

EXAM SCHEDULING AND STUDENT PERFORMANCE

Giorgio Di Pietro

Westminster Business School and IZA

ABSTRACT

This paper uses data from a single School of a UK university to estimate the impact of a switch from end-of-semester to end-of-year final exams on student performance. The identification strategy exploits the fact that while the timing of final exams changed, the timing of mid-term exams remained the same. Estimates are based on a difference-in-differences methodology that compares final and mid-term exam scores after the switch, with final and mid-term exam scores before the switch. The empirical findings suggest that the shift of final exams to the end of the academic year had a negative effect on student achievement. Many changes in higher education are often done on the basis of financial, administrative or ideological considerations, underestimating the magnitude of their effect on student learning and student performance.

Keywords: exam scheduling, student performance, difference-in-differences (DiD)

JEL classification numbers: I20, I21

I. INTRODUCTION

Most studies of student performance look at the role played by family background, peers, teachers and schools. However, much less attention is given to the influence exerted by aspects of examination arrangements. In an attempt to address this gap in the literature, this paper uses data from a School of a UK university to examine the impact of a switch from end-of-semester to end-of-year final exams on student performance in a two-semester study programme. This means that for first-semester subjects students were now assessed several months after the end of teaching, rather than immediately afterwards. Our identification strategy relies on the fact that while the timing of final exams changed, the timing of mid-term exams remained the same. Hence our intuition consists of comparing final and mid-term exam scores after the reform, with final and mid-term exam scores before the reform. In essence, this is a difference-in-differences (DiD) framework that hinges on the assumption that without the movement of final exams to the end of the academic year the difference between final and mid-term exam scores would have

Correspondence: Giorgio Di Pietro, Department of Economics and Quantitative Methods, Westminster Business School, 35 Marylebone Road, London NW1 5LS, UK. Tel: 0044(0)2035066665; Fax: 0044(0)2079115839; Email: G.D.I.Pietro@westminster.ac.uk. I would like to thank two anonymous reviewers for their very helpful comments. However, all remaining errors are solely my own.

been the same as that observed when final exams for each subject were taking place at the end of the semester.

The reform was unexpected as the decision to implement it was taken in spring 2005 following management and staff discussions that began in autumn 2004. This means that, as the change in the scheduling of final exams could not have been anticipated by students and their families, it can be interpreted as an exogenous shock. Though final exams are moved to the end of the academic year, this is still divided into two teaching semesters (i.e., the autumn semester and the spring semester) of twelve weeks each¹ and all subjects remain semester-long. The number of subjects taken by students per semester has also remained the same as well as instructional time² and the duration of final exams.

Before the reform, there were two final exam sessions of two weeks each, one before the start of the second semester (i.e., January) for those subjects taken during the first semester and one after the end of the second semester (i.e., June) for those subjects taken during this period. With the reform, all the final exams were moved to the end of the academic year (i.e., during a three-week period in May). This had two implications. First, given the delay between the end of classes for first-semester subjects and the corresponding final exams, review sessions for these subjects shifted from the end of the first semester (i.e., December) to the end of April.³ However, one should note that the total amount of review time did not change. Second, the reform made the academic year end slightly earlier.⁴ Students may benefit from this as a longer summer break enhances their opportunity to have a job, do an internship or even travel.⁵ On the other hand, however, if they do not utilize this additional time, there is the risk that students will be exposed to a depreciation of their knowledge and skills.

In addition to a change in the allocation of final exams, the reform brought about a change in the School's student recruitment policy. It established that undergraduate students can only enrol at the beginning of the first semester in September, thus eliminating the recruitment of students at the start of the second semester. Although having two student entry points per academic year has some advantages,⁶ in practice this may not always turn out to be a profitable choice. Whilst the start-date flexibility imposes additional costs because of the higher organizational burden, the number of students choosing to enrol in the second semester is often relatively small, especially in undergraduate programmes. We will return to this issue in the next section.

This paper is related to two strands of literature. The first strand analyses the effect of different types of academic calendar on achievement at school (see, for instance, Merino, 1983; Shields and Oberg, 1999; Graves, 2010) and university (see, for instance, Coleman *et al.*, 1984) levels.

¹ Though the length of the second semester did not vary, the reform changed its timing. As final exams no longer take place at the end of the first semester, the second semester starts in January and runs until the beginning of April. Before the reform, because of final exams in January, the second semester used to begin in February and finish at the beginning of May.

² Each full-time student typically attends four subjects per semester and each subject involves 3 hours of class time (1.5 hours lecture and 1.5 hours seminar) per week.

³ For a small number of subjects (only first-year subjects) the assessment procedure was revised with a view to reducing the weight of the final exam score in the overall subject assessment. The assessment schemes for the subjects examined here, however, remained the same.

⁴ There is approximately a one-month difference. Before the reform students took their final exams for second-semester subjects in June and got their academic results in July. After the reform students take all their final exams in May and get their academic results in June. The start of the academic year did not change.

⁵ The earlier end of the academic year may also bring benefits for graduates and academics. Graduates are able to start searching for a job at the very beginning of the summer period and this, in turn, may give them a better chance for employment. Academic staff members have more time to conduct research.

⁶ Offering the possibility to enrol at the beginning of the second semester may play an important role in attracting international students, mature students and those from poor backgrounds who may have missed the chance of enrolling over the summer because they were not clear what to do.

Our study is part of the same broad research agenda as it investigates the impact of timing and organization of learning on student performance. The second strand of literature related to our paper deals with high-stake assessment. This term refers to tests or exams whose results are crucial for the people who have to take them and thus may cause great concern and anxiety among the takers (Casbarro, 2004). In this study we measure academic performance through the scores received by students in several third-year subjects. Given that these scores have a lot of weight in final degree classification, students are more likely to suffer from considerable test anxiety. When students perceive the evaluative threat to be high or when success or failure are particularly relevant, it is well documented that test anxiety is very high and this may decrease performance (see, for instance, Hancock, 2001).

The remainder of the paper is as follows. Section II outlines the rationale for the reform and explains how the change in the scheduling of final exams is expected to impact on student performance. Section III describes the data and reports summary statistics. Section IV presents and discusses the results of the empirical analysis. Section V concludes.

II. BACKGROUND

In this paper, we look at the experience of a School of a UK university that in the 2005/06 academic year changed undergraduate student recruitment policy and shifted all final exams for undergraduate students to the end of the academic calendar. This School, which has a large proportion of students from 'non-traditional' backgrounds, had to introduce these changes quickly following a decision made by the university's Academic Council in spring 2005.

The reform appears to be primarily the result of financial rather than educational considerations. This is especially clear when one compares the undergraduate programme with the postgraduate one. In previous years, the benefits resulting from recruiting undergraduate students at the beginning of the second semester turned out to be lower than the corresponding administrative costs. However, this did not occur at postgraduate level as the demand for entry was quite strong even in January.⁷ As a consequence, while undergraduate students were no longer recruited at the start of the second semester, postgraduate students continued to be able to enrol at the beginning of each semester.

The decision to stop recruiting undergraduate students at the beginning of the second semester has been accompanied by a change in the scheduling of final exams. All undergraduate final exams were moved to the end of the academic year since a final exam session at the end of the first semester was no longer needed to allow students to graduate in three years. The elimination of final exams at the end of the first semester was also justified on the grounds that these were causing a higher drop-out rate among first-year students. It was observed that a proportion of first-year students tended not to return to university after the Christmas break because they were worried about final exams taking place in January. Additionally, it is also possible that some first-year students have decided to discontinue their studies as a result of their poor performance in the final exams taken at the end of the first semester.

We now turn our attention to the expected impact on student performance of the switch of final exams from the end of the first semester to the end of the academic year. *A priori*, it is unclear what net effect this change will have as both positive and negative arguments can be made.

⁷ Enrolment at postgraduate level in the second semester is mainly driven by international students who make up more than three-quarters of postgraduate students. By contrast, the proportion of international students within the undergraduate programme is significantly smaller (i.e., about 15 percent).

On the one hand, two theoretical considerations suggest that the movement of final exams to the end of the academic year may actually improve student performance. First, delaying final exams until the end of the academic year may provide students with time for proper digestion of material and understanding. Second, taking all the final exams at the end of the academic year offers students the possibility of creating their own 'learning calendars' to best meet their personal, professional and academic needs.

Not only can arguments be made to show that the shift of final exams to the end of the academic year is beneficial to student performance, but, seen from another perspective, one can also point to the fact that taking final exams at the end of the first semester is already having negative consequences on academic progress and achievement. As outlined above, there is anecdotal evidence suggesting that many students do not view the Christmas break as a time to study for final exams as they tend to think of it as a time to relax and stay with their family and friends. This attitude does not make students perform well in exams at the end of the first semester. Additionally, because before the reform there were two relatively long final exam periods, it was believed by some in the 'Learning and Teaching' community that too much time and importance was being attached to assessment, often at the expense of more meaningful learning activities. Finally, over-assessment may reduce the time available for reflection. Learning is expected to encourage students to consider new ideas and assimilate them, thus producing deeper learning. Without the opportunity for reflection, most academic concepts are only incorporated into the short-term memory (Bailey, 1999).

On the other hand, however, a number of arguments support the proposition that the shift of final exams to the end of the academic year may negatively affect student performance. To start with, the concentration of final exams in a short period of time⁸ may be associated with cognitive switching costs that increase the probability of students making a mistake. Something similar occurs with multitasking. When workers have to deal with too many projects simultaneously, tasks pile up, productivity declines and errors multiply. A recent study by Aral *et al.* (2007) finds that the relationship between multitasking and productivity follows an inverted U shape such that, beyond an optimum, increasing multitasking is associated with a decrease in both project completion rates and revenue generation.

It is also possible that the concentration of final exams may increase the level of stress on students (Shirom, 1986) and hence may have a negative impact on their performance. They may perceive that taking several exams in a short period of time is either challenging (positive) or threatening (negative) and both these emotions may produce stress. Although there are some types of emotions that have a positive impact on certain types of memory, there is robust neurological evidence showing that stress may have a detrimental effect on memory formation and learning (see, for instance, Erk *et al.*, 2003). In addition to the neurological evidence, the adverse effect of stress on learning is also found in several educational studies. In these studies stress is often measured through anxiety on the assumption that the latter is a debilitating reaction to the former. A number of works (Fredericks and Mundy, 1969; Westerman, *et al.*, 1986; Cecchini and Friedman, 1987) investigate the impact of anxiety on educational achievement among dental students. They generally support the hypothesis that there is a negative relationship between anxiety and academic performance. A similar result is obtained by Ross *et al.* (2006) who conclude that worrying about money makes students underperform at university.

There are also concerns that the scheduling of all final exams at the end of the academic year may have a detrimental effect on student learning and subsequent student academic performance, especially in those subjects studied in the first semester. Students may be less ready to learn

⁸ Students can face a maximum of eight exams in three weeks.

and may be unable to maintain their motivation throughout the first semester as they are aware that the final exams for the first-semester subjects will take place some months after the end of classes. Furthermore, the long delay between studying the material and final exams may also have a negative impact on learning, given that students are likely to spend valuable time revisiting old lessons. In fact, while for some this is an opportunity to reconsider concepts and theories they have not fully understood first time round, for others it represents extra time needed to review the material they have forgotten.

III. DATA

We employ administrative record data on six successive cohorts of third-year undergraduate students. This period comprises three academic years before the reform (i.e., 2002/03, 2003/04 and 2004/05) and three academic years after the reform (i.e., 2005/06, 2006/07 and 2007/08). We use individual-level information on final exam scores in three different subjects (i.e., Subject A, Subject B and Subject C) attended by the students in the first semester. These subjects were chosen as, during the six academic years under examination, they did not experience any significant changes with respect to the content of their syllabus, teaching staff and methods (i.e., same teachers, same books, same exercises, etc.). The grading criteria, the final exam format as well as the weight of the final exam in the overall assessment have also remained constant over the entire period. Additionally, there was no predetermined distribution of final exam scores. As observed by Bandiera *et al.* (2010), in the UK, unlike North America, academics have no incentives to manipulate the exam scores in order to increase student numbers or to raise their own student evaluations. Such a manipulation would even be difficult given that examination scripts are double blind marked by two faculty members and one external examiner. This lends credibility to our measure of outcome as it suggests that final exam scores are likely to reflect true student performance.⁹ Given all this, it is reasonable to assume that the final exam scores received by students before the reform are comparable with those received by their peers after the reform.

Each subject comprises two pieces of assessment. Whilst the second piece is the final exam discussed above, the first piece of assessment is a mid-term exam taking place during the teaching period.¹⁰ The mid-term exam and the final exam scores account for 30 and 70 percent of the overall assessment, respectively. These two assessments are independent in the sense that performing poorly in the first assessment does not prevent students from taking the final exam. Additionally, the mid-term exam score has no influence on the final exam score. Students must pass both assessments in order to pass the subject.

Our record data are quite rich in the quality of information they provide on student characteristics. The individual-level indicators included in our analysis as control variables are: gender, nationality, ethnicity, mode of study (full or part time), age, disability status, family background (proxied by occupation of parent or guardian), and pre-university qualifications (proxied by UCAS tariff points¹¹).

⁹ As shown in the Appendix, the distribution of final exam scores (grouping together scores for all three subjects) is quite different before and after the reform.

¹⁰ For all three selected subjects, the first piece of assessment has the same format, the same weight and the same testing regime across all the academic years we are looking at. Hence not only final exam scores but also mid-term exam scores are comparable before and after the reform.

¹¹ Tariff points are computed by UCAS (Universities and Colleges Admissions Service) to indicate academic equivalence across different academic qualifications (see wwwucas.ac.uk). The system of UCAS tariff points is used by universities and colleges for making offers to applicants.

TABLE 1
Descriptive statistics

Variables	End-of-academic year final exam (A)		End-of-semester final exam (B)		Difference between (A) and (B)	
	Mean	SD	Mean	SD	Mean	SE
Final exam score	51.04	15.58	54.38	13.12	-3.34***	0.74
Mid-term exam score	60.28	13.06	59.55	11.65	0.73	0.63
Female	0.54	0.50	0.53	0.50	0.1	0.25
Age (years)	23.27	2.67	23.95	3.28	-0.68***	0.15
Part-time student	0.04	0.19	0.06	0.23	-0.02*	0.01
Disability status	0.03	0.17	0.01	0.12	0.02*	0.007
<i>Ethnicity</i>						
White	0.31	0.46	0.40	0.49	-0.09***	0.02
Indian/Pakistani/ Bangladeshi	0.29	0.45	0.23	0.42	0.06***	0.02
Black	0.08	0.28	0.08	0.27	0.007	0.01
Chinese	0.06	0.23	0.06	0.24	-0.004	0.01
Other Asian	0.06	0.24	0.06	0.23	0.001	0.01
Others or unknown	0.21	0.40	0.18	0.38	0.03	0.02
<i>Nationality</i>						
UK	0.54	0.50	0.53	0.50	0.01	0.03
Other EU	0.25	0.43	0.30	0.46	-0.05**	0.02
Non EU	0.22	0.41	0.17	0.38	0.05***	0.02
<i>UCAS tariff score</i>						
0-150	0.04	0.19	0.06	0.23	-0.02	0.01
151-299	0.42	0.49	0.31	0.46	0.11***	0.02
≥300	0.10	0.30	0.07	0.25	0.03**	0.01
Missing	0.44	0.50	0.57	0.50	-0.13***	0.03
<i>Occupation of parent or guardian</i>						
Managerial or professional occupations	0.23	0.42	0.07	0.25	0.16***	0.02
Other occupations	0.28	0.45	0.07	0.25	0.21***	0.02
Missing	0.49	0.50	0.86	0.34	-0.37***	0.02
<i>Subjects</i>						
Subject A	0.20	0.40	0.19	0.39	0.01	0.02
Subject B	0.36	0.48	0.36	0.48	-0.003	0.02
Subject C	0.44	0.50	0.45	0.50	-0.01	0.03
Class size	122.55	37.07	107.93	31.66	14.62***	1.76
Number of ob.	804		737			

***Significant at the 1% level.

**Significant at the 5% level.

*Significant at the 10% level.

Our final sample includes 1541 observations.¹² Within the sample, the proportion of students who attended the university before the reform (i.e., 47.83 percent) is very similar to that of students who attended the university after the reform (i.e., 52.17 percent). Table 1 compares the characteristics of the students who took the final exams at the end of the semester with those

¹² Students who are on an exchange programme have been excluded from the final sample.

who took the final exams at the end of the academic year. These descriptive statistics appear to support the hypothesis that the movement of final exams to the end of the academic year has negatively affected student performance. The final exam score across the three subjects under examination averages 54.38 points (out of 100) before the reform, whereas the corresponding figure is 51.04 points after the reform.

Unfortunately, we do not have information on UCAS tariff scores for many students.¹³ Similarly, there is also missing information on many students' family background. However, in order to maximize the sample size, we keep observations with missing values on parental occupation and UCAS tariff scores, and create dummy variables indicating missing status for these variables. The proportion of students of Indian/Pakistani/Bangladeshi origin is statistically higher during the period in which final exams took place at the end of the academic year relative to the period during which final exams occurred at the end of the first semester. The opposite occurs for the proportion of students of White origin. Additionally, the proportion of students who have a UCAS tariff score of either 151–299 or ≥ 300 is statistically higher after the reform than before the reform. While the groups used before and after the reform differ on a number of observable characteristics, this is not problematic for identification in our DiD approach unless these different groups would systematically have different relationships between the mid-term and final exam scores in the absence of the reform.

IV. ANALYSIS

To examine the effect of the movement of final exams to the end of the academic year on student performance, we employ a DiD approach. Our strategy relies on using mid-term exam score as a control group, and compare how the final exam score changed after the reform relative to the control group. Whilst the reform changed the scheduling of final exams, this did not occur with the mid-term exams. Especially for first-semester subjects,¹⁴ there are no reasons to believe that the reform has had an effect on the score students receive on the mid-term exam given that its timing has remained unchanged. Since final and mid-term exams take place in the same academic year, one would expect all time effects to be common across these two pieces of assessment, i.e., time-varying unobservables influencing mid-term exam scores are likely to be similar to those affecting final exam scores before and after the reform.

Figure 1 illustrates this approach and previews our main result. In this figure, the solid lines represent average final and mid-term exam scores (grouping the three subjects together) between academic years 2002/03 and 2007/08. Though these two time-series appear to be almost perfectly parallel during the period 2002/03–2004/05, from 2005/06, which is the year the reform was implemented, they start to diverge with the average final exam score falling more sharply than the average mid-term exam score. If the usual DiD assumption of parallel trends is met (i.e., in the absence of the reform, the average final exam score would have continued to follow a trend parallel to that of the average mid-term exam score), a dotted line can be drawn. This represents the trend in average final score that would have been observed if the reform had not been implemented. Therefore the average effect of the reform on student performance can be estimated as the difference between final and mid-term exam scores after the reform,

¹³ This is because many students (including international students) enter university without traditional UK qualifications. Of students for whom information on UCAS tariff scores is missing, 78.11 percent are of non-UK nationality and 21.89 percent are UK nationals.

¹⁴ For first-semester subjects the mid-term exam, which typically takes the form of an essay or an in-class test, has occurred between the end of November and the beginning of December in the academic years after and before the reform.

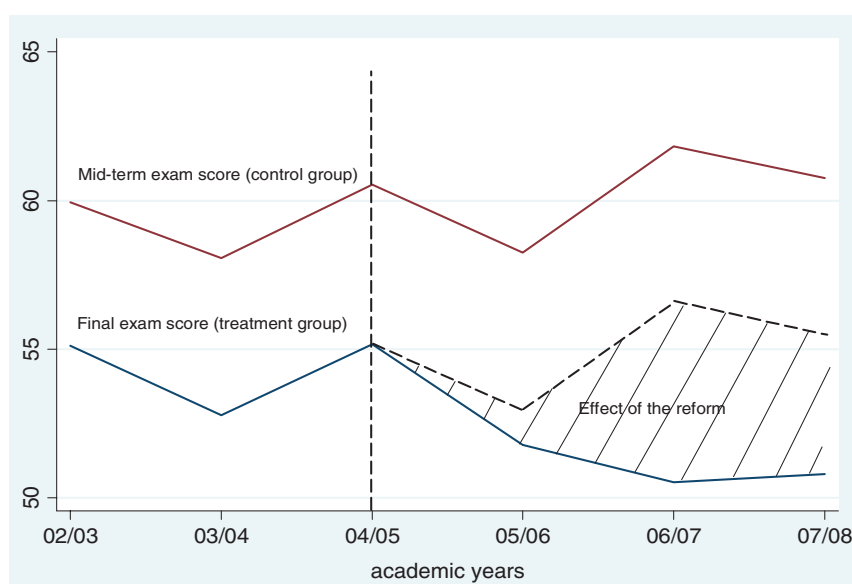


Fig. 1. Average final and mid-term exam scores, academic years 2002/03-2007/08

The solid lines represent actual trends in average final and mid-term exam scores. The dotted line represents the trend in average final exam score that one would have observed if the reform had not been implemented (assuming that the assumption of the parallel trends is met, i.e., in the absence of the reform, the average final exam score would have continued to follow a trend parallel to that of the average mid-term exam score). The average effect of the reform on student performance is computed as the difference between final and mid-term exam scores after the reform, relative to same difference before the reform. The dashed area shows this effect.

TABLE 2

Average final and mid-term exam scores by timing of final exams (standard errors in brackets)

	End-of-academic year final exam (A)	End-of-semester final exam (B)	Difference between (A) and (B)
Final exam score	51.036 (0.549)	54.379 (0.483)	-3.343*** (0.737)
Mid-term exam score	60.281 (0.461)	59.553 (0.429)	0.728 (0.633)
Difference	-9.245 (0.545)	-5.175 (0.454)	-4.070*** (0.716)

***Significant at the 1% level.

**Significant at the 5% level.

*Significant at the 10% level.

relative to the same difference before the reform. This effect is represented by the dashed area in Figure 1.

While Figure 1 is suggestive of a negative effect of the reform on academic achievement, statistical techniques need to be used to determine whether this relationship is significant. To this purpose, Table 2 reports the average final and mid-term exam scores (grouping the three subjects together) before and after the reform. In both cases, the average final exam score is lower than the average mid-term exam score. However, this difference is larger after the reform than before the reform. The difference in these differences can be interpreted as the causal effect of moving final exams to the end of the academic calendar, under the assumption that in the

absence of the reform, the difference between the final and mid-term exam scores would have been the same as that observed when final exams took place at the end of the first semester. Table 2 shows that students who studied at university after the reform received, on average, 4.07 points less in their final exam than those who studied before the reform.

In order to check whether the above result is robust to the inclusion of control variables, the following model is estimated:

$$y_{ijt}^{FE} - y_{ijt}^{JT} = \beta_0 + \beta_1' X_{ijt} + \alpha_0 K_{ijt} + \alpha_1 Z_{jt} + \gamma_j + v_{ijt} \quad (1)$$

where y_{ijt}^{FE} and y_{ijt}^{JT} represent the final and mid-term exam scores received by student i in subject j in academic year t , respectively; X is a vector of student characteristics that are thought to influence student performance; K is a dummy variable equal to 1 for students who took the final exams at the end of the academic year and 0 otherwise; Z represents class size; γ is a set of subject dummies; and v_{ijt} is an error term. The main parameter of interest is α_0 . It captures the average difference between the final and mid-term exam scores after and before the change in the scheduling of final exams. In so doing, any common trends, which will affect both final and mid-term exam scores, get differenced out.

The first column of Table 3 depicts our DiD estimates¹⁵ once we control for student characteristics, class size and subject fixed effects.¹⁶ The estimates indicate that students who attended university after the reform scored, on average, about 3.64 points less in their final exam than students who attended university before the reform.¹⁷ This result is reassuring as it shows that, although most of our control variables are likely to impact on academic achievement, their inclusion does not considerably change our estimates of the reform's effect on a student's performance.

In an attempt to check the robustness of this finding, we estimate our DiD model on subsets of the data. Given that the number of missing values for some of our explanatory variables may cause some concern, we remove from the sample students for whom we do not have information on either parental occupation or UCAS tariff scores. The estimates for this smaller sample, which are depicted in the second column of Table 3, show that the effect of the reform on student performance is quite similar to the one reported in the first column of Table 3. In the third and fourth columns of Table 3 we include only White and UK students, respectively. The rationale for this is that these subgroups may be more similar in terms of unobserved factors affecting student performance. These estimates are also broadly consistent with our previous findings.¹⁸

One of the limitations of this study is that we are unable to disentangle the effect of the change in the scheduling of final exams from that caused by changes in the academic calendar (e.g., a shorter break between semesters and a change in the timing of the second semester). However, not only are these minor changes (there is no change in important aspects of the

¹⁵ Throughout the paper we report estimates in which standard errors are not clustered at subject and academic year levels. Nichols and Schaffer (2007) argue that with a small number of clusters (less than 50) inference using the cluster-robust estimator may be incorrect more often than not. One may, however, still note that, although clustering at both subject and academic year levels (following the approach of Cameron *et al.*, 2006) leads to a higher standard error for our reform dummy, the results (available upon request) still imply that we cannot accept the null hypothesis of no effect (i.e. $\alpha_0 = 0$) at the 5% significance level in all the specifications used in this study.

¹⁶ Whilst it should be noted that the R-squared is rather low, it is close to the value achieved in studies attempting to model undergraduate student performance (see, for instance, Chevalier *et al.*, 2009).

¹⁷ In an attempt to check the robustness of the estimates depicted in Table 3, we estimate our model for each subject separately. The results, which are not reported here but are available upon request, show that the reform had a statistically significant negative effect on final exam score for all the subjects.

¹⁸ Additional analysis shows that the effect of the reform on final exam scores does not differ across gender, other EU nationals and non-EU nationals (these estimates are available upon request).

TABLE 3
DiD estimates (robust standard errors in brackets); pooled model

<i>Independent variables</i>	<i>(2) Only students who do not have missing information on either family background or UCAS tariff scores</i>			
	<i>(1) All students</i>		<i>(3) Only White students</i>	<i>(4) Only UK students</i>
Constant	-6.951 (4.436)	-1.229 (6.491)	3.147 (7.021)	-7.102 (7.208)
End-of-academic year final exam	-3.635*** (0.835)	-4.062*** (1.137)	-5.054*** (1.444)	-3.924*** (1.252)
Female	2.381*** (0.730)	2.515** (1.014)	2.737** (1.278)	3.111*** (1.048)
Age (years)	-0.001 (0.142)	-0.060 (0.219)	-0.498** (0.227)	0.008 (0.241)
Part-time student	3.281 (2.298)	4.512 (4.388)	3.543 (4.540)	5.480* (3.213)
Disability status	-2.635 (2.712)	-4.904 (4.586)	1.998 (4.080)	-1.595 (3.732)
<i>Ethnicity – Reference group is White</i>				
Indian/Pakistani/ Bangladeshi	-0.998 (1.147)	-0.939 (1.403)	–	-0.801 (1.394)
Black	-2.147 (1.375)	-3.389 (1.691)	–	-1.684 (1.814)
Chinese	1.059 (1.785)	0.789 (2.359)	–	-0.892 (2.599)
Other Asian	-2.854 (1.740)	-2.867 (2.381)	–	-2.292 (2.451)
Others or unknown	0.080 (0.956)	0.524 (1.554)	–	-0.562 (1.776)
<i>Nationality – Reference group is Non-EU</i>				
UK	-0.349 (1.172)	0.004 (1.588)	-0.130 (2.098)	–
Other EU	-1.064 (1.079)	-0.890 (2.135)	-0.636 (1.700)	–
<i>UCAS tariff score – Reference group is ≥ 300</i>				
0–150	1.165 (2.211)	-0.046 (2.154)	-1.923 (3.284)	0.516 (2.447)
151–299	-0.386 (1.283)	-1.282 (1.166)	-1.177 (2.331)	-1.302 (1.409)
Missing	0.273 (1.362)	–	0.572 (2.154)	-2.111 (1.717)
<i>Occupation of parent or guardian – Reference group is Managerial or professional occupations</i>				
Other occupations	-1.641 (1.409)	-1.851 (1.164)	-0.259 (2.897)	1.071 (1.806)
Missing	0.018 (1.187)	–	0.221 (2.145)	1.769 (1.691)

(Continued)

TABLE 3
(Continued)

Independent variables	(2) Only students who do not have missing information on either family background or UCAS tariff scores			
	(1) All students	(3) Only White students	(4) Only UK students	
<i>Subjects – Reference group is Subject C</i>				
Subject A	0.908 (1.954)	–2.374 (2.591)	3.211 (3.019)	–0.855 (2.812)
Subject B	1.141 (1.111)	–1.231 (1.452)	2.379 (1.820)	–1.002 (1.521)
Class size	0.007 (0.020)	–0.007 (0.027)	0.016 (0.031)	0.007 (0.030)
R-squared	0.039	0.049	0.052	0.049
Number of ob.	1,541	917	538	826

***Significant at the 1% level.

**Significant at the 5% level.

*Significant at the 10% level.

academic calendar such as the number of teaching weeks and the length of semesters), but there are also no obvious reasons as to why one should expect them to significantly affect student performance.

We now perform three additional sensitivity checks in an attempt to address sample selection issues. First, in the first column of Table 4 we restrict our sample to those students for whom Subjects A, B and C were compulsory. This is to tackle the possibility that the change in the scheduling of final exams could have influenced subject choice of students, and unobserved characteristics driving this choice could be correlated with final exam performance. The estimates for this subsample confirm the negative effect associated with the movement of final exams to the end of the academic year. Second, the change in student recruitment policy triggered by the reform could have introduced a selection issue and bias our estimates if students starting university at the beginning of the second semester have systematically different unobserved characteristics that influence educational performance. To address this concern, those students who, in the pre-reform period, enrolled at university in January are excluded from the sample. These additional estimates, which are reported in the second column of Table 4, confirm that the reform has a statistically significant detrimental influence on student performance. Third, since, as outlined in Section 2, the scheduling of final exams at the end of the first semester may cause a higher drop-out rate among first-year students, the reform may lead to a different third-year student population. To account for this possibility, the 2007/08 student cohort (the first for which this possibility occurs) is removed from our sample. We keep in the sample the 2005/06 and 2006/07 student cohorts as the reform occurred while these students were in their second and third year, respectively, and hence they will have faced final exams taking place at the end of the first semester during their first year of study. The estimates, which are shown in the third column of Table 4, indicate that the shift of final exams to the end of the academic year still has a significant negative effect on student performance.

Our identification strategy relies on the assumption that students do not change their level of effort on the mid-term exam in response to the change in the scheduling of final exams.¹⁹ To test

¹⁹ On the one hand, the reform may discourage students from putting a lot of effort into preparation for mid-term exams for first semester subjects given that the corresponding final exams will take place a few months later. On the other hand, however, we focus our attention on third-year students and assessment in

TABLE 4
DiD estimates (robust standard errors in brackets); pooled model

<i>Independent variables</i>	<i>(1) Only students for which Subjects A, B and C were compulsory</i>	<i>(2) Only students who enrolled at university at the start of the first semester in the pre-reform period and all students who attended university in the post-reform period</i>	<i>(3) Only students who enrolled at university between academic years 2002/03 and 2006/07</i>
Constant	-11.334 (8.184)	-5.424 (4.495)	-5.837 (4.583)
End-of-academic year final exam	-4.799*** (1.586)	-3.552*** (0.834)	-3.491*** (0.976)
Female	3.311** (1.327)	2.229*** (0.746)	1.651** (0.777)
Age (years)	-0.001 (0.255)	-0.014 (0.146)	-0.064 (0.145)
Part-time student	0.656 (4.492)	3.416 (2.703)	2.977 (2.197)
Disability status	-6.828** (3.218)	-3.646 (2.798)	-1.146 (4.056)
<i>Ethnicity – Reference group is White</i>			
Indian/Pakistani/ Bangladeshi	0.112 (2.094)	-1.138 (1.185)	-1.439 (1.241)
Black	-2.677 (2.217)	-2.038 (1.413)	-1.603 (1.434)
Chinese	-1.792 (3.417)	2.760 (1.708)	1.126 (1.822)
Other Asian	-1.584 (2.963)	-3.086* (1.829)	-3.459* (1.968)
Others or unknown	1.988 (1.740)	-0.048 (0.974)	0.401 (1.054)
<i>Nationality – Reference group is Non-EU</i>			
UK	0.811 (2.153)	-0.453 (1.208)	-1.397 (1.232)
Other EU	-1.592 (2.146)	-1.009 (1.103)	-1.269 (1.157)
<i>UCAS tariff score – Reference group is ≥ 300</i>			
0–150	5.008 (3.604)	2.016 (2.258)	1.102 (2.268)
151–299	-1.757 (2.138)	-0.073 (1.271)	-1.037 (1.473)
Missing	2.754 (2.486)	0.763 (1.363)	-1.215 (1.502)

(Continued)

TABLE 4
(Continued)

Independent variables	(2) Only students who enrolled at university at the start of the first semester in the pre-reform period and all students who attended university in the post-reform period		(3) Only students who enrolled at university between academic years 2002/03 and 2006/07
	(1) Only students for which Subjects A, B and C were compulsory		
<i>Occupation of parent or guardian – Reference group is Managerial or professional occupations</i>			
Other occupations	–1.469 (2.086)	–1.832 (1.411)	–1.173 (1.620)
Missing	1.007 (1.786)	–0.010 (1.186)	1.219 (1.326)
<i>Subjects – Reference group is Subject C</i>			
Subject A	2.413 (3.721)	–0.117 (1.984)	2.551 (2.078)
Subject B	1.680 (2.349)	1.233 (1.136)	1.323 (1.174)
Class size	0.025 (0.040)	–0.004 (0.020)	0.017 (0.021)
R-squared	0.060	0.046	0.037
Number of ob.	696	1,449	1,286

***Significant at the 1% level.

**Significant at the 5% level.

*Significant at the 10% level.

for this assumption, for each academic year in the post-reform period we compare the observed average mid-term exam score with a prediction calculated using a simple decomposition procedure (Oaxaca, 1973). More precisely, this method is employed to compute what would have been the average mid-term exam score if students who attended university after the reform behaved as those who studied when final exams were taking place at the end of the semester.²⁰ The results of this exercise show that for each one of the academic years in the post-reform period there is no statistically significant difference between the observed average mid-term exam score and our prediction.²¹ Such a finding supports the credibility of our identification strategy.

Next we test the robustness of our estimated treatment effect by estimating placebo effects for two pre-reform years. Two mis-coded reform dummies (i.e., D0304 and D0405) are constructed by artificially assuming that the change in the scheduling of final exams took place in the

subjects in the third year has a lot of weight in final degree classification, one would expect these students to always give their best effort regardless of the timing of the final exams. Additionally, in order to pass each subject, students have to get a minimum score on the final exam as well as on the mid-term exam.

²⁰ Our prediction for the average mid-term exam score is computed as $\delta'_1 \bar{X}_{ijt}$, where \bar{X}_{ijt} is a set of mean characteristics only for those students who studied in the post-reform period, and δ'_1 is a vector of estimated coefficients obtained by regressing the mid-term exam score only for those students who attended university in the pre-reform period against the above characteristics. Whilst \bar{X}_{ijt} varies across each academic year of the post-reform period (i.e., 2005/06, 2006/07 and 2007/08), δ'_1 is the same.

²¹ These results are available upon request.

TABLE 5
Robustness test with placebo regressions

	(1)	(2)
Constant	-6.864 (4.503)	-6.599 (4.565)
End-of-academic year final exam	-3.695*** (0.931)	-3.696*** (0.932)
D0405	0.139 (1.129)	0.410 (1.274)
D0304		-0.546 (1.160)
R-squared	0.0393	0.0394

The dependent variable is the difference between final and mid-term exam scores.

The sample size is 1541. The models comprise age, class size, gender, disability status, part-time status, binary indicators for ethnicity (5), UCAS score (3), nationality (2), subject (2) and occupation of parent or guardian (2). Robust standard errors are given in parentheses.

***Significant at the 1% level.

academic years 2003/04 and 2004/05, respectively. Table 5 presents the results from the placebo regression models where, following the approach of Olsson (2009), the placebo effects are gradually added. The first column of Table 5 shows the DiD results from a specification that comprises D0405 in addition to all the other explanatory variables including the true reform dummy. In the second column of Table 5 we add D0304 to our independent variables. The placebo effects turn out not to be relevant as the only statistically significant coefficient is the one associated with the true reform dummy. These findings suggest that the effect occurred in the academic year 2005/06 is not random.

Given that our previous empirical findings consistently indicate that the shift of final exams to the end of the academic year has led to a decrease in student performance, we attempt to explore whether this effect can be attributed to an increase in stress levels due to the concentration of final exams in a short period of time. One way to address this issue is to include a measure for student stress among the explanatory variables of our model. If lower student performance is due to increased stress at the end of the year, then one would expect the negative effect of the reform dummy to disappear or to considerably decrease in magnitude once we control for this additional variable. Unfortunately, we do not have a good indicator for student stress. The best available proxy, however, is a dummy taking the value 1 if Mitigating Circumstances (MCs) have been granted to a student, and 0 otherwise. Students are encouraged to submit an application for MCs if illness or unforeseen circumstances unavoidably and significantly affect their performance in assessment. This application is considered by the Mitigating Circumstances Board that normally meets within one month of the circumstances occurring and communicates its decision to the student and to the Assessment Board.²² Our assumption is that those students genuinely suffering from high levels of stress would have applied and obtained MCs.

Table 6 reports the DiD estimates once we include MCs among the explanatory variables. Whilst the first column of Table 6 presents estimates of a basic specification, in the second column of Table 6 we include all the control variables. Although, in line with our expectations, MCs do exert a negative influence on final exam performance, one may note that the inclusion of the MCs dummy has no significant effect on either the size or the statistical significance of the dummy related to the change in the scheduling of final exams. Although this finding should be interpreted cautiously given our weak proxy for student stress, it is at variance with the hypothesis that the decrease in student performance caused by the reform is explained by high student stress levels.

²² Scores are not changed as a result of accepted MCs claims, but typically the students have the possibility of retaking the assessment without a cap on the result.

TABLE 6
DiD estimates (robust standard errors in brackets); pooled model

<i>Independent variables</i>	<i>(1)</i>	<i>(2)</i>
Constant	−5.123*** (0.511)	−6.950 (4.427)
End-of-academic year final exam	−4.052*** (0.708)	−3.565*** (0.838)
Female		2.472*** (0.327)
Age (years)		0.005 (0.141)
Part-time student		3.203 (2.293)
Disability status		−2.402 (2.676)
MCs	−3.277** (1.701)	−3.435** (1.743)
<i>Ethnicity – Reference group is White</i>		
Indian/Pakistani/Bangladeshi		−0.934 (1.145)
Black		−2.224 (1.372)
Chinese		0.992 (1.775)
Other Asian		−2.762 (1.746)
Others or unknown		0.044 (0.957)
<i>Nationality – Reference group is Non-EU</i>		
UK		−0.298 (1.167)
Other EU		−1.143 (1.080)
<i>UCAS tariff score – Reference group is ≥ 300</i>		
0–150		1.269 (2.212)
151–299		−0.365 (1.286)
Missing		0.265 (1.370)
<i>Occupation of parent or guardian – Reference group is Managerial or professional occupations</i>		
Other occupations		−1.583 (1.408)
Missing		0.181 (1.196)
<i>Subjects – Reference group is Subject C</i>		
Subject A	−0.402 (1.152)	0.747 (1.966)
Subject B	0.518 (0.747)	1.131 (1.113)
Class size		0.006 (0.020)
R-squared	0.024	0.042
Number of ob.	1,541	1,541

***Significant at the 1% level.

**Significant at the 5% level.

*Significant at the 10% level.

VI. CONCLUSIONS

This paper exploits a reform undertaken by a School of a UK university to examine how, in a two-semester study programme, a change from an end-of-semester final examination format to an end-of-academic year final examination format affects student performance. This means that for first-semester subjects students experienced several months' delay between the end of classes and final exams instead of taking the latter just after the former.

Using a difference-in-differences framework, we find that the movement of final exams to the end of the academic year had a statistically significant negative effect on student performance. This result is robust to different specifications and sample selection criteria. Such a detrimental effect may be driven by two factors. First, students may suffer from high levels of stress because of the concentration of final exams in a short period of time. Second, a long gap between the

end of teaching and final exams may also decrease performance levels. Our study provides weak evidence suggesting the observed negative effect on academic achievement is unlikely to operate through high stress levels.

While our study is based on the analysis of data from one institution only, two considerations suggest the possible generalizability of our results. First, we find that the negative effect of the reform does not differ by the student's nationality and ethnicity. As argued by Bandiera *et al.* (2010), this means that our findings would probably hold in other higher education institutions whose students come from different countries and different ethnic groups, but have an otherwise similar structure of education. Second, our results are consistent across different academic subjects. This suggests that the change in the scheduling of final exams exerts a negative influence on the performance of students who have been exposed to a variety of teaching methods and materials.

These findings may have important policy implications. Changes in examination arrangements should ideally be tested for their impact on student performance before they are introduced. Many changes in higher education are driven not by student learning considerations, but by other reasons such as financial and administrative convenience. Sutmpt and Tymon (2002) have already warned educators about this risk. Specifically, they argue that 'the timing of student assessment must be driven by learning objectives, and not by faculty convenience'.

Finally, the adverse effect on final exam scores in third-year subjects triggered by the reform may have significant repercussions on students' labour market prospects. Several studies (see, for instance, Smith *et al.*, 2000) suggest that final degree classification, which is mainly determined by the grades received by students in the last year of university, is a crucial determinant of UK graduate first-destination outcomes. Therefore graduates who attended universities where final exams take place at the end of the academic year may face lower employment and earnings opportunities.

REFERENCES

- Aral, S., Brynjolfsson, E. and Van Alstyne, M. (2007). 'Information, technology and information worker productivity: task level evidence', *Working Paper 13172*, NBER.
- Bailey, P. D. (1999). 'Assessment of chemistry degrees', *University Chemistry Education*, 3, pp. 64–67.
- Bandiera, O., Larcinese V. and Rasul, I. (2010). 'Heterogeneous class size effects: new evidence from a panel of university students', *Economic Journal*, 120, pp. 1365–98.
- Cameron, A. C., Gelbach, J. B. and Miller, D. L. (2006). 'Robust inference with multi-way clustering', *Technical Working Paper 327*, NBER.
- Casbarro, J. (2004). 'Reducing anxiety in the era of high-stakes testing', *Principals*, 83, pp. 36–38.
- Cecchini, J. J. and Friedman, N. (1987). 'First-year dental students: relationship between stress and performance', *International Journal of Psychosomatics*, 34, pp. 17–19.
- Chevalier, A., Gibbons S., Thorpe, A., Snell, M. and Hoskins, S. (2009). 'Students' academic self-perception', *Economics of Education Review*, 28, pp. 716–27.
- Coleman, D., Bolte, J. R. and Franklin, L. (1984). 'Academic calendar change impact on enrollment patterns and instructional outcomes', *Research in Higher Education*, 20, pp. 155–66.
- Erk, S., Kiefer, M., Grothe, J., Wunderlich, A. P., Spitzer, M. and Walther, H. (2003). 'Emotional context modulates subsequent memory effect', *Neuroimage*, 18, pp. 439–47.
- Fredericks, M. A. and Mundy, P. (1969). 'Relations between social class, stress-anxiety responses, academic achievement and professional attitudes of dental students', *Journal of Dental Education*, 33, pp. 377–84.

- Graves, J. (2010). 'The academic impact of year-round school calendars: a response to school overcrowding', *Journal of Urban Economics*, 67, pp. 378–91.
- Hancock, D. (2001). 'Effects of text anxiety and evaluative threat on students' achievement and motivation', *Journal of Educational Research*, 94, pp. 284–90.
- Merino, B. (1983). 'The impact of year-round schooling: a review', *Urban Education*, 18, pp. 298–316.
- Nichols, A. and Schaffer, M. (2007). *Clustered standard errors in STATA*, United Kingdom STATA Users' Group Meeting.
- Oaxaca, R. L. (1973). 'Male-female wage differentials in urban labor markets', *International Economic Review*, 14, pp. 693–709.
- Olsson, M. (2009). 'Employment protection and sickness absence', *Labour Economics*, 16, pp. 208–14.
- Ross, S., Cleland, J. and Macleod, M. (2006). 'Stress, debt and undergraduate medical student performance', *Medical Education*, 40, pp. 584–89.
- Shields, M. C. and Oberg, S. L. (1999). 'What can we learn from the data? Towards a better understanding of the effects of multitrack year round schooling', *Urban Education*, 34, pp. 125–54.
- Shirom, A. (1986). 'Students' stress', *Higher Education*, 15, pp. 667–76.
- Smith, J., Naylor, R. A. and McKnight A. A. (2000) 'Graduate employment outcomes and university performance measures', *Economic Journal*, 110, pp. F382–F411.
- Suttmpt, S.A. and Tymon, W. G. (2002). 'Interdisciplinary learning in an EMBA: making it happen', *Journal of Executive Education*, 1, pp. 12–23.
- Westerman, G. H., Grandy, T. G., Lupo, J.V. and Mitchell, R. E. (1986). 'Relationship of stress and performance among five-year dental students', *Journal of Dental Education*, 50, pp. 264–67.

APPENDIX: DISTRIBUTION OF FINAL EXAM SCORE BY TIMING OF FINAL EXAMS

