University of New South Wales
School of Economics

Honours Thesis

Badges: A Randomized Controlled Trial[#]

Author:
Mitchell KAZMIERCZAK
Student ID: z3459686

Supervisors:
A/Prof. Isabella DOBRESCU
A/Prof. Alberto MOTTA

Economics (Honours)
AND
Bachelor of Commerce (Finance)
AND
Bachelor of Engineering (Mechanical) (Honours)

22nd November, 2019

[#]This work was approved by the Human Ethics Research Committee at University of New South Wales (20190365). All errors remain our own.

Declaration

I hereby declare that the content of this thesis is my own work and that, to the best of my knowledge, it contains no material published or written by any other author or authors, except where acknowledged. This thesis has not been submitted for award of any other degree or diploma at the University of New South Wales or any other educational institution.

……

Mitchell Kazmierczak

22nd November, 2019

Acknowledgements

Firstly, I would like to thank Isa and Alb for giving me the opportunity to complete this thesis. I know the year hasn't always been easy (most of the time), but I am immensely grateful for the months of work I know you both put in to help this all come together.

To Roxanne, your continued belief in me is what made this submission possible. Always being prepared to drop everything to help me with my grammar and sentence structure is something I am very grateful for.

To Dan, for always reminding me that it'll be done or I'll be dead but either way, it'll be over. Your inspirational words are what made this thesis possible. I think I made it through this one with less than 10 years added to my appearance, so I'll take that as a win.

Lastly to my dad, I guess this thesis was one way to get out of spending time with you on your birthday.

Abstract

We study the impact of micro credentials on educational outcomes. We conduct this by offering private (achievements) and public (badges) micro credentials to university students and track their overall grade and by mark by component in a principles of microeconomics course. With our experimental design, behavioural changes in student interaction towards course material as a response to treatment are simply disentangled. We find that students subject to the public treatment scored 0.19 standard deviations (SDs) higher in the course than those receiving no micro credential offering. IV estimates suggest that those interacting with the public treatment had their course mark increase by 0.22 SDs on average. Using a subsample analysis, we find that the impact is most pronounced in females (0.43 SDs), domestic students (0.19 SDs), those enrolled in a business program (0.22 SDs) and high performing students (75th percentile, 0.29 SDs). The mechanism behind these findings involves students improving their precision when attempting quizzes and practice questions, that is getting a similar number of questions correct while making less mistakes in these attempts.

Table of Contents

Introduction

Digital technologies are rapidly becoming a bigger part of everyday life. Once such technology is micro credentials. These are smaller scale recognition for achieving specific performance milestones across various dimensions. As such they offer the flexibility to recognise achievements in many different forms at many different scale and they will soon be common place alongside larger milestones as the pace of information continues to increase.

Generally micro-credentials take form through online "badges" these are electronic, block chain backed and instantly verifiable against the awarding organisation. While currently only established in information technologies, being recognised by firms such as Microsoft, Adobe and IBM the micro credential market is beginning to expand into other areas. Global micro-accreditation market size is massive and growing quickly. A report by Markets and Markets estimated the size of the digital badges market to be US$65.0 MIL in 2017and is predicted to grow at a compounding annual growth rate of 19.8% to US$205.6 MIL by 2023. With this the introduction of micro credentials into educational institutions is inevitable. The signalling value of education is indisputable, but micro-credentials can also offer information otherwise inaccessible to the labour market. Correctly designed micro credentials can provide information about whether an individual is a fast learner, a team player, an analytical thinker or competitive. From this we ask a simple question how do micro-credentials impact academic behaviour & performance? Do students have demand for these signalling devices and in what way does this shape a students choices?

This work presents experimental evidence on the impact of offering different types of micro credentials on academic performance. To this effect we take advantage of a randomised control trial (RCT henceforth) setting involving first year students at a large, research intensive university.

The RCT occurred in Term 2 2019 and involved the students of a large Principles of Microeconomics course at a large research intensive, selective university. As part of these tests, student were randomly assigned into three groups, one control (who were not offered any micro credentials) and two treatments (whom were offered all micro credentials). The first treatment group, known as the private treatment where students could privately observe

and claim micro credentials (which we refer to as achievements). The second treatment group or public treatment group had the same access as the private treatment except had the additional opportunity to make their achievements public (which we refer to as badges).

We report four main sets of results. First, we find that students in the public treatment scored 0.19 standard deviations (SDs henceforth) higher in the course relative to control students. This result is largely driven by an increase in 0.28 SDs for public treatment students in the final exam compared to their control group peers. this effect is comparable in magnitude to employing teachers that are 2 SDs above the average or reducing class sizes by 20%. While already significant we move beyond these intent-to-treat estimates and using the random assignment as an instrumental variable for treatment take up (determined by the claiming of at least one micro credential). This treatment-on-treated analysis provides an increased effect size of roughly 10% for public compared against control in the overall course and final exam respectively.

Second, we also explore heterogenous treatment effects by conducting a subsample analysis along several dimensions including gender, international status and those studying towards a business degree. Key findings include significant positive effects for both treatments compared to control in several outcomes. In particular females, in the private and public treatment experience an increase of 0.294 SDs and 0.430 SDs in the overall course respectively when compared to females in the control group. The results for the overall course are also significant when compared against males in the same treatment group. We also find significant results in the overall course for domestic students and those studying towards a business degree within the public treatment.

Third we examine the heterogeneity of ITT effects across the grade distribution. We use the unconditional quantile regression (UQR henceforth) as it provides more interpretable results since the method marginalises the effect over the distributions of other covariates in the model. Our analysis shows that all students benefited from the public treatment for the final exam and the overall course. The exception to this is students in Q25 which have a positive but statistically insignificant impact on the overall course despite being the most impacted by the treatment in the final exam. As a robustness check we also conduct conditional quantile regressions (CQR henceforth) and find the same trend as with the UQR except that the overall course is now also marginally significant for low performing students.

Finally, we also investigate the mechanism driving our results and find our findings to be consistent with the competitiveness theory as discussed in (Azmat and Iriberri, 2010). Given the opportunity to acquire micro credentials all students, regardless of the position within the grade distribution, increase their level of effort which in turn provides better academic performance. We explore two sets of measures that could capture a higher amount of effort. The first is related to spending more time engaging with the online materials, while the second is related to higher effort intensity for the same time amount spend. In order to examine these potential avenues we utilise software logs for the online components of the course. Investigating spending more time as effort we run regressions along a number of dimensions including the number of log ins to the online assignment, number of log ins to the online course material (which includes the textbook), number of page views on the online course material, videos watched (these are embedded in the textbook) and time spent in the online assignment. We find that along each of these dimensions we find no significant results in any pairwise comparison between the control, private treatment and public treatment. In addressing the second effort style, we consider measures that show the precision of a students work within the online components of the course. To this end we consider the number of passes and fails for quizzes (part of the online assignment grade allocation) and academia (practice questions) in the online assignment. The quizzes show significantly less fails for both treatments compared with control but is unable to distinguish between the two treatments while the impact of the treatments is localised to the public treatment. Expanding the analysis to the academia questions we see that the public treatment fails significantly less questions than private treatment. As such while we see that both treatments increase their effort through greater precision when attempting the quizzes. Only public treatment students extend this increased precision through to the academia practice questions. It is this sustained higher effort level and precision in both question formats that leads to a significant impact in course performance.

To sum up, offering micro credentials, despite not being related to student's final course grades, has a positive and sizeable impact on their academic performance, both on average and across the entire grade distribution. Moreover, this type of instruments also has the added benefit of equipping students with additional signalling devices. This could potentially

bring substantial advantages on the labour market, that is increasingly focusing on professional skills that go beyond those conveyed by a university degree.

We contribute to the literature of educational interventions along four dimensions. First, our experimental design allows for the clean identification of behavioural response to micro credential offering. Second, we are the first to provide evidence that offering micro credentials can be effective in higher education as a tool to improve academic performance. Specifically, we find that public micro credentials (badges) have a positive impact on both final exam scores and overall course grades. Our third contribution is to examine how micro credential characteristics vary student behaviour by analysis alternative methods of micro credential provision to students. Finally, we provide insight to the mechanism through which micro credentials translate into academic performance. Our findings suggest that when micro credentials are available students place more effort into their learning through higher precision in conducting their work rather than working longer. This greater effort leads to increase human capital accumulation and as a result significantly improves results in the final exam and overall course.

While the literature on micro credentials as an educational intervention is virtually unexplored. We can draw some insight form pervious studies as to where they fit in with other related areas. Signalling theory, commonly used to explain how the labour market determines the quality of a candidate. Previous studies shown that college prestige is sufficient to signal ability and graduation provide a near perfect signal (Arcidiacono et al., 2010, MacLeod et al., 2017). However, with the increased demand for individuals with soft skills the labour market is looking for signals beyond ability (Bacolod et al., 2009). Other amalgamated theories such as sorting models we combined signalling and human capital theories provide a clear way for badges to be significant inputs to both models. We add to this line of literature by investigating how to students value their signal and do they demand additional signalling devices?

This thesis next discusses our RCT and introduces its context. From this we present the data and our empirical approach, followed by a discussion of our findings and the driving mechanism. We then conclude by discussion future research areas.

1. Randomized Controlled Trial

1.1. Environment

The RCT we are examining occurred in Term 2 2019 and involved the students of a large Principles of Microeconomics course at a large research intensive, selective university. The course runs in every teaching period with between 3,500 and 4,000 students annually. Over a 10-week teaching period students attend two 1.5-hour lectures. There are no structured tutorial classes that students will attend. If they so desire, however, students are able to join independent 1.5-hour study groups that are offered weekly and where they can ask questions about any issues/exercises they find problematic. Neither lectures nor study group participation are compulsory. Lectures are delivered by academic staff, while the study groups are run by teaching assistants (known as 'tutors'). All lectures and study groups take place on the same campus. Students are randomly allocated to a lecture stream and have no prior knowledge of the staff allocation. They are free to attend any and as many lectures and study groups they want during each term week, including none of them. All instructors (i.e., lecturers and tutors) teach and provide guidance using the same teaching material, including textbook, course notes, slides and revision exercises (with standardized solutions), all of which are also available online through the course website. Finally, there is an active discussion board associated with the course, which is open to all students and all instructors and is used to post comments that ask or answer questions related to course administrative matters and/or course content.

From 2015 onwards, an educational software was adopted as part of the course material. The software provides an extensive set of exercises and questions available for students to practice their understanding of the course concepts, and that include links to the course textbook. The textbook covers all the topics traditionally taught in a standard Principles of Microeconomics course, from the Principle of Comparative Advantage to Externalities and Public goods. The exercises provided via the educational software focus on different economic topics corresponding to each one of the 10 textbook chapters, and are grouped as such. While students are encouraged to complete the content as it is released, they may work through the exercises in any order at any time during the term. There are two types of exercises that students can complete. The first is quiz-type exercises displayed in a traditional

multiple-choice format (MC henceforth), with such questions drawn from an extensive database of approximately 500 questions. The second type of exercises asks them to apply the course concepts in an experiential learning setting. Correctly solving a certain number of either type of exercise by the end of the teaching period (i.e., completing the online assignment) was worth 20% of the overall course grade. Partial completion of the assignment resulted in students receiving a corresponding proportional grade.

In addition to this term-long online assessment, the course structure also included an invigilated mid-term exam in week 6 that contained a mix of essay and multiple-choice questions and was worth 25% of the overall grade. The remaining 55% of the course grade was allocated via an invigilated final exam, taking place at the end of the teaching period and consisting exclusively of multiple-choice questions. The exam questions for both invigilated exams are decided by the course coordinator who draws the corresponding questions from a pre-existing database of uniformly difficult questions. Casual markers are responsible for marking essay questions in the mid-term exam; marking is double blind and follows a strict set of marking guidelines provided by the course coordinator with rigorous consistency checks conducted at all stages of the marking process. All multiple-choice questions - in both the mid-term and final exam - are machine graded.

1.2.Treatments

As mentioned, students completed the online assignment by correctly solving MC questions and experiential problems. Correctly solving exercises gained points, while incorrect answers detracted points. Doing certain exercises or reaching certain points thresholds was associated with being awarded specific ‘achievements’, i.e., digital rewards given to students for achieving specific milestones in the assignment. Once one became eligible for an achievement, that achievement was made available online and the student could claim it by clicking a button. Once claimed, the achievement was added to a rewards list accessible by clicking a ‘rewards’ button. The lowest level achievements were calibrated to be easily attainable by students of any ability, while the highest level of achievements was extremely difficult to achieve. Gaining an achievement had no bearing on the grade associated with the online assignment.

To test whether making achievements available had any impact of students' behaviour, a (platform-wide) RCT was conducted varying the way in which achievements were made available. Students were randomly assigned – based on the last number of their student ID – at the beginning of the term into one of three groups, namely (i) control; (ii) private achievements; and (iii) public achievements (i.e., badges). Students in the Control group did not have access to achievements, although they were notified when reaching the threshold associated with them. Students in the (private) Achievements group had achievements made available to them when they reached the associate milestones; once claimed, the achievements were added to their rewards list, which was available only to them. Finally, compared to the Achievements group, students in the Badges group had the additional opportunity to make their achievements public (primarily on Facebook and Linkedin).

Students were not aware of the nature of these tests and did not know that they were being treated. Furthermore, the course discussion board that the three groups could access were also separated. Additionally, none of the instructors or markers were involved in running any of the tests and were not aware of the treatment group that each student was assigned to. This allows us to cleanly identify the impact of demand for signalling devices (through private or public performance achievements) on student academic performance in the online assessment and more broadly the course. In what follows, we will attempt to do so, as well as investigate the potential for heterogeneous treatment effect and the mechanism behind our results.

2. Data and Empirical Analysis

2.1. Data and Descriptive Statistics

The data we use in this analysis comes from university administrative records, course performance records, software logs (for the online assignment) and the course discussion board. Our data contains a sample of 645 students, from 33 countries taking the Principles of Microeconomics course during Term 2 2019. As mentioned, while students are randomly assigned to classes when they enrol, they are effectively free to attend whichever physical classes they choose to or none at all.

Table 1 presents the key descriptive statistics for the cohort at a student level. A quick glance to the figures reveals that students are on average only slightly older than 19 years old, with a 40:60 split between females and males. Additionally, over 70% of students are domestic, while the remaining majority comes from Asia (around 27%). Slightly more than 60% of students are undertaking a Business degree, with an additional 22% studying degrees in Science, Technology, Engineering & Mathematics (STEM henceforth).

For ease of comparison, student performance in each of the course assessments in scaled out of 10. We see that students performed well in the online assignment with an average score of 9.4 out of 10, while in both invigilated exams the average score was close to 5.8 out of 10. Overall students scored 6.5 out of 10 for the course, with a standard deviation of 1.2.

Table 2 provides the balance checks between the control and each treatment and then between the two treatments, based on the student characteristics available in our data. These characteristics include students' age, gender, degree undertaken, international student status, ethnicity and previous academic ability. The first two columns of the table provide the mean and standard deviation for the specified treatment group and the following two provide the same descriptive statistics for the Control group. The final two columns show the difference in means and the associated standard error. Our results show that none of the observable characteristics available in our data are statistically significant at conventional levels between any of the three groups in a pairwise comparison. We can therefore conclude that the randomization was successful.

One remark regarding our measure of previous ability. For domestic students, we use their Australian Tertiary Admission Rank (ATAR hereafter) high school score to capture previous academic ability. ATAR represents a student's percentile rank within their high school cohort and it is computed based on a combination of their score in the final year assessments and in a common final exam across their best five subjects (or equivalent). The international students enrolled in this course, however, come from 32 different countries and have usually completed high school in their country of origin. As their international high school graduation and international university admission exams are of different scale and difficulty, the university where the experiment was conducted only uses this information for admission into its programs and does not maintain it. To proxy academic ability for this sub-sample, we will use the grade point average (GPA henceforth) they achieved during the previous term (i.e., Term 1 2019) when available.

In all our main analysis, we will control for all student characteristics with the exception of prior ability. This is due to the fact that our prior ability indicator is based on multiple data sources and balanced across our three groups. We do conduct, however, several robustness checks that account for this additional element (see Appendix A).

3. Empirical Methodology

3.1. Course Effects

Our identification strategy involves comparing the outcomes of students that share similar characteristics but whom are given different micro-credential treatments. We will start by examining whether the various randomly assigned treatments had any impact on student's performance in each of their assessment items as well as in the aggregated results. Specifically, our outcome variables are related to performance in the online assignment, in the invigilated week 6 and final exams separately, the weighted average score of the two invigilated exams and the overall course mark.

This information provides a good measure of student learning and performance over the term for various reasons.

First, there is no systematic way an instructor could have shifted the performance scores for a particular treatment since they never knew which treatment a student was assigned to. Second, both exams are close-book and invigilated which makes them an objective measure of student attainment and capability. Third, the week 6 exam was marked to strict marking guidelines by markers who were unaware of the experiment. Marking was also peer-reviewed by senior staff members to ensure consistency with and adherence to the guidelines. Fourth, the final exam marking is fully computerised, leaving no potential for subjectivity. Finally, exam grades are not adjusted, re-weighted or scaled in any manner such that they are a true impression of a student's absolute performance during the course.

This results in the baseline estimating equation taking the form:

$$Y_{i,d} = \alpha + \beta Treatment_{i,d} + \gamma X_{i,d} + u_{i,d}$$

where $Y_{i,d}$ is the score achieved by student i enrolled in degree d, in any of the course assessments, the weighted average exam score or the overall course grade. The dummy variable $Treatment_{i,d}$ takes the value of one if a student is a member of one of the treatment groups or zero for students belonging to the Control group. $X_{i,d}$ captures individual student's characteristics (i.e. age, gender, international status, dummies for country of birth groups, and a dummy for those students undertaking a business degree). $u_{i,d}$ is the error term.

We note that no control is used for instructor or tutor effects given the random assignment to treatments. Additionally, no control is used for tutorial classes given students' random attendance to their classes (lectures or study groups) of choice.

One final note regarding potential spillover effects between various experimental groups due to students potentially attending the same classes. While we cannot fully exclude this possibility, there are several elements that considerably mitigate this issue. First, note that the treatments only influenced their online assignment and on this dimension students in each treatment group were given identical (except for the treatment) yet completely segregated versions of the online assignment and course website, so that no spillover could occur via the online course components. Second, neither lectures nor study groups ever discussed the online assignment (which contained the treatments) except for a short administrative presentation prior to the release of the online assignment. Finally, as we mentioned already, students were free to attend (or not) any lecture and/or study group

during the term and this high mobility minimised any discussions related to topics that were not scheduled or not directly related to academic material. We thus conclude that any spillover or contamination is extremely unlikely.

3.2. The Mechanism

In order to determine what is driving students' response to the provision of micro-credentials, we use additional data from the software developers' logs and the course discussion board. Both sources of data allow us to construct proxies of student effort. This effort may be either direct through greater involvement with the online assignment, or indirect through peer-to-peer discussions using the course discussion board. A key component to these logs that capture student progression and involvement with the course, is the number and nature of attempts for two central elements of the online assignment. The first element is the number of successful and failed for a traditional, multiple-choice format quiz. Note that (i) if a student failed a quiz they would receive a new, random set of questions drawn from the software database on their next attempt (with no penalty to the grade awarded), and (ii) students could successfully pass each quiz only once. These quizzes constitute one of two ways to achieve marks in the online component of the course. The second element is the number of passes and fails on specific (extra) questions in the software database (known as the academia questions in the online assignment). Students had (i) access to the full database of questions that the quizzes draw on, and (ii) unlimited attempts at all academia questions whether they passed or failed such a question. As academia questions were not worth any marks, students had no incentive to answer these questions (especially given they could continue to attempt the quizzes until they passed them and gained the associated marks). Therefore, attempting the academia questions is purely a voluntary, formative assessment. To see whether the offering of either treatment type produces a difference in the way students interact with the online assignment, we will run a modified version of our baseline estimating equation using the number of successful and failed attempts in either the quizzes or academia questions as our outcome variables.

4. Results

4.1. Baseline Estimates

We begin the results section by discussing the estimates of the impact of micro-credential provision on student performance in each course assessment, weighted average exam score and overall course grade. Results on the intention-to-treat are presented in Table 3. Column (1) shows the effect on the online assignment, columns (2) and (3) show estimates for the week 6 and final exams, column (4) for the weighted average of those exams and column (5) for the overall course grade. Each column shows the standardized performance ($\mu = 0, \sigma = 1$) to allow for easier interpretation and comparability across studies.

The first row presents the effect of the provision of achievements (private treatment) compared with control, while the remaining rows present the impact estimates for the provision of badges compared with control and finally the provision of badges compared with the provision of achievements. The estimates suggest the badges treatment effect to be positive and statistically significant in the final exam, weighted exam and overall course. Additionally, the impact of badges treatment compared to achievements is also positive and statistically significant for the final exam. The associated coefficient indicates that students in the badges treatment outperform their peers in the control group by 0.28 SDs and outperform their peers in the achievements treatment by 0.16 SDs in the final exam. This largely drives an increase in overall course performance of 0.19 SDs for the badges treatment compared to the control group. While the estimates for the effect on the online assignment, week 6 exam and remaining estimates for the achievement treatment are not statistically significant we can descriptively see that they are all negligible or positive. This implies that we cannot definitively conclude that the achievement treatment has a positive impact on student academic outcomes. However, we can say that the badges treatment performs better against the control group than the achievement treatment against the control where results are positive but insignificant. This paired with the lack of significance between the two treatments for the overall course performance could suggest that the achievements treatment may have had a small positive impact that we could not significantly detect in our data. Intuitively, this can be explained by the private achievements treatment somewhat representing a goal setting nudge. These types of nudge are mixed in impact and are highly dependant on the

implementation and the outcome assessed; even when positive the effect is normally restricted to a particular subgroup (Clark et al., 2017) (van Lent and Souverijn, 2017). We hence see that it is only with the added incentive of a badge that students change their behaviour sufficiently to realise the impact of the public treatment.

The course effect of the badges treatment (0.19 SDs) is outstanding. Specifically, it is of comparable magnitude to being taught by a teacher between 1.5 and 2 SDs above the average (Hanushek et al., 2005; Chetty et al., 2014) or to reducing the class size by 20% (Angrist and Lavy, 1999; Krueger, 1999). In contrast to these interventions, however, the offering of micro credentials has a significantly lower cost to implement, and with the rapid digitalisation of education it is also increasingly feasible to use, being easy to tailor in order to follow a particular course design or lesson planning. Finally, micro credentials offer students an opportunity to acquire additional signalling devices for the job market and while college graduation is considered a near perfect signal (Arcidiacono et al., 2010), micro credentials give those achieving them an opportunity to improve their signal, especially for junior students looking to gain their initial experience.

4.2. Heterogeneity

While these results themselves have already substantial policy implications, it is important to investigate any possible variation in effect across different categories of students. To this effect, we split the sample along various observable dimensions and use our baseline estimating equation for each subsample. Results are reported as standardized performance and are presented in Table 5, Table 6 and Table 7.

First, we look at the treatment effect by gender as presented in Table 5. The estimates show positive and significant results for females for both treatments in a number of assessments. Specifically, we see the effect of the badges treatment for females rise from 0.28 SDs to 0.47 SDs in the final exam and from 0.19 SDs to 0.43 SDs in the course overall, a considerable increase in impact. Females in this treatment also have a significant and positive effect for the week 6 exam of 0.34 SDs. While smaller than the overall course badges impact, females also experience a positive significant impact in the achievements treatment of 0.29 SDs; there is no significance difference between the treatments as with the full sample analysis. The

significance of these results is notable considering that females only make up about 40% of the sample, which increases the standard errors considerably in this subsample analysis. For males, the impact of both treatments on final exam performance remains positive, but is now statistically insignificant. Interestingly, the impact of the badges treatment for males on the overall course becomes negligible when compared to the size of the standard errors. We speculate that this is due in part to the negative but statistically insignificant effect of the badges treatment for males on the online assignment. A negative coefficient, meaning a lower mark could be considered a proxy for less effort invested in the online assignment. This lower effort would also mean less interaction with the treatment than their females counterparts, and hence an insignificant impact on their final exam leading to a negligible (and insignificant) effect of badge treatment on male's overall course grade. Additionally, we looked at the difference in impact between males and females within the same treatment group. We found that all significant sub sample results were also significant between males and females within the same treatment, except for females' the final exam in the public treatment compared with control.

Second, a substantial proportion of our sample are international students (29%). We report the treatment effects by international status in Table 6 using the same baseline estimating equation for our domestic and international subsamples. As with the original analysis, we find a positive and significant impact of the badges treatment on the final exam for both subsamples (0.26 SDs for domestic, 0.33 SDs for international), as well as a beneficial effect on the average exam score and overall course grade but for domestic students only. Interestingly, the point estimates for international students are larger than those for domestic students in both treatments for the final exam, average exam and overall course. However, many of these estimates for international students are not statistically significant, likely due to the relatively lower number of observations in that subsample.

Third, we divide students into those that are studying towards a business degree and those enrolled in other programs, presenting the results in Table 7. The estimates suggest that those studying a business degree are more receptive to the badges treatment, with positive significant results for the final exam, average exam and overall course. The effect size is similar for the final exam and increases from 0.19 SDs to 0.22 SDs when just looking at the business subsample. Additionally, we note that the effect for non-business students in both treatments

is positive but not statistically significant, potentially due to the smaller fraction of non-business students (~30% of the total sample).

Finally, the heterogenous effect split along age was not analysed as the sample could not be partitioned in such a way that both subsamples had a balanced – even in some relatively small measure – number of observations.

4.3. Non-linearities

The baseline estimating equation has so far assumed that the impact of the treatment is linear in nature. However, there is no reason to make this assumption and the effect may indeed vary across the grade distribution. It may be the case that low performing students benefit from the treatments while high performing students are unaffected or at the very least impacted differently.

To investigate this potential issue we allow for non-linear treatment effects by running both unconditional quantile regressions. We estimate the impact of the unconditional distribution of grades as follows:

$$Y_{Quant\theta} = \alpha_\theta + \beta_\theta Treatment_i + \gamma_\theta X_i + \varepsilon_{i,\theta}$$

where θ represents the 25th, 50th and 75th quartiles or each decile $\theta \in [0,1]$ where appropriate. Table 8 presents the quartile regressions for performance in the week 6 exam in columns (1)-(3), the final exam in columns (4)-(6), the weighted average exam in columns (7)-(9) and the overall course in columns (10)-(12). Figures X,Y,Z plot the β_θ coefficients for the decile regressions and the 95% confidence interval. In all regressions, the outcome variable is again the standardized performance. The decile analysis computes standard errors using bootstrapping with 500 replications. $\varepsilon_{i,\theta}$ is the error term.

We use the unconditional quantile regression as it provides more interpretable results since the method marginalises the effect over the distributions of other covariates in the model. Given our education policy context, it is preferable to have these more interpretable results to better look at the heterogeneity across our set distributions. As the conditional quantile regression is more widespread method, we also run these models as a robustness check.

The unconditional quantile regression analysis supports our previous findings related to the positive and statistically significant impacts on the final exam, average exam and overall

course for the badges treatment. We also observe that the estimates for this treatment are largest for lower percentiles. Specifically for the final exam the effect estimate is 0.38 SDs, 0.33 SDs and 0.29 SDs for Q25, Q50 and Q75 respectively. A similar pattern is observed for the overall course with an effect estimate of 0.25 SDs, 0.24 SDs and 0.17 SDs for Q25, Q50 and Q75 respectively. (Note, however, that the Q75 estimate is not statistically significant). Figures 1-4 support this pattern illustrating that the badge effect estimates are largest for middle quantiles, especially those in Q30 and Q40. These lower quantiles also represent the majority of the significant results. Across all badge estimates, we can observe a smaller - or even negative - impact on the top and bottom quantiles, albeit none are significant. Finally, none of the achievement quantiles are significant at the 95% level.

The same analysis was re-run instead using the conditional quantile regression model. Table 9 presents the quartile regressions and Figure 5-8 plot the β_θ coefficients for the decile regressions and the 95% confidence interval. The quartile regression estimates are similar in absolute value to the unconditional quartile regression, with most shifting by no more than 10% of their original value. We do, however, observe more statistically significant values such as the Q75 for the overall course mark. All significant estimates support the same argument that lower quartiles tend to experience a larger effect and that badges have a larger impact than achievements.

4.4. IV estimates of treatment-response relationship

All results presented to this point are intent-to-treat estimates. A total of 52 students, however, did not interact with the treatment being offered. We now present LATE estimates of the impact of the treatments on those that actually interacted with them by claiming such achievements. We estimate the treatment response relationship between claiming micro credentials (as opposed to not claiming any) and our outcomes using

$$Y_{i,d} = \alpha + \beta Claim_{i,d} + \gamma X_{i,d} + \eta_{i,d}$$

where $Y_{i,d}$ is defined as previously, and the dummy variable $Claim_{i,d}$ takes the value of one if a student claimed at least one micro credential or zero if they did not (which is true for all control students). $X_{i,d}$ refers to an individual student's characteristics (i.e. age, gender, international status, dummies for country of birth groups and a dummy for those students undertaking a business degree), and $\eta_{i,d}$ is the error term. Since micro credential claims may

be endogenous to expected gains, we instrument for $Claim_{i,d}$ with the randomized allocation to the treatment.

The IV estimates shown in Table 4 suggest that, on average, public treatment students that claimed at least one of the micro credentials they achieved outperformed their peers in the control group by roughly 0.32 SDs in the final exam, 0.24 SDs in the average exam and most importantly 0.22 SDs in the overall course. These estimates identify the average casual response (ACR) of the treatment which 'captures a weighted average of casual responses to a unit change in treatment (in our case, whether a student claimed any of their achieved micro credentials), for those whose treatment status is affected by the instrument' (Angrist and Imbens, 1995).

Further discussion of these IV estimates is dependant on the mechanism that explains why treatments did not have 100% compliance. It is possible that those that did not claim any of the micro credentials failed to do so simply because they did not know how or misunderstood the information presented. This would then mean that the IV estimates provide information of the average treatment effect on the treated. Alternatively, some students may have understood how to claim the micro credentials but preferred not to or did not see any value in claiming them, even while perhaps seeing them. In this case, the interpretation of the IV estimates is more complex.

Irrespective of this, we can definitively say that we find significant results for all our academic performance measures, as with our intent-to-treat analysis. Additionally, we also see that all the point estimates have increased by roughly 10% compared to the intent-to-treat ones. This suggests that if we can identify why the non-compliers exist - due to lack of information or understanding of how to access the treatments - it may be possible to increase compliance rate by providing improved information about the micro credentials. This could then further drive up the positive effect on outcomes when considering the entire cohort (intent-to-treat), which is important as generally better students could improve the value of the signal they provide to the labour market through their degree.

5. Mechanism

Our findings clearly show that students exposed to the public treatment outperform control students in the final exam, average exam and most importantly, the overall course. This

suggests that students have a demand for these badges that they can use to enhance the accuracy of their signal to the labour market and potentially distinguish themselves from their peers that would otherwise provide a similar signal. Furthermore, this effect appears to be significant in impacting student outcomes irrespective of a student's grade position within the cohort. This demand is thus a way to gain a competitive edge and can provide incentives regardless of a student's ability, consistent with a model where competitive preferences induce everyone to exert more effort.

While we are not the first to use nudges, this intervention does so via micro credentials. Up to our best knowledge, we are the first to manage to measure the impact of micro credentials and give direct evidence as to the mechanism that is driving the results.

We offer two possible suggestions as to how students could be increasing their effort, hence human capital accumulation in response to the treatment. First, students could simply be spending more time working on the online assignment and more generally engaging with the course material, i.e., they could increase the amount of effort they exert in the course. Second, students could increase the level of precision or focus during the (same amount of) time they spend studying for the course, i.e., they could increase the intensity of effort they exert in the course .

To proxy the first type of effort we use the software logs data along several dimensions that would indicate increased effort through time spent on the course. These include the number of log ins into the online assignment, number of log ins to access the online course material (via the textbook), number of textbook page views, number of videos watched (as they are embedded in the textbook), and total time spent in the online assignment. On none of these dimensions do we find significant results, in any pairwise comparison between the control, private treatment and public treatment. We conclude that the impact of our treatment, and hence mechanism, is not the result of this type of effort.

To proxy the second type of effort we again use the software logs data, this time related to measures that show the precision of a student's work. To this end we consider the number of successful and failed attempts for quizzes and academia questions in the online assignment. Given the random (and balanced) allocation of students to treatment groups, we should expect no difference in either of these dimensions. Additionally, students had unlimited

attempts to pass a quiz (and gained the associated marks) so they should not care about failing the quiz any number of times. And academia questions were not incentivized at all (i.e., brought no course marks whatsoever). Part of the micro credential offering, however, included achievements and badges for successfully completing a certain number of quizzes (there were a total of 10 quizzes for the entire course) or academia questions, as well as achievements and badges awarded for the precision of the associated attempts, such as passing a quiz on the first attempt or remaining in the top 10 for their academia score determined by both the number of questions completed and attempts to pass that question. This means that any behavioural difference observed between our treatment groups will most likely be due to the treatment they were subjected to.

Table 10 presents our regression results for (i) the standardised number of successful quiz attempts at the final due date of the online assignment in columns (1-3), and at the end of week 6 in columns (4-6), and (ii) the standardised number of failed quiz attempts at the final due date in columns (7-9) and at the week 6 in columns (10-12). Specifications (1), (4), (7) and (10) use the baseline model, while specifications (2), (5), (8) and (11) include no controls and specifications (3), (6), (9) and (12) include an additional control for prior student ability. The baseline specification shows that there is no significant impact on the number of quizzes passed between either treatment and control or between the treatments. This is unsurprising since students can only pass each quiz once, and since they are worth course marks it is anticipated that all students will complete a rather similar number of quizzes. We also see that the number of quiz fails is significantly lower for both treatments compared with the control group, both at the end of week 6 and at the quizzes due date. Students in the private treatment failed 0.18 SDs fewer quizzes by the week 6, which grew to 0.26 SDs fewer fails by the due date. Students in the public treatment failed 0.26 SDs fewer quizzes by the week 6, which grew slightly to 0.27 SDs fewer fails by the due date. While we do not see a significant difference in the number of quizzes failed between the two treatments, we can descriptively observe that the point estimates are greater for the public than for the private treatment at both points in time. The other specifications yield the same significant results and follow the same trend, hence supporting the robustness of our baseline results. This provides evidence that the effort intensity channel is at play, with students from both treatments reacting to the provision of micro credentials by increasing the precision of their work, hence making fewer

mistakes to qualify for the accuracy style micro credentials. This greater precision in learning then boosts their course performance compared to other students who spend a similar amount of time preparing for assessments. We note however that the analysis is unable to distinguish between the two treatments, while the impact of the treatments is localised to the public treatment.

In order to address this problem, we expand our analysis to investigate the students performance related to the academia questions. Table 11 presents our regression results for (i) the standardised number of successful academia question attempts at the final due date of the online assignment in columns (1-3), and at the end of week 6 in columns (4-6), and (ii) the standardised number of failed academia question attempts at the final due date in columns (7-9) and at the end of week 6 in columns (10-12). Specifications (1), (4), (7) and (10) use the baseline model, while specification (2), (5), (8) and (11) include no controls and specifications (3), (6), (9) and (12) include an additional control for prior student ability. The baseline specification shows that, similar to the quizzes, we find no significant impact on the number of academia questions passed between either treatment and control or between the treatments. Unlike the quizzes, however, we now observe a very different trend in the question fail results. The baseline specification shows a decrease of 0.12 SDs in academia question fails by the end of week 6, which grows to a fall of 0.21 SDs by the end of the study period. Neither treatment has significant results against the control group. The other specifications yield the same significant results and follow the same trend, confirming the robustness of the baseline results. This provides the missing link to the effort intensity (i.e., precision) channel when combined with the quiz analysis. While both the private and public treatment induce greater effort in the form of improved precision on the quizzes, only the public treatment extends this to the voluntary academia questions. While not significant, we can descriptively see that students in the public treatment fail less academia questions than their control counterparts at both the end of week 6 and the start of the exam period.

These results combined may be the result of private treatment students being satisfied with only meeting the requirements of micro credentials that require less effort or can be attained in conjunction with other goals, such as course marks. As private treatment students cannot share their achievements, and with the achievements not given any real course value, it is not unrealistic to assume that a student will only want to maximise their number of achievements

regardless of their type. In contrast, public treatment students may be more interested in meeting the requirements of more challenging or complex micro credentials, specifically those with a badge associated with it. This is consistent with our behavioural analysis of improved student precision. More than this, it shows that students are willing to place more effort in achieving specific micro credentials, i.e., that there is a demand for them. This behaviour then in turn leads to greater human capital accumulation and generates higher performance in the course for public treatment students.

To sum up, compared to control students both treatment groups improve the precision of their quiz attempts by having significantly fewer fails. However, only public treatment students extend this increased precision through to the extra academia practice questions. It is this sustained higher intensity of effort via increased work precision in both question formats that leads to a significant impact on course performance. As an additional robustness check, we also run the IV specification for both mechanisms and present estimates in Table 19 and Table 20. This analysis shows significant results in the same direction as the intent-to-treat regressions, and similar to the main results, we see an increase in the effect size of roughly 10%; all trends in the results remain the same. We also suggest that while the significant results presented here are robust, many of the descriptively interesting but statistically insignificant results are being lost due to potential power issues with the size of our sample. Further research is required to examine if these are random or significant but too small to detect in this study.

6. Robustness

To check the robustness of our results, we re-run the intent-to treat and treatment-on-treated analysis using two additional specifications. The first includes none of the controls used in the baseline analysis. The second includes the full suite of controls (i.e., age, gender, degree undertaken, international student status, ethnicity), as well as our measure of previous academic performance to account for prior ability. As mentioned, using ATAR scores would severely restrict the sample size as this information only available for domestic students. Hence, missing previous academic ability values (for international students) are

taken as GPA attained in the previous term when available. Constructing the proxy for prior ability in this way follows (Dobrescu et al., 2019) and yields non-missing values for 0.16% of our sample. Table 12 and Table 13 present the estimates for the intention-to-treat and treatment-on-treated respectively for both types of specifications. We observe that the estimates remain very similar (i.e., estimates change by less than 0.01 SDs), with the same level of significance and the key trends remaining unchanged. That is, the final exam, average exam and overall course scores remain statistically higher for students belonging to the public treatment when compared to control students.

The same two additional set of specifications were also run for each of the heterogenous analyses to check their robustness, with results presented in Table 14-16. For our gender analysis, both specifications in Table 14 show similar point estimates as with the main analysis. Some marginally significant results are not significant in the robustness checks, but the key significant results are also the same, i.e.,, females in the public treatment experience a positive effect for all outcomes (except the online assignment) compared to females in the control group. As for international status results, Table 15 shows that all significant results are robust, with students in the public treatment experiencing positive effects compared to control students for (i) the final exam regardless of international status, and (ii) the average exam and overall course for domestic students. Finally, in Table 16 we see that all significant results for field of study remain robust too: Public treatment students undertaking a business degree experience positive effects in the final exam, average exam and overall course when compared with business control students. Additionally, in the first specification the same comparison is also marginally significant for the week 6 exam.

To provide further evidence as to the robustness of the results we conduct a randomized treatments type of robustness test using the original specification. These new placebo treatments then do not reflect a students' real treatment assignment and allow us to test if a similar of pattern of results can be found as in the main analysis. We did this by ranking observations using a new random variable that divided students into the three groups. If the placebo treatments are found to be significant determinates of performance, this may indicate that students are reacting to confounding factors (not perfectly coinciding with the real treatments) that lead to the increased performance. If a correlation exists between the confounding factors and the real treatment assignment, then the placebo treatments would

also detect some of these effects. We conduct this analysis by using the baseline estimating equation from the main analysis but substituting the real treatments for our randomly generated placebo treatments. The results presented in Table 21, show no robust statistically significant differences from any placebo treatment group in any of our outcomes. We thus conclude that our findings are unlikely to be the result of simultaneous effects other than the treatment prescribed to each group of students.

General comment: Throughout your results, try to come up with arguments as to why you're seeing the patterns you're seeing, try to propose at least 2 such 'economic stories' for each results section (main results, heterogeneity, and mechanism)

7. Conclusion

Micro credentials are becoming an increasingly widespread signalling instrument on the labour market. What is the demand for such instruments however? In other words, to what extent are they shaping academic performance and student behaviour and how early can such effects come into play?

To answer these questions, we take advantage of a randomized controlled trial involving hundreds of university students who were (randomly) assigned to two treatments allowing them to obtain micro credentials by satisfying a range of requirements within their online assignment. Two types of micro credentials were made available: public and private. Specifically, some students were exposed to micro credentials in the form of "achievements", which could be obtained within the online assignment and viewed only from within the software (private treatment). Others were provided the same set of micro credentials, but some of them were in the form of "badges" that, once obtained, could be verified and then shared online (public treatment).

This study is among the first to provide direct evidence on the impact of micro credential offerings and investigate the mechanism driving the results. We find that offering public micro credentials (or badges) to students had a significant impact on their academic performance, with the overall course grade increasing by a sizeable 0.19 SDs. Estimates show that these average positive effects extend to the entire ability distribution, with both low and high

performing students benefitting. There are several heterogeneity dimensions to our effects, with females, domestic and business students experiencing an additional boost in academic performance.

From a policy perspective, our findings show that a well-designed and implemented technology enabled signalling device can produce considerable gains in course performance over a short time frame and with little interference with current course structure. The implications are thus substantial, as improving educational attainment and performance is a priority for all policy makers and practitioners. Traditionally focus has been placed on a variety of interventions including class size, hiring better teachers, financial and non-financial incentives, benchmarking, etc. While all these interventions have known and significant impacts in small trials, they tend to be very costly and are uncertain on a larger scale. Ours is an intervention that largely circumvents these shortcomings.

That being said, the study of any educational intervention is challenging and our investigation into the impact of badges is no exception. This is due to confounding factors leading to noisy results with mechanism that can only be speculated due to entanglement with other factors or the possible presence of externalities. Several experimental design choices allow, however, for potential avenues of further research. In our experiment we chose more complex achievements to become badges in the public treatment. First, converting all achievements to badges may yield some interesting results that could provide interesting insights into the balance between demand and saturation. Second, our study offered a total of 107 possible micro credentials for the treatments, covering many different types of achievement (e.g., benchmark, effort, competitive style requirements). Isolating each type of micro credentials to specific treatments could help isolate the types of badges students best respond to. Third, although positive and significant (robust) results are found based on a well-defined mechanism that can explain why the effect might occur, our intervention was only delivered to a single cohort of students. While no significant effect was found on the time spent engaging with the intervention during the term, we have not looked beyond that. And we also do not know how our estimates might change if badges become available in all courses a student undertakes. Finally, it is also important to look further into the mechanism that generates the impact of our intervention. Our results suggest that students increase their effort intensity by increasing the precision of their work even when there is no mark incentive:

Specifically, while both treatments failed quizzes fewer times than the control, only the public treatment extended this precision to the extra academia questions. With a larger sample, however, we may be able to find more specifics of the mechanism - in particular the relationship between student behaviour in the private and public treatments.

References

ANGRIST, J. D. & IMBENS, G. W. 1995. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association,* 90**,** 431-442.

ARCIDIACONO, P., BAYER, P. & HIZMO, A. 2010. Beyond Signaling and Human Capital: Education and the Revelation of Ability. *American Economic Journal: Applied Economics,* 2**,** 76-104.

AZMAT, G. & IRIBERRI, N. J. J. O. P. E. 2010. The importance of relative performance feedback information: Evidence from a natural experiment using high school students. 94**,** 435-452.

BACOLOD, M., BLUM, B. S. & STRANGE, W. C. J. J. O. E. G. 2009. Urban interactions: Soft skills versus specialization. 9**,** 227-262.

CLARK, D., GILL, D., PROWSE, V. & RUSH, M. 2017. Using goals to motivate college students: Theory and evidence from field experiments. National Bureau of Economic Research.

DOBRESCU, L. I., FARAVELLI, M., MEGALOKONOMOU, R. & MOTTA, A. 2019. Rank Incentives and Social Learning: Evidence from a Randomized Controlled Trial. Institute of Labor Economics (IZA).

MACLEOD, W. B., RIEHL, E., SAAVEDRA, J. E. & URQUIOLA, M. J. A. E. J. A. E. 2017. The big sort: College reputation and labor market outcomes. 9**,** 223-61.

VAN LENT, M. & SOUVERIJN, M. 2017. Goal setting and raising the bar: A field experiment. Tinbergen Institute Discussion Paper.

Main Tables and Figures

Table 1: Descriptive Statistics

Variable	Mean	SD	Min.	Max.	N
Panel A: Student Characteristics					
Age	19.488	1.712	17	43	645
Male	0.592	0.492	0	1	645
Undertaking Business Degree	0.707	0.456	0	1	645
International student status	0.290	0.454	0	1	645
COB: Australia	0.515	0.500	0	1	645
COB: Other Oceania	0.012	0.111	0	1	645
COB: Europe	0.017	0.130	0	1	645
COB: Asia	0.434	0.496	0	1	645
COB: Americas	0.011	0.104	0	1	645
COB: Africa	0.011	0.104	0	1	645
Previous Academic Ability: ATAR / Prev GPA	86.067	17.139	0	107.7	644
ATAR Score	95.433	6.892	60.5	107.7	448
Previous Term GPA	69.034	12.033	0	94.5	639
Panel B: Performance Indicators					
Assignment Grade	9.406	1.673	0.000	10.000	645
Week 6 Exam Grade	5.805	1.809	0.600	9.800	640
Final Exam Grade	5.756	1.251	2.245	8.776	645
Avg. (Weighted) Exam Grade	5.777	1.312	2.043	8.971	640
Overall Course Grade	6.490	1.198	1.921	9.177	645

Notes: The classification of the country of birth (COB) follows the Standard Australian Classification of Countries, 2011. The Oceania group includes Oceania countries other than Australia. ATAR (Australian Tertiary Admission Rank) score denotes a student's ranking relative to his/her peers when completing secondary education. Previous Term GPA is Term 1 2019 GPA for students enrolled at the university before the intervention semester (Term 2 2019).

Table 2: Balance Tests for Treatment and Control Groups

	Treatment Group		Control Group		Difference	
Variable	Mean	SD	Mean	SD	Diff.	SE
Panel A: Private vs Control						
Age	19.508	(1.216)	19.546	(2.312)	0.038	(0.185)
Male	0.621	(0.487)	0.559	(0.498)	-0.061	(0.048)
Undertaking Business Degree	0.708	(0.456)	0.727	(0.446)	0.02	(0.044)
International student status	0.303	(0.461)	0.268	(0.444)	-0.034	(0.044)
COB: Australia	0.492	(0.501)	0.559	(0.498)	0.067	(0.049)
COB: Other Oceania	0.005	(0.072)	0.014	(0.116)	0.009	(0.01)
COB: Europe	0.015	(0.123)	0.014	(0.116)	-0.002	(0.012)
COB: Asia	0.467	(0.5)	0.4	(0.491)	-0.067	(0.049)
COB: Americas	0.01	(0.101)	0.005	(0.067)	-0.006	(0.008)
COB: Africa	0.01	(0.101)	0.009	(0.095)	-0.001	(0.01)
Previous Academic Ability	85.666	(15.798)	86.43	(16.693)	0.765	(1.603)
Previous Term GPA	68.541	(11.585)	68.767	(11.433)	0.226	(1.139)
ATAR Score if (No GPA Data)	94.805	(6.825)	95.634	(7.017)	0.829	(0.819)
Panel B: Public vs Control						
Age	19.417	(1.367)	19.546	(2.312)	0.128	(0.178)
Male	0.6	(0.491)	0.559	(0.498)	-0.041	(0.047)
Undertaking Business Degree	0.687	(0.465)	0.727	(0.446)	0.04	(0.043)
International student status	0.3	(0.459)	0.268	(0.444)	-0.032	(0.043)
COB: Australia	0.491	(0.501)	0.559	(0.498)	0.068	(0.047)
COB: Other Oceania	0.017	(0.131)	0.014	(0.116)	-0.004	(0.012)
COB: Europe	0.022	(0.146)	0.014	(0.116)	-0.008	(0.013)
COB: Asia	0.439	(0.497)	0.4	(0.491)	-0.039	(0.047)
COB: Americas	0.017	(0.131)	0.005	(0.067)	-0.013	(0.01)
COB: Africa	0.013	(0.114)	0.009	(0.095)	-0.004	(0.01)
Previous Academic Ability	86.062	(18.65)	86.43	(16.693)	0.368	(1.673)
Previous Term GPA	69.701	(12.946)	68.767	(11.433)	-0.934	(1.156)
ATAR Score if (No GPA Data)	95.761	(6.833)	95.634	(7.017)	-0.127	(0.78)
Panel C: Public vs Private						
Age	19.508	(1.216)	19.546	(2.312)	0.038	(0.185)
Male	0.621	(0.487)	0.559	(0.498)	-0.061	(0.048)
Undertaking Business Degree	0.687	(0.465)	0.708	(0.456)	0.021	(0.045)
International student status	0.303	(0.461)	0.268	(0.444)	-0.034	(0.044)
COB: Australia	0.492	(0.501)	0.559	(0.498)	0.067	(0.049)
COB: Other Oceania	0.005	(0.072)	0.014	(0.116)	0.009	(0.01)
COB: Europe	0.015	(0.123)	0.014	(0.116)	-0.002	(0.012)
COB: Asia	0.467	(0.5)	0.4	(0.491)	-0.067	(0.049)
COB: Americas	0.01	(0.101)	0.005	(0.067)	-0.006	(0.008)
COB: Africa	0.013	(0.114)	0.009	(0.095)	-0.004	(0.01)
Previous Academic Ability	86.062	(18.65)	85.666	(15.798)	-0.396	(1.694)
Previous Term GPA	69.701	(12.946)	68.541	(11.585)	-1.16	(1.21)
ATAR Score if (No GPA Data)	95.761	(6.833)	94.805	(6.825)	-0.957	(0.801)

Notes: Each panel reports differences in pre-determined characteristics for students in the Private (Panel A) and Public Treatment group (Panel B) vs. the Control group, respectively. Panel C reports differences in pre-determined characteristics for students between the Public and Private Treatments. The last two columns report the difference in means and the corresponding standard error of the difference, respectively.

Table 3: Intent-to-treat: Treatment Effects of Micro-Credentials on Academic Performance

	Assignment	Week 6 Exam	Final Exam	Average Exam	Overall Course
	(1)	(2)	(3)	(4)	(5)
Private vs Control	0.091	-0.003	0.123	0.068	0.104
	(0.089)	(0.096)	(0.097)	(0.096)	(0.095)
Observations	415	412	415	412	415
Public vs Control	-0.002	0.087	0.280***	0.213**	0.191**
	(0.098)	(0.092)	(0.092)	(0.092)	(0.093)
Observations	450	445	450	445	450
Public vs Private	-0.099	0.092	0.159*	0.147	0.088
	(0.094)	(0.097)	(0.095)	(0.096)	(0.095)
Observations	425	423	425	423	425
Student Characteristics	✓	✓	✓	✓	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variable in column (1) is the standardized exam grade for the online assignment, while column (2) and (3) refer to the same outcome formulations for the Week 6 mid-term exam and the final exam respectively. Column (4) presents the weighted average of these series over both exams. Column (5) uses the same outcome formulation for the overall course grade. In all specifications we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether one is enrolled in a Business degree. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table 4: Treatment-on-treated: Treatment Effects of Micro-Credentials on Academic Performance

	Assignment	Week 6 Exam	Final Exam	Average Exam	Overall Course
	(1)	(2)	(3)	(4)	(5)
Private vs Control	0.098	-0.003	0.132	0.073	0.112
	(0.094)	(0.102)	(0.103)	(0.102)	(0.100)
Observations	415	412	415	412	415
Public vs Control	-0.003	0.099	0.317***	0.240**	0.216**
	(0.110)	(0.103)	(0.103)	(0.103)	(0.103)
Observations	450	445	450	445	450
Public vs Private	-0.112	0.104	0.181*	0.166	0.1
	(0.106)	(0.109)	(0.106)	(0.107)	(0.107)
Observations	425	423	425	423	425
Student Characteristics	✓	✓	✓	✓	✓

Notes: Each row presents estimates from separate IV regressions. Treatment group students (private, public and public for each row respectively) who were randomly assigned to the treatment but did not claim any of their relevant micro credentials are marked as claiming zero as are all student in the comparison group (control, control and private in each row respectively). Each regression instruments the number of micro credentials claimed with the randomized allocation to the treatment. The dependent variable in column (1) is the standardized grade for the online assignment, while column (2) and (3) refer to the same outcome formulations for the Week 6 mid-term exam and the final exam respectively. Column (4) presents the weighted average of these series over both course exams. Column (5) uses the same outcome formulation for the overall course mark. In all specifications we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether one is enrolled in a Business degree. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table 5: Heterogeneous Treatment Effects on Academic Performance: Gender

	Assignment		Week 6 Exam		Final Exam		Average Exam		Overall Course	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Private vs Control	*-0.026*	*0.271***	*-0.159*	*0.242**	0.048	0.221	-0.058	0.249*	*-0.023*	*0.294***
	(0.119)	(0.137)	(0.128)	(0.146)	(0.132)	(0.143)	(0.130)	(0.140)	(0.129)	(0.138)
Observations	244	171	241	171	244	171	241	171	244	171
Public vs Control	-0.126	0.169	*-0.072*	*0.338***	0.163	0.469***	*0.063*	*0.454****	*0.037*	*0.430****
	(0.122)	(0.165)	(0.121)	(0.145)	(0.124)	(0.143)	(0.123)	(0.145)	(0.122)	(0.148)
Observations	261	189	257	188	261	189	257	188	261	189
Public vs Private	-0.105	-0.074	0.097	0.106	0.124	0.247*	0.127	0.209	0.071	0.146
	(0.132)	(0.130)	(0.125)	(0.157)	(0.125)	(0.148)	(0.126)	(0.150)	(0.127)	(0.147)
Observations	259	166	258	165	259	166	258	165	259	166
Student Characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variable in columns (1)-(2) is the standardized grade for the online assignment, while columns (3)-(4) and (5)-(6) refer to the same outcome formulations for the Week 6 mid-term exam and the final exam respectively. Columns (7)-(8) presents the weighted average of these series over both course exams. Columns (9)-(10) uses the same outcome formulation for the overall course mark. Italicised effects represent significant differences between the subsample component estimates within a treatment. In all specifications we include controls for age, dummies for countries of birth groups, a dummy denoting whether one is enrolled in a Business degree. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table 6: Heterogeneous Treatment Effects on Academic Performance: International Status

	Assignment		Week 6 Exam		Final Exam		Average Exam		Overall Course	
	Domestic	International	Domestic	International	Domestic	International	Domestic	International	Domestic	International
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Private vs Control	0.161	-0.031	-0.005	0.008	0.09	0.222	0.057	0.108	0.095	0.15
	(0.108)	(0.150)	(0.118)	(0.165)	(0.116)	(0.171)	(0.117)	(0.163)	(0.114)	(0.168)
Observations	297	118	296	116	297	118	296	116	297	118
Public vs Control	0.05	-0.134	0.103	0.095	0.256**	0.333*	0.217**	0.222	0.185*	0.216
	(0.118)	(0.166)	(0.107)	(0.180)	(0.108)	(0.177)	(0.108)	(0.179)	(0.108)	(0.182)
Observations	322	128	319	126	322	128	319	126	322	128
Public vs Private	-0.119	-0.118	0.104	0.067	0.178	0.114	0.167	0.104	0.093	0.058
	(0.110)	(0.184)	(0.118)	(0.166)	(0.112)	(0.176)	(0.114)	(0.173)	(0.113)	(0.177)
Observations	297	128	295	128	297	128	295	128	297	128
Student Characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variable in columns (1)-(2) is the standardized grade for the online assignment, while columns (3)-(4) and (5)-(6) refer to the same outcome formulations for the Week 6 mid-term exam and the final exam respectively. Columns (7)-(8) presents the weighted average of these series over both course exams. Columns (9)-(10) uses the same outcome formulation for the overall course mark. Italicised effects represent significant differences between the subsample component estimates within a treatment. In all specifications we include controls for age, gender, a dummy denoting whether one is enrolled in a Business degree. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table 7: Heterogeneous Treatment Effects on Academic Performance: Field of Study

	Assignment		Week 6 Exam		Final Exam		Average Exam		Overall Course	
	Non-Buss.	Buss.	Non-Buss.	Buss.	Non-Buss.	Buss.	Non-Buss.	Buss.	Non-Buss.	Buss.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Private vs Control	0.049	0.119	0.019	0.005	0.103	0.157	0.031	0.106	0.113	0.125
	(0.186)	(0.100)	(0.169)	(0.117)	(0.190)	(0.114)	(0.179)	(0.115)	(0.187)	(0.111)
Observations	117	298	115	297	117	298	115	297	117	298
Public vs Control	-0.17	0.051	-0.049	0.131	0.256	0.284**	0.117	0.244**	0.092	0.219**
	(0.223)	(0.107)	(0.169)	(0.111)	(0.175)	(0.112)	(0.171)	(0.113)	(0.188)	(0.109)
Observations	132	318	129	316	132	318	129	316	132	318
Public vs Private	-0.251	-0.069	-0.121	0.194	0.13	0.179	*0.042*	*0.202**	-0.06	0.148
	(0.205)	(0.103)	(0.159)	(0.122)	(0.164)	(0.116)	(0.160)	(0.119)	(0.170)	(0.116)
Observations	129	296	128	295	129	296	128	295	129	296
Student Characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variable in columns (1)-(2) is the standardized grade for the online assignment, while columns (3)-(4) and (5)-(6) refer to the same outcome formulations for the Week 6 mid-term exam and the final exam respectively. Columns (7)-(8) presents the weighted average of these series over both course exams. Columns (9)-(10) uses the same outcome formulation for the overall course mark. Italicised effects represent significant differences between the subsample component estimates within a treatment. In all specifications we control for age, gender and dummies for countries of birth groups. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table 8: Heterogeneous Treatment Effects on Academic Performance: Quantile Regression RIFREG

	Week 6 Exam			Final Exam			Average Exam			Overall Course		
	Q25	Q50	Q75	Q25	Q50	Q75	Q25	Q50	Q75	Q25	Q50	Q75
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Private vs Control	0.05	-0.079	0.067	0.098	0.095	0.073	0.132	0.068	0.177	0.208	0.096	0.108
	(0.126)	(0.121)	(0.122)	(0.108)	(0.120)	(0.129)	(0.115)	(0.169)	(0.193)	(0.135)	(0.139)	(0.123)
Observations	412	412	412	445	445	445	415	415	415	450	450	450
Public vs Control	0.118	0.007	0.14	0.340**	0.204*	0.229***	0.189*	0.231*	0.241*	0.195	0.230*	0.291*
	(0.200)	(0.115)	(0.135)	(0.141)	(0.117)	(0.088)	(0.112)	(0.123)	(0.127)	(0.149)	(0.139)	(0.152)
Observations	412	412	412	445	445	445	415	415	415	450	450	450
Public vs Private	0.092	0.071	0.053	0.086	0.102	0.129	0.047	0.166	0.068	0.057	0.158	0.059
	(0.128)	(0.106)	(0.157)	(0.115)	(0.107)	(0.089)	(0.108)	(0.106)	(0.113)	(0.127)	(0.127)	(0.117)
Observations	412	412	412	445	445	445	415	415	415	450	450	450
Student Characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Each row presents estimates from separate unconditional quantile regressions. The dependent variable in columns (1)-(3) and (4)-(6) is the standardized exam grade for the Week 6 mid-term exam and the final exam respectively. Columns (7)-(9) presents the weighted average of these series over both course exams. Columns (10)-(12) uses the same outcome formulation for the overall course mark. In all specifications we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether one is enrolled in a Business degree. Standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table 9: Heterogeneous Treatment Effects on Academic Performance: STD Quantile Regression SQREG

	Week 6 Exam			Final Exam			Average Exam			Overall Course		
	Q25	Q50	Q75	Q25	Q50	Q75	Q25	Q50	Q75	Q25	Q50	Q75
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Private vs Control	-0.002	-0.068	0.055	0.245*	0.082	0.163	0.118	0.118	0.094	0.182*	0.149	0.094
	(0.160)	(0.142)	(0.159)	(0.147)	(0.159)	(0.130)	(0.141)	(0.179)	(0.140)	(0.109)	(0.169)	(0.162)
Observations	412	412	412	445	445	445	415	415	415	450	450	450
Public vs Control	0.204	0.021	0.157	0.454***	0.359***	0.326***	0.309*	0.309*	0.154	0.240*	0.259**	0.182**
	(0.140)	(0.118)	(0.162)	(0.144)	(0.127)	(0.103)	(0.171)	(0.182)	(0.100)	(0.143)	(0.125)	(0.087)
Observations	412	412	412	445	445	445	415	415	415	450	450	450
Public vs Private	0.063	0.083	0.055	0.163	0.163	0.21	0.221*	0.167	0.189	0.123	0.12	0.199
	(0.236)	(0.135)	(0.165)	(0.208)	(0.152)	(0.180)	(0.130)	(0.136)	(0.129)	(0.112)	(0.094)	(0.124)
Observations	412	412	412	445	445	445	415	415	415	450	450	450
Student Characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Each row presents estimates from separate conditional quantile regressions. The dependent variable in columns (1)-(3) and (4)-(6) is the standardized exam grade for the Week 6 mid-term exam and the final exam respectively. Columns (7)-(9) present the weighted average of these series over both course exams. Columns (10)-(12) uses the same outcome formulation for the overall course mark. In all specifications we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether one is enrolled in a Business degree. Standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table 10: Mechanism – Online Quiz – Intent-to-Treat

	Quiz Passes						Quiz Fails					
	Due Date			Week 6 Exam			Due Date			Week 6 Exam		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Private vs Control	0.003	-0.006	0.004	-0.001	-0.009	0.001	-0.257***	-0.257***	-0.253**	-0.183*	-0.184*	-0.178*
	(0.099)	(0.101)	(0.099)	(0.105)	(0.103)	(0.105)	(0.098)	(0.097)	(0.098)	(0.107)	(0.105)	(0.107)
Observations	402	402	401	370	370	369	402	402	401	370	370	369
Public vs Control	0.017	0.025	0.019	-0.036	-0.041	-0.032	-0.266***	-0.266***	-0.264***	-0.263**	-0.270***	-0.259**
	(0.094)	(0.094)	(0.094)	(0.102)	(0.101)	(0.103)	(0.098)	(0.096)	(0.098)	(0.102)	(0.101)	(0.103)
Observations	437	437	436	401	401	400	437	437	436	401	401	400
Public vs Private	0.031	0.032	0.03	-0.034	-0.032	-0.037	0.005	-0.009	0.004	-0.083	-0.086	-0.084
	(0.098)	(0.100)	(0.098)	(0.105)	(0.104)	(0.105)	(0.098)	(0.098)	(0.098)	(0.103)	(0.100)	(0.103)
Observations	415	415	415	371	371	371	415	415	415	371	371	371
Student Characteristics	✓	×	✓	✓	×	✓	✓	×	✓	✓	×	✓
Ability	×	×	✓	×	×	✓	×	×	✓	×	×	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variable in column (1) is the standardized number of quizzes passed for the online assignment up to the due date, while column (2) refers to the same outcome formulations except up to the week 6 exam only. Column (3) is the standardized number of quizzes failed for the online assignment up to the due date, while column (4) refers to the same outcome formulations except up to the week 6 exam only. In all specifications we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether one is enrolled in a Business degree. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table 11: Mechanism – Academia Questions– Intent-to-Treat

	Question Passes						Question Fails					
	Start of Exam Period			Week 6 Exam			Start of Exam Period			Week 6 Exam		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Private vs Control	0.075	0.095	0.075	0.065	0.083	0.065	0.096	0.115	0.094	0.042	0.056	0.041
	(0.087)	(0.087)	(0.087)	(0.078)	(0.077)	(0.078)	(0.091)	(0.090)	(0.091)	(0.079)	(0.078)	(0.079)
Observations	404	404	403	404	404	403	403	403	402	403	403	402
Public vs Control	-0.054	-0.045	-0.054	0.005	0.01	0.006	-0.12	-0.111	-0.12	-0.076	-0.07	-0.076
	(0.080)	(0.081)	(0.081)	(0.075)	(0.074)	(0.076)	(0.078)	(0.078)	(0.078)	(0.075)	(0.074)	(0.075)
Observations	442	442	441	441	441	440	441	441	440	440	440	439
Public vs Private	-0.128	-0.141	-0.128	-0.065	-0.073	-0.065	-0.214***	-0.226***	-0.213***	-0.122*	-0.126*	-0.121*
	(0.085)	(0.085)	(0.085)	(0.077)	(0.077)	(0.077)	(0.081)	(0.082)	(0.081)	(0.072)	(0.071)	(0.072)
Observations	418	418	418	417	417	417	418	418	418	417	417	417
Student Characteristics	✓	×	✓	✓	×	✓	✓	×	✓	✓	×	✓
Ability	×	×	✓	×	×	✓	×	×	✓	×	×	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variable in column (1) is the standardized number of academia questions passed for the online assignment up to the beginning of the exam period, while column (2) refers to the same outcome formulations except up to the week 6 exam only. Column (3) is the standardized number of academia questions failed for the online assignment up to the beginning of the exam period, while column (4) refers to the same outcome formulations except up to the week 6 exam only. In all specifications we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether one is enrolled in a Business degree. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Appendix A: Additional Figures and Tables

Table 12: Intent-to-treat Robustness Check: Treatment Effects of Micro-Credentials on Academic Performance

	Assignment		Week 6 Exam		Final Exam		Average Exam		Overall Course	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Private vs Control	0.108	0.093	0.008	-0.013	0.138	0.117	0.084	0.059	0.121	0.098
	(0.092)	(0.089)	(0.098)	(0.093)	(0.099)	(0.094)	(0.099)	(0.092)	(0.097)	(0.091)
Observations	415	414	412	411	415	414	412	411	415	414
Public vs Control	0.011	-0.001	0.091	0.077	0.284***	0.277***	0.219**	0.204**	0.198**	0.188**
	(0.099)	(0.098)	(0.095)	(0.089)	(0.094)	(0.089)	(0.095)	(0.089)	(0.095)	(0.089)
Observations	450	449	445	444	450	449	445	444	450	449
Public vs Private	-0.097	-0.099	0.082	0.087	0.146	0.159*	0.135	0.141	0.077	0.088
	(0.095)	(0.092)	(0.099)	(0.091)	(0.096)	(0.089)	(0.097)	(0.089)	(0.097)	(0.088)
Observations	425	425	423	423	425	425	423	423	425	425
Student Characteristics	×	✓	×	✓	×	✓	×	✓	×	✓
Ability	×	✓	×	✓	×	✓	×	✓	×	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variable in columns (1)-(2) is the standardized grade for the online assignment, while columns (3)-(4) and (5)-(6) refer to the same outcome formulations for the Week 6 mid-term exam and the final exam respectively. Columns (7)-(8) present the weighted average of these series over both course exams. Columns (9)-(10) uses the same outcome formulation for the overall course mark. In specifications (1), (3), (5), (7), (9) no controls are used while specifications (2), (4), (6), (8), (10) include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether one is enrolled in a Business degree and prior academic ability. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table 13: Treatment-on-treated Robustness Check: Treatment Effects of Micro-Credentials on Academic Performance

	Assignment		Week 6 Exam		Final Exam		Average Exam		Overall Course	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Private vs Control	0.115	0.1	0.009	-0.014	0.148	0.126	0.09	0.063	0.129	0.105
	(0.097)	(0.094)	(0.105)	(0.098)	(0.106)	(0.099)	(0.105)	(0.097)	(0.104)	(0.096)
Observations	415	414	412	411	415	414	412	411	415	414
Public vs Control	0.012	-0.001	0.102	0.087	0.322***	0.314***	0.247**	0.230**	0.224**	0.213**
	(0.112)	(0.109)	(0.106)	(0.099)	(0.106)	(0.100)	(0.106)	(0.099)	(0.107)	(0.099)
Observations	450	449	445	444	450	449	445	444	450	449
Public vs Private	-0.11	-0.112	0.093	0.098	0.166	0.180*	0.152	0.16	0.087	0.1
	(0.109)	(0.104)	(0.111)	(0.102)	(0.108)	(0.100)	(0.109)	(0.099)	(0.109)	(0.098)
Observations	425	425	423	423	425	425	423	423	425	425
Student Characteristics	×	✓	×	✓	×	✓	×	✓	×	✓
Ability	×	✓	×	✓	×	✓	×	✓	×	✓

Notes: Each row presents estimates from separate IV regressions. Treatment group students (private, public and public for each row respectively) who were randomly assigned to the treatment but did not claim any of their relevant micro credentials are marked as claiming zero as are all student in the comparison group (control, control and private in each row respectively). Each regression instruments the number of micro credentials claimed with the randomized allocation to the treatment. The dependent variable in columns (1)-(2) is the standardized grade for the online assignment, while columns (3)-(4) and (5)-(6) refer to the same outcome formulations for the week 6 mid-term exam and the final exam respectively. Columns (7)-(8) present the weighted average of these series over both course exams. Columns (9)-(10) uses the same outcome formulation for the overall course mark. In specifications (1), (3), (5), (7), (9) no controls are used while specifications (2), (4), (6), (8), (10) include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether one is enrolled in a Business degree and prior academic ability. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table 14: Alternate Heterogeneous Treatment Effects on Academic Performance: Gender

	Assignment				Week 6 Exam				Final Exam			
	Male		Female		Male		Female		Male		Female	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A												
Private vs Control	-0.017	-0.022	0.275*	0.268*	-0.171	-0.162	0.208	0.226	0.033	0.051	0.23	0.204
	(0.115)	(0.118)	(0.141)	(0.137)	(0.127)	(0.121)	(0.149)	(0.143)	(0.130)	(0.126)	(0.146)	(0.140)
Observations	244	243	171	171	241	240	171	171	244	243	171	171
Public vs Control	-0.118	-0.134	0.18	0.168	-0.064	-0.111	0.261*	0.339**	0.164	0.144	0.413***	0.467***
	(0.119)	(0.121)	(0.167)	(0.165)	(0.122)	(0.114)	(0.144)	(0.141)	(0.123)	(0.115)	(0.139)	(0.141)
Observations	261	260	189	189	257	256	188	188	261	260	189	189
Public vs Private	-0.101	-0.122	-0.096	-0.067	0.107	0.054	0.052	0.127	0.131	0.097	0.184	0.261*
	(0.131)	(0.130)	(0.134)	(0.127)	(0.126)	(0.113)	(0.157)	(0.152)	(0.125)	(0.114)	(0.145)	(0.144)
Observations	259	259	166	166	258	258	165	165	259	259	166	166
Student Characteristics	×	✓	×	✓	×	✓	×	✓	×	✓	×	✓
Ability	×	✓	×	✓	×	✓	×	✓	×	✓	×	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variable in columns (1)-(4) is the standardized grade for the online assignment, while columns (5)-(8) and (9)-(12) refer to the same outcome formulations for the Week 6 mid-term exam and the final exam respectively. In specifications (1), (3), (5), (7), (9), (11) no controls are used, while in specifications (2), (4), (6), (8), (10), (12) we include controls for age, dummies for countries of birth groups, a dummy denoting whether one is enrolled in a Business degree and prior academic ability. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

	Average Exam				Overall Course			
	Male		Female		Male		Female	
	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Panel B								
Private vs Control	-0.073	-0.059	0.240*	0.231*	-0.034	-0.021	0.288**	0.277**
	(0.129)	(0.123)	(0.144)	(0.136)	(0.127)	(0.123)	(0.143)	(0.134)
Observations	241	240	171	171	244	243	171	171
Public vs Control	0.064	0.02	0.383***	0.455***	0.045	0.015	0.373***	0.428***

	(0.123)	(0.114)	(0.141)	(0.142)	(0.123)	(0.111)	(0.143)	(0.146)
Observations	257	256	188	188	261	260	189	189
Public vs Private	0.137	0.081	0.143	0.229	0.078	0.04	0.086	0.162
	(0.127)	(0.112)	(0.149)	(0.145)	(0.128)	(0.112)	(0.145)	(0.142)
Observations	258	258	165	165	259	259	166	166
Student Characteristics	×	✓	×	✓	×	✓	×	✓
Ability	×	✓	×	✓	×	✓	×	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variable in columns (13)-(16) and (17)-(20) is the weighted average of the series over both course exams and the overall course mark respectively. In specifications (13), (15), (17), (19) no controls are used while in specifications (14), (16), (18), (20) we include controls for age, dummies for countries of birth groups, a dummy denoting whether one is enrolled in a Business degree and prior academic ability. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table 15: Alternate Heterogeneous Treatment Effects on Academic Performance: International Status

	Assignment				Week 6 Exam				Final Exam			
	Domestic		International		Domestic		International		Domestic		International	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A												
Private vs Control	0.162	0.167	-0.041	-0.061	0.015	0.009	0.027	-0.054	0.113	0.105	0.232	0.145
	(0.112)	(0.109)	(0.150)	(0.157)	(0.120)	(0.113)	(0.170)	(0.159)	(0.119)	(0.114)	(0.174)	(0.157)
Observations	297	296	118	118	296	295	116	116	297	296	118	118
Public vs Control	0.056	0.055	-0.116	-0.164	0.131	0.09	0.031	0.049	0.293***	0.257**	0.298*	0.283*
	(0.122)	(0.118)	(0.163)	(0.166)	(0.108)	(0.105)	(0.186)	(0.157)	(0.110)	(0.107)	(0.175)	(0.151)
Observations	322	321	128	128	319	318	126	126	322	321	128	128
Public vs Private	-0.106	-0.122	-0.076	-0.099	0.116	0.081	0.004	0.094	0.18	0.171	0.066	0.144
	(0.112)	(0.110)	(0.183)	(0.169)	(0.120)	(0.118)	(0.168)	(0.135)	(0.114)	(0.110)	(0.174)	(0.140)
Observations	297	297	128	128	295	295	128	128	297	297	128	128
Student Characteristics	×	✓	×	✓	×	✓	×	✓	×	✓	×	✓
Ability	×	✓	×	✓	×	✓	×	✓	×	✓	×	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variable in columns (1)-(4) is the standardized grade for the online assignment, while columns (5)-(8) and (9)-(12) refer to the same outcome formulations for the Week 6 mid-term exam and the final exam respectively. In specifications (1), (3), (5), (7), (9), (11) no controls are used, while in specifications (2), (4), (6), (8), (10), (12) we include controls for age, gender, a dummy denoting whether one is enrolled in a Business degree and prior academic ability. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

	Average Exam				Overall Course			
	Domestic		International		Domestic		International	
	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Panel B								
Private vs Control	0.082	0.073	0.127	0.034	0.115	0.111	0.161	0.072
	(0.120)	(0.113)	(0.169)	(0.149)	(0.118)	(0.111)	(0.169)	(0.156)
Observations	296	295	116	116	297	296	118	118
Public vs Control	0.255**	0.207*	0.171	0.171	0.217*	0.186*	0.179	0.162

	(0.109)	(0.106)	(0.181)	(0.149)	(0.111)	(0.105)	(0.180)	(0.151)
Observations	319	318	126	126	322	321	128	128
Public vs Private	0.174	0.145	0.045	0.135	0.102	0.083	0.018	0.09
	(0.116)	(0.114)	(0.173)	(0.131)	(0.116)	(0.110)	(0.174)	(0.132)
Observations	295	295	128	128	297	297	128	128
Student Characteristics	×	✓	×	✓	×	✓	×	✓
Ability	×	✓	×	✓	×	✓	×	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variable in columns (13)-(16) and (17)-(20) is the weighted average of the series over both course exams and the overall course mark respectively. In specifications (13), (15), (17), (19) no controls are used while in specifications (14), (16), (18), (20) we include controls for age, gender, a dummy denoting whether one is enrolled in a Business degree and prior academic ability. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table 16: Alternate Heterogeneous Treatment Effects on Academic Performance: Field of Study

	Assignment				Week 6 Exam				Final Exam			
	Non-Buss.		Buss.		Non-Buss.		Buss.		Non-Buss.		Buss.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A												
Private vs Control	0.081	0.057	0.121	0.122	0.086	-0.078	-0.009	0.013	0.126	0.02	0.151	0.167
	(0.196)	(0.192)	(0.102)	(0.100)	(0.184)	(0.159)	(0.114)	(0.113)	(0.193)	(0.183)	(0.114)	(0.111)
Observations	117	116	298	298	115	114	297	297	117	116	298	298
Public vs Control	-0.1	-0.165	0.065	0.052	-0.055	-0.093	0.183*	0.132	0.228	0.241	0.327***	0.285***
	(0.222)	(0.222)	(0.105)	(0.106)	(0.172)	(0.163)	(0.109)	(0.107)	(0.173)	(0.168)	(0.111)	(0.108)
Observations	132	131	318	318	129	128	316	316	132	131	318	318
Public vs Private	-0.181	-0.217	-0.055	-0.079	-0.141	-0.095	0.192	0.175	0.102	0.177	0.176	0.159
	(0.203)	(0.195)	(0.104)	(0.103)	(0.165)	(0.148)	(0.118)	(0.114)	(0.173)	(0.152)	(0.113)	(0.110)
Observations	129	129	296	296	128	128	295	295	129	129	296	296
Student Characteristics	×	✓	×	✓	×	✓	×	✓	×	✓	×	✓
Ability	×	✓	×	✓	×	✓	×	✓	×	✓	×	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variable in columns (1)-(4) is the standardized grade for the online assignment, while columns (5)-(8) and (9)-(12) refer to the same outcome formulations for the Week 6 mid-term exam and the final exam respectively. In specifications (1), (3), (5), (7), (9), (11) no controls are used, while in specifications (2), (4), (6), (8), (10), (12) we include controls for age, gender, dummies for countries of birth groups and prior academic ability. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

	Average Exam				Overall Course			
	Non-Buss.		Buss.		Non-Buss.		Buss.	
	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
Panel B								
Private vs Control	0.083	-0.06	0.097	0.114	0.155	0.028	0.117	0.136
	(0.190)	(0.169)	(0.114)	(0.111)	(0.196)	(0.181)	(0.110)	(0.107)
Observations	115	114	297	297	117	116	298	298
Public vs Control	0.095	0.074	0.296***	0.245**	0.094	0.076	0.267**	0.219**

	(0.174)	(0.166)	(0.111)	(0.108)	(0.186)	(0.177)	(0.107)	(0.105)
Observations	129	128	316	316	132	131	318	318
Public vs Private	0.012	0.071	0.199*	0.182	-0.06	-0.003	0.15	0.125
	(0.170)	(0.147)	(0.115)	(0.111)	(0.176)	(0.150)	(0.112)	(0.108)
Observations	128	128	295	295	129	129	296	296
Student Characteristics	×	✓	×	✓	×	✓	×	✓
Ability	×	✓	×	✓	×	✓	×	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variable in columns (13)-(16) and (17)-(20) is the weighted average of the series over both course exams and the overall course mark respectively. In specifications (13), (15), (17), (19) no controls are used while in specifications (14), (16), (18), (20) we include controls for age, gender, dummies for countries of birth groups and prior academic ability. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table 17: Alternate Heterogeneous Treatment Effects on Academic Performance: Quantile Regression RIFREG

	Week 6 Exam						Final Exam					
	Q25	Q25	Q50	Q50	Q75	Q75	Q25	Q25	Q50	Q50	Q75	Q75
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A												
Private vs Control	0.066	0.037	-0.055	-0.09	0.063	0.061	0.118	0.099	0.1	0.083	0.088	0.063
	(0.126)	(0.128)	(0.122)	(0.119)	(0.123)	(0.128)	(0.107)	(0.107)	(0.122)	(0.117)	(0.132)	(0.126)
Observations	412	412	412	412	412	412	445	445	445	445	445	445
Public vs Control	0.148	0.096	0.015	-0.003	0.136	0.13	0.349**	0.339**	0.203*	0.195*	0.223**	0.225**
	(0.201)	(0.206)	(0.116)	(0.114)	(0.138)	(0.133)	(0.141)	(0.136)	(0.116)	(0.114)	(0.090)	(0.087)
Observations	411	411	411	411	411	411	444	444	444	444	444	444
Public vs Private	0.074	0.086	0.051	0.067	0.06	0.045	0.07	0.086	0.09	0.102	0.123	0.129
	(0.128)	(0.123)	(0.107)	(0.104)	(0.158)	(0.150)	(0.114)	(0.111)	(0.107)	(0.103)	(0.091)	(0.087)
Observations	411	411	411	411	411	411	444	444	444	444	444	444
Student Characteristics	×	✓	×	✓	×	✓	×	✓	×	✓	×	✓
Ability	×	✓	×	✓	×	✓	×	✓	×	✓	×	✓

Notes: Each row presents estimates from separate unconditional quantile regressions. The dependent variable in columns (1)-(6) and (7)-(12) is the standardized exam for the Week 6 mid-term exam and the final exam respectively. In specifications (1), (3), (5), (7), (9), (11) no controls are used, while in specifications (2), (4), (6), (8), (10), (12) we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether one is enrolled in a Business degree and prior academic ability. Standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

	Average Exam						Overall Course					
	Q25	Q25	Q50	Q50	Q75	Q75	Q25	Q25	Q50	Q50	Q75	Q75
	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
Panel B												
Private vs Control	0.152	0.143	0.085	0.051	0.19	0.159	0.231*	0.205	0.119	0.069	0.117	0.099
	(0.116)	(0.113)	(0.172)	(0.166)	(0.197)	(0.187)	(0.136)	(0.131)	(0.141)	(0.136)	(0.125)	(0.120)
Observations	415	415	415	415	415	415	450	450	450	450	450	450
Public vs Control	0.192*	0.183*	0.239*	0.193	0.231*	0.232*	0.202	0.195	0.247*	0.222	0.281*	0.285*
	(0.111)	(0.108)	(0.123)	(0.120)	(0.131)	(0.124)	(0.149)	(0.144)	(0.138)	(0.137)	(0.156)	(0.151)

Observations	414	414	414	414	414	414	449	449	449	449	449	449
Public vs Private	0.039	0.042	0.15	0.161	0.06	0.063	0.039	0.056	0.143	0.157	0.056	0.059
	(0.108)	(0.103)	(0.107)	(0.102)	(0.115)	(0.109)	(0.127)	(0.121)	(0.128)	(0.123)	(0.119)	(0.114)
Observations	414	414	414	414	414	414	449	449	449	449	449	449
Student Characteristics	×	✓	×	✓	×	✓	×	✓	×	✓	×	✓
Ability	×	✓	×	✓	×	✓	×	✓	×	✓	×	✓

Notes: Each row presents estimates from separate unconditional quantile regressions. The dependent variable in columns (13)-(18) and (19)-(24) is the weighted average of the series over both course exams and the overall course mark respectively. In specifications (13), (15), (17), (19), (21), (23) no controls are used, while in specifications (14), (16), (18), (20), (22), (24) we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether one is enrolled in a Business degree and prior academic ability. Standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table 18: Alternate Heterogeneous Treatment Effects on Academic Performance: STD Quantile Regression SQREG

	Week 6 Exam						Final Exam					
	Q25	Q25	Q50	Q50	Q75	Q75	Q25	Q25	Q50	Q50	Q75	Q75
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A												
Private vs Control	0.055	0.068	-0.055	-0.035	0.111	0.022	0.163	0.233	0.163*	0.076	0.163	0.073
	(0.201)	(0.151)	(0.148)	(0.078)	(0.203)	(0.111)	(0.139)	(0.151)	(0.099)	(0.113)	(0.133)	(0.144)
Observations	412	412	412	412	412	412	445	445	445	445	445	445
Public vs Control	0.166	0.095	0	-0.101	0.166*	0.162	0.326***	0.394***	0.163	0.245**	0.163	0.156
	(0.229)	(0.173)	(0.082)	(0.133)	(0.099)	(0.126)	(0.090)	(0.145)	(0.114)	(0.106)	(0.111)	(0.096)
Observations	411	411	411	411	411	411	444	444	444	444	444	444
Public vs Private	0.111	0.076	0.055	0.022	0.055	0.043	0.163	0.151	0	0.095	0	0.132
	(0.174)	(0.172)	(0.138)	(0.160)	(0.168)	(0.170)	(0.151)	(0.155)	(0.117)	(0.109)	(0.145)	(0.142)
Observations	411	411	411	411	411	411	444	444	444	444	444	444
Student Characteristics	×	✓	×	✓	×	✓	×	✓	×	✓	×	✓
Ability	×	✓	×	✓	×	✓	×	✓	×	✓	×	✓

Notes: Each row presents estimates from separate conditional quantile regressions. The dependent variable in columns (1)-(6) and (7)-(12) is the standardized exam for the Week 6 mid-term exam and the final exam respectively. In specifications (1), (3), (5), (7), (9), (11) no controls are used while in specifications (2), (4), (6), (8), (10), (12) we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether one is enrolled in a Business degree and prior academic ability. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

	Average Exam						Overall Course					
	Q25	Q25	Q50	Q50	Q75	Q75	Q25	Q25	Q50	Q50	Q75	Q75
	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
Panel B												
Private vs Control	0.142	0.2	0.095	0.09	0.19	0.053	0.229	0.176	0.093	0.186	0.157	0.092
	(0.101)	(0.151)	(0.135)	(0.131)	(0.143)	(0.112)	(0.143)	(0.121)	(0.148)	(0.123)	(0.100)	(0.109)
Observations	415	415	415	415	415	415	450	450	450	450	450	450
Public vs Control	0.202*	0.169	0.239	0.122	0.237	0.133*	0.250**	0.204	0.270***	0.19	0.208*	0.108

	(0.112)	(0.182)	(0.152)	(0.167)	(0.147)	(0.076)	(0.120)	(0.151)	(0.096)	(0.171)	(0.113)	(0.123)
Observations	414	414	414	414	414	414	449	449	449	449	449	449
Public vs Private	0.059	0.09	0.144	0.044	0.047	0.139*	0.022	0.106	0.177	0.05	0.051	0.107
	(0.115)	(0.141)	(0.100)	(0.128)	(0.165)	(0.075)	(0.108)	(0.166)	(0.145)	(0.132)	(0.175)	(0.101)
Observations	414	414	414	414	414	414	449	449	449	449	449	449
Student Characteristics	×	✓	×	✓	×	✓	×	✓	×	✓	×	✓
Ability	×	✓	×	✓	×	✓	×	✓	×	✓	×	✓

Notes: Each row presents estimates from separate conditional regressions. The dependent variable in columns (13)-(18) and (19)-(24) is the weighted average of the series over both course exams and the overall course mark respectively. In specifications (13), (15), (17), (19), (21), (23) no controls are used while in specifications (14), (16), (18), (20), (22), (24) we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether one is enrolled in a Business degree and prior academic ability. Standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table 19: Mechanism – Online Quiz – Effect-on-Treated

	Quiz Passes						Quiz Fails					
	Due Date			Week 6 Exam			Due Date			Week 6 Exam		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Private vs Control	0.003	-0.007	0.005	-0.001	-0.009	0.001	-0.271***	-0.271***	-0.268***	-0.189*	-0.191*	-0.185*
	(0.104)	(0.107)	(0.104)	(0.107)	(0.107)	(0.107)	(0.103)	(0.103)	(0.103)	(0.109)	(0.109)	(0.109)
Observations	402	402	401	370	370	369	402	402	401	370	370	369
Public vs Control	0.02	0.028	0.022	-0.038	-0.045	-0.035	-0.299***	-0.300***	-0.297***	-0.284***	-0.292***	-0.279**
	(0.105)	(0.106)	(0.105)	(0.109)	(0.109)	(0.109)	(0.110)	(0.109)	(0.110)	(0.109)	(0.109)	(0.109)
Observations	437	437	436	401	401	400	437	437	436	401	401	400
Public vs Private	0.035	0.035	0.033	-0.037	-0.035	-0.04	0.006	-0.01	0.005	-0.089	-0.093	-0.091
	(0.109)	(0.112)	(0.109)	(0.112)	(0.112)	(0.112)	(0.109)	(0.110)	(0.109)	(0.110)	(0.108)	(0.110)
Observations	415	415	415	371	371	371	415	415	415	371	371	371
Student Characteristics	✓	×	✓	✓	×	✓	✓	×	✓	✓	×	✓
Ability	×	×	✓	×	×	✓	×	×	✓	×	×	✓

Notes: Each row presents estimates from separate IV regressions. Treatment group students (private, public and public for each row respectively) who were randomly assigned to the treatment but did not claim any of their relevant micro credentials are marked as claiming zero as are all student in the comparison group (control, control and private in each row respectively). Each regression instruments the number of micro credentials claimed with the randomized allocation to the treatment. The dependent variable in column (1) is the standardized number of quizzes passed for the online assignment up to the beginning of the exam period, while column (4) refers to the same outcome formulations except for up to the week 6 exam only. Column (7) is the standardized number of quizzes failed for the online assignment up to the beginning of the exam period, while column (10) refers to the same formulations except for up to the week 6 exam only. In these specifications we use controls for age, gender, dummies for country of birth groups, a dummy denoting whether one is enrolled in a Business degree. For columns (2), (5), (8), (11) we remove these specifications. For columns (3), (6), (9) and (12) we again include the specifications, along with an additional specification for prior academic ability. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table 20: Mechanism – Academia Questions– Effect-on-Treated

	Question Passes						Question Fails					
	Start of Exam Period			Week 6 Exam			Start of Exam Period			Week 6 Exam		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Private vs Control	0.079	0.101	0.079	0.068	0.087	0.069	0.102	0.121	0.099	0.044	0.059	0.043
	(0.091)	(0.092)	(0.091)	(0.081)	(0.081)	(0.081)	(0.095)	(0.095)	(0.094)	(0.082)	(0.082)	(0.082)
Observations	404	404	403	404	404	403	403	403	402	403	403	402
Public vs Control	-0.061	-0.051	-0.061	0.006	0.012	0.006	-0.136	-0.126	-0.136	-0.086	-0.079	-0.086
	(0.090)	(0.091)	(0.090)	(0.084)	(0.084)	(0.084)	(0.087)	(0.088)	(0.087)	(0.083)	(0.084)	(0.083)
Observations	442	442	441	441	441	440	441	441	440	440	440	439
Public vs Private	-0.144	-0.159*	-0.144	-0.073	-0.082	-0.074	-0.242***	-0.255***	-0.241***	-0.137*	-0.142*	-0.137*
	(0.095)	(0.096)	(0.095)	(0.086)	(0.086)	(0.086)	(0.091)	(0.093)	(0.091)	(0.080)	(0.080)	(0.080)
Observations	418	418	418	417	417	417	418	418	418	417	417	417
Student Characteristics	✓	×	✓	✓	×	✓	✓	×	✓	✓	×	✓
Ability	×	×	✓	×	×	✓	×	×	✓	×	×	✓

Notes: Each row presents estimates from separate IV regressions. Treatment group students (private, public and public for each row respectively) who were randomly assigned to the treatment but did not claim any of their relevant micro credentials are marked as claiming zero as are all student in the comparison group (control, control and private in each row respectively). Each regression instruments the number of micro credentials claimed with the randomized allocation to the treatment. The dependent variable in column (1) is the standardized number of academia questions passed for the online assignment up to the beginning of the exam period, while column (4) refers to the same outcome formulations except for up to the week 6 exam only. Column (7) is the standardized number of academia questions failed for the online assignment up to the beginning of the exam period, while column (10) refers to the same formulations except for up to the week 6 exam only. In these specifications we use controls for age, gender, dummies for country of birth groups, a dummy denoting whether one is enrolled in a Business degree. For columns (2), (5), (8), (11) we remove these specifications. For columns (3), (6), (9) and (12) we again include the specifications, along with an additional specification for prior academic ability. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

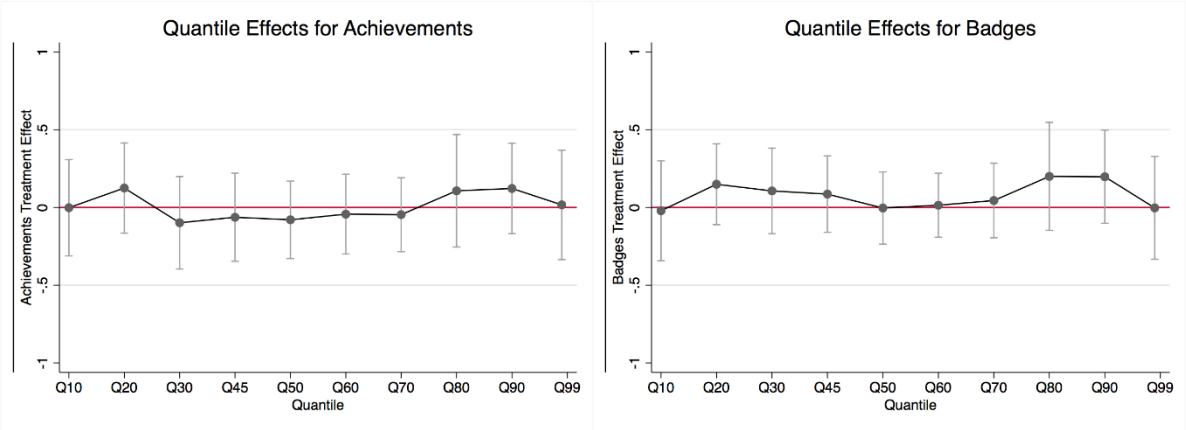
Table 21: Robustness Check: Placebo Treatment Intent-to-treat

	Assignment	Week 6 Exam	Final Exam	Average Exam	Overall Course
	(1)	(2)	(3)	(4)	(5)
Private vs Control	-0.058	0.134	0.144	0.144	0.13
	(0.107)	(0.097)	(0.094)	(0.095)	(0.096)
Observations	428	424	428	424	428
Public vs Control	0.061	0.018	-0.036	-0.021	0.011
	(0.092)	(0.096)	(0.096)	(0.096)	(0.096)
Observations	440	436	440	436	440
Public vs Private	0.116	-0.08	-0.143	-0.123	-0.085
	(0.099)	(0.092)	(0.094)	(0.093)	(0.092)
Observations	422	420	422	420	422
Student Characteristics	✓	✓	✓	✓	✓

Notes: Each row presents estimates from separate OLS regressions. Treatment allocation is placebo and determined by a random variable. The dependent variable in column (1) is the standardized grade for the online assignment, while column (2) and (3) refer to the same outcome formulations for the Week 6 mid-term exam and the final exam respectively. Column (4) presents the weighted average of these series over both exams. Column (5) uses the same outcome formulation for the overall course grade. In all specifications we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether one is enrolled in a Business degree. Robust standard errors are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

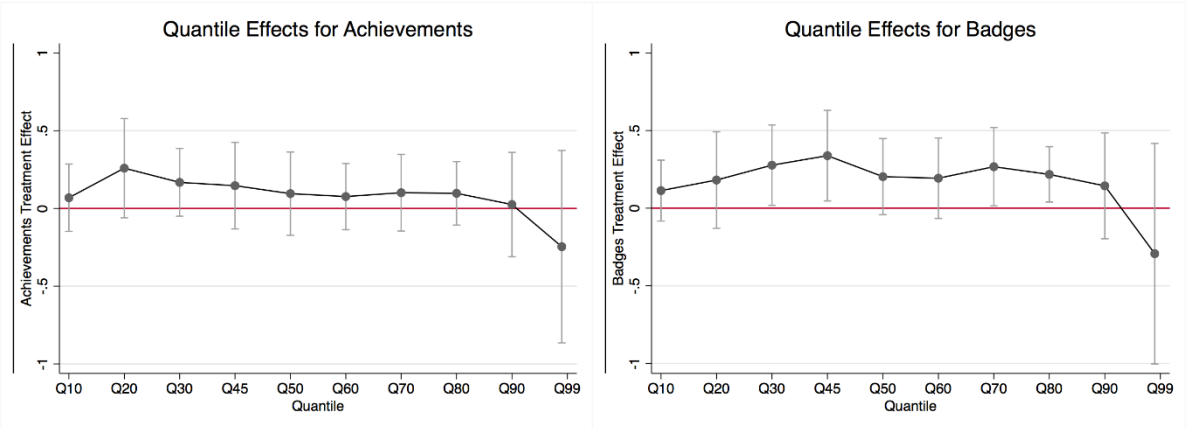
Graphs

Figure 1: Decile regression estimates of the effect of achievements and badges treatments for the week 6 exam



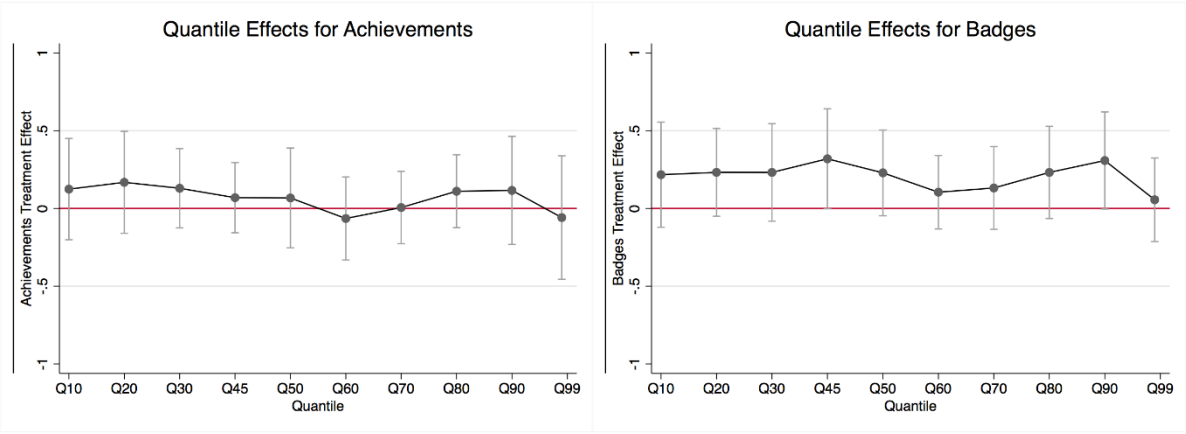
Notes: This figure presents the estimated decile effects (marginal effects) of achievements and badges on the standardized grades for the Week 6 exam grade, and the associated 95% confidence interval. The decile regressions are unconditional on students' age, gender, dummies for country of birth and a dummy denoting whether one is enrolled in a business degree. We use bootstrapped standard errors with 500 repetitions. The left plot presents the decile effects for the achievement treatment; the right plot presents the decile effect for the badges treatment.

Figure 2: Decile regression estimates of the effect of achievements and badges treatments for the final exam



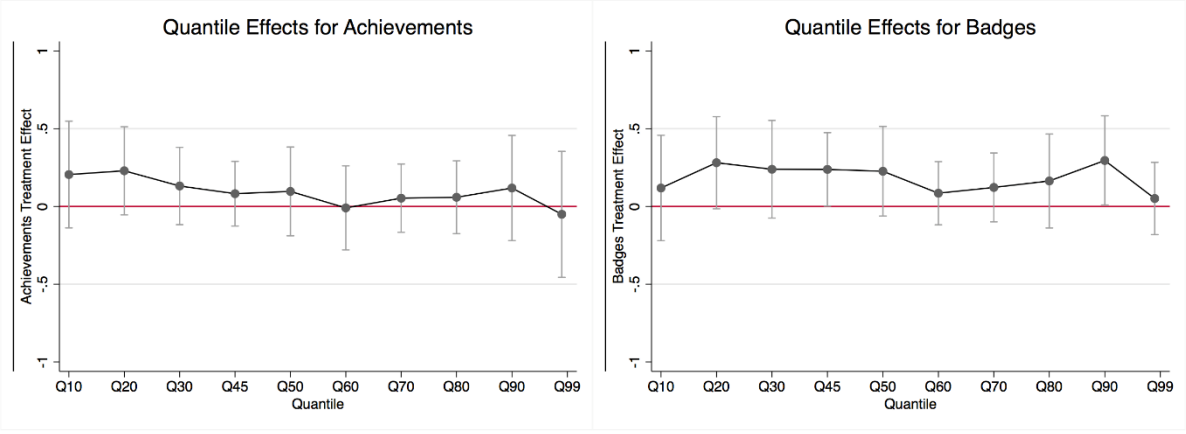
Notes: This figure presents the estimated decile effects (marginal effects) of achievements and badges on the standardized grades for the final exam grade, and the associated 95% confidence interval. The decile regressions are unconditional on students' age, gender, dummies for country of birth and a dummy denoting whether one is enrolled in a business degree. We use bootstrapped standard errors with 500 repetitions. The left plot presents the decile effects for the achievement treatment; the right plot presents the decile effect for the badges treatment.

Figure 3: Decile regression estimates of the effect of achievements and badges treatments for the average exam



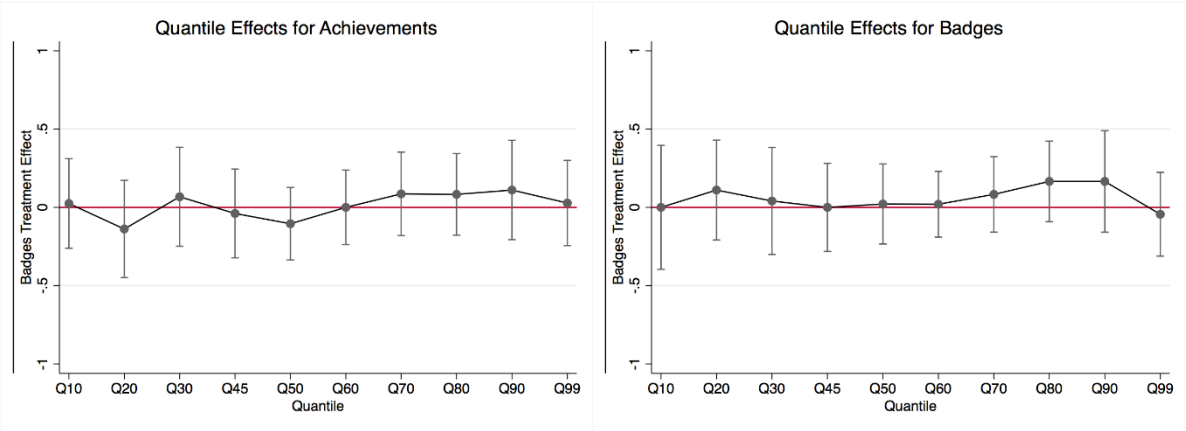
Notes: This figure presents the estimated decile effects (marginal effects) of achievements and badges on the standardized grades for the average exam grade, and the associated 95% confidence interval. The decile regressions are unconditional on students' age, gender, dummies for country of birth and a dummy denoting whether one is enrolled in a business degree. We use bootstrapped standard errors with 500 repetitions. The left plot presents the decile effects for the achievement treatment; the right plot presents the decile effect for the badges treatment.

Figure 4: Decile regression estimates of the effect of achievements and badges treatments for the overall course



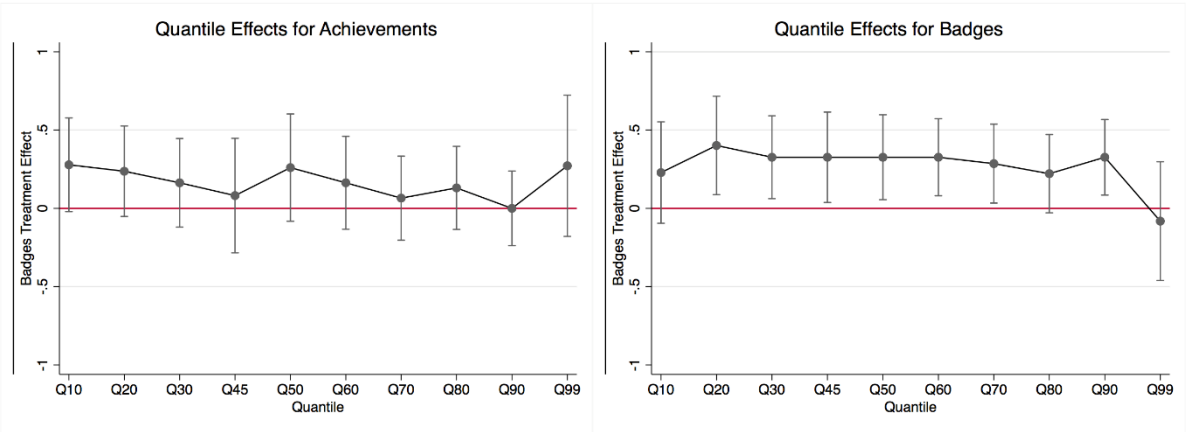
Notes: This figure presents the estimated decile effects (marginal effects) of achievements and badges on the standardized grades for the overall course grade, and the associated 95% confidence interval. The decile regressions are unconditional on students' age, gender, dummies for country of birth and a dummy denoting whether one is enrolled in a business degree. We use bootstrapped standard errors with 500 repetitions. The left plot presents the decile effects for the achievement treatment; the right plot presents the decile effect for the badges treatment.

Figure 5: Decile regressions estimates of the effect of achievement and badges treatments for the week 6 exam



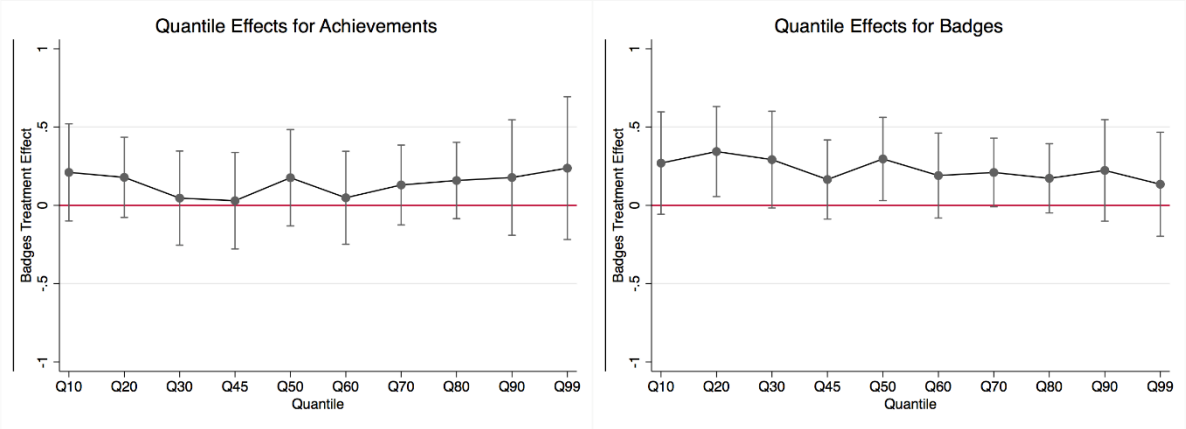
Notes: This figure presents the estimated decile effects (marginal effects) of achievements and badges on the standardized grades for the Week 6 exam grade, and the associated 95% confidence interval. The decile regressions are conditional on students' age, gender, dummies for country of birth and a dummy denoting whether one is enrolled in a business degree. We use bootstrapped standard errors with 500 repetitions. The left plot presents the decile effects for the achievement treatment; the right plot presents the decile effect for the badges treatment.

Figure 6: Decile regression estimates of the effect of achievements and badges treatments for the final exam



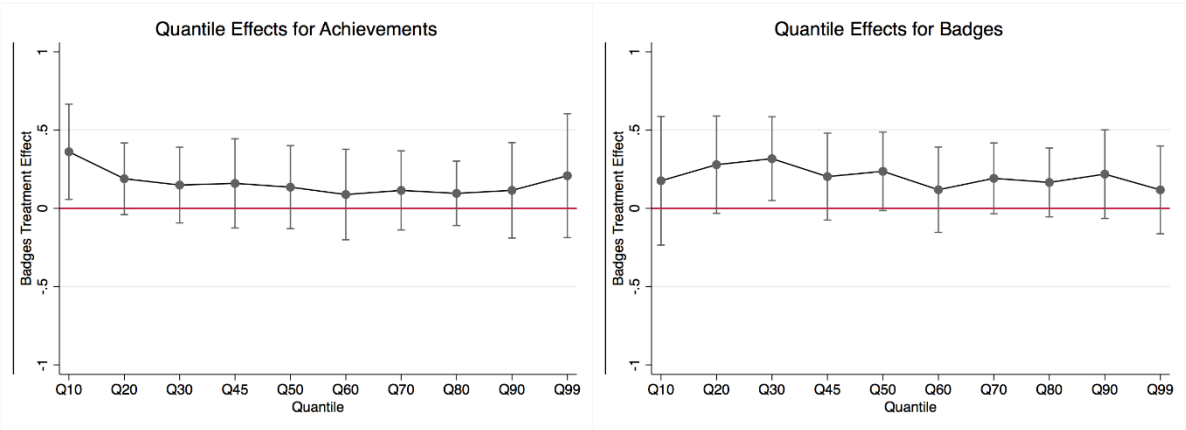
Notes: This figure presents the estimated decile effects (marginal effects) of achievements and badges on the standardized grades for the final exam grade, and the associated 95% confidence interval. The decile regressions are conditional on students' age, gender, dummies for country of birth and a dummy denoting whether one is enrolled in a business degree. We use bootstrapped standard errors with 500 repetitions. The left plot presents the decile effects for the achievement treatment; the right plot presents the decile effect for the badges treatment.

Figure 7: Decile regression estimates of the effect of achievements and badges treatments for the average exam



Notes: This figure presents the estimated decile effects (marginal effects) of achievements and badges on the standardized grades for the average exam grade, and the associated 95% confidence interval. The decile regressions are conditional on students' age, gender, dummies for country of birth and a dummy denoting whether one is enrolled in a business degree. We use bootstrapped standard errors with 500 repetitions. The left plot presents the decile effects for the achievement treatment; the right plot presents the decile effect for the badges treatment.

Figure 8: Decile regression estimates of the effects of achievements and badges treatments for the overall course



Notes: This figure presents the estimated decile effects (marginal effects) of achievements and badges on the standardized grades for the overall course grade, and the associated 95% confidence interval. The decile regressions are conditional on students' age, gender, dummies for country of birth and a dummy denoting whether one is enrolled in a business degree. We use bootstrapped standard errors with 500 repetitions. The left plot presents the decile effects for the achievement treatment; the right plot presents the decile effect for the badges treatment.