Evening_g + \beta_2 Statistics_e + \beta_3 EvStatistics_{ge} + \varepsilon_{ige}$. Difference is significantly different from 0 using a t-test at 0.01***; 0.05**; 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. 'Observations' represents the total number of exams in each specification, with each student taking two exams, one in each subject. The Difference-in-differences estimate is the coefficient β_3 in the following regression: $Grade_{ige} = \alpha + \beta_1 Evening_g + \beta_2 Statistics_e + \beta_3 EvStatistics_{ge} + \varepsilon_{ige}$. Difference is significantly different from 0 using a t-test at 0.01***; 0.05**; 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***	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. 'Observations' in the regressions represents the total number of exams in each specification, with each student taking two exams, one in each subject. 'Observations' in the percentile comparisons represents number of students. The Difference-in-differences estimate is the coefficient β_3 in the following regression: $Grade_{ige} = \alpha + \beta_1 Evening_g + \beta_2 Statistics_e + \beta_3 EvStatistics_{ge} + \varepsilon_{ige}$. Difference is significantly different from 0 using a t-test at 0.01***; 0.05**; 0.10*. Standard errors in parentheses.

Table A1: Student Study Habit Survey

	Morning Group (Control) (n = 72)				Evening Group (Treatment) (n =102)			
Age	23.68				25.33			
Gender	Female	Male			Female	Male		
	0.389 (28)	0.611 (44)			0.461 (47)	0.539 (55)		
Employed	Full-time	Part-time	No		Full-time	Part-time	No	
	0.111 (8)	0.722 (52)	0.167 (12)		0.804 (82)	0.157 (16)	0.039 (4)	
Do you take time off before exams?	Yes	No			Yes	No		
	.833 (50)	0.167 (10)			0.857 (84)	0.143 (14)		
Would more days before exam increase scores?	Yes	No	Maybe		Yes	No	Maybe	
	0.903 (65)	0.028 (2)	0.069 (5)		0.843 (86)	0.059 (6)	0.098 (10)	
Study throughout semester?	Always	Sometimes	No	Tougher courses	Always	Sometimes	No	Tougher courses
	0.347 (25)	0.403 (29)	0.097 (7)	0.153 (11)	0.402 (41)	0.412 (42)	0.118 (12)	0.069 (7)
How often do you study during final exam period?	Every day	Every day (except Sabbath)	Almost every day	Every few days	Every day	Every day (except Sabbath)	Almost every day	Every few days
	0.778 (56)	0.125 (9)	0.083 (6)	0.014 (1)	0.657 (67)	0.069 (7)	0.245 (25)	0.029 (3)

Notes: Table displays the proportion and number of students who gave each answer (in parentheses). The full text of the questions and answers is in the appendix.