Fun in Education*

L.I. Dobrescu† R. Holden‡ A. Motta§ C.Y. Wong¶

March 14, 2018

Abstract

The Beckerian view that 'fun' can be productive has new fertile applications in education, where interactive activities such as videogame tasks (consumption-intensive) are being used alongside standard instructions (work-intensive). The implications of this transformation for equity, efficiency and academic performance have gone largely unexplored, both theoretically and empirically. This paper presents the first estimates of the causal effect of a videogame task on students' academic choices and performance and offers a simple model to flesh out potential mechanisms and policy implications. Our empirical results are consistent with a model where low-performing students benefit more from consumption-intensive education, owing to an increase in the productivity of leisure hours and a decrease in the likelihood of procrastination. The impacts are large, with students who had access to this simple videogame task scoring 9.05% higher in their final exam. They also take more courses on related subjects, being more likely to graduate with degrees in the same field.

*We are grateful to Dimitris Christelis, David Figlio, Krzysiek Karbownik, Sujata Visaria and seminar and conference participants to many institutions for helpful discussions. We also thank Kate Benett, Laura Castrique, Michelle Florance, Claudio Sissa and Chong Eng Tay for outstanding support with the administrative records. Holden acknowledges support from the Australian Research Council (ARC) Future Fellowship FT130101159.

†UNSW Business School and CEPAR, dobrescu@unsw.edu.au.

‡UNSW Business School, richard.holden@unsw.edu.au.

§UNSW Business School, motta@unsw.edu.au.

¶University of Technology Sydney, chunyee.wong@chere.uts.edu.au.

1 Introduction

In his seminal work on time allocation, Becker (1965) formalized the idea that *fun* is not necessarily unproductive. A considerable amount of sleep, food and even play indirectly contribute to our ability to work. The same applies to a business lunch, a good diet or relaxation. Based on this observation, he proposed to categorize commodities depending on their relative contribution to consumption and work.

Pure-consumption would be a limiting commodity that does not contribute to work at all. Intermediated commodities would simultaneously contribute to both work and consumption, and pure-work would occupy the opposite side of the spectrum.

In education, fun is becoming an integral part of the learning experience. Interactive activities such as videogames (traditionally a consumption-intensive commodity) are currently being used alongside standard modes of instruction (traditionally a work-intensive commodity) and are set to continue to do so even more in the future (Cowen and Tabarrok, 2014; FAS, 2006).

These advancements in educational design and methods highlight two main points of departure from Becker's model: (i) not only consumption-intensive instructions have the potential to contribute indirectly to work (by increasing the productivity of study-hours), they can also do so directly (by substituting traditional instructions altogether); (ii) given that educators can choose between alternative approaches, the work- and consumption-intensiveness of the learning experience is an important design choice.

This transformation of education raises many interesting questions that add to the larger debate on *computer-aided learning* (Banerjee et al., 2007; Muralidharan, Singh, and Ganimian, 2016), *incentives in education* (Fryer, Devi, Holden, 2016), *digital learning* (Donovan et al., 2006; Figlio et al., 2013), *learning outcomes* (Angrist et al., 2013; Fryer, 2014; Chetty et all., 2014a and 2014b; Abdulkadiroğlu et al., 2014a; Dobbie and Fryer, 2015; Figlio et al., 2015), and *equity* (Fryer and Katz, 2013; Fryer and Levitt, 2013; Dobbie and Fryer, 2014; Abdulkadiroğlu et al., 2014b; Figlio et al., 2016): What is the optimal educational design? Does consumption-intensive education impact high- and low-performing students differently? Is there an underlying trade-off between equity and efficiency when education becomes more consumption-intensive? Can consumption-intensive instructions benefit those environments, such as on-line learning, that are afflicted by last minute cramming and engagement issues?

A natural reaction to this research agenda is to question whether education can indeed be transformed into a commodity that is, at once, both work- and consumption-intensive. If the educator could just ramp up both contributions and make education fun and productive — say, by using effective videogame tasks — the policy implications would be remarkable. Surprisingly, even though videogame tasks have been around for a while and have already been labelled in various circles either as a panacea or as a terrible danger, our literature review suggests that there is virtually no clean (causal) evidence to support any policy statement either way.[1]

In this paper, we address the lack of empirical evidence by first developing a custom-made videogame task. We then run an experiment to obtain causal evidence of its impact on the academic performance of a large student sample in a research-intensive selective university. Next, we develop a simple theoretical model to clarify the underlying mechanisms and flesh out some general policy implications. Finally, we turn to the long-run effects of our intervention by examining its impact on one's subsequent academic career.

We start with the canonical labour model where a student decides how much

[1] Up to our best knowledge, only Dobrescu et al. (2015) evaluates the effectiveness of educational videogames in economics via lab experiments that contrast game-playing with traditional textbook learning. They find no evidence that playing the videogame led to lower exam scores than reading a textbook, in either multi-choice or essay questions. These findings, however, are based on a relatively small student sample that played the game or read the textbook for only 1 hour before being tested, which may present issues related to the treatment conditions being exogenously assigned rather than endogenously self-selected by students (i.e., a rational learning model would predict that students optimally allocate more time to studying when the available material is a more engaging game rather than a less enjoyable textbook). Beyond educational videogames, several studies focus on computer-based simulations, but none presents evidence from clean experiments comparing alternative pedagogies. And even in these cases, findings are generally mixed (Sitzmann, 2011; Wouters et al., 2013).

time to spend on a pure-work activity (study) and on a pure-consumption activity (leisure). We assume that student's academic performance (i.e., the student's test score) improves with the number of hours spent studying. We then consider a cross-section of students that are otherwise identical, except for their relative preferences for academic performance and leisure. This setup implies that students who care more about academic performance end up devoting more time to studying and so, they obtain higher marks.

Next, we examine the impact of a policy that increases the work-intensiveness of traditional instructions, effectively making study-hours more productive. This policy reduces the relative price of studying vis-à-vis leisure and ultimately improves the academic performance of all types of students by increasing the amount of study-hours. However, for a large class of utility functions, we show that this policy benefits high-performing students more than low-performing ones, and therefore widens the educational gap. The intuition is particularly straightforward if one specializes the utility function to Cobb-Douglas: high-performing students spend a larger fraction of their day studying and so, they will benefit relatively more from an increase in the productivity of study-hours.

We then examine a second policy that increases the work-intensiveness of leisure, i.e., the hours spent on leisure now positively affect a student's marks.[2] This policy changes the endowment: the effective number of daily hours increases because the student can now engage in an activity that contributes to both work and consumption. Contrary to the first policy, this one reduces the relative price of leisure. The ensuing increase in the number of (now productive) leisure-hours induces an increase in academic performance across the board. This time, however, low-performing students benefit more than high-performing students, because they devote a larger fraction of their day to leisure.

A first implication of the model is that an increase in the work-intensiveness of both leisure and studying is required for a policy to be equitable. Any other policy would benefit specific sub-groups differentially. Keeping this in mind, it

[2]Same results would be obtained if one were to interpret this as an increase in consumption-intensiveness of study-hours; in our model, the difference between the two policies is merely a matter of semantics.

is possible to carefully design a policy that is equivalent to a lump-sum transfer — essentially a parallel shift to the right of the budget constraint, with no price distortion and therefore no substitution effect.

The theory also offers some testable predictions. If an intervention makes leisure productive — for example, by introducing a videogame task — we must observe positive treatment effects across the board (owing to work-intensiveness of productive leisure) and larger effects for low-performing students (owing to consumption-intensiveness of productive leisure). If the videogame task succeeds in making leisure productive, then we must also observe that students play it even if the game is un-incentivized, anonymous, and covers material that is already available in traditional forms (such as the textbook and slides). By the same token, we should observe no last minute cramming in its usage.

In a subsequent section of the paper we explore a dynamic version of our model where academic performance is not realized instantly — it is instead revealed at a later date when the student sits an exam. Hence, the cost of studying (i.e., forgone leisure hours) is incurred immediately, whereas the reward (i.e., the mark) is accrued in the future. If some form of present-bias is allowed, this framework is well known (Laibson, 1997; Fischer, 2001) to give rise to time-inconsistent behaviors whereby students procrastinate studying. Our model predicts that students who are less subject to the need for instant gratification outperform students suffering from present-bias. A switch to consumption-intensive instructions makes present-biased students behave as if they were less time-inconsistent because part of the learning experience — the part that is carried out via leisure — is not procrastinated. Hence, as in the static version of the model, low-performing students should benefit more: they share the same educational gains high-performing students experience, *plus* they benefit from the reduction in procrastination. Unlike the static model though, the dynamic version offers an additional testable prediction: the later the exam date, the more the time-inconsistent students should gain from consumption-intensive instructions. This is because procrastination is more of a problem when the study period is longer.

In order to test the predictions of the model, a videogame task was especially crafted to capture the basic elements of a videogame while maintaining full consistency with the course material. Specifically, using the Comparative Advantage chapter from the recommended course textbook, we: (i) gave the model a visual representation using 3D graphics, (ii) allowed students to make economic decisions by controlling an agent in the model, (iii) designed the objectives of the game to match the exercises that appear in the main text of the chapter, (iv) presented the traditional (textbook) definitions via a media library, and (v) created a score system to capture whether students made the correct choices.

To assess its impact on academic performance, we use administrative data on the student population enrolled in a "Principles of Microeconomics" course taught at a large selective research-intensive university. This is one of the largest and most diverse courses: in a typical year, enrollments count more than 2,300 students, coming from 54 countries (and speaking 66 languages). The period under investigation covers four semesters over the academic period 2012-2013. Our treatment group consists of all students enrolled in the course in Semester 1 2013, while our control includes all those taking the course in Semester 1 2012; the only difference between these semesters is that in Semester 1 2013 the videogame module was also made available. Of course, our approach could be problematic if the 2012 and 2013 cohorts were systematically different. To rule out this possibility, we control for a number of observable student characteristics and we run a series of placebo tests using various invigilated assessments (the game covered specific examinable material) and further exploiting the fact that neither Semester 2 2012 nor Semester 2 2013 students had access to the game - see Section 3.4.) During both our treatment and control semesters an independent lecturer-in-charge (taking no part in this research) served as the course administrator. The team of lecturers and their student allocations were also virtually identical. The game was made available (without previous notice) during Week 2 of classes, it covered half of the content taught in that week and was offered in addition to the pre-existing (2012) teaching materials. Neither the students nor the lecturer-in-charge knew that the videogame task would be offered in Semester 1 2013.

Having this framework in place, we test the following theoretical predictions:

(i) *The videogame is engaging even in the absence of explicit incentives (con-sumption-intensiveness of productive leisure).* To test this hypothesis, we took a conservative approach and stacked the cards against the videogame task. We did so by designing our experiment in such a way that completing the videogame task was completely un-incentivized, voluntary and anonymous: the game was made publicly available, but playing was optional, it brought no marks, no login credentials were required and students could play it anytime at their own convenience. This design heavily disadvantaged the videogame relative to traditional online instruction, both in terms of usage and potential impact, as optional material is usually disregarded by most students and the lack of a deadline is likely to encourage last minute cramming. Despite all this, we find no evidence of last minute cramming or lack of engagement. If anything, there seems to be a *premiere effect*: the game usage spiked in the very first days after its release, well ahead of the first mid-term exam date. Usage was also substantial, with aggregate server data showing that only in the first six weeks of deployment, the videogame module was roughly completed 1,300 times for a total of over 1,900 hours of game-play.[3] This suggests that a considerable proportion of students might have played the game at least to some degree.

(ii) *The videogame generates positive treatment effects across the board (work-intensiveness of productive leisure) and larger effects for low-performing students (consumption-intensiveness of productive leisure).* We begin by comparing the average marks from Semester 1 2012 (the control group) and Semester 1 2013 (the treatment group). There are three invigilated exams that we can use to compare marks across years: the Week-5 and Week-9 mid-term exams (based on essay questions) and the final exam (based on machine-graded multiple choice questions). The questions for all exams where drawn from a preexisting database of uniformly difficult questions.

The videogame covered roughly 20% and 10% of the assessable material for the Week-5 mid-term and the final exam, respectively. After controlling for a number of observables, including lecturer and tutor/marker fixed effects, we find that students who have been exposed to the videogame attained 8.35% and 9.05% higher mean scores in their Week-5 mid-term and final exam respectively, com-

[3]Completing the videogame task (including all videos) was designed to take 1 hour on average.

pared to those who did not have access to the game. This positive result holds for both men and women, and appears stronger for students enrolled in Economics & Commerce or Science, Technology, Engineering and Mathematics (STEM). Overall, these impacts are remarkable given that the videogame covered only a fraction of the overall assessable material and that it was non-compulsory material. They also suggest that playing the game might have generated persistent dynamic complementarities (Cunha and Heckman, 2007) or perhaps that playing the game had a differential effect depending on the type of exam questions.

Our approach could be problematic however if the 2012 and 2013 cohorts, despite being remarkably similar in terms of observable characteristics, were systematically different in some other way. If that were the case, our result could be driven by the fact the 2013 students were simply better than their 2012 counterparts. To rule out this possibility, we look at the Semester 1 results from the Week-9 mid-term exam. The videogame task covered none of the assessable material for this test; in fact, the model presented in the game (Ricardo's model of Comparative Advantage) has almost nothing in common with the Demand and Supply model that forms the backbone of the Week-9 mid-term exam. Arguably then, if students were better in Semester 1 2013 relative to Semester 1 2012, they should have scored higher marks in the Week-9 exam as well. Our results do not support this hypothesis, as the difference in average marks appears to be statistically insignificant. As a further robustness check, we compare the marks of the students enrolled in Semester 2 2012 and Semester 2 2013. Both groups of students were not exposed to the videogame. Again, if there were a general cohort effect, we would observe that Semester 2 2013 students performed better than their 2012 counterparts at least in some of the assessments. This conjecture is not supported by the data, as we do not find robust and statistically significant differences in any of the three exams (i.e., Week-5, Week-9 or final exam).

An additional concern with our results is that they could be driven by the very nature of the essay questions as a form of assessment: it is hard to mark essay questions consistently and it is possible that some confounding factors remain even after controlling for the marker fixed effect. This criticism, however, would not apply to our results concerning the final exam, which was solely based on multiple-choice questions graded by a machine.

We next attempt to unpack some of the heterogeneity we see in our estimates. First, our positive game effects appear remarkably robust across various subsamples. They are particularly strong for the final exam of female students and, as mentioned, for students studying Economics & Commerce or STEM. Second, we find asymmetries in the exam scores responses to this new learning tool at different points of the students' performance distribution that are consistent with the prediction of the model: the positive effect of the game is large and robust for all score quantiles, but particularly so for lower score quantiles.

(iii) *The later the exam date the larger the treatment effect for low-performing students both in absolute terms and relative to high-performing students.* We find that the positive effect of the game features an increasing pattern for the score quantiles of the Week-5 mid-term, and a decreasing pattern for the final exam score. This confirms the hypothesis that low-performing students benefit more when the study period is longer, while the relationship is reversed for high-performing students. It is worth noticing that the final exam covers the same material as the Week-5 mid-term (i.e., the material from Week 1 to 3) plus all the remaining content up to Week 12.

(iv) *The videogame benefits those environments, such as on-line learning, that are afflicted by last minute cramming and engagement issues.* Our results confirm this conjecture. A discussion of our contribution to the literature on on-line learning is postponed to Section 4.

(v) *The videogame benefits face-to-face learning environment.* To assess whether videogames could also enhance live lectures, in Semester 1 2013 we randomly assigned an additional treatment in which the videogame was used during one of the Week 2 face-to-face lectures as a full substitute for the usual power point slides. This condition yielded no significant estimates suggesting that the videogame task in its current state was unable to enhance live lectures.

(vi) *The videogame increases students' attitudes toward Economic subjects and degrees.* We finally turn to the log-run impacts of our intervention, by examining the academic performance during the following semesters, until graduation. We are interested in whether the videogame task affected subsequent course choices and scores (overall and in related courses), degree field or graduation timing for those with access to it. Given the task's limited scope, we expect no long-lasting effects. This is generally true with two noteworthy exceptions: we find that treated students end up taking one more Economics course than their control counterparts, with most of the effect coming from them having roughly 10% higher chances to complete an Economics & Commerce degree.

The remainder of the paper is organized as follows: Section 2 presents our model. In Section 3 we describe the videogame task and the data, discuss our empirical methodology and main results, and conduct several robustness checks and subsample analysis. In Section 4 we develop a dynamic version of the model with time-inconsistency and discuss the implications of our findings for on-line learning. Section 5 concludes.

2 A Simple Model

Consider a textbook labour model where a student chooses how many hours to allocate to studying (S) and leisure (L). The *academic production function* is given by $M = c_s S + c_l L$, where c_s and c_l capture the relative contribution of leisure and study to the student's mark (M). To begin with, consider the case where $c_l = 0$, which corresponds to the canonical model where leisure does not contribute to work. For the sake of simplicity, also assume that $c_s = 1$ so that each additional hour of study increases the student's mark by 1 point.

With $\overline{L}$ denoting the maximum amount of available hours, the budget constraint is

$$M = \overline{L} - L.$$

The student's utility function $f(M, L)$ is well-behaved and increasing in both marks and leisure. In Becker's terminology, our assumptions on c_s and c_l, and our choice of utility function imply that studying is a *pure-work* commodity (i.e., it does not contribute to utility directly) and leisure is a *pure-consumption* commodity (i.e., it does not contribute to work). Panel A in Figure 1 depicts the optimal marks (M_{LP}

and M_{HP}) for two students (respectively, low- and high-performing) that are otherwise identical, except for their relative preferences for marks and leisure.[4]

2.1 Policy: Increase in Work-Intensiveness of Studying

A policy that increases the work-intensiveness of studying is effectively making study-hours more productive so that $M = c_s S$, with $c_s > 1$. The budget constraint becomes

$$M = c_s(\overline{L} - L).$$

Panel B in Figure 1 presents the impact of this policy on both low- and high-performing students for Cobb-Douglas utility. High-performing students benefit more from this policy because they devote a larger fraction of their time-budget to studying. The same results would be obtained for a large class of standard utility functions.

2.2 Policy: Increase in Work-Intensiveness of Leisure

A policy that increases the work-intensiveness of leisure makes leisure productive so that $M = S + c_l L$, where $0 < c_l < 1$ captures the contribution of leisure to work.[5] If $c_l = 1$, leisure contributes to work as much as study-hours do. The budget constraint becomes

$$M = \overline{L} - (1 - c_l)L.$$

Panel C in Figure 1 presents the impact of this policy on both low- and high-performing students for Cobb-Douglas utility. Low-performing students benefit

[4] In the benchmark version of the model, we look at a cross section of students that are otherwise identical except for their preferences. It is quite straightforward to extend our model to allow for more heterogeneity — for example, in the form of varying abilities. Qualitatively, our results would remain unchanged, while the analysis would encompass a larger number of sub-groups, each one defined by different distributions of students traits.

[5] In this section we consider a policy aimed at increasing the work-intensiveness of leisure. Same results would obtain if one were to interpret this policy as increasing the consumption-intensiveness of study-hours instead. In our model, the difference between such policies is merely a matter of semantics.

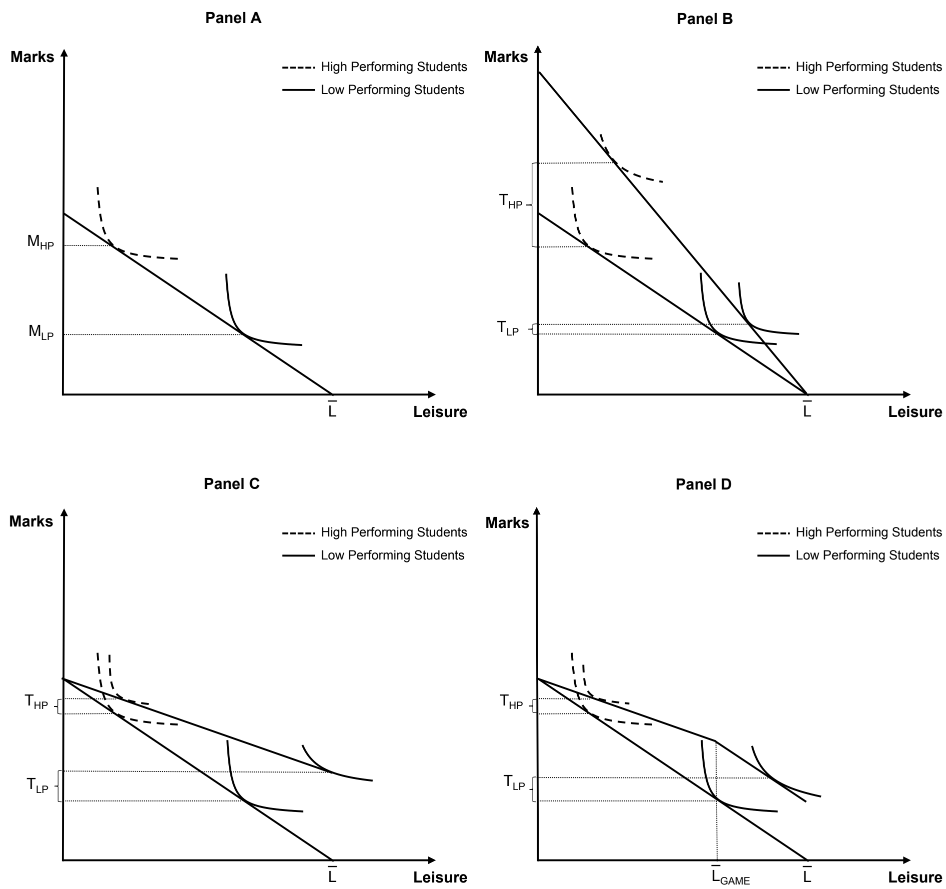

Figure 1 – Labor supply for low- & high-performing students in the absence of any intervention (*Panel A*); Treatment effects of a policy that increases the work-intensiveness of study (*Panel B*) and the work-intensiveness of leisure — with (*Panel D*) and without (*Panel C*) binding upper bound on productive leisure hours.

more from this policy because they devote a larger fraction of their time to leisure. As before, this results is robust to a large class of utility functions.

In the context of this policy, it is natural to imagine that there might be an upper bound to the number of hours that one can devote to "productive" leisure. For example, the re-playability of a videogame task might be limited. This would make it infeasible for a student to devote all her leisure hours to productive leisure. Denote this upper bound by $\overline{L}_{game}$. When $L > \overline{L}_{game}$ the budget constraint is amended as follows:

$$M = \overline{L} + \overline{L}_{game} - L$$

Panel D in Figure 1 depicts the optimal allocation of hours with a binding upper bound on productive leisure hours. Our results are qualitatively unchanged.

2.3 Testable Predictions

If an intervention (such as the introduction of a videogame task) successfully increases the work-intensiveness of leisure then (i) it should generate positive treatment effects for both low- and high- performing students (owing to work-intensiveness of leisure); (ii) the treatment effects should be larger for low-performing students (owing to consumption-intensiveness of leisure); (iii) students should engage with the videogame task even in the absence of explicit incentives (owing to consumption-intensiveness of leisure).

2.4 Policy Implications: Equity vs Efficiency

Our model suggests that any policy will disproportionally benefit specific subgroups of students unless it increases the work-effectiveness of *both* leisure and study. This also implies that a budget-constrained policy maker whose objective is to maximizes the aggregate academic gains, will likely to do so at the expenses of equity. Hence, the usual trade-off applies.

Panel A in Figure 2 presents an alternative policy that is carefully designed to equalize the treatment effects across student groups. A quick inspection re-

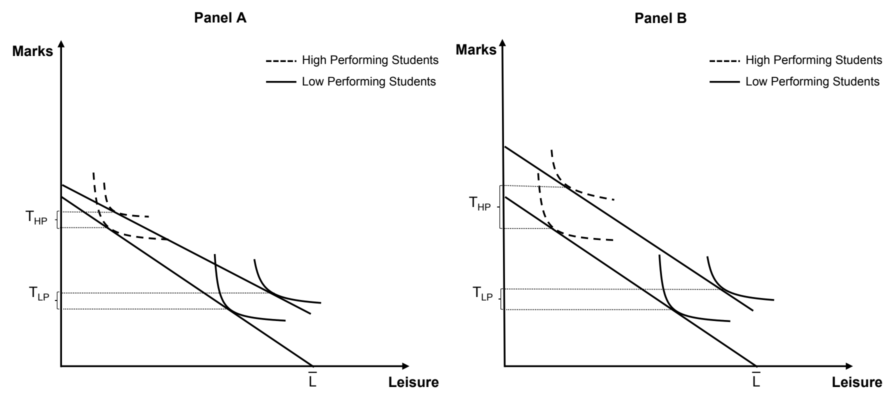

Figure 2 – A policy that increases the work-intensiveness of both study and leisure and (i) generates equal treatment effects (*Panel A*), (ii) is equivalent to a lump-sum transfer (*Panel B*) with no price distortions.

veals that this can be achieved by increasing work-effectiveness of both leisure and study. By manipulating work-effectiveness this way, one could also design a policy that essentially translates into a lump-sum transfer (Figure 2 Panel B) with no price distortions.

Just one final remark. So far we maintained that productive leisure is a perfect substitutes for regular leisure activities (i.e., the contribution to consumption is exactly the same). However, they are imperfect substitute for regular study-hours (i.e., the contribution to work is smaller, owing to $c_l < 1$). One could interpret the model differently and stipulate that c_l is a reduced-form solution to the underlying optimization problem where a student allocates hours between regular and productive leisure, and the two are no longer perfect substitutes. Then, c_l represents the fraction of leisure-hours that a student would optimally allocate to consumption-intensive instructions — and the latter contributes to work exactly as much as study-hours do.

3 'Fun' and Academic Performance

3.1 The Videogame Task

This section briefly presents the educational videogame task deployed in Semester 1 2013. A short video is also available online.[6]

Content-wise, the videogame module covers the content from Chapter 2 ("Comparative Advantage") of the prescribed course textbook (Frank, Jennings and Bernanke, 2012). Past experiences with Principles of Microeconomics students suggest that this chapter is particularly difficult to understand, as reflected by systematically low past exam scores associated with this topic. The reason is that grasping the corresponding concepts requires familiarity with a rich array of tools, ranging from diagrams to basic math. Such concepts are, however, essential to establish important economic notions such as comparative advantage, absolute advantage and opportunity cost. To ensure that the content is adequately represented, the videogame task (i) presents the same concepts as the textbook and in exactly the same order, and (ii) is based on numerical examples that are similar to the ones offered in the textbook.

The game contains six levels, plus a tutorial. Students are required to master these levels in a certain order via a system of compulsory 'primary game objectives' (i.e., students must complete the primary objectives specified in one level in order to access the next levels). However, upon completing a level, students can replay it at will, in any order and at any time. All the instructions necessary to play the game and use it as a learning tool are offered in a tutorial.

Each level is presented as a videogame-like map, which features a certain number of places, goods and agents. By exploring the map, students can get a sense of what an economic model represents and what are the basic underlying assumptions. For instance, they can observe the agents as they leave their homes in the morning and get to work. Through labor, they produce certain quantities of goods depending on their productivities. Whether these are the optimal quantities for the economy depends, for instance, on whether they specialize correctly. At night

[6] https://www.dropbox.com/s/6asy0ohmwx1q0ec/Field%20Experiment%20Video.mp4?dl=0

(i.e., when the time constraint binds) the agents go back home and sleep, the transition being punctuated by a simulated day-cycle.

In playing the videogame, the student first selects one production location (by clicking on it). A standard (textbook) graph with one good on the x-axis and the other on the y-axis is displayed in the top-right corner of the screen. This graph is updated in real time, allowing students to see the combination of the two goods produced until that point in time and observe the changes as they occur. By playing this level students realize that certain combinations of goods are attainable, while others are not (i.e., they discover the production possibility curve, hereafter PPC).

Subsequently, another character is introduced. This is a 'non-playing character' (or NPC) controlled by the computer with whom the student can trade. The trade offer must be low enough to be acceptable to the NPC but high enough to be profitable. By playing this part, students learn the basics of bargaining. For a profitable deal to take place, however, the student must also specialize according to his/her 'comparative advantage': she should specialize in the productive activity she is relatively more efficient at. Hence, by offering the right terms of trade and specializing in the right activity, everyone can consume a combination of goods to the right and above the PPC (i.e., a combination of goods that was not attainable in the absence of trade and specialization).

Aside from game-play, the videogame task features a media library accessible at any time by the player. This media library contains the definitions of the basic economic concepts and includes some videos. The library definitions are the game-equivalent of the concepts highlighted at the margin of the textbook. The videos consist of voice recordings and screen annotations. Most of the concepts are conveyed via game play, with some crucial concepts being reinforced by the videos.

Finally, we note that completing the videogame task (i.e., watching all the videos and solving the full set of levels) was designed to take, on average, 1 hour (Dobrescu et al., 2015).

3.2 Data

Course Structure. Our sample consists of all the students enrolled in a Principles of Microeconomics course taught at a large research-intensive university during the academic years 2012 and 2013. As we mentioned, the course itself is one of the largest in the university, with over 2,300 students enrolled annually. It is a typical course with (i) 2 hours of live lectures and 1 hour of tutorials per week, (ii) 12 teaching weeks in a semester, and (iii) all course material (lecture notes and power point slides, tutorial material, etc.) also provided online. In any academic year, the course is offered in both Semester 1 and Semester 2, with first semester enrollments being substantially higher than in the second one (see Table 1). Lectures are usually delivered by academic staff, while tutorials are taught by teaching assistants known as 'tutors' in the Australian tertiary education system.

Because of the large number of students enrolled in a semester (i.e., roughly between 600 and 1,800), the student pool is divided, for teaching purposes, into several groups called streams. Depending on room-size constraints, a stream can count anywhere between 200 and 500 students. Each lecturer is assigned to one or more streams, while the lecturer-in-charge and the tutor-in-charge are the overall administrators of the course. In 2012 and 2013 the lecturer-in-charge and tutor-in-charge did not change. The team of lecturers was also by and large the same (see Table 1): for instance, all four lecturers who taught in Semester 1 2012 also taught in Semester 1 2013. Notably, each of them had roughly the same number of students as the previous year.[7] The team of tutors changed more substantially, as one would expect. Overall, there were 42 tutors employed between Semester 1 2012 and Semester 2 2013. Together, they taught 82 (27) and 80 (28) tutorials in Semester 1 (2) 2012 and Semester 1 (2) 2013, respectively; each tutor held 2 to 6 tutorials per semester. All tutors and lecturers used the same teaching material, including the slides and the tutorial questions (with solutions provided by the tutor-in-charge) and comprehensive consistency checks were established to ensure that every student would benefit from a similar learning experience.

[7] The only difference in Semester 1 2013 is Lecturer A joining the course and taking a fraction of students from Lecturer E.

The assessment structure of the course contains i) online multiple-choice quizzes and hand-in assignments, typically completed over one week, ii) two invigilated mid-terms (in Week 5 and Week 9 of each semester), and iii) one invigilated final exam, containing multiple-choice questions.[8] In both 2012 and 2013, the exam papers for the mid-term and final examinations were created by the lecturer-in-charge, who drew the corresponding questions from the same pre-existing database of uniformly difficult assessments. The tutors marked the essay questions, with strict marking consistency checks in place. A machine automatically graded the multiple-choice questions of the final exams.

In terms of the timeline, we focus on four semesters across two academic years, 2012 and 2013. We note that the homogeneity of the course in this period is quite remarkable. First, the material taught in 2012 (both semesters) is identical to the one taught in 2013, both content-wise and as far as the sequence of topics is concerned. Second, the structure of the two mid-terms and of the final exam also remained unchanged.[9] Third, the same lecturer-in-charge and the same tutor-in-charge were employed in all four semesters. Finally, the set of lecturers who taught in a semester remained consistent across years.

The difference between the four semesters we analyze was that in Semester 1 2013 the course material also included the videogame module. Neither the students nor the lecturer-in-charge knew that this task would be offered. The game was announced and made available during Week 2 of classes, it covered the first half of the content taught in that week and completely substituted the traditional material (i.e., power point lecture slides) in four of the six existing streams. In the remaining two streams the videogame was used only for the initial 20 minutes, while the rest of the time the concepts were presented using standard slides. Playing the game outside class was completely optional, required no login credentials and yielded no marks. Students could access the game online and play

[8]Semester 1 2012 final exam also contained two short answer questions. In the remainder of the paper we are going to consider only the multiple-choice questions component to facilitate the comparison with the Semester 1 2013 final exam.

[9]The mid-term (essay) questions and the final exam (multiple-choice) questions were equivalent across all semesters both in terms of allocated solving time and weight in the final course grade.

anytime during the semester. Hence, playing the videogame was fully voluntary, un-incentivized and anonymous.

Data and Descriptive Statics. The student-specific data we use in this study comes from the university administrative records for 2012-2017. Given that quizzes and assignments are completed at home with full access to course materials, we will focus our analysis on the two mid-term scores and on the final exam mark as objective measures of academic performance.[10] To ease the comparison across semesters, we normalize both mid-terms and final exam scores to lie between 0 and 10. Additionally, we have data on (i) the number of times a student repeated the course after Semester 1 2012,[11] (ii) all the course scores achieved during the eight (available) semesters that followed the one in which they entered our sample, (iii) whether and when they graduated, and iv) the field of the degree they completed, as well as their 'intended' field as reported at the start of their academic career.

Besides these cognitive learning indicators, our data covers several demographics characteristics such as age, gender, country of birth, and whether a student is international or domestic. We also know whether they are enrolled full-time or part-time, the type of and stage within their degree and, since most students take the Principles of Microeconomics course in their first enrollment semester (or year), we use the Australian Tertiary Admission Rank (ATAR) to account for previous academic performance.[12]

Our full sample size includes the entire population of students enrolled in the course in Semesters 1 and 2 in 2012 and 2013. We then exclude 2.24% of the sample, representing those with a missing course grade or with a missing or zero

[10] We note that none of the researchers involved in the creation and deployment of the videogame task participated in any way in the marking process.

[11] Retaking the course is more likely to happen in Semester 2 than in Semester 1; given the high number of students enrolled in Semester 1 who, upon failing, will try again next term, this is not surprising.

[12] ATAR denotes a student's high-school ranking relative to his/her peers when completing secondary education and it is the main criterion for entry into most undergraduate programs in Australia. It is derived from a single aggregate score as the sum of the four best subjects that the student completed at Year 12 standard added to 10% of the sum of the weakest two subjects (for a total of six subjects).

course mark (i.e., those who withdrew) and those with a missing or zero mark in both Week-9 mid-term and final exam (which is compulsory for all those enrolled at the end of the census). We thus end up with a total sample of 4,794 observations across all four semesters (see Table 1).

Table 2A.1 describes the student characteristics in the control (Semester 1 2012) and treatment (Semester 1 2013) group. Ideally, the student characteristics should be perfectly balanced across these two groups. It turns out that most of them are: there are around 1,800 students in each group (1,770 and 1,812 in 2012 and 2013, respectively); most students are roughly 19 years old, with very high ATAR scores (roughly 94 out of 100); around 56% are males and more than 90% are enrolled full-time; roughly 80% are initially undecided about their field of study and only around 21% graduate 'in time', with scores of 6.69-6.78 (out of 10) and after taking about three Economics courses on average. Some characteristics, however, appear to be statistically different (see Table 2B.1), but the corresponding discrepancies are quantitatively small: compared to Semester 1 2013, Semester 1 2012 students were slightly more likely to be international (3.6%)[13] and born in Asia (4.1%). They were also a bit less likely to come from the Americas (1%) and Australia (3.1%). Finally, there is a negligible difference in age (1 month) and ATAR (0.4%). In our analysis, we will control for student characteristics, as well as for other demographic variables.

Although our main specification involves the comparison of Semester 1 in 2012 and 2013, we are still going to use Semester 2 data as a robustness check. Table 2A.2 presents the descriptive statistics for Semester 2 students in 2012 and 2013. The characteristics are well balanced across these two groups. There are roughly 600 students in each year (588 and 624 in 2012 and 2013, respectively); most students are 19.5 years old, with average ATAR of 92 and around 60% are males. More than 80% are enrolled full-time and roughly 40% of the students are international. There are only two statistically significant differences between 2012 and 2013, with Semester 2 2012 students being slightly more likely to be males (by 5.3%) and to have been born in Africa & Middle East (by 1.6%) (see Table 2B.2).

[13] We note, however, that the overall proportion of international students is not particularly high in Semester 1 in either 2012 or 2013, with only one in four students coming from abroad.

One note regarding the number of course retakes. This number is counted with respect to Semester 1 2012 (pre-Semester 1 2012 data was not available). Hence, by construction, the students in all the other semesters appear to be more likely to have previously taken the course. This turns out not to be a major concern as (i) our baseline results remain unchanged when we exclude the repeating group from the analysis, and (ii) the Semester 2 placebo tests exclude those repeating the course anyway (see Section 3.4).[14]

We now turn to our outcome variables. To ease the comparison across the four semesters, Figures 1A-C in the Appendix show the distribution and mean of the marks corresponding to the two mid-term exams (Figure 1A for Week 5 and Figure 1B for Week 9) and to the final exam (Figure 1C). A quick glance at these figures reveals several interesting facts: First, there are no significant differences in students' marks in either the mid-terms or the final exam in Semester 2. This is unsurprising given that the Semester 2 student pool in 2012 is very similar to the one in 2013, and that both the course content and the instructing team remained the same from one year to the next. Interestingly, the same does not apply to Semester 1. Indeed, the average score in Week-5 mid-term appears roughly 7.0% higher in Semester 1 2013 than in Semester 1 2012, while the final exam score differs by 13.5%. Week-9 mid-term scores, however, remain basically unchanged, students achieving on average 6.86 points (out of 10) in 2012 and 6.54 points in 2013 (an insignificant decrease). This is not unexpected, given that the game covered the material tested in the Week-5 mid-term and the final exam, but not the material tested in the Week-9 mid-term (which covered Week 6 to 8 topics).

When considering the rest of their academic career, we note that course composition includes 7.9% more Economics courses for Semester 1 2013 students compared to their 2012 counterparts. The latter also opt less frequently for Economics & Commerce degrees, albeit no distinct differences in marks (overall or in related - Economics & Commerce - courses) or 'speed' of graduation were present.

All in all, our descriptive statistics seem to point towards a strong positive as-

[14]Because at least some of Semester 2 2013 students were exposed to the game in Semester 1 2013, the repeating group can potentially confound the results by inflating the performance of Semester 2 2013 students vis-à-vis their 2012 counterparts.

sociation between introducing the videogame task as a course resource and academic choices and performance. In the remainder of the paper we investigate the extent to which this strong association can be interpreted as a causal effect.

3.3 Empirical Methodology and Results

The main outcome of interest is whether the availability of the videogame resulted in any improvement in learning outcomes. Learning was measured via three main cognitive learning variables, namely the marks achieved in the two mid-terms and in the final examination. We prefer these cognitive indicators as measure of academic performance because, unlike the course grade, they do not incorporate additional non-cognitive (attendance) and cognitive (online quizzes or home assignments) outcomes.[15] As for our main forcing variable, we can only study the effect of the videogame becoming an available resource (rather than the effect of individual videogame usage) as we did not match individual IP addresses to specific student ids by using a log-in system. This was a deliberate design choice: we wanted to stack the cards against the videogame by making it completely unincentivized, voluntary and anonymous. That being said, the videogame server did routinely collect data on student use and completion of the videogame task. The aggregate data collected shows in fact extensive usage (see below), which in turn makes it reasonable for the game to have influenced course performance. To complete the videogame, students had to go through all the teaching material and correctly solve all corresponding exercises. Hence, completing it implies a full understanding of the concepts presented therein, which are a crucial part of the examinable material in the Week-5 mid-term and the final exam.

Methodologically, given that (i) the introduction of the videogame as a course resource was completely exogenous,[16] (ii) the videogame was relevant for the ma-

[15] Attendance might be endogenous and students can collaborate with their peers on online quizzes or home assignments (even for questions that presented them with different parameter sets). Hence, we consider the overall course mark as a less-precise indicator of academic performance than the supervised (in-class) exams.

[16] The timing of the deployment (Week 2) was dictated solely by the date when the videogame task was finalized to include all the required educational features, and the corresponding topic was scheduled to be presented in class according to the course syllabus.

terials examinable in Week-5 mid-term and in the final exam, but not for the Week-9 mid-term, and (iii) except for the introduction of the game, the course topics and structure, as well as the admin and lecturer team remained unchanged, we employ a treatment effect approach and estimate the following model:

$$Score_{it} = \beta_0 + \beta_1 Y13_{it} + \boldsymbol{X}'_{\boldsymbol{it}}\boldsymbol{\beta_2} + \boldsymbol{L}'_{\boldsymbol{it}}\boldsymbol{\beta_3} + \boldsymbol{T}'_{\boldsymbol{it}}\boldsymbol{\beta_4} \tag{1}$$

where $Score_{it}$ denotes either the score of the Week-5 mid-term or the final exam score for student i in Semester 1 of Year t (where t = 2012 or 2013), and Y13 is the dummy indicator of the treatment group (i.e., students in Semester 1 of Year 2013). The set of covariates $\boldsymbol{X_i}$ includes student i's individual characteristics: age, gender, country of birth, whether a student is international or not, whether she is enrolled full-time or part-time, and the number of times a student repeated the course (since Semester 1 2012). To account for instructor quality, we include (i) a set of lecturer indicators $\boldsymbol{L}$, to capture any potential teaching-style differences that might bias our results, and (ii) a set of tutor indicators $\boldsymbol{T}$, to account for potential teaching- and marking-style differences. Additionally, some specifications also include a variable denoting in-class exposure to the videogame, as showed by the number of minutes a student was presented with the game (the entire class as opposed to only the first 20 minutes). Finally, we also control for prior academic performance (via the variable denoting high-school ATAR score).[17] In all specifications, we present robust standard errors clustered at the individual level.

Last minute cramming and lack of engagement. Despite the fact that playing the game was un-incentivized, voluntary and anonymous, our data show no evidence of last minute cramming or lack of engagement. If anything there was a premier effect: the videogame's usage spiked in the very first days after its release, well ahead of the mid-term exam date. Usage was substantial, with aggregate server data showing that only in the first six weeks of deployment, the (1 hour) 'Comparative Advantage' module was fully completed roughly 1,300 times, for a

[17] The vast majority of students in our sample are in their first university year and do not have a GPA equivalent score yet.

total of over 1,900 hours of game-play, suggesting that a considerable proportion of Semester 1 2013 students might have played the game at least to some degree (see Figure 1D in Appendix). (Unfortunately, usage data was not collected for the remaining part of the semester based on the assumption that students would only play the game in view of the Week-5 mid-term. Our results on the final exam suggest that students might have instead kept playing beyond the mid-term date.)

Performance on Week-5 mid-term and final exam. Table 3 and Table 4 present the results from estimating model (1) for the Week-5 mid-term exam and the final exam respectively, under different specifications. A quick glance at specification (1) in Table 3 shows that the average score of the first mid-term test for the treatment group is 4.74% higher than for the control group. And remarkably, the final exam score is 8.20% higher in Semester 1 2013 than in Semester 1 2012 (see Table 4). These results are robust to the inclusion of individual characteristics such as age, gender, country of birth, whether one is enrolled full time or is an international student and the number of course repeats (specification 2).

A legitimate source of concern with our empirical strategy pertains to the presence of potential lecturer and tutor/marker effects, or to the possibility that our results might differ by in-class game exposure. The estimates in our baseline specification (4) in Table 3 and Table 4 indicate that after controlling for such potentially confounding factors the treatment effect becomes even larger, with the students who could access the game achieving 8.35% (9.05%) higher marks in their first mid-term (final exam) than the students who could not use the game.[18] The in-class game exposure treatment, on the other hand, yielded no robust significant impacts, suggesting that the videogame task in its current state was not able to enhance live lectures.

Two remarks are in order here. First, note that these results are quite sizable despite the videogame covering only small proportions of the assessable material. For instance, they are of comparable magnitude (i) to being taught by contingent

[18] This considerable increase in the treatment effect for the mid-term is not entirely surprising since the mid-term is marked by tutors; hence, including the tutor fixed effect explains an important part of the variation in the mid-term marks.

(teaching) faculty rather than tenure-track/tenured staff (Figlio et al., 2015), or (ii) to being taught by instructors 5.3 to 6.7 standard deviations above the average of perceived effectiveness as evaluated by students at the end of the course (Hoffmann and Oreopoulos, 2009), or (iii) to reducing (mean) class size by 16% (Bandiera et al., 2010).

Second, also note that the game impact appears to be stronger for the final exam than for the mid-term, even though the game covered less assessable material for the former (10%) compared to the latter (20%). The structure of these two tests is however quite different,[19] which makes it difficult to directly (and meaningfully) compare the game impact between these two assessments. To address this issue, we express the estimated effects in terms of units of scores standard deviation in Table 5. We find treatment effects of 0.468 and 0.587 standard deviations for the Week-5 mid-term and the final exam scores, respectively. Hence, even when we take into account the differences in score distributions, the impact of the game is indeed stronger on the final exam performance than on the mid-term scores.

This differential effect could be due to the type of exam questions or the timing of game availability (i.e., while in preparing for the mid-term students had only two weeks to use the game, by the final exam date the game was available for several more weeks). But it might also suggest that playing the game has generated persistent dynamic complementarities (Cunha and Heckman, 2007). One way to test this hypothesis involves examining the game effect on the final exam score achieved in the subset of questions related to topics covered in the videogame vs. the score attained in game-irrelevant questions. We present results in Table 4A. While the game impact appears significantly stronger for the game-relevant questions compared to the game-irrelevant ones, the effect is present for both types of final exam questions. This suggests that (prior) skills acquired while playing the game can affect one's productivity at subsequent (game-unrelated) stages, perhaps because of the cumulative nature of knowledge.

Turning to individual characteristics, it appears that age and number of times

[19]Week-5 mid-term contains only essay questions; the final exam included only multiple-choice questions.

one repeats the course have a negative (and significant) impact on both the mid-term and final exam scores, in almost all specifications. Gender and part-time/full-time status remain systematically insignificant in all Week-5 specifications, but yield a positive coefficient in the final exam ones. Interestingly, for both sexes, the enrolment status effect appears to be driven mostly by low quantile students (see Section 3.3). The ethnic background of students also plays a role in shaping academic performance. This is not surprising given that the Principles of Microeconomics course is highly diverse ethnically (with 54 countries represented and 66 languages spoken).

Finally, it also seems that among the relatively high number of instructors teaching the course, only one lecturer has a positive and significant effect on students' mid-term outcomes, although no lecturer effect is present for final exam. The same also applies to tutors, with over half of the tutors (out of 42) having a significant effect on students' Week-5 mid-exam marks but only 10 of them being associated with a significant coefficient in the final exam specifications. This might be due to the fact that tutors not only teach students in their tutorials, but they also mark their mid-term exams.

Performance on Week-5 mid-term and final exam: Quantile results. In assessing the effectiveness of a new learning tool, the effect on the lower tail of the academic performance distribution (conditional on student characteristics) may be of more interest than the effects on the mean of the distribution. For instance, previous findings by Figlio et al. (2013) suggest that high performing students may be less affected by the lecture format than weaker students who may need more interactive learning. Keeping this in mind, we wanted to check if the game impact varied for the high- compared to the low-performing students. To do so, we conducted a quantile regression analysis and show the results from estimating the 25%, 50% and 75% percentile for Week-5 mid-term and final exam in Table 6 (the estimated coefficients are also shown in Figures 2A-C in the Appendix).

First, note that the results on the overall (Week-5 mid-term and final exam) scores show a positive treatment effect across quantiles, with the game leading to 9.36% higher scores for the low-performing students and 5.99% for the high-performing ones, a statistically significant difference. This is in line with the static model predictions related to low-performing students benefitting more from consumption-intensive instructions than their high-performing counterparts.

Interestingly, this effect has an increasing pattern with the score quantiles of Week-5 mid-term, and a decreasing pattern for the final exam score. This suggests that better students benefit more in the shorter-term, while lower achieving students gain more in the log run. These results are consistent with the theoretical prediction of the dynamic model presented in Section 3. Assuming that low-performing students are more likely to be hyperbolic discounters, one should observe a larger treatment effect overall for low-performing students, with a particular focus on the final exam mark. The later the exam date, the more the hyperbolic discounting students should gain because procrastination is more of a problem when the study period is longer. This is the pattern we observe in our treatment effects, with a gradient across quantiles — positive for the Week-5 mid-term and negative for the final exam — that is consistent with the theoretical predictions.

In both cases, however, unlike Figlio et al. (2013), we find that the game availability has substantially benefitted all students, strong and weak, with the highest positive effects concentrated in the lower part of the distribution for the final exam and in the upper part for the Week-5 mid-term. Surprisingly little empirical work has been done using quantile regressions to evaluate the effect of various learning tools on student performance. Our findings aim to fill this gap.

Long-run results. Our results so far showed that the videogame task has substantially improved course performance. Can however this positive effect extend to one's academic career beyond the course (or the semester) in which the task was used?

To tease out the potential long-run impact associated with our intervention, we study the academic choices and performance of the students in our main sample during the eight semesters that followed the relevant one (i.e., Semester 1 2012 for the control students, and Semester 1 2013 for treated students). Specifically, we start by examining (i) the number of all other Economics courses completed after the relevant semester, (ii) the mean score achieved in all courses completed after the relevant semester, (iii) the mean score achieved in all other Economics & Commerce courses completed after the relevant semester, and (iv) whether a student graduated 'in time' (i.e., within the term prescribed by their degree).[20]

To do so, we re-run our main specifications (e.g., specifications (1)-(4) from Table 4) on a sample that excludes those who repeated the course in both relevant semesters (0.65% of the sample). When examining the outcome variables (i)-(iii), we further exclude 3.57% of the sample for whom the relevant semester is their final one before graduating, while 'graduation' models exclude Cross-Institutional Undergraduate and Non-award Program enrolees who lack graduation status (1.63% of the sample).

Results are presented in Tables 14 and 15. A quick glance reveals that among variables (i)-(iv), the only relatively robust[21] treatment effects are present in the specifications involving the number of other Economics courses taken. No effects are further present for course scores (overall or in Economics & Commerce courses)[22]or graduating 'in time'. This is not surprising given the scope of the videogame task that students had access to, which may have been too limited to have a long-lasting impact on their entire academic career, but could have influenced subsequent course choices.

The findings related to taking other Economics courses warrant however further investigation. Table 14 estimates imply that having access to the videogame task led students to enroll in and complete roughly one more Economics course than the control group. This is a considerable figure in the context of our intervention, and we next attempt to investigate what could have prompted this result.

There are two potential channels that can lead to more Economics courses being taken: students can either decide to opt for (or change into) degrees that requires more Economics courses or they can simply take more such courses as electives (i.e., without changing their field). To test the first channel, we use an in-

[20]As Principles of Microeconomics is an undergraduate course, graduating from the associated program implies being awarded a bachelor degree.

[21]The only exception is specification (2) in Table 14, with a treatment coefficient p-value of 0.14.

[22]No significant effects are present for Economics courses in isolation either. Results are available from the authors upon request.

dicator denoting whether the degree completed is in Economics & Commerce. Re-running our standard specifications using this variable as outcome yields robust and significant estimates, with those exposed to our intervention having roughly 10% higher chances of completing an Economics & Commerce degree (see Table 14). For the second channel, we use a sample that excludes those enrolled in Economics & Commerce programs and re-run our analysis to find positive but statistically insignificant effects. We therefore conclude that the first channel is at play, with our intervention increasing the 'appetite' for Economics, despite not having any effect on subsequent academic performance.

3.4 Robustness checks

As a first robust test, we exclude from our sample students who repeated the course in either 2012 or 2013 and re-run our baseline specification (4). Results are presented in Table 3 and Table 4 as specification (5). A quick glance at the new estimates reveals that excluding repeating students leaves the treatment effects largely unchanged (compared to when the entire population is used).

To account for prior academic performance, we also run a specification that includes the ATAR score. The estimates in specification (6) in Table 3 and Table 4 show that accounting for previous achievements leaves the game effect unchanged for the Week-5 mid-term, but it increases its magnitude for the final exam. We note, however, that such effects are identified only from the sample of students who have attended high-school in Australia, with a large proportion of international students being excluded from the estimation. Another way of interpreting these estimates is as suggestive evidence that the positive treatment effect of the game is stronger among local than among international students in the final exam.

Next, we also wanted to take a closer look at the symmetry of our results between men and women. To this effect, we split the sample by sex and re-run specifications (4) from Table 3 and Table 4, showing the resulting estimates in Table 7. Overall, we find a considerable positive (and significant) effect of the game on both types of scores (mid-term and final exam) for both men and women. Comparing these estimates with the general sample ones, we see that the effect of the game for females is 0.082 points (out of 10) lower for the mid-term but 0.126 points higher for the final exam. This is a surprising result as the literature on games suggests that men learn better through games than women, because women (i) have higher computer anxiety, lower computer self-efficacy and less favorable computer attitudes than men (Cooper and Kugler, 2008); (ii) have less interest in digital games and less game-related knowledge, and play less frequently (Lucas and Sherry, 2004); (iii) are less tempted to defer learning and cram for the exam, while men's attraction for competitive games (Vorderer et al., 2006) might help them stay engaged with the material; and (iv) are less competitive - and performing worse in competitive environments - than men (Croson and Gneezy, 2009; Gneezy et al., 2003; Niederle and Vesterlund, 2007, 2011). In our case, such a stronger effect for females in final exam might be due to them being more conscientious and using optional material (in this case, the game) more than men, who are known to engage less with non-compulsory tasks (Woodfield et al., 2006).

Given the interesting age effects found in our baseline results, we also wanted to investigate more this potential source of heterogeneity. We note that in all Table 3 and Table 4 specifications, the age coefficient (when significant) is negative, implying that older students perform worse in the exams than younger students. But what is the role of the game? Previous studies indicate that older students have lower computer literacy levels or prefer the 'chalk and talk' learning style that they grew up with (Prensky, 2001; Gee, 2003), which might mean that a game could be less beneficial for their learning. In our sample, however, there is only a minority (roughly 1%) of 'mature' students (25+ years old) with almost 6 in 7 students being 19 or younger. As a result, we split the sample between students with ages below 19 vs. 19+ and re-run our baseline specification (4) (in Table 3 and Table 4). The estimates in Table 8 show that the game has indeed a robust and sizeable positive effect on mid-term and final exam scores for students aged 19 and below. For those older than 19, however, we find no significant impact on the mid-term score but a positive effect for the final exam. These findings suggest that younger generations might respond better to non-traditional teaching approaches like videogames, selection effects notwithstanding.

Furthermore, we also wanted to check whether the game had a different learning impact by field of study. To this effect, we grouped our student population in three categories: (i) Economics & Commerce (Econ & Comm), (ii) Science, Technology, Engineering and Mathematics (STEM), and (iii) Others (e.g., Arts, Education, Media, Social Work, etc.), and re-run our baseline specifications (see Table 9). We find that the game had a strong, positive and sizable, effect on both the Econ & Comm and STEM students' scores, but we find no (robust) effect for the remaining group. Indeed, after controlling for individual and class characteristics, the game impact stays positive but becomes statistically insignificant for the 'Others' group, which may due to the relatively small size and high heterogeneity in the fields of study in this sub-sample.

As an additional robustness check, we also employ a non-parametric method (i.e., propensity score matching — PSM henceforth) to evaluate the game average treatment effect. Despite the fact that introducing the videogame module in Semester 1 2013 course curriculum was completely random (and so, impossible to anticipate by students), it might still be possible for the students in the treatment and control groups to be systematically different. The propensity score analysis allows us to eliminate the statistical bias caused by such potential systemic difference and facilitates the comparison between treatment and control groups who are demographically similar to each other. Three matching methods (one-to-one matching, radius matching and kernel) were applied in this analysis. In doing so, we use specification (4) in Table 3 and Table 4 (excluding in-class game exposure, the lecturer and tutor dummy indicators)[23] to estimate the propensity scores of our student sample. We show our findings in Table 10 and Table 11. Table 10 presents the balancing measures for our variables of interest between the treatment and control groups, before and after the matching. The figures indicate that there is a good balance between the treatment and control groups after matching (across all the three matching methods), with the after-matching standardized percentage bias being generally quite low (ranging from 0% to 9.3%, in absolute value for all

[23]Since number of retakes, in-class game exposure time and most of the lecturer and tutor dummy variables can predict the treatment on the student sample (or being in the control group) perfectly, they are excluded from the logistic regression model used to calculate the propensity score. Using specification (5) to estimate the propensity score leaves our results unchanged.

variables).

Table 11 presents the average treatment effects of the videogame module on the treated for Week-5 mid-term and for the final exam. For each assessment, the estimates of the average treatment effects are quite robust among different matching methods. The effect of the game is around 0.44 points (out of 10) for Week-5 mid-term and around 0.81 points for the final exam. These estimates are statistically significant and, importantly, they are comparable to the estimates obtained in the baseline analysis. Overall, our PSM results indicate that, even after we adjust for potential differences in individual characteristics between the treatment and control groups, we still observe a positive and significant effect of the videogame on students' Week-5 and final exam scores.

Finally, we also extend our long-term analysis to include four additional summer semesters. This does not affect our results, which is unsurprising given that these semesters offer very few courses, to a very limited number of students. Additionally, when looking at graduation timing, we also run our baseline specifications on a narrower sample that further excludes those for whom the relevant semester is not their first one. Doing so circumvents any comparability issues related to potential changes in degree duration, and leaves our results unchanged.[24]

Placebo Tests. As discussed in Section 1, our approach could be problematic if the 2012 and 2013 student cohorts were systematically different based on some unobservable characteristics. If that were the case, our result could be driven by the fact the 2013 students were simply better than their 2012 counterparts.

To rule out this possibility we conduct a placebo test to check if the game had any effect on Week-9 mid-term scores. Since the videogame only covered only 20% of the material tested in Week-5 mid-term, this placebo analysis effectively provides counterfactual evidence on the effect of the game if no effect is detected among the 'to-be-treated' group without the treatment. A quick glance at Table 12A shows that our specifications produce no robust placebo effects: the corre-

[24]Results are available from the authors upon request. We also note that there are no significant differences between the balance results pertaining to the main sample and those for this sub-sample.

sponding placebo estimates become insignificant after considering individual and class characteristics, which reinforces our confidence in the identification strategy we employ.

As a further robustness check, in Tables 12B - 12D we also compare the Week-5 mid-term, the Week-9 mid-term and the final exam marks for (non-repeating) students in Semester 2 2012 and Semester 2 2013. We note that no students in these groups were exposed to the videogame. And as expected, we do not find any robust statistically significant differences for any of these exams, suggesting that indeed there were no general cohort effects. If there were any effects, we would have observed that Semester 2 2013 students performed better than their Semester 2 2012 counterparts in at least some of these assessments.

Difference-in-Difference. The last robustness check we conducted is the difference-in-difference (DID) analysis. Since the game only covers topics examined in the Week-5 mid-term (but not in the Week-9 mid-term), we can compare the difference in marks between the Week-5 and Week-9 mid-terms, among the treatment group (Semester 1 2013) and the control group (Semester 1 2012). The results in Table 13 specification (3) indicate that compared to the control group, the treatment group's average mark in the Week-5 mid-term exam is 0.795 points (out of 10) significantly higher than that in the Week-9 mid-term exam. Since Week-5 mid-term and Week-9 mid-term exams are different topics-wise, the result from this analysis does not provide an intuitive measure of the game impact. However, it provides evidence in support of the positive effect of the game, as we do not only compare the difference in the marks of the affected exam (Week-5 mid-term) between the treatment and control groups, but we also use the marks from the unaffected exam (Week-9 mid-term) as baseline across the treatment and control groups.

4 Extensions and Discussion

4.1 Dynamic Model with Time-Inconsistency

In our simple static model, we considered an academic task that required a one-time action. Academic assignments are usually structured differently: they require small actions to be completed over time. In the case of a mid-term exam, marks are accrued after a study period that can potentially span months.

In this section, we extend our model to consider this kind of structure. This extension will allows us to capture academic procrastination — a well documented phenomenon in education (Fischer, 2001). Procrastination can manifest itself as last minute cramming, in which case most of the work is performed close to the deadline. Or it can exhibit itself in missed deadlines and abandoned tasks.

Consider an assessment taking place T days from now. The mark is a function of the total work (including productive leisure) completed by the deadline. Hence, our production function becomes

$$M_T = \sum_{i=0}^{T-1} c_s S_t + \sum_{i=0}^{T-1} c_l L_t.$$

For the sake of simplicity, we use a log utility function (separable in leisure and marks) and we set $c_s = 1$. The time-t utility function is given by

$$U_t = ln(L_t) + \sum_{i=0}^{T-t-1} \beta\delta^i ln(L_{t+1}) + \beta\delta^{T-t}\alpha ln(M_T).$$

This model features Laibson's (1997) quasi-hyperbolic discounting. At any given point in time, the student is relatively impatient (i.e., *he* discounts tomorrow's utility by $\beta\delta$), but he expects to be less impatient the next day (i.e., he expects his future self to discount utility by δ). When tomorrow comes, the future self does not live up to expectations, and dynamic inconsistency arises.

Following Fischer (1999), we assume that students are aware of this inconsistency and solve the problem by backward induction. It is useful to think of M_t

as the total stock of available time not yet used in (unproductive) leisure that is passed on to self t. The problem reduces to solving a modified Bellman equation.[25] Using a constant relative risk aversion utility function — such as our log utility — simplifies the analysis substantially. In each period t, there is a unique optimal amount of leisure given by $\mu_t M_t$, where

$$\mu_t = \frac{1}{c_l[1 - \beta + \sum_{i=0}^{T-t-1} \beta\delta^i + (\beta\delta)^{T-t}\alpha]}.$$

Figure 3 shows the optimal number of hours of study for a given set of parameters.[26] Panel A and B presents two alternative deadlines: 84 days (final exam) and 28 days (mid-term exam).

In the absence of consumption-intensive instructions ($c_l = 0$), the model generates an increasing pattern that represents a typical example of last minute cramming (labelled "Control"). This pattern is more accentuated for the final exam, where full procrastination occurs in the first couple of months. This is due to the fact that the deadline is further ahead for the final exam compared to the mid-term exam, a situation that makes procrastination more appealing. It is worth noting that in this kind of setup, present and future selves would strictly benefit from a uniform increase in hours of study across all periods. In other words, the solution depicted in Figure 3 is not Pareto Optimal.

When consumption-intensive instructions are introduced ($c_l > 0$), a fraction of leisure hours becomes productive. This induces students to spend more time on leisure, but it also increases their *effective* hours of study. The new (effective) study pattern is also depicted in Figure 3 (labelled "Treatment"). The positive treatment effect is particularly strong when full procrastination is the alternative. In that situation, the student would have had no work completed in the absence of consumption-intensive instructions. By introducing such instructions, a fraction of leisure hours becomes productive, effectively contributing to work. When full procrastination ends, the treatment effect is somewhat smaller because the student increases the number of (now productive) leisure hours at the expense of

[25]Notes available from the authors upon request.

[26]The problem was solved numerically with $\beta = .8$, $\alpha = 10000$, $\delta = 1$, $c_l = .5$, $\overline{L} = 16$.

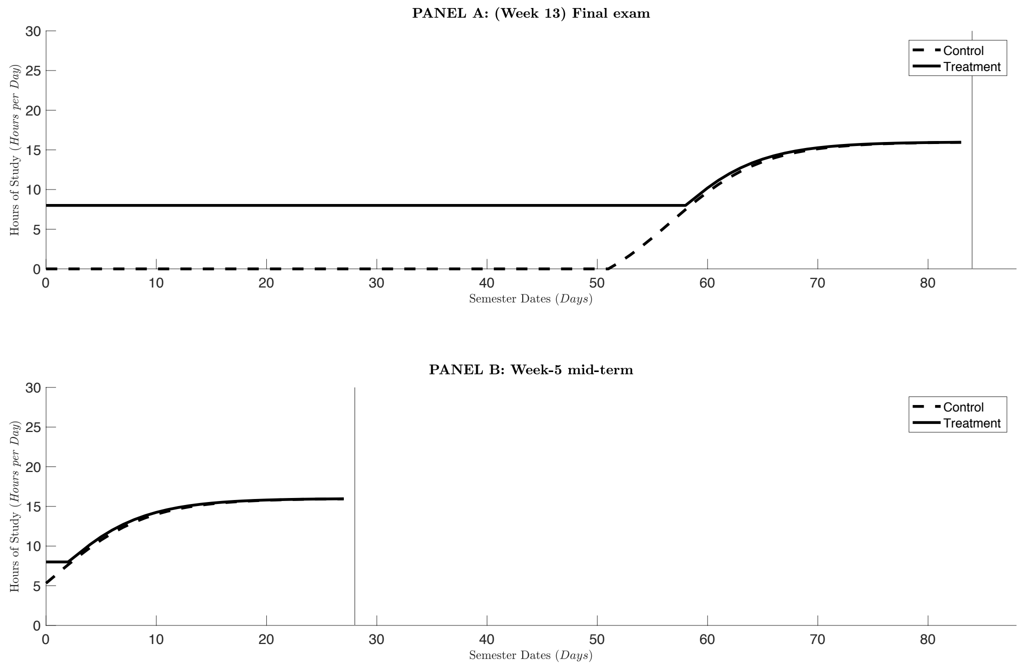

Figure 3 – Consumption-Intensive Education with Hyperbolic Discounting.

study hours, canceling out some of the treatment effect. These patterns suggest that consumption-intensive instructions have the potential to reduce procrastination and increase total hours of study — leaving the students strictly better off. Although students' preferences are unchanged, consumption-intensive instructions make students behave as if they were qualitatively less time-inconsistent.

4.1.1 Testable Predictions

In both Panel A and Panel B of Figure 3 the treatment effects can be gauged by looking at the area comprised between the two curves. A quick inspection reveals that time-inconsistent students gain more from consumption-intensive instructions when the deadline is further away and procrastination more of a concern. This theoretical prediction is consistent with our empirical results: assuming that low-performing students are more likely to be hyperbolic discounters, one should observe a larger treatment effect overall for low-performing students, with a particular focus on the final exam mark. This is the pattern we observe in our treatment effects, with a gradient across quantiles — positive for the Week-5 mid-term and negative for the final exam — that is consistent with the theoretical predictions.

4.2 Consumption-Intensive Education and Online Learning

Higher education faces an unprecedented demand for skills. In the U.S., the college wage premium has steadily increased since 1980, suggesting that the supply of educated labor has not kept pace with demand (Goldin and Katz, 2008). The resulting gap is not surprising given how labor-intensive the higher education industry is, with the cost per student outpacing inflation and promising to continue to do so in the years ahead (Baum et al., 2013; Baumol and Bowen, 1966). The state of U.S. public funding has also played a considerable role. Not only inflation-adjusted state appropriations have failed to rise in recent years, they have in fact declined by 16% since 2007 after a long stagnation period commencing in 1990 (Baum and Ma, 2014).

In an attempt to do more with less, the search for innovations that could 'bend the cost curve' has taken center stage (Deming et al., 2015; Bowen et al., 2014). Online education promises to do just that, by reducing labor costs through larger class size and less face-to-face interaction (Bowen, 2012). From sophisticated online courses to Massive Open Online Courses (MOOCs), the adoption of online classes has exploded, especially in the past few years (Christensen and Eyring, 2011; Cowen and Tabarrok, 2014). For instance, of all U.S. undergraduates seeking a degree in 2013, 11.1% were enrolled in fully online programs, while another 25% took at least one course online (Deming et al., 2015). Interestingly, this phenomenon does not involve just community and junior colleges: almost all institutions with more than 15,000 students offer Internet classes. This includes 10 of the largest 4-year colleges and universities in the U.S., some with more than 10,000 students enrolled in at least one online class per term (Figlio et al., 2013).

This profound transformation of higher education prompts the question of whether online instruction can fulfill its promise of reducing costs while achieving the desired learning objectives. Early evidence suggests that online learning could indeed significantly cut higher education costs (Deming et al., 2015). This might come, however, at the expense of educational quality and outcomes. Experimental evidence on the effectiveness of online courses is mixed, lack of engagement and last minute cramming being the main culprits of students' poor performance compared with more traditional approaches (Donovan et al., 2006).

Early randomized trials found no difference in student achievement (Bowen et al., 2014; Figlio et al. 2013; Chou, 2012), but two recent studies show that switching from live to online delivery has a negative impact on final grades in introductory economics classes (Couch et al., 2014; Joyce et al., 2014). And, perhaps as a consequence of this quality drop, Deming et al. (2014) find experimental evidence that employers are less likely to contact job applicants with degrees from online institutions.

Given how crucial education is in all aspects of life,[27] these mixed results have lent urgency to designing innovations that can render online education more effective, with a particular focus on engagement. To this effect, videogame tasks have been heralded as a possible solution (FAS, 2006; Cowen and Tabarrok, 2014). Because of their interactive nature, they can successfully foster engagement. And people learn more and better when they enjoy the learning process and feel motivated (Bell and Kozlowski, 2008).

Despite their learning potential and rapid adoption rates, empirical evidence on the impact of educational videogames on students' academic performance is virtually non-existent. Our results fill this gap and seem to directly attest to the effectiveness of videogame tasks for university learning. And with the surging need to deliver high-standard services within increasingly tight budgets, videogames could play a crucial role in online learning.

[27]Education has long been showed to have crucial (and long-lasting) effects on one's life: those who miss out, for instance, have more chances to be unemployed or underemployed, and are more likely to become trapped in poverty (Nickell, 1979; Leighton and Mincer, 1982; Mincer, 1991; Card, 2001; Farber, 2004; Riddell and Song, 2011). This has dramatic consequences in terms of health, marriage, parenting, social isolation and intergenerational equity (Lochner and Moretti, 2004; Milligan et al., 2004; Grossman, 2005; Lleras-Muney, 2005; Oreopoulos and Salvanes, 2011, and references therein).

5 Conclusions

Our study provides the first causal estimates of the impact of a simple videogame task on students' academic performance. To this purpose, in Semester 1 2013, we designed and deployed an online interactive learning tool to all the students enrolled in a large Principles of Microeconomics course at a major research-intensive university. Anonymized server data showed high game usage, both at the intensive and extensive margin, and no sign of last minute cramming.

Next, we use administrative data and assess the impact of the game on the first mid-term and the final exam scores. The material tested and the tests' structure were equivalent to the previous year, which provides a unique opportunity to evaluate the effect of gamification on student learning. We find that the videogame task substantially improved students' academic performance in both these assessments, generating 8.35% and 9.05% higher mid-term and final exam scores, respectively. There are no major differences by sex, but we do find stronger effects for those enrolled in STEM and Economics & Commerce. Finally, our quantile estimates also show significant heterogeneity in the effect of the game across the student performance distribution, with low-performing students benefitting the most. This is consistent with a model where low-performing students gain more from consumption-intensive education, owing to an increase in the productivity of leisure hours and a decrease in the likelihood of procrastination.

In the long-run, we find that those with access to the game completed roughly one more course in Economics during their academic career. This effect is due to choices related to their field of study, with our intervention leading to about 10% higher chances to opt for (and complete) Economics & Commerce degrees.

Overall, these results are directly comparable to those from the recent literature on improving the efficacy of higher education, but importantly, they are associated with an intervention that comes at zero marginal cost, while all the interventions studied so far (hiring teaching dedicated staff, improving instructor quality, reducing class size, etc.) are significantly more costly. Given, however, the small scale of the material covered by the videogame task deployed and the heterogeneity of the game effects, more experimentation is needed before drawing a definite conclusion regarding the superiority of a fully gamified online course compared to traditional online (or live) courses. Disentangling the channels through which educational videogames may enhance academic outcomes (via cognitive processes and/or learning motivation) is also an important task that we hope to address in future research.

References

[1] Abdulkadiroğlu, A., Angrist, J. & Pathak, P. (2014b). The elite illusion: Achievement effects at Boston and New York exam schools. Econometrica 82(1): 137-196.

[2] Abdulkadiroğlu, A., Angrist, J. D., Hull, P. D. & Pathak, P. A. (2014a). Charters without lotteries: Testing takeovers in New Orleans and Boston NBER Working Paper No.20792.

[3] Angrist, J.D., Cohodes, S.R., Dynarski, S.M., Pathak, P.A. & Walters, C.R. (2016). Stand and deliver: Effects of Boston's charter high schools on college preparation, entry and choice. Journal of Labor Economics 34(2): 275-318.

[4] Angrist, J.D., Pathak, P.A. & Walters, C.R. (2013). Explaining charter school effectiveness. American Economic Journal: Applied Economics 5(4): 1-27.

[5] Bandiera, O., Larcinese, V. & Rasul, I. (2010). Heterogeneous class size ffects: New evidence from a panel of university students. Economic Journal 120(549): 1365-1398.

[6] Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., Shotland, M. and Walton, M., 2016. Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of *Teaching at the Right Level* in India (No. w22746). National Bureau of Economic Research.

[7] Baum, S., Kurose, C. & McPherson, M. (2013). An overview of American higher education. The Future of Children 23(1): 17-39.

[8] Baum, S. & Ma, J. (2014). Trends in college pricing. Princeton, NJ: The College Board.

[9] Baumol, W.J. & Bowen, W.G. (1966). Performing arts: The economic dilemma; A study of problems common to theater, opera, music and dance. New York: The Twentieth Century Fund.

[10] Becker, G.S. (1965). A theory of time allocation. Economic Journal 75(299): 493-517.

[11] Bell, B. & Kozlowski, S. (2008). Active learning: Effects of core training design elements on self-regulatory processes, learning, and adaptability. Journal of Applied Psychology 93, 296-316.

[12] Bowen, W.G., Chignos, M.M., Lack, K.A. & Nygren, T.I. (2014) Interactive learning online at public universities: Evidence from a six-campus randomized trial, Journal of Policy Analysis and Management 33(1), 94-111.

[13] Bowen, W.G. (2012). The 'cost disease' in higher education: Is technology the answer? Lectures presented at the Tanner Lectures on Human Values at Stanford University, Stanford, CA.

[14] Card, D. (2001). Estimating the return to schooling: progress on some persistent econometric problems. Econometrica 69, 1127-1160.

[15] Chetty, R., Friedman, J. N. & Rockoff, J. E. (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. American Economic Review 104(9): 2593-2632.

[16] Chetty, R., Friedman, J. N. & Rockoff, J. E. (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. American Economic Review 104(9): 2633-2679.

[17] Chou, P.N. (2012). Effect of students' self-directed learning abilities on online learning outcomes: Two exploratory experiments in electronic engineering. International Journal of Humanities and Social Science 2(6), 172-179.

[18] Christensen, C.M. & Eyring, H.J. (2011). The innovative university: Changing the DNA of higher education from the inside out. San Francisco: Jossey-Bass.

[19] Couch, K., Harmon, O.R. & Alpert, W. (2014). Online, blended and classroom teaching of Economics Principles: A randomized experiment. Paper presented at the Committee on Economic Education Conference on Teaching and Research in Economic Education, Washington, DC.

[20] Cowen, T. & Tabarrok, A. (2014). The industrial organization of online education. American Economic Review: Papers & Proceedings 104(5): 519-522.

[21] Cooper, J. & Kugler, M.B. (2008). The digital divide: The role of gender in human computer interactions. In A. Sears & J.A. Jacko (Eds.), The human-computer interaction handbook. Lawrence Erlbaum Associates, 763-775.

[22] Croson, R. & Gneezy, U. (2009). Gender differences in preferences. Journal of Economic Literature 47(2), 448-474.

[23] Cunha, F. & Heckman, J. (2007). The technology of skill formation. American Economic Review 97(2): 31-47.

[24] Deming, D.J., Goldin, C., Katz, L.F. & Yuchtman, N. (2015). Can online learning bend the higher education cost curve? American Economic Review: Papers & Proceedings 105(5): 496-501.

[25] Deming, D.J., Yuchtman, N., Abulafi, A., Goldin, C. & Katz, L.F. (2014). The value of postsecondary credentials in the labor market: An experimental study. NBER Working Paper No. 20528.

[26] Dobbie, W. & Fryer Jr, R.G. (2015). The medium-term impacts of high-achieving charter schools. Journal of Political Economy 123(5): 985-1037.

[27] Dobbie, W. & Fryer Jr, R.G. (2014). The impact of attending a school with high-achieving peers: Evidence from New York City exam schools. American Economic Journal: Applied Economics 6(3).

[28] Dobrescu, L.I., Greiner, B. & Motta, A. (2015). Learning economics concepts through game-Play: An experiment. International Journal of Education Research 69, 23-37.

[29] Donovan, C., Figlio, D.N. & Rush, M. (2006). Cramming: The effects of school accountability on college-bound students. NBER Working Paper No. 12628.

[30] Farber, H.S. (2004). Job loss in the United States, 1981 to 2001. Research in Labor Economics 23, 69-117.

[31] Federation of American Scientists (FAS). (2006). Report: Summit on educational games: Harnessing the power of video games for learning. Washington, D.C.

[32] Figlio, D.N., Rush, M. & Yin, L. (2013). Is it live or is it internet? Experimental estimates of the effects of online instruction on student learning. Journal of Labor Economics 31(4), 763-784.

[33] Figlio, D. N., Schapiro, M. O., & Soter, K. B. (2015). Are tenure track professors better teachers? Review of Economics and Statistics, 97(4), 715-724.

[34] Figlio, D., Karbownik, K., Roth, J., & Wasserman, M. (2016). School quality and the gender gap in educational achievement. The American Economic Review, 106(5), 289-295.

[35] Fischer, C. (2001) Read this paper later: procrastination with time-consistent preferences, Journal of Economic Behavior & Organization, Volume 46, Issue 3, November 2001, Pages 249-269.

[36] Fischer, C. (1999) Read this paper even later: procrastination with time-inconsistent preferences, Discussion Paper 99-20.

[37] Frank, R., Jennings, S. & Bernanke, B. (2012). Principles of Microeconomics, 3rd Edition, McGraw Hill, Australia.

[38] Fryer Jr, R.G., Devi, T & Holden, R.T. (2016). Vertical versus horizontal incentives in education: Evidence from randomized trials. NBER Working Paper No. 17752.

[39] Fryer Jr, R.G. (2014). Injecting charter school best practices into traditional public schools: Evidence from field experiments. Quarterly Journal of Economics 129 (3): 1355-1407.

[40] Fryer Jr, R.G. & Katz, L.F. (2013). Achieving escape velocity: Neighborhood and school interventions to reduce persistent inequality. American Economic Review 103(3): 232-237.

[41] Fryer Jr, R.G. & Levitt, S.D. (2013). Testing for racial differences in the mental ability of young children. American Economic Review 103(2): 981-1005.

[42] Gee, J.P. (2003). What video games have to teach us about learning and literacy. United States of America: Palgrave MacMillan.

[43] Gneezy, U., Niederle, M. & Rustichini, A. (2003). Performance in competitive environments: Gender differences. Quarterly Journal of Economics 118(3): 1049-1074.

[44] Goldin, C. & Katz, L.F. (2008). The race between education and technology. Cambridge, MA: Belknap Press of Harvard University.

[45] Grossman, M. (2005). Education and non-market outcomes. NBER Working Paper No. 11582.

[46] Hoffmann, F. & Oreopoulos, P. (2009). Professor qualities and student achievement. Review of Economics and Statistics 91(1): 83-92.

[47] Joyce, T.J., Crockett, S., Jaeger, D.A., Altindag, O. & O'Connell, S.D. (2014). Does classroom time matter? A randomized field experiment of hybrid and traditional lecture formats in economics. NBER Working Paper 20006.

[48] Muralidharan, K., Singh, A. and Ganimian, A.J., (2016). Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India (No. w22923). National Bureau of Economic Research.

[49] Laibson, D. (1997), Golden Eggs and Hyperbolic Discounting, The Quarterly Journal of Economics Vol. 112, No. 2, pp. 443-477

[50] Leighton, L. & Mincer, J. (1982). Labor turnover and youth unemployment. In R.B. Freeman & D.A. Wise (Eds.), The youth labor market problem: Its nature, causes, and consequences. University of Chicago Press, 235-276.

[51] Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States. Review of Economic Studies 72, 189-221.

[52] Lochner, L. & Moretti, E. (2004) The effect of education on crime: evidence from prison inmates, arrests and self-reports. American Economic Review 94, 155-189.

[53] Lucas, K. & Sherry, J. L. (2004). Sex differences in video game play: A communication-based explanation. Communication Research 31(5), 499-523.

[54] Milligan, K., Moretti, E. & Oreopoulos, P. (2004). Does education improve citizenship: evidence from the U.S. and U.K. Journal of Public Economics 88, 1667-1695.

[55] Mincer, J. (1991). Education and unemployment. NBER Working Paper No. 3838.

[56] Nickell, S. (1979). Estimating the probability of leaving unemployment. Econometrica 47, 1249-1266.

[57] Niederle, M. & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? Quarterly Journal of Economics 122(3), 1067-1101.

[58] Niederle, M. & Vesterlund, L. (2011). Gender and competition. Annual Review of Economics 3(1), 601-630.

[59] Oreopoulos, P. & Salvanes, K. (2011). Priceless: The Nonpecuniary Benefits of Schooling. Journal of Economic Perspectives 25(1), 159-184.

[60] Prensky, M. (2001). Digital game-based learning. New York: McGraw-Hill.

[61] Riddell, W.C. & Song, X. (2011). The impact of education on unemployment incidence and re-employment success: Evidence from the U.S. labour market. Labour Economics 18(4), 453-463.

[62] Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. Personnel Psychology 64(2), 489-528.

[63] Vorderer, P., Hartmann, T. & Klimmt, C. (2006). Explaining the enjoyment of playing video games: The role of competition. In D. Marinelli, ed., 'ICEC Conference Proceedings 2003: Essays on the Future of Interactive Entertainment', Carnegie Mellon University Press, Pittsburgh, PA, pp. 107-120.

[64] Wouters, P., van Nimwegen, C., van Oostendorp, H. & van der Spek, E.D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. Journal of Educational Psychology 105 2, 249-265.

[65] Woodfield, R., Jessop,D. & L.McMillan (2006). Gender differences in undergraduate attendance rates. Studies in Higher Education 31(1), 1-22.

Table 1. Lecturer allocation, by year and semester

Teaching period	Number of students taught by						Total
	Lecturer A	Lecturer B	Lecturer C	Lecturer D	Lecturer E	Lecturer F	
2012, Semester 1	0	0	290	501	689	290	1,770
2012, Semester 2	0	332	256	0	0	0	588
2013, Semester 1	288	0	319	627	269	309	1,812
2013, Semester 2	0	335	289	0	0	0	624
Total	288	667	545	501	958	580	*4,794*

Notes: The table presents the number of students taught by a lecturer in each of the four semesters that we analyse.

Table 2A.1: Descriptive statistics for Semester 1

	Year 2012					Year 2013				
	Obs.	Mean	Std. Dev.	Min	Max	Obs.	Mean	Std. Dev.	Min	Max
Male	1770	0.56	0.50	0.00	1.00	1812	0.56	0.50	0.00	1.00
Age	1770	18.72	2.09	16.00	66.00	1812	18.59	1.65	16.00	37.00
Full-time status	1770	0.93	0.25	0.00	1.00	1812	0.94	0.24	0.00	1.00
International status	1770	0.26	0.44	0.00	1.00	1812	0.22	0.42	0.00	1.00
COB: Australia	1770	0.53	0.50	0.00	1.00	1812	0.56	0.50	0.00	1.00
COB: Other Oceania	1770	0.01	0.11	0.00	1.00	1812	0.01	0.10	0.00	1.00
COB: Europe	1770	0.02	0.15	0.00	1.00	1812	0.02	0.14	0.00	1.00
COB: Asia	1770	0.41	0.49	0.00	1.00	1812	0.37	0.48	0.00	1.00
COB: Americas	1770	0.01	0.09	0.00	1.00	1812	0.02	0.14	0.00	1.00
COB: Africa & Middle East	1770	0.02	0.13	0.00	1.00	1812	0.02	0.14	0.00	1.00
ATAR score	1321	93.73	5.80	43.00	99.95	1346	94.16	5.40	60.60	99.95
Times repeated the course after S1 2012	1770	0.00	0.00	0.00	0.00	1812	0.03	0.17	0.00	2.00
Game exposure	1770	0.00	0.00	0.00	0.00	1812	47.70	18.46	20.00	60.00
FOS: Others	1770	0.10	0.29	0.00	1.00	1812	0.07	0.26	0.00	1.00
FOS: STEM	1770	0.17	0.38	0.00	1.00	1812	0.16	0.36	0.00	1.00
FOS: Econ & Commerce	1770	0.73	0.44	0.00	1.00	1812	0.77	0.42	0.00	1.00
IFOS: Others	1770	0.02	0.15	0.00	1.00	1812	0.02	0.15	0.00	1.00
IFOS: STEM	1770	0.15	0.36	0.00	1.00	1812	0.13	0.34	0.00	1.00
IFOS: Econ & Commerce	1770	0.03	0.18	0.00	1.00	1812	0.03	0.17	0.00	1.00
IFOS: Unknown	1770	0.78	0.41	0.00	1.00	1812	0.80	0.40	0.00	1.00
IFOS: Undeclared	1770	0.02	0.12	0.00	1.00	1812	0.02	0.13	0.00	1.00
Week-5 mid-term score	1759	6.83	1.66	0.00	10.00	1798	7.31	1.87	0.00	10.00
Week-9 mid-term score	1769	6.86	1.71	0.00	10.00	1812	6.54	2.06	0.00	10.00
Final exam score	1770	6.06	1.58	0.00	9.71	1790	6.88	1.39	2.40	9.80
Number of all other Economics courses	1679	3.05	3.96	0.00	20.00	1729	3.29	4.02	0.00	24.00
Completed an Econ & Commerce degree	1770	0.73	0.44	0.00	1.00	1812	0.77	0.42	0.00	1.00
Mean score in all other courses	1675	6.77	1.07	0.00	9.09	1726	6.88	1.1	0.00	9.56
Mean score in all other Econ & Commerce courses	1511	6.69	1.13	0.00	9.21	1598	6.78	1.17	0.00	9.60
Graduated 'in time'	1720	0.21	0.40	0.00	1.00	1758	0.22	0.41	0.00	1.00

Notes: COB refers to the country of birth. The classification of the country of birth follows the Standard Australian Classification of Countries, 2011, Version 2.3 (http://www.abs.gov.au/AUS STATS/abs@.nsf/DetailsPage/1269.02011?OpenDocument). The Other Oceania group includes Oceania countries other than Australia. Australian Tertiary Admission Rank (ATAR) score denotes a student's ranking relative to his/her peers when completing secondary education. Game exposure refers to the time (in minutes) that a lecturer introduced the game to students during class. FOS denotes the field of study of their (completed) degree, and STEM refers to Science, Technology, Engineering & Mathematics. IFOS captures the intended field of study as reported by the students at the beginning of their academic career. To ease comparison, all scores for Week-5 and Week-9 mid-terms, and final exam are showed on a 0 – 10 scale. "All other courses" and "all other Econ & Commerce courses" refer to courses taken after being included in the control (Semester 1 2012) or treatment (Semester 1 2013) for the subsequent eight semesters.

Table 2A.2: Descriptive statistics for Semester 2

	Year 2012					Year 2013				
	Obs.	Mean	Std. Dev.	Min	Max	Obs.	Mean	Std. Dev.	Min	Max
Male	588	0.61	0.49	0.00	1.00	624	0.56	0.50	0.00	1.00
Age	588	19.57	1.96	17.00	33.00	624	19.48	2.72	17.00	58.00
Full-time status	588	0.82	0.38	0.00	1.00	624	0.80	0.40	0.00	1.00
International status	588	0.41	0.49	0.00	1.00	624	0.41	0.49	0.00	1.00
COB: Australia	588	0.41	0.49	0.00	1.00	624	0.40	0.49	0.00	1.00
COB: Other Oceania	588	0.01	0.07	0.00	1.00	624	0.01	0.11	0.00	1.00
COB: Europe	588	0.04	0.19	0.00	1.00	624	0.04	0.21	0.00	1.00
COB: Asia	588	0.51	0.50	0.00	1.00	624	0.51	0.50	0.00	1.00
COB: Americas	588	0.01	0.11	0.00	1.00	624	0.02	0.14	0.00	1.00
COB: Africa & Middle East	588	0.02	0.15	0.00	1.00	624	0.01	0.09	0.00	1.00
ATAR score	322	91.56	7.12	63.30	99.95	355	91.95	6.78	46.65	99.90
Times repeated the course after S1 2012	588	0.11	0.32	0.00	1.00	624	0.07	0.28	0.00	2.00
Game exposure	588	0.00	0.00	0.00	0.00	624	0.00	0.00	0.00	0.00
FOS: Others	588	0.25	0.43	0.00	1.00	624	0.18	0.38	0.00	1.00
FOS: STEM	588	0.35	0.48	0.00	1.00	624	0.35	0.48	0.00	1.00
FOS: Econ & Commerce	588	0.40	0.49	0.00	1.00	624	0.47	0.50	0.00	1.00
Week-5 mid-term score	578	6.55	2.12	0.00	10.00	620	6.81	1.94	0.75	10.00
Week-9 mid-term score	588	6.50	2.46	0.00	10.00	604	6.55	2.03	0.00	10.00
Final exam score	588	5.86	1.66	0.00	9.71	611	5.90	1.36	2.00	9.00

Notes: COB refers to the country of birth. The classification of the country of birth follows the Standard Australian Classification of Countries, 2011, Version 2.3 (http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1269.02011?OpenDocument). The Other Oceania group includes Oceania countries other than Australia. Australian Tertiary Admission Rank (ATAR) score denotes a student's ranking relative to his/her peers when completing secondary education. Game exposure refers to the time (in minutes) that a lecturer introduced the game to student during class. FOS denotes the field of study of their (completed) degree, and STEM refers to Science, Technology, Engineering & Mathematics. To ease the comparison, all scores for Week 5 and Week 9 mid-terms, as well as for the final exam are showed on a 0 – 10 scale.

Table 2B.1: Difference in personal characteristics between the treatment and control groups in Semester 1

<table>
<tr><th></th><th></th><th></th><th></th><th></th><th>Ha: diff < 0</th><th>Ha: diff != 0</th><th>Ha: diff > 0</th></tr>
<tr><th></th><th>Difference</th><th>Obs</th><th>S.E.</th><th>T-stat</th><th>Pr(T < t)</th><th>Pr(|T| > |t|)</th><th>Pr(T > t)</th></tr>
<tr><td>Male</td><td>0.002</td><td>3582</td><td>0.017</td><td>0.123</td><td>0.549</td><td>0.902</td><td>0.451</td></tr>
<tr><td>Age</td><td>0.123**</td><td>3582</td><td>0.063</td><td>1.963</td><td>0.975</td><td>0.050</td><td>0.025</td></tr>
<tr><td>Full-time status</td><td>-0.005</td><td>3582</td><td>0.008</td><td>-0.659</td><td>0.255</td><td>0.510</td><td>0.745</td></tr>
<tr><td>International status</td><td>0.036**</td><td>3582</td><td>0.014</td><td>2.506</td><td>0.994</td><td>0.012</td><td>0.006</td></tr>
<tr><td>COB: Australia</td><td>-0.031*</td><td>3582</td><td>0.017</td><td>-1.879</td><td>0.030</td><td>0.060</td><td>0.970</td></tr>
<tr><td>COB: Other Oceania</td><td>0.000</td><td>3582</td><td>0.004</td><td>0.075</td><td>0.530</td><td>0.941</td><td>0.470</td></tr>
<tr><td>COB: Europe</td><td>0.002</td><td>3582</td><td>0.005</td><td>0.335</td><td>0.631</td><td>0.738</td><td>0.369</td></tr>
<tr><td>COB: Asia</td><td>0.041**</td><td>3582</td><td>0.016</td><td>2.515</td><td>0.994</td><td>0.012</td><td>0.006</td></tr>
<tr><td>COB: Americas</td><td>-0.010***</td><td>3582</td><td>0.004</td><td>-2.652</td><td>0.004</td><td>0.008</td><td>0.996</td></tr>
<tr><td>COB: Africa & Middle East</td><td>-0.001</td><td>3582</td><td>0.004</td><td>-0.280</td><td>0.390</td><td>0.779</td><td>0.610</td></tr>
<tr><td>IFOS: Other</td><td>0.000</td><td>3582</td><td>0.005</td><td>-0.006</td><td>0.498</td><td>0.995</td><td>0.502</td></tr>
<tr><td>IFOS: STEM</td><td>0.018</td><td>3582</td><td>0.012</td><td>1.584</td><td>0.943</td><td>0.113</td><td>0.057</td></tr>
<tr><td>IFOS: Econ & Commerce</td><td>0.004</td><td>3582</td><td>0.006</td><td>0.698</td><td>0.757</td><td>0.485</td><td>0.243</td></tr>
<tr><td>IFOS: Unknown</td><td>-0.021</td><td>3582</td><td>0.014</td><td>-1.551</td><td>0.060</td><td>0.121</td><td>0.940</td></tr>
<tr><td>IFOS: Undeclared</td><td>-0.001</td><td>3582</td><td>0.004</td><td>-0.311</td><td>0.378</td><td>0.756</td><td>0.622</td></tr>
<tr><td>ATAR score</td><td>-0.432**</td><td>2667</td><td>0.217</td><td>-1.992</td><td>0.023</td><td>0.046</td><td>0.977</td></tr>
</table>

Notes: COB refers to the country of birth. IFOS captures the intended field of study reported by the students at the beginning of their academic career. Australian Tertiary Admission Rank (ATAR) score denotes a student’s ranking relative to his/her peers when completing secondary education. IFOS captures the intended field of study as reported by the students at the beginning of their academic career. STEM refers to Science, Technology, Engineering & Mathematics. * P<0.10, ** P<0.05, *** P<0.01.

Table 2B.2: Difference in personal characteristics between the students in Semester 2 2012 and Semester 2 2013

	Difference	Obs	S.E.	T-stat	Ha: diff < 0 Pr(T < t)	Ha: diff != 0 Pr(|T| > |t|)	Ha: diff > 0 Pr(T > t)
Male	0.053*	1212	0.028	1.871	0.969	0.062	0.031
Age	0.084	1212	0.137	0.613	0.730	0.540	0.270
Full-time status	0.017	1212	0.023	0.748	0.773	0.454	0.227
International status	0.001	1212	0.028	0.046	0.518	0.963	0.482
COB: Australia	0.004	1212	0.028	0.149	0.559	0.881	0.441
COB: Other Oceania	-0.006	1212	0.005	-1.176	0.120	0.240	0.880
COB: Europe	-0.006	1212	0.012	-0.498	0.309	0.618	0.691
COB: Asia	-0.001	1212	0.029	-0.028	0.489	0.977	0.511
COB: Americas	-0.007	1212	0.007	-1.026	0.153	0.305	0.847
COB: Africa & Middle East	0.016**	1212	0.007	2.215	0.987	0.027	0.013
ATAR score	-0.394	677	0.535	-0.737	0.231	0.461	0.769

Notes: COB refers to the country of birth. Australian Tertiary Admission Rank (ATAR) score denotes a student's ranking relative to his/her peers when completing secondary education. * P<0.10, ** P<0.05, *** P<0.01.

Table 3: Results on the game impact on Week-5 mid-term scores in Semester 1

	Week-5 mid-term					
	(1)	(2)	(3)	*(4)*	(5)	(6)
Year 2013	0.474***	0.448***	0.535***	0.835***	0.831***	0.827***
(Treatment group)	(0.059)	(0.057)	(0.200)	(0.220)	(0.222)	(0.239)
Male		0.085	0.085	0.081	0.067	0.061
		(0.059)	(0.059)	(0.057)	(0.057)	(0.060)
Age		-0.152***	-0.151***	-0.148***	-0.147***	-0.101***
		(0.028)	(0.028)	(0.029)	(0.029)	(0.032)
Full-time		0.012	-0.014	-0.011	0.008	0.083
		(0.136)	(0.137)	(0.133)	(0.135)	(0.151)
International student		-0.540***	-0.534***	-0.526***	-0.526***	-0.383***
		(0.097)	(0.097)	(0.094)	(0.095)	(0.148)
No. of retakes		-0.609**	-0.574**	-0.534**		-0.423
		(0.274)	(0.273)	(0.266)		(0.280)
Country of birth						
Other Oceania		-0.338	-0.337	-0.349	-0.308	-0.215
		(0.285)	(0.285)	(0.265)	(0.267)	(0.238)
Europe		0.298	0.303	0.243	0.243	0.093
		(0.223)	(0.225)	(0.215)	(0.215)	(0.259)
Asia		-0.256***	-0.255***	-0.249***	-0.254***	-0.223***
		(0.080)	(0.080)	(0.077)	(0.078)	(0.077)
Americas		-0.436	-0.435	-0.439	-0.435	-0.154
		(0.276)	(0.273)	(0.290)	(0.291)	(0.417)
Africa & Middle East		-0.264	-0.273	-0.213	-0.201	-0.141
		(0.227)	(0.227)	(0.217)	(0.215)	(0.225)
Game exposure			-0.004	-0.005	-0.005	-0.007*
			(0.004)	(0.004)	(0.004)	(0.004)
Lecturer C			0.169	0.196	0.175	0.093
			(0.180)	(0.178)	(0.180)	(0.197)
Lecturer D			0.431**	0.544***	0.523***	0.465**
			(0.181)	(0.180)	(0.183)	(0.198)
Lecturer E			0.156	0.182	0.163	0.048
			(0.155)	(0.155)	(0.158)	(0.173)
ATAR						0.091***
						(0.007)
Constant	6.832***	9.863***	9.622***	9.187***	9.188***	-0.257
	(0.040)	(0.546)	(0.569)	(0.589)	(0.597)	(0.952)
Tutor dummies	×	×	×	✓	✓	✓
N	3557	3557	3557	3557	3516	2649
R-sq	0.018	0.098	0.102	0.185	0.185	0.245

Notes: The classification of the country of birth follows the Standard Australian Classification of Countries, 2011, Version 2.3 (http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1269.02011?OpenDocument). The Other Oceania group includes Oceania countries other than Australia (which is the omitted base group). Robust standard errors are reported in parentheses below the estimates. * P<0.10, ** P<0.05, *** P<0.01.

Table 4: Results on the game impact on final exam scores in Semester 1

	Final exam					
	(1)	(2)	(3)	*(4)*	(5)	(6)
Year 2013	0.820***	0.834***	0.629***	0.905***	0.907***	0.981***
(Treatment group)	(0.049)	(0.049)	(0.179)	(0.209)	(0.210)	(0.226)
Male		0.527***	0.529***	0.517***	0.515***	0.542***
		(0.050)	(0.050)	(0.050)	(0.051)	(0.052)
Age		-0.082***	-0.082***	-0.080***	-0.080***	0.011
		(0.016)	(0.016)	(0.016)	(0.017)	(0.031)
Full-time		0.367***	0.364***	0.363**	0.374***	0.364**
		(0.138)	(0.140)	(0.142)	(0.145)	(0.162)
International student		0.081	0.082	0.089	0.083	0.222*
		(0.078)	(0.078)	(0.077)	(0.078)	(0.130)
No. of retakes		-0.983***	-0.962***	-0.958***		-0.094
		(0.186)	(0.183)	(0.189)		(0.221)
Country of birth						
Other Oceania		0.499***	0.495***	0.519***	0.560***	0.528***
		(0.184)	(0.184)	(0.184)	(0.185)	(0.164)
Europe		0.375**	0.375**	0.392**	0.395**	0.360**
		(0.165)	(0.164)	(0.161)	(0.162)	(0.164)
Asia		-0.033	-0.033	-0.044	-0.037	-0.027
		(0.068)	(0.068)	(0.067)	(0.068)	(0.064)
Americas		-0.150	-0.148	-0.188	-0.183	-0.267
		(0.228)	(0.226)	(0.226)	(0.226)	(0.304)
Africa & Middle East		-0.144	-0.148	-0.19	-0.197	-0.103
		(0.171)	(0.170)	(0.168)	(0.172)	(0.163)
Game exposure			0.004	0.004	0.004	0.000
			(0.004)	(0.004)	(0.005)	(0.005)
Lecturer C			-0.009	-0.007	-0.017	0.029
			(0.082)	(0.083)	(0.084)	(0.087)
Lecturer D			0.011	0.027	0.013	0.095
			(0.143)	(0.147)	(0.149)	(0.158)
Lecturer E			0.057	0.048	0.035	0.024
			(0.113)	(0.115)	(0.118)	(0.126)
Lecturer F			0.042	0.044	0.041	0.041
			(0.082)	(0.084)	(0.085)	(0.089)
ATAR						0.129***
						(0.006)
Constant	6.056***	6.940***	6.908***	7.019***	7.017***	-6.904***
	(0.037)	(0.343)	(0.361)	(0.399)	(0.403)	(0.914)
Tutor dummies	✗	✗	✗	✓	✓	✓
N	3560	3560	3560	3560	3520	2652
R-sq	0.071	0.120	0.121	0.135	0.134	0.345

Notes: The classification of the country of birth follows the Standard Australian Classification of Countries, 2011, Version 2.3 (http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1269.02011?OpenDocument). The Other Oceania group includes Oceania countries other than Australia (which is the omitted base group). Robust standard errors are reported in parentheses below the estimates. * P<0.10, ** P<0.05, *** P<0.01.

Table 4A: Results on the game impact on final exam scores in Semester 1, by question relevance

	Relevant questions						Irrelevant questions					
	(1)	(2)	(3)	*(4)*	(5)	(6)	(1)	(2)	(3)	*(4)*	(5)	(6)
Year 2013	1.318***	1.353***	1.126***	1.239***	1.179***	1.579***	0.845***	0.858***	0.645***	0.947***	0.950***	1.014***
(Treatment group)	(0.098)	(0.098)	(0.400)	(0.454)	(0.457)	(0.523)	(0.050)	(0.049)	(0.181)	(0.210)	(0.212)	(0.228)
Male		0.817***	0.825***	0.799***	0.783***	0.740***		0.488***	0.490***	0.479***	0.477***	0.510***
		(0.101)	(0.101)	(0.102)	(0.102)	(0.116)		(0.050)	(0.050)	(0.050)	(0.051)	(0.053)
Age		-0.071***	-0.070***	-0.070***	-0.067***	0.041		-0.082***	-0.083***	-0.080***	-0.080***	0.012
		(0.027)	(0.027)	(0.026)	(0.026)	(0.061)		(0.017)	(0.017)	(0.017)	(0.017)	(0.031)
Full-time		0.206	0.182	0.154	0.143	0.314		0.367***	0.367***	0.368***	0.381***	0.366**
		(0.231)	(0.233)	(0.237)	(0.239)	(0.291)		(0.138)	(0.140)	(0.141)	(0.144)	(0.160)
International student		0.350**	0.359**	0.344**	0.276*	0.397		0.059	0.06	0.069	0.065	0.192
		(0.157)	(0.157)	(0.158)	(0.158)	(0.269)		(0.079)	(0.079)	(0.078)	(0.079)	(0.130)
No. of retakes		-1.313***	-1.283**	-1.272**		-0.966		-0.969***	-0.949***	-0.944***		-0.065
		(0.502)	(0.502)	(0.509)		(0.658)		(0.187)	(0.185)	(0.190)		(0.221)
Country of birth												
Other Oceania		0.708*	0.691*	0.686*	0.625	0.862**		0.475***	0.472***	0.497***	0.542***	0.499***
		(0.411)	(0.407)	(0.407)	(0.412)	(0.368)		(0.181)	(0.180)	(0.182)	(0.182)	(0.162)
Europe		-0.718*	-0.719*	-0.750*	-0.733*	-0.853		0.440***	0.440***	0.453***	0.456***	0.427***
		(0.419)	(0.420)	(0.419)	(0.418)	(0.543)		(0.163)	(0.162)	(0.159)	(0.159)	(0.161)
Asia		-0.071	-0.071	-0.075	-0.047	-0.033		-0.034	-0.034	-0.045	-0.039	-0.033
		(0.136)	(0.136)	(0.137)	(0.137)	(0.143)		(0.069)	(0.069)	(0.068)	(0.069)	(0.065)
Americas		-0.559	-0.556	-0.539	-0.507	-1.006		-0.14	-0.138	-0.182	-0.178	-0.248
		(0.510)	(0.509)	(0.515)	(0.515)	(0.820)		(0.227)	(0.226)	(0.225)	(0.225)	(0.302)
Africa & Middle East		0.055	0.048	0.018	-0.067	-0.05		-0.165	-0.169	-0.214	-0.218	-0.115
		(0.337)	(0.336)	(0.342)	(0.362)	(0.368)		(0.176)	(0.176)	(0.173)	(0.178)	(0.172)
Game exposure			0.004	0.007	0.008	-0.006			0.005	0.004	0.004	0
			(0.010)	(0.010)	(0.010)	(0.011)			(0.004)	(0.005)	(0.005)	(0.005)
Lecturer C			0.063	0.027	-0.007	0.152			-0.017	-0.008	-0.019	0.028
			(0.178)	(0.183)	(0.185)	(0.209)			(0.083)	(0.084)	(0.085)	(0.088)
Lecturer D			0.024	-0.056	-0.089	0.21			0.002	0.029	0.016	0.088
			(0.317)	(0.330)	(0.334)	(0.374)			(0.144)	(0.149)	(0.151)	(0.160)
Lecturer E			-0.01	-0.034	-0.063	0.135			0.06	0.051	0.039	0.02
			(0.271)	(0.281)	(0.285)	(0.318)			(0.115)	(0.117)	(0.120)	(0.127)
Lecturer F			0.108	0.081	0.081	0.165			0.037	0.045	0.042	0.042
			(0.178)	(0.183)	(0.185)	(0.209)			(0.083)	(0.085)	(0.086)	(0.089)
ATAR						0.115***						0.128***
						(0.011)						(0.006)
Constant	6.633***	7.259***	7.199***	7.229***	7.219***	-5.917***	5.986***	6.895***	6.869***	7.002***	7.001***	-6.895***
	(0.067)	(0.585)	(0.655)	(0.716)	(0.710)	(1.780)	(0.037)	(0.348)	(0.367)	(0.406)	(0.411)	(0.919)
Tutor dummies	✗	✗	✗	✓	✓	✓	✗	✗	✗	✓	✓	✓
N	3554	3554	3554	3554	3514	2646	3554	3554	3554	3554	3514	2646
R-sq	0.048	0.073	0.074	0.081	0.08	0.125	0.074	0.119	0.121	0.136	0.135	0.342

Notes: The “Relevant questions” (“Irrelevant questions”) specifications implement specification (4) in Table 4 using as dependent variable the final exam score (out of 10) that students achieved considering only the questions that were (not) pertinent to the videogame module deployed in Week 2. Robust standard errors are reported in parentheses below the estimates. * P<0.10, ** P<0.05, *** P<0.01.

Table 5: Game treatment effect, by standard deviation

	Specification	$\widehat{\beta_1}$	Std. dev. of exam scores	Treatment effect by std. dev.
Week-5 mid-term	(1)	0.474***	1.785	0.266***
		(0.059)		(0.033)
	(2)	0.448***	1.785	0.251***
		(0.057)		(0.032)
	(4)	0.835***	1.785	0.468***
		(0.220)		(0.123)
Final exam	(1)	0.820***	1.541	0.532***
		(0.049)		(0.032)
	(2)	0.834***	1.541	0.541***
		(0.049)		(0.032)
	(4)	0.905***	1.541	0.587***
		(0.209)		(0.136)

Notes: The table shows the OLS estimates of the treatment effect ($\widehat{\beta_1}$) obtained from specifications (1), (2) and (4) of Table 3 and Table 4. * P<0.10, ** P<0.05, *** P<0.01.

Table 6: Results from the quantile analysis for Semester 1

Quantile	Week-5 mid-term			Final exam			Joint (Week-5 & final exam) score		
	25% (A1)	50% (A2)	75% (A3)	25% (B1)	50% (B2)	75% (B3)	25% (A1)	50% (A2)	75% (A3)
Year 2013	0.548	0.578**	0.737***	1.112***	0.763***	0.443	0.936***	0.766***	0.599**
(Treatment group)	(0.370)	(0.250)	(0.245)	(0.299)	(0.265)	(0.307)	(0.249)	(0.245)	(0.241)
Male	0.107	0.065	-0.022	0.520***	0.527***	0.544***	0.267***	0.306***	0.271***
	(0.086)	(0.059)	(0.054)	(0.075)	(0.065)	(0.061)	(0.065)	(0.055)	(0.055)
Age	-0.204***	-0.147***	-0.127***	-0.075***	-0.089***	-0.089***	-0.149***	-0.141***	-0.104***
	(0.037)	(0.030)	(0.032)	(0.024)	(0.018)	(0.018)	(0.023)	(0.028)	(0.025)
Full-time	0.193	-0.039	-0.087	0.360	0.000	-0.019	0.201	-0.079	-0.083
	(0.254)	(0.161)	(0.122)	(0.220)	(0.174)	(0.136)	(0.178)	(0.155)	(0.118)
International student	-0.409**	-0.677***	-0.472***	0.031	0.129	0.046	-0.233**	-0.264***	-0.300***
	(0.159)	(0.100)	(0.099)	(0.113)	(0.104)	(0.088)	(0.110)	(0.092)	(0.092)
No. of retakes	-0.788	-0.257	-0.424	-1.113***	-0.905**	-0.803***	-0.862**	-0.687***	-0.800***
	(0.495)	(0.340)	(0.313)	(0.236)	(0.398)	(0.288)	(0.350)	(0.253)	(0.193)
Country of birth									
Other Oceania	-0.548	-0.489	-0.191	0.767**	0.177	0.511	-0.100	0.097	0.125
	(0.334)	(0.373)	(0.265)	(0.307)	(0.273)	(0.331)	(0.325)	(0.248)	(0.189)
Europe	0.400	0.314	0.380**	0.361*	0.077	0.436	0.404**	0.309	0.369*
	(0.355)	(0.218)	(0.178)	(0.213)	(0.233)	(0.272)	(0.198)	(0.201)	(0.197)
Asia	-0.376**	-0.109	-0.155**	0.001	-0.045	-0.019	-0.129	-0.083	-0.059
	(0.152)	(0.080)	(0.075)	(0.100)	(0.095)	(0.079)	(0.095)	(0.073)	(0.074)
Americas	-0.935*	-0.198	-0.198	-0.126	-0.345	-0.200	-0.627*	-0.258	-0.255
	(0.532)	(0.343)	(0.328)	(0.365)	(0.278)	(0.247)	(0.337)	(0.298)	(0.252)
Africa & Middle East	0.043	-0.050	-0.200	-0.118	0.000	-0.311	-0.319	-0.055	-0.283*
	(0.295)	(0.210)	(0.181)	(0.246)	(0.228)	(0.195)	(0.349)	(0.186)	(0.153)
Game exposure	-0.005	-0.002	-0.002	-0.002	0.008	0.012*	-0.003	-0.002	0.003
	(0.006)	(0.004)	(0.004)	(0.006)	(0.006)	(0.007)	(0.005)	(0.005)	(0.005)
Lecturer C	0.136	0.215	0.260	0.074	0.002	-0.055	0.129	0.078	0.023
	(0.287)	(0.203)	(0.218)	(0.123)	(0.113)	(0.111)	(0.107)	(0.102)	(0.100)
Lecturer D	0.604**	0.469**	0.476**	0.182	-0.063	-0.169	0.407**	0.317*	0.197
	(0.289)	(0.203)	(0.216)	(0.212)	(0.192)	(0.215)	(0.184)	(0.180)	(0.185)
Lecturer E	0.181	0.188	0.258	0.097	-0.043	0.000	0.192	0.120	0.120
	(0.257)	(0.180)	(0.204)	(0.173)	(0.158)	(0.176)	(0.147)	(0.149)	(0.154)
Lecturer F				0.106	0.045	0.090	0.104	0.100	0.065
				(0.129)	(0.118)	(0.110)	(0.109)	(0.102)	(0.098)
Constant	9.463***	9.374***	9.902***	5.959***	7.605***	8.678***	7.977***	8.869***	8.983***
	(0.805)	(0.618)	(0.716)	(0.639)	(0.471)	(0.424)	(0.514)	(0.602)	(0.522)
Tutor dummies	✓	✓	✓	✓	✓	✓	✓	✓	✓
N		3557			3560			3579	
R-sq	0.1369	0.1089	0.109	0.094	0.101	0.0864	0.109	0.094	0.101

Notes: Joint (Week-5 & final exam) score is the average of Week-5 mid-term and Final exam marks (out of 10). The classification of the country of birth follows the Standard Australian Classification of Countries, 2011, Version 2.3 (http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage /1269.02011?OpenDocument). The Other Oceania group includes Oceania countries other than Australia (which is the omitted base group). Robust standard errors are estimated via bootstrap and reported in parentheses below the estimates. * P<0.10, ** P<0.05, *** P<0.01.

Table 7: Results on the game impact on Semester 1 scores, by gender

	Females		Males	
	Week-5 mid-term	Final exam	Week-5 mid-term	Final exam
Year 2013	0.753**	1.031***	0.889***	0.879***
(Treatment group)	(0.365)	(0.326)	(0.276)	(0.278)
Age	-0.196***	-0.091***	-0.122***	-0.066***
	(0.032)	(0.027)	(0.032)	(0.019)
Full-time	-0.278	0.056	0.208	0.599***
	(0.196)	(0.189)	(0.176)	(0.199)
International student	-0.410***	0.231*	-0.599***	-0.048
	(0.135)	(0.119)	(0.126)	(0.104)
No. of retakes	-1.047***	-0.944***	-0.131	-0.944***
	(0.388)	(0.257)	(0.324)	(0.277)
Country of birth				
Other Oceania	-0.342	0.595**	-0.441	0.558**
	(0.334)	(0.289)	(0.433)	(0.232)
Europe	0.221	0.377	0.229	0.501**
	(0.376)	(0.274)	(0.255)	(0.195)
Asia	-0.280**	-0.037	-0.238**	-0.056
	(0.117)	(0.104)	(0.102)	(0.090)
Americas	-0.792**	-0.697**	-0.106	0.109
	(0.335)	(0.337)	(0.403)	(0.297)
Africa & Middle East	-0.131	-0.485*	-0.298	-0.122
	(0.469)	(0.262)	(0.242)	(0.210)
Game exposure	-0.006	-0.001	-0.005	0.006
	(0.006)	(0.007)	(0.005)	(0.006)
Lecturer C	0.099	0.071	0.246	-0.059
	(0.292)	(0.127)	(0.227)	(0.111)
Lecturer D	0.399	0.121	0.643***	0.003
	(0.297)	(0.230)	(0.229)	(0.196)
Lecturer E	0.015	0.188	0.292	-0.042
	(0.261)	(0.178)	(0.194)	(0.153)
Lecturer F		0.149		-0.011
		(0.130)		(0.113)
Constant	10.715***	7.588***	8.224***	6.893***
	(0.697)	(0.568)	(0.704)	(0.539)
Tutor dummies	✓	✓	✓	✓
N	1554	1553	2003	2007
R-sq	0.190	0.106	0.209	0.147

Notes: The classification of the country of birth follows the Standard Australian Classification of Countries, 2011, Version 2.3 (http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1269. 02011?OpenDocument). The Other Oceania group includes Oceania countries other than Australia (which is the omitted base group). Robust standard errors are reported in parentheses below the estimates. * P<0.10, ** P<0.05, *** P<0.01.

Table 8: Results on the game impact on Semester 1 scores, by age

	Week-5 mid-term		Final exam	
	<=19	>19	<=19	>19
Year 2013	0.977***	0.584	0.937***	0.935**
(Treatment group)	(0.245)	(0.524)	(0.238)	(0.449)
Male	0.100	-0.101	0.544***	0.364***
	(0.061)	(0.147)	(0.055)	(0.126)
Full-time	0.034	0.248	0.500**	0.256
	(0.172)	(0.204)	(0.195)	(0.203)
International student	-0.727***	0.160	-0.018	0.502***
	(0.107)	(0.214)	(0.092)	(0.183)
No. of retakes	-1.067***	0.337	-1.191***	-0.919***
	(0.325)	(0.384)	(0.241)	(0.286)
Country of birth				
Other Oceania	-0.146	-3.150***	0.595***	-0.642
	(0.249)	(1.191)	(0.187)	(0.717)
Europe	0.033	0.160	0.251	0.162
	(0.291)	(0.365)	(0.200)	(0.302)
Asia	-0.181**	-0.650***	0.027	-0.392*
	(0.083)	(0.248)	(0.072)	(0.212)
Americas	-0.136	-0.845*	-0.054	-0.526
	(0.374)	(0.480)	(0.248)	(0.421)
Africa & Middle East	-0.026	-0.866*	-0.116	-0.651*
	(0.257)	(0.486)	(0.209)	(0.372)
Game exposure	-0.008*	0.005	0.002	0.010
	(0.004)	(0.010)	(0.005)	(0.010)
Lecturer C	0.200	0.314	0.045	-0.157
	(0.198)	(0.429)	(0.094)	(0.201)
Lecturer D	0.611***	0.287	0.065	-0.108
	(0.199)	(0.453)	(0.168)	(0.327)
Lecturer E	0.145	0.308	0.070	0.026
	(0.172)	(0.361)	(0.133)	(0.246)
Lecturer F			0.045	0.019
			(0.095)	(0.197)
Constant	6.388***	6.421***	5.359***	5.864***
	(0.299)	(0.642)	(0.311)	(0.496)
Tutor dummies	✓	✓	✓	✓
N	2876	681	2878	682
R-sq	0.160	0.174	0.127	0.202

Notes: The classification of the country of birth follows the Standard Australian Classification of Countries, 2011, Version 2.3 (http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1269.02011?OpenDocument). The Other Oceania group includes Oceania countries other than Australia (which is the omitted base group). Robust standard errors are reported in parentheses below the estimates. * P<0.10, ** P<0.05, *** P<0.01.

Table 9: Results on game impact on Semester 1 scores, by field of study

	Week-5 mid-term			Final exam		
	Econ & Comm	STEM	Others	Econ & Comm	STEM	Others
Year 2013	0.768***	0.982*	1.364	0.887***	1.232**	0.980
(Treatment group)	(0.255)	(0.552)	(0.848)	(0.246)	(0.498)	(0.734)
Male	0.128**	0.147	0.228	0.573***	0.331**	0.436**
	(0.063)	(0.196)	(0.211)	(0.055)	(0.147)	(0.195)
Age	-0.134***	-0.105**	-0.052*	-0.108***	0.000	0.005
	(0.029)	(0.047)	(0.031)	(0.022)	(0.039)	(0.031)
Full-time	-0.028	0.321	-0.178	0.456**	-0.048	0.395
	(0.171)	(0.315)	(0.296)	(0.185)	(0.240)	(0.453)
International student	-0.447***	-0.709***	0.315	0.123	-0.162	1.077***
	(0.107)	(0.235)	(0.360)	(0.094)	(0.196)	(0.311)
No. of retakes	-0.565*	-0.136	-1.103	-1.068***	-0.669**	0.002
	(0.329)	(0.520)	(0.921)	(0.242)	(0.309)	(0.641)
Country of birth						
Other Oceania	-0.281	-0.738		0.487**	0.344	
	(0.282)	(0.706)		(0.211)	(0.348)	
Europe	0.071	0.244	0.024	0.252	1.678***	0.579
	(0.254)	(1.243)	(0.553)	(0.178)	(0.536)	(0.575)
Asia	-0.144*	-0.563***	-0.855**	0.015	-0.085	-0.491
	(0.086)	(0.206)	(0.387)	(0.075)	(0.185)	(0.316)
Americas	-0.376	-0.745	-0.751	-0.608	0.273	0.010
	(0.570)	(0.862)	(0.575)	(0.441)	(0.683)	(0.579)
Africa & Middle East	0.067	-0.821*	0.147	-0.300	0.043	-0.023
	(0.272)	(0.490)	(0.553)	(0.217)	(0.387)	(0.538)
Game exposure	-0.005	-0.007	-0.008	0.002	0.003	-0.003
	(0.004)	(0.010)	(0.015)	(0.005)	(0.011)	(0.018)
Lecturer C	0.295	-0.169	0.543	0.086	-0.100	-0.289
	(0.199)	(0.505)	(0.795)	(0.093)	(0.223)	(0.345)
Lecturer D	0.614***	0.271	0.912	0.080	-0.104	0.453
	(0.203)	(0.503)	(0.817)	(0.169)	(0.372)	(0.594)
Lecturer E	0.267	-0.359	0.711	0.104	-0.01	-0.078
	(0.174)	(0.425)	(0.727)	(0.131)	(0.307)	(0.469)
Lecturer F				0.076	0.193	0.057
				(0.094)	(0.228)	(0.374)
Constant	8.959***	7.996***	6.610***	7.451***	6.192***	3.805***
	(0.604)	(1.293)	(1.248)	(0.485)	(0.970)	(1.445)
Tutor dummies	✓	✓	✓	✓	✓	✓
N	2680	582	295	2683	581	296
R-sq	0.154	0.263	0.291	0.139	0.199	0.260

Notes: The classification of the country of birth follows the Standard Australian Classification of Countries, 2011, Version 2.3 (http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1269.02011?OpenDocument). The Other Oceania group includes Oceania countries other than Australia (which is the omitted base group). Robust standard errors are reported in parentheses below the estimates. * P<0.10, ** P<0.05, *** P<0.01.

Table 10: Balancing test between the treated and control groups, by matching method

		Week-5 mid-term						Final exam					
		One-to-one		Radius		Kernel		One-to-one		Radius		Kernel	
		% bias	p>|t|	% bias	p>|t|	% bias	p>|t|	% bias	p>|t|	% bias	p>|t|	% bias	p>|t|
Male	Unmatched	-0.1	0.974	-0.1	0.974	-0.1	0.974	0.0	0.993	0.0	0.993	0.0	0.993
	Matched	0.9	0.788	-0.8	0.813	0.0	0.991	0.9	0.788	-0.6	0.847	0.1	0.976
Age	Unmatched	-6.0	0.072	-6.0	0.072	-6.0	0.072	-6.9	0.039	-6.9	0.039	-6.9	0.039
	Matched	2.4	0.379	1.9	0.489	2.3	0.416	2.7	0.329	0.1	0.969	2.4	0.378
Full-time	Unmatched	2.2	0.515	2.2	0.515	2.2	0.515	4.2	0.207	4.2	0.207	4.2	0.207
	Matched	0.7	0.836	-1.7	0.594	-2.4	0.455	3.0	0.365	-2.2	0.482	-4.0	0.197
International student	Unmatched	-8.1	0.016	-8.1	0.016	-8.1	0.016	-8.5	0.011	-8.5	0.011	-8.5	0.011
	Matched	1.2	0.718	3.7	0.245	0.3	0.929	1.3	0.687	2.5	0.447	0.0	0.989
Country of birth													
Oceania	Unmatched	-0.2	0.944	-0.2	0.944	-0.2	0.944	-0.1	0.972	-0.1	0.972	-0.1	0.972
	Matched	1.6	0.620	-1.8	0.612	-1.0	0.768	1.6	0.620	-1.4	0.677	-0.9	0.788
Europe	Unmatched	-1.1	0.743	-1.1	0.743	-1.1	0.743	-1.3	0.690	-1.3	0.690	-1.3	0.690
	Matched	1.9	0.543	1.4	0.660	1.0	0.759	1.9	0.538	1.5	0.640	1.8	0.569
Asia	Unmatched	-8.3	0.013	-8.3	0.013	-8.3	0.013	-8.4	0.012	-8.4	0.012	-8.4	0.012
	Matched	-1.5	0.654	-1.8	0.580	-4.0	0.224	-0.7	0.835	-3.1	0.347	-4.2	0.205
Americas	Unmatched	8.5	0.011	8.5	0.011	8.5	0.011	8.7	0.010	8.7	0.010	8.7	0.010
	Matched	0.0	1.000	0.0	1.000	9.3	0.004	0.0	1.000	0.0	1.000	8.8	0.008
Africa & Middle East	Unmatched	1.8	0.584	1.8	0.584	1.8	0.584	0.3	0.935	0.3	0.935	0.3	0.935
	Matched	5.9	0.055	8.4	0.004	3.2	0.328	4.2	0.170	0.1	0.987	-0.3	0.940

Notes: One-to-one matching matches nearest neighbour, controls for identical propensity scores. Radius matching involves matching within the radius of 0.2 standard deviation of logit of propensity score (0.028 for Week 5 mid-term, 0.030 for final exam). Kernel matching is a matching by gaussian kernel. "% bias" denotes the standardised percentage bias, which is the percentage difference in the sample means between the treated and control (full or matched) groups as a percentage of the square root of the average of the sample variances in the treated and control groups. "p>|t|" represents the p-value of the t-tests for equality of means in the two samples before and after matching.

Table 11: Results from propensity score matching

Matching method	Treated	Controls	Difference	S.E.
Week-5 mid-term				
One-to-one	7.305	6.854	0.452***	0.061
Radius	7.305	6.866	0.440***	0.060
Kernel	7.305	6.866	0.440***	0.059
Final exam				
One-to-one	6.877	6.049	0.828***	0.053
Radius	6.877	6.062	0.815***	0.050
Kernel	6.877	6.072	0.805***	0.050

Notes: * P<0.10, ** P<0.05, *** P<0.01.

Table 12A: Results on the game impact on Week-9 mid-term scores in Semester 1

	Week-9 mid-term					
	(1)	(2)	(3)	*(4)*	(5)	(6)
Year 2013	-0.319***	-0.328***	-0.125	-0.158	-0.167	-0.289
(Placebo group)	(0.063)	(0.062)	(0.220)	(0.248)	(0.251)	(0.259)
Male		0.061	0.061	0.047	0.044	0.044
		(0.063)	(0.063)	(0.062)	(0.062)	(0.064)
Age		-0.125***	-0.128***	-0.141***	-0.139***	-0.176***
		(0.032)	(0.033)	(0.032)	(0.032)	(0.037)
Full-time		0.184	0.190	0.197	0.207	0.351*
		(0.159)	(0.160)	(0.158)	(0.158)	(0.182)
International student		-0.427***	-0.425***	-0.463***	-0.461***	-0.443***
		(0.106)	(0.106)	(0.103)	(0.103)	(0.157)
No. of retakes		-0.905***	-0.921***	-0.942***		-0.126
		(0.284)	(0.286)	(0.291)		(0.291)
Country of birth						
Other Oceania		0.176	0.182	0.128	0.214	0.143
		(0.212)	(0.211)	(0.197)	(0.181)	(0.174)
Europe		0.461**	0.479**	0.515**	0.515**	0.670***
		(0.231)	(0.233)	(0.238)	(0.238)	(0.234)
Asia		-0.156*	-0.151*	-0.122	-0.129	-0.006
		(0.085)	(0.085)	(0.083)	(0.083)	(0.080)
Americas		-0.511	-0.501	-0.459	-0.459	0.202
		(0.354)	(0.354)	(0.353)	(0.353)	(0.443)
Africa & Middle East		-0.028	-0.029	0.061	0.052	-0.036
		(0.254)	(0.254)	(0.246)	(0.252)	(0.278)
Game exposure			-0.002	-0.002	-0.001	-0.000
			(0.004)	(0.004)	(0.004)	(0.005)
Lecturer C			0.033	0.040	0.047	-0.051
			(0.229)	(0.228)	(0.231)	(0.240)
Lecturer D			0.265	0.169	0.172	0.098
			(0.199)	(0.201)	(0.204)	(0.211)
Lecturer E			0.167	0.176	0.178	0.017
			(0.173)	(0.176)	(0.181)	(0.186)
Lecturer F			0.243	0.163	0.165	0.035
			(0.199)	(0.199)	(0.203)	(0.210)
ATAR						0.117***
						(0.008)
Constant	6.858***	9.159***	8.988***	8.331***	8.298***	-2.149**
	(0.041)	(0.615)	(0.644)	(0.661)	(0.667)	(1.091)
Tutor dummies	×	×	×	✓	✓	✓
N	3581	3581	3581	3581	3539	2666
R-sq	0.007	0.057	0.058	0.135	0.132	0.261

Notes: The classification of the country of birth follows the Standard Australian Classification of Countries, 2011, Version 2.3 (http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1269.02011?OpenDocument). The Other Oceania group includes Oceania countries other than Australia (which is the omitted base group). Robust standard errors are reported in parentheses below the estimates. * P<0.10, ** P<0.05, *** P<0.01.

Table 12B: Results on the game impact on Week-5 mid-term scores in Semester 2

	Week-5 mid-term					
	(1)	(2)	(3)	*(4)*	(5)	(6)
Year 2013	0.260**	0.223*	0.215*	0.232	0.209	-0.190
(Placebo group)	(0.117)	(0.115)	(0.115)	(0.226)	(0.237)	(0.256)
Male		-0.045	-0.045	0.014	0.048	0.084
		(0.118)	(0.118)	(0.115)	(0.122)	(0.139)
Age		-0.088***	-0.086***	-0.086***	-0.090***	-0.007
		(0.023)	(0.023)	(0.020)	(0.021)	(0.055)
Full-time		0.036	0.051	0.016	0.013	0.073
		(0.175)	(0.175)	(0.175)	(0.178)	(0.239)
International student		-0.342*	-0.389**	-0.360**	-0.418**	-0.389
		(0.175)	(0.175)	(0.171)	(0.181)	(0.284)
No. of retakes		-0.615***	-0.628***	-0.558***		-0.259
		(0.168)	(0.167)	(0.163)		(0.204)
Country of birth						
Other Oceania		-0.583	-0.601	-0.363	-0.462	-0.296
		(0.515)	(0.497)	(0.472)	(0.502)	(0.572)
Europe		0.970***	0.977***	0.986***	1.093***	0.861*
		(0.319)	(0.320)	(0.316)	(0.325)	(0.490)
Asia		-0.463**	-0.461**	-0.431**	-0.357*	-0.194
		(0.180)	(0.180)	(0.175)	(0.184)	(0.171)
Americas		0.292	0.285	-0.028	0.019	1.997**
		(0.495)	(0.496)	(0.498)	(0.504)	(0.872)
Africa & Middle East		0.054	0.051	0.269	0.269	-0.098
		(0.517)	(0.513)	(0.494)	(0.496)	(0.544)
Lecturer B			-0.258**	-0.272**	-0.261**	-0.513***
			(0.116)	(0.114)	(0.121)	(0.134)
ATAR						0.120***
						(0.011)
Constant	6.551***	8.674***	8.787***	8.245***	8.295***	-4.045**
	(0.088)	(0.492)	(0.487)	(0.526)	(0.550)	(1.607)
Tutor dummies	×	×	×	✓	✓	✓
N	1198	1198	1198	1198	1099	671
R-sq	0.004	0.070	0.073	0.146	0.136	0.298

Notes: The classification of the country of birth follows the Standard Australian Classification of Countries, 2011, Version 2.3 (http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1269.02011?OpenDocument). The Other Oceania group includes Oceania countries other than Australia (which is the omitted base group). Robust standard errors are reported in parentheses below the estimates. * P<0.10, ** P<0.05, *** P<0.01.

Table 12C: Results on the game impact on Week-9 mid-term scores in Semester 2

	Week-9 mid-term					
	(1)	(2)	(3)	*(4)*	(5)	(6)
Year 2013	0.050	-0.004	-0.012	-0.098	-0.124	-0.176
(Placebo group)	(0.130)	(0.129)	(0.129)	(0.267)	(0.278)	(0.309)
Male		-0.076	-0.076	-0.070	-0.154	0.198
		(0.134)	(0.134)	(0.132)	(0.139)	(0.170)
Age		-0.077**	-0.076**	-0.074**	-0.074**	-0.029
		(0.032)	(0.032)	(0.031)	(0.032)	(0.065)
Full-time		0.010	0.024	-0.042	0.011	-0.085
		(0.182)	(0.182)	(0.180)	(0.181)	(0.260)
International student		-0.360**	-0.407**	-0.376**	-0.383*	-0.917**
		(0.183)	(0.184)	(0.183)	(0.197)	(0.358)
No. of retakes		-0.667***	-0.679***	-0.678***		-0.309
		(0.209)	(0.211)	(0.214)		(0.295)
Country of birth						
Other Oceania		-0.952	-0.970	-1.003	-1.023	-0.840
		(0.759)	(0.759)	(0.749)	(0.840)	(1.017)
Europe		1.569***	1.580***	1.510***	1.561***	0.338
		(0.328)	(0.330)	(0.319)	(0.335)	(0.694)
Asia		0.201	0.204	0.192	0.181	0.383**
		(0.189)	(0.189)	(0.185)	(0.200)	(0.194)
Americas		1.592***	1.587***	1.434***	1.439***	3.143**
		(0.442)	(0.443)	(0.466)	(0.472)	(1.311)
Africa & Middle East		0.131	0.130	0.018	0.017	-0.687
		(0.527)	(0.536)	(0.524)	(0.528)	(0.634)
Lecturer B			-0.247*	-0.293**	-0.316**	-0.337**
			(0.131)	(0.132)	(0.139)	(0.163)
ATAR						0.125***
						(0.014)
Constant	6.503***	8.095***	8.208***	8.305***	8.340***	-4.071**
	(0.101)	(0.669)	(0.656)	(0.702)	(0.729)	(2.013)
Tutor dummies	×	×	×	✓	✓	✓
N	1192	1192	1192	1192	1092	664
R-sq	0.000	0.036	0.039	0.08	0.076	0.240

Notes: The classification of the country of birth follows the Standard Australian Classification of Countries, 2011, Version 2.3 (http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1269.02011?OpenDocument). The Other Oceania group includes Oceania countries other than Australia (which is the omitted base group). Robust standard errors are reported in parentheses below the estimates. * P<0.10, ** P<0.05, *** P<0.01.

Table 12D: Results on the game impact on final exam scores in Semester 2

	Final exam					
	(1)	(2)	(3)	*(4)*	(5)	(6)
Year 2013	0.039	0.005	-0.002	0.274*	0.270	0.133
(Placebo group)	(0.088)	(0.087)	(0.086)	(0.155)	(0.165)	(0.179)
Male		0.318***	0.319***	0.322***	0.279***	0.346***
		(0.087)	(0.086)	(0.086)	(0.092)	(0.110)
Age		-0.052***	-0.050***	-0.046***	-0.049***	0.107**
		(0.016)	(0.016)	(0.016)	(0.017)	(0.044)
Full-time		-0.051	-0.035	-0.070	-0.066	0.087
		(0.127)	(0.127)	(0.126)	(0.128)	(0.160)
International student		0.060	0.015	-0.013	-0.083	-0.446*
		(0.124)	(0.126)	(0.127)	(0.135)	(0.249)
No. of retakes		-0.931***	-0.944***	-0.885***		-0.726***
		(0.131)	(0.131)	(0.137)		(0.151)
Country of birth						
Other Oceania		0.180	0.161	0.110	-0.131	0.620
		(0.487)	(0.481)	(0.490)	(0.602)	(0.556)
Europe		0.911***	0.920***	0.909***	1.028***	0.734
		(0.241)	(0.240)	(0.244)	(0.248)	(0.668)
Asia		-0.010	-0.008	0.007	0.064	0.169
		(0.129)	(0.129)	(0.130)	(0.139)	(0.124)
Americas		0.026	0.019	0.091	0.173	1.332*
		(0.409)	(0.417)	(0.405)	(0.407)	(0.774)
Africa & Middle East		0.153	0.150	0.139	0.139	0.026
		(0.314)	(0.317)	(0.320)	(0.320)	(0.370)
Lecturer B			-0.250***	-0.238***	-0.241***	-0.154
			(0.087)	(0.088)	(0.093)	(0.105)
ATAR						0.103***
						(0.008)
Constant	5.857***	6.769***	6.878***	6.780***	6.861***	-5.641***
	(0.068)	(0.338)	(0.333)	(0.391)	(0.415)	(1.230)
Tutor dummies	×	×	×	✓	✓	✓
N	1199	1199	1199	1199	1097	673
R-sq	0.000	0.068	0.075	0.089	0.058	0.308

Notes: The classification of the country of birth follows the Standard Australian Classification of Countries, 2011, Version 2.3 (http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1269.02011?OpenDocument). The Other Oceania group includes Oceania countries other than Australia (which is the omitted base group). Robust standard errors are reported in parentheses below the estimates. * P<0.10, ** P<0.05, *** P<0.01.

Table 12E: Results on the game impact on final exam scores in Semester 2, by question relevance

	Relevant questions						Irrelevant questions					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
Year 2013	-0.586***	-0.582***	-0.596***	-0.192	-0.258	-0.34	0.174**	0.143*	0.135	0.417***	0.413**	0.293
(Treatment group)	(0.220)	(0.220)	(0.220)	(0.371)	(0.390)	(0.462)	(0.087)	(0.086)	(0.086)	(0.153)	(0.163)	(0.180)
Male		0.521**	0.522**	0.490**	0.541**	0.381		0.296***	0.297***	0.302***	0.259***	0.337***
		(0.223)	(0.223)	(0.223)	(0.231)	(0.304)		(0.087)	(0.086)	(0.086)	(0.091)	(0.110)
Age		-0.070*	-0.066*	-0.066	-0.071*	0.161		-0.050***	-0.048***	-0.044***	-0.046***	0.114***
		(0.038)	(0.038)	(0.041)	(0.041)	(0.122)		(0.016)	(0.016)	(0.016)	(0.017)	(0.044)
Full-time		0.255	0.284	0.309	0.352	0.352		-0.045	-0.029	-0.068	-0.066	0.096
		(0.307)	(0.309)	(0.312)	(0.316)	(0.515)		(0.126)	(0.125)	(0.124)	(0.127)	(0.159)
International student		0.312	0.225	0.156	0.31	-0.944		0.048	0	-0.023	-0.115	-0.445*
		(0.320)	(0.322)	(0.321)	(0.336)	(0.635)		(0.123)	(0.124)	(0.126)	(0.134)	(0.249)
No. of retakes		-1.208***	-1.233***	-1.211***		-0.724		-0.915***	-0.929***	-0.863***		-0.731***
		(0.398)	(0.396)	(0.402)		(0.538)		(0.130)	(0.130)	(0.136)		(0.147)
Country of birth												
Other Oceania		1.405	1.371	1.405	1.263	0.702		0.191	0.172	0.112	-0.132	0.609
		(0.929)	(0.919)	(0.958)	(1.028)	(1.281)		(0.461)	(0.456)	(0.465)	(0.576)	(0.526)
Europe		0.949*	0.966*	0.996*	1.006*	0.571		0.901***	0.911***	0.897***	1.030***	0.738
		(0.569)	(0.562)	(0.553)	(0.552)	(0.926)		(0.244)	(0.242)	(0.246)	(0.250)	(0.674)
Asia		-0.281	-0.276	-0.27	-0.403	-0.118		0.02	0.022	0.037	0.103	0.182
		(0.332)	(0.331)	(0.331)	(0.346)	(0.356)		(0.127)	(0.127)	(0.128)	(0.138)	(0.124)
Americas		-1.143	-1.156	-0.976	-1.069	2.366**		0.06	0.053	0.119	0.21	1.320*
		(0.902)	(0.927)	(0.918)	(0.926)	(0.977)		(0.405)	(0.414)	(0.401)	(0.402)	(0.764)
Africa & Middle East		1.968***	1.962***	2.004***	1.951***	2.254***		0.086	0.083	0.075	0.075	-0.049
		(0.596)	(0.593)	(0.605)	(0.602)	(0.366)		(0.311)	(0.312)	(0.315)	(0.315)	(0.360)
Lecturer B			-0.469**	-0.438*	-0.419*	-0.541*			-0.260***	-0.248***	-0.259***	-0.168
			(0.219)	(0.223)	(0.231)	(0.290)			(0.086)	(0.087)	(0.092)	(0.104)
ATAR						0.153***						0.100***
						(0.023)						(0.008)
Constant	6.609***	7.514***	7.715***	7.236***	7.383***	-10.800***	5.714***	6.579***	6.690***	6.599***	6.675***	-5.638***
	(0.197)	(0.834)	(0.847)	(0.966)	(0.989)	(3.382)	(0.067)	(0.338)	(0.331)	(0.390)	(0.413)	(1.208)
Tutor dummies	×	×	×	✓	✓	✓	×	×	×	✓	✓	✓
N	1192	1192	1192	1192	1093	671	1192	1192	1192	1192	1093	671
R-sq	0.006	0.034	0.038	0.047	0.046	0.138	0.003	0.067	0.074	0.09	0.059	0.3

Notes: The "Relevant questions" ("Irrelevant questions") specifications implement specification (4) in Table 4 using as dependent variable the final exam score (out of 10) that students achieved considering only the questions that were (not) pertinent to the videogame module deployed in Week 2. Robust standard errors are reported in parentheses below the estimates. * P<0.10, ** P<0.05, *** P<0.01.

Table 13: Results from difference-in-differences analysis in Semester 1

	Week-9 vs. Week-5 mid-terms		
	(1)	(2)	(3)
Year 2013	-0.319***	-0.337***	-0.059
	(0.063)	(0.062)	(0.197)
Week-5 mid-term	-0.026	-0.028	-0.029
	(0.049)	(0.049)	(0.049)
Y2013*Wk-5 mid-term	0.792***	0.794***	0.795***
	(0.071)	(0.071)	(0.071)
Male		0.073	0.064
		(0.049)	(0.048)
Age		-0.138***	-0.144***
		(0.029)	(0.029)
Full-time		0.099	0.096
		(0.122)	(0.121)
International student		-0.483***	-0.494***
		(0.084)	(0.082)
No. of retakes		-0.758***	-0.740***
		(0.202)	(0.204)
Country of birth			
Other Oceania		-0.081	-0.109
		(0.209)	(0.189)
Europe		0.380**	0.381**
		(0.181)	(0.180)
Asia		-0.206***	-0.184***
		(0.067)	(0.065)
Americas		-0.473*	-0.450
		(0.278)	(0.283)
Africa & Middle East		-0.145	-0.079
		(0.201)	(0.196)
Game exposure			-0.004
			(0.003)
Lecturer C (Wk 5)			0.180
			(0.154)
Lecturer C (Wk 9)			-0.068
			(0.086)
Lecturer D (Wk 5)			0.357**
			(0.155)
Lecturer E (Wk 5)			0.180
			(0.136)
Constant	6.858***	9.523***	8.771***
	(0.041)	(0.544)	(0.576)
Tutor dummies	✗	✗	✓
N	7138	7138	7138
R-sq	0.022	0.084	0.133

Notes: The classification of the country of birth follows the Standard Australian Classification of Countries, 2011, Version 2.3 (http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1269.02011?OpenDocument). The Other Oceania group includes Oceania countries other than Australia (which is the omitted base group). Robust standard errors are reported in parentheses below the estimates. * P<0.10, ** P<0.05, *** P<0.01.

Table 14: Results on the game impact on other Economics courses taken and degree completed

	Number of all other Economics courses				Completed an Econ & Commerce degree			
	(1)	(2)	(3)	(*4*)	(1)	(2)	(3)	(*4*)
Year 2013	0.240*	0.202	1.126**	1.003*	0.039***	0.029**	0.115**	0.099*
(Treatment group)	(0.137)	(0.136)	(0.537)	(0.598)	(0.014)	(0.013)	(0.052)	(0.059)
Male		0.191	0.197	0.166		-0.103***	-0.104***	-0.103***
		(0.137)	(0.137)	(0.138)		(0.014)	(0.014)	(0.014)
Age		-0.318***	-0.315***	-0.325***		-0.070***	-0.069***	-0.067***
		(0.043)	(0.043)	(0.043)		(0.014)	(0.014)	(0.013)
Full-time		0.463	0.446	0.321		0.100***	0.098***	0.098***
		(0.323)	(0.322)	(0.322)		(0.032)	(0.033)	(0.033)
International Student		-0.302	-0.289	-0.288		-0.065**	-0.065**	-0.066**
		(0.197)	(0.197)	(0.199)		(0.027)	(0.027)	(0.026)
Country of birth								
Other Oceania		0.013	-0.043	-0.193		-0.008	-0.013	-0.015
		(0.671)	(0.681)	(0.675)		(0.054)	(0.055)	(0.055)
Europe		0.793	0.745	0.782		-0.001	-0.005	-0.004
		(0.633)	(0.633)	(0.634)		(0.053)	(0.053)	(0.053)
Asia		-0.248	-0.246	-0.256		-0.008	-0.009	-0.009
		(0.181)	(0.181)	(0.181)		(0.018)	(0.018)	(0.018)
Americas		-1.759***	-1.794***	-1.795***		-0.315***	-0.320***	-0.320***
		(0.377)	(0.392)	(0.428)		(0.068)	(0.069)	(0.070)
Africa & Middle East		-0.317	-0.302	-0.281		-0.104*	-0.100*	-0.108*
		(0.549)	(0.548)	(0.551)		(0.055)	(0.054)	(0.055)
Game exposure			-0.023*	-0.021*			-0.002	-0.002
			(0.013)	(0.013)			(0.001)	(0.001)
Lecturer C			0.621***	0.544**			0.055**	0.056**
			(0.238)	(0.244)			(0.022)	(0.023)
Lecturer D			0.745*	0.678			-0.005	-0.001
			(0.423)	(0.439)			(0.040)	(0.041)
Lecturer E			0.513	0.348			-0.019	-0.012
			(0.352)	(0.365)			(0.032)	(0.033)
Lecturer F			0.213	0.115			-0.042*	-0.038
			(0.239)	(0.247)			(0.023)	(0.023)
Constant	3.051***	8.594***	7.860***	7.865***	0.736***	2.036***	2.023***	2.071***
	(0.097)	(0.877)	(0.924)	(1.006)	(0.011)	(0.255)	(0.251)	(0.247)
Tutor dummies	×	×	×	✓	×	×	×	✓
N	3408	3408	3408	3408	3536	3536	3536	3536
R-sq	0.001	0.03	0.032	0.045	0.002	0.164	0.171	0.19

Notes: “Number of all other Economics courses” refers to economics courses taken after being included in the control (Semester 1 2012) or treatment (Semester 1 2013) for the subsequent eight semesters, while “Completed an Econ & Commerce degree” is an indicator denoting whether one has graduated with an Econ & Commerce degree. Robust standard errors are reported in parentheses below the estimates. * P<0.10, ** P<0.05, *** P<0.01.

Table 15: Results on the game impact on other courses scores - overall and in Econ & Commerce courses, and graduation timing

	Mean score in all other Econ & Commerce courses				Mean score in all other courses				Graduated 'in time'			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Year 2013	0.083**	0.067*	-0.349*	-0.209	0.106***	0.076**	-0.235	-0.107	0.012	0.024*	-0.042	-0.052
(Treatment group)	(0.041)	(0.040)	(0.180)	(0.203)	(0.037)	(0.036)	(0.160)	(0.179)	(0.014)	(0.013)	(0.051)	(0.057)
Male		-0.003	0.000	0.006		-0.043	-0.043	-0.037		-0.036***	-0.035***	-0.035***
		(0.040)	(0.040)	(0.040)		(0.036)	(0.036)	(0.036)		(0.013)	(0.013)	(0.013)
Age		-0.098***	-0.098***	-0.099***		-0.064***	-0.065***	-0.065***		-0.008*	-0.008*	-0.008*
		(0.019)	(0.020)	(0.020)		(0.019)	(0.020)	(0.019)		(0.004)	(0.004)	(0.004)
Full-time		0.202	0.171	0.173		0.297*	0.286	0.296*		0.144***	0.145***	0.148***
		(0.205)	(0.207)	(0.204)		(0.176)	(0.177)	(0.173)		(0.019)	(0.019)	(0.019)
International Student		-0.391***	-0.394***	-0.396***		-0.335***	-0.335***	-0.336***		0.321***	0.321***	0.320***
		(0.068)	(0.068)	(0.068)		(0.062)	(0.062)	(0.062)		(0.023)	(0.023)	(0.023)
Country of birth												
Other Oceania		0.143	0.154	0.195		0.065	0.077	0.118		0.080	0.080	0.085
		(0.132)	(0.131)	(0.136)		(0.136)	(0.136)	(0.138)		(0.068)	(0.069)	(0.068)
Europe		0.404**	0.415**	0.434***		0.384***	0.399***	0.422***		0.078	0.078	0.090*
		(0.171)	(0.167)	(0.167)		(0.142)	(0.140)	(0.140)		(0.051)	(0.051)	(0.050)
Asia		-0.119**	-0.119**	-0.113**		-0.169***	-0.170***	-0.169***		0.048***	0.047***	0.047***
		(0.053)	(0.053)	(0.053)		(0.048)	(0.048)	(0.048)		(0.017)	(0.017)	(0.017)
Americas		0.359***	0.352***	0.377***		0.339***	0.353***	0.350***		0.062	0.063	0.067
		(0.127)	(0.126)	(0.130)		(0.115)	(0.113)	(0.109)		(0.074)	(0.073)	(0.072)
Africa & Middle East		0.145	0.135	0.124		0.208*	0.200*	0.185		-0.044	-0.045	-0.041
		(0.156)	(0.158)	(0.157)		(0.117)	(0.116)	(0.116)		(0.050)	(0.050)	(0.050)
Game exposure			0.010**	0.009**			0.007**	0.006**			0.001	0.001
			(0.004)	(0.004)			(0.003)	(0.003)			(0.001)	(0.001)
Lecturer C			-0.152**	-0.146*			-0.145**	-0.147**			-0.018	-0.016
			(0.075)	(0.077)			(0.067)	(0.069)			(0.023)	(0.023)
Lecturer D			-0.179	-0.189			-0.122	-0.137			-0.061	-0.052
			(0.135)	(0.141)			(0.121)	(0.127)			(0.040)	(0.041)
Lecturer E			-0.138	-0.155			-0.096	-0.113			-0.027	-0.008
			(0.117)	(0.123)			(0.105)	(0.110)			(0.034)	(0.034)
Lecturer F			-0.062	-0.061			-0.019	-0.022			-0.014	-0.007
			(0.076)	(0.078)			(0.068)	(0.070)			(0.023)	(0.023)
Constant	6.695***	8.427***	8.625***	8.823***	6.774***	7.850***	7.995***	8.106***	0.206***	0.137*	0.174*	0.111
	(0.029)	(0.356)	(0.371)	(0.391)	(0.026)	(0.347)	(0.359)	(0.372)	(0.010)	(0.080)	(0.086)	(0.092)
Tutor dummies	✗	✗	✗	✓	✗	✗	✗	✓	✗	✗	✗	✓
N	3109	3109	3109	3109	3401	3401	3401	3401	3478	3478	3478	3478
R-sq	0.001	0.066	0.07	0.085	0.002	0.065	0.068	0.082	0	0.142	0.143	0.156

Notes: "Mean score in all courses" denotes the average mark achieved in all courses taken after being included in the control (Semester 1 2012) or treatment (Semester 1 2013) for the subsequent eight semesters. Graduated 'in time' captures whether one graduated within the specified duration of her degree; specifications (9)-(10) use a narrower sample that further excludes those for whom the relevant semester is not their first one. Robust standard errors are reported in parentheses below the estimates. * P<0.10, ** P<0.05, *** P<0.01.

A Appendix

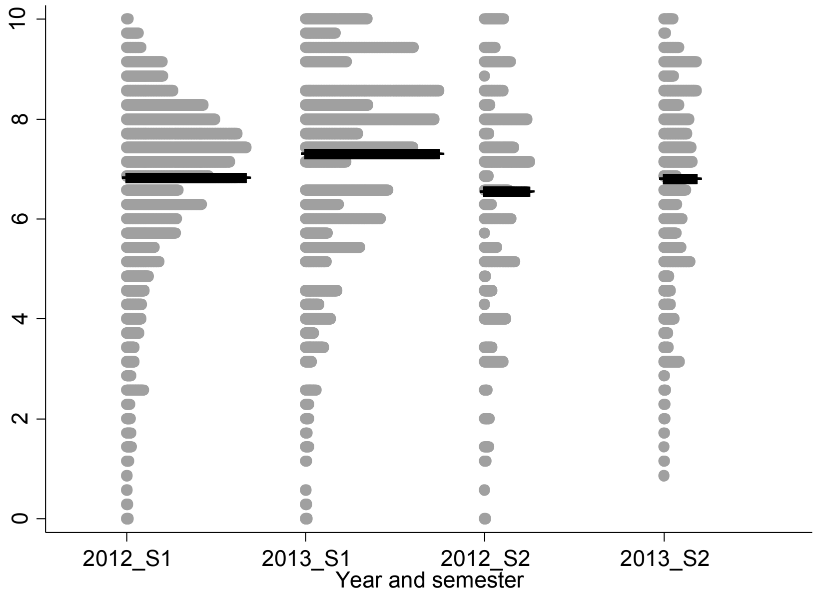

Figure 1A. : Distribution and mean of scores for the Week-5 mid-term

Notes: The actual maximum scores for the Week-5 mid-term in Semester 1 2013 and Semester 2 2013 are 15 and 20, respectively. The scores for these two semesters are converted to 0 – 10 scale for easy comparison.

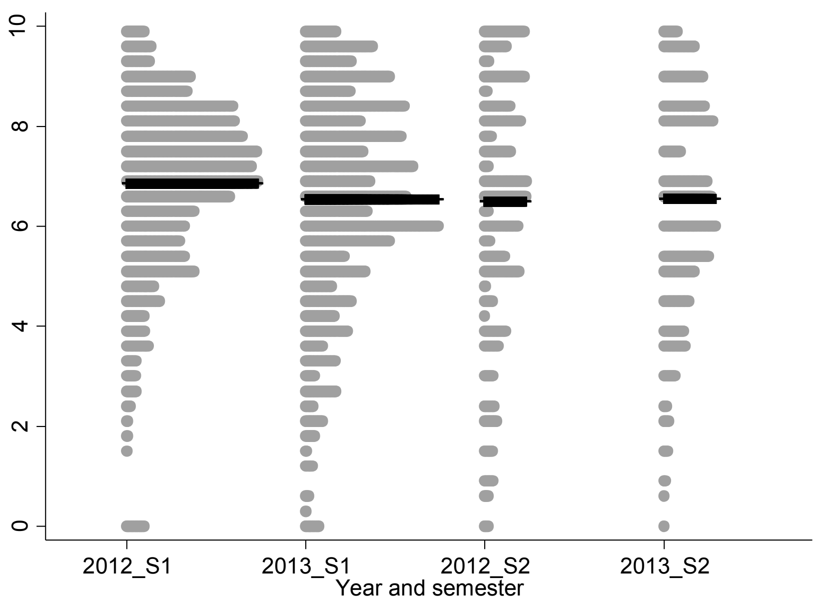

Figure 1B. : Distribution and mean of scores for the Week-9 mid-term

Notes: The actual maximum scores for the Week-9 mid-term in all the semesters are 20. The scores for these tests are converted to 0 – 10 scale for easy comparison.

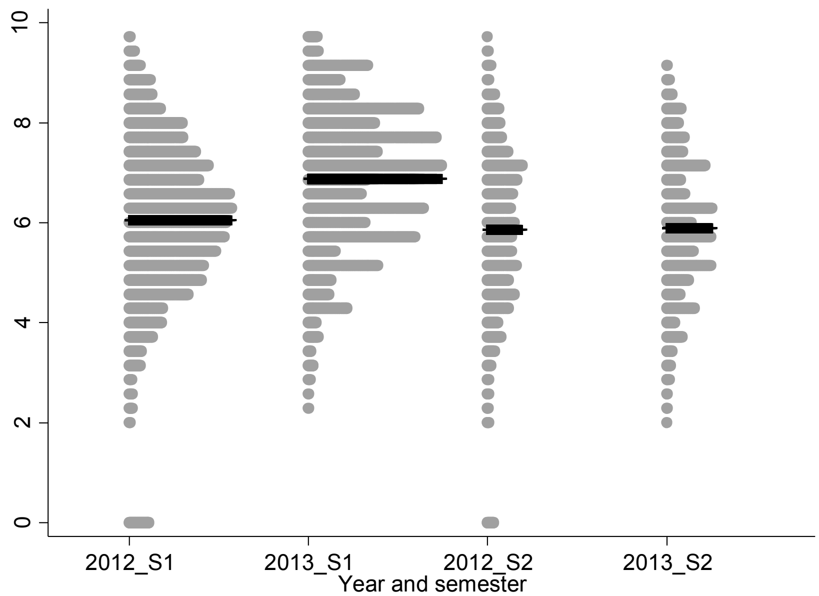

Figure 1C. : Distribution and mean of scores for the final exam

Notes: The actual maximum scores for the final exam in 2012 and 2013 are 65 and 50, respectively. The scores for these tests are converted to 0 – 10 scale for easy comparison.

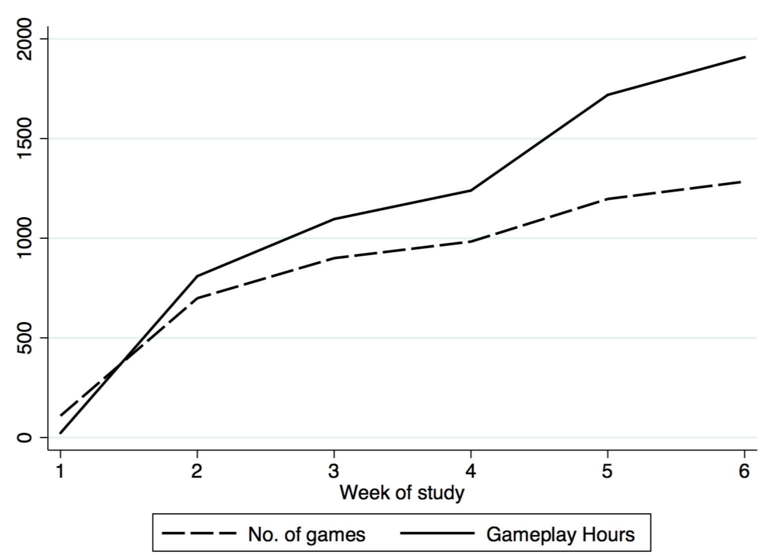

Figure 1D. : Aggregate game usage statistics

Notes: Data was provided via the EApps server established to allow students to access and play the game online. No identifiers were collected due to privacy rules. One complete play-through of the video game module was designed to take approcimately1 hour if no mistakes were made.

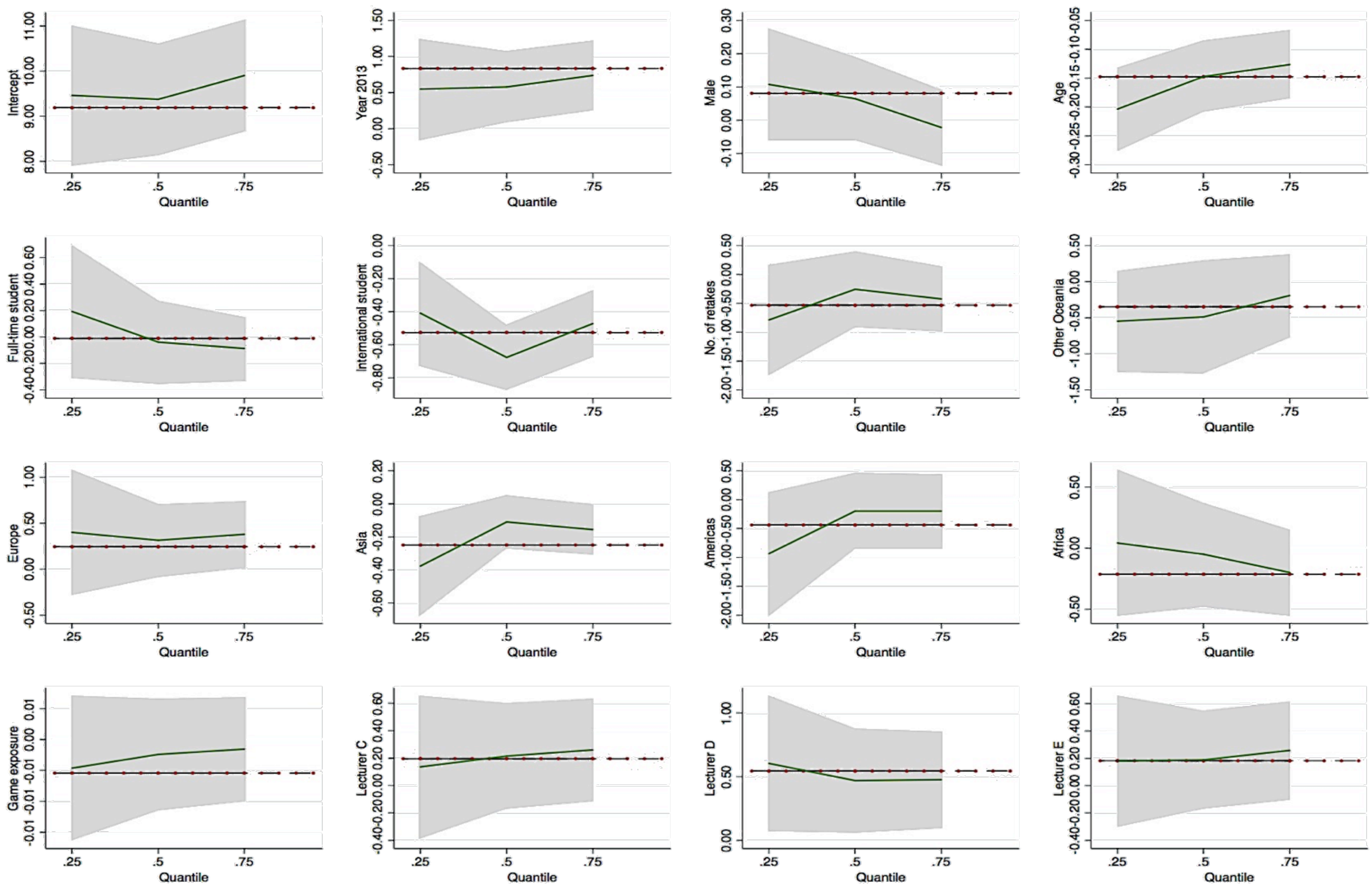

Figure 2A. : Quantile regression coefficients for Week-5 mid-term scores

Notes: Each panel shows a specific quantile regression coefficient with the corresponding confidence interval, with the actual coefficient represented by the horizontal black line.

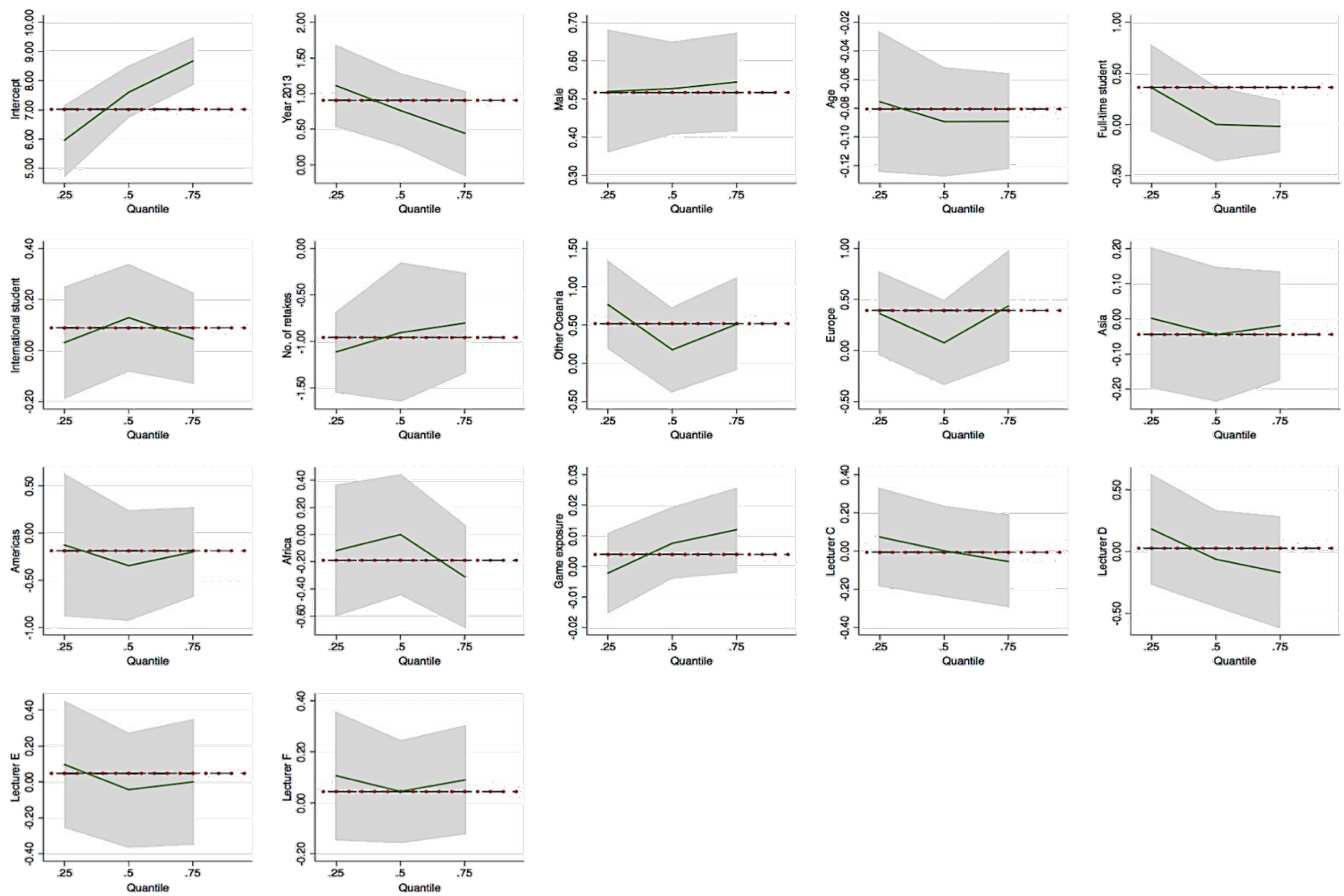

Figure 2B. : Quantile regression coefficients for the final exam scores

Notes: Each panel shows a specific quantile regression coefficient with the corresponding confidence interval, with the actual coefficient represented by the horizontal black line.

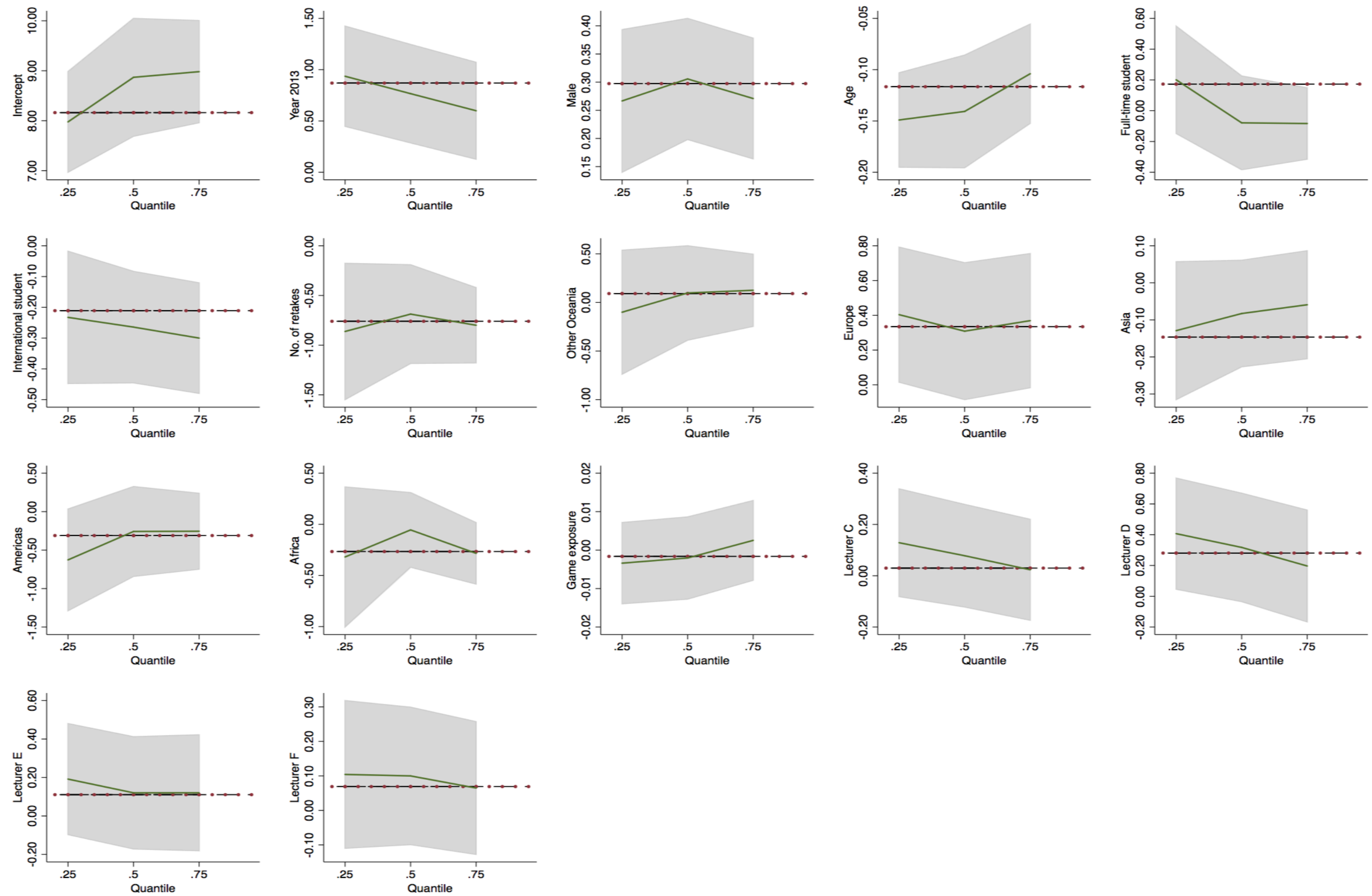

Figure 2C. : Quantile regression coefficients for Joint (Week-5 & final exam) scores

Notes: Each panel shows a specific quantile regression coefficient with the corresponding confidence interval, with the actual coefficient represented by the horizontal black line.