

Luehrmann, Melanie; Chevalier, Arnaud; Dolton, Peter

**Conference Paper**

## "Making it count": Evidence from a Field Experiment on Assessment Rules, Study Incentives and Student Performance

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Effort Compensation, No. C07-V3

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Luehrmann, Melanie; Chevalier, Arnaud; Dolton, Peter (2013) : "Making it count": Evidence from a Field Experiment on Assessment Rules, Study Incentives and Student Performance, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Effort Compensation, No. C07-V3, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/79795>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# “Making it count”

## Evidence from a Field Study on Assessment Rules, Study Incentives and Student Performance

Arnaud Chevalier

(Royal Holloway, University of London, CEE, LSE, SFI and IZA)

Peter Dolton

(University of Sussex and CEP, LSE and IZA)

Melanie Lührmann

(Royal Holloway, University of London, IFS and MEA)

### Abstract

Improving students' performance in (higher) education has become increasingly important in the agenda of policy-makers and academic institutions. Yet, little is known about students' production function. This paper examines a quasi-experiment in which we encourage students to study continuously by setting weekly incentives to engage in online tests (quizzes). Our identification strategy exploits i) weekly variation in incentives for students to determine their impact on student effort, and ii) cross-cohort experimental variation to determine their performance effects. We find that assessment incentives strongly encourage quiz participation but that tournament incentives have discouraging effects on effort. Overall, incentives result in an increase in quiz participation - by up to 50% over the year - through the resulting increase in continuous studying. The average grade improvement amounts to 0.25 of a standard deviation (or around 5%), an order of magnitude comparable to the impact of costlier financial incentives. Assessment incentives are most effective for students below median ability, and thus help increase pass rates and reduce grade gaps.

**JEL Code:** I23, D20

**Keywords:** Incentive, Feedback, Effort, Higher Education

**Acknowledgement:** We wish to acknowledge the help of the administration staff at RHUL and thank Eleftherios Giovanis for research assistance. We are grateful to Macke Raymond, Ed Leuven, Hessel Osterbeek, Su Dynarski, John Bound, Santiago Oliveros and seminar participants at CESifo 2012, RES 2013, Sussex, Michigan, City University, Nuremberg, SOFI Stockholm, and NIESR for comments. This project was partially financed by the RHUL Faculty Initiative Fund.

**Address for Correspondence:** Arnaud Chevalier, Royal Holloway, University of London, Email: [arnaud.chevalier@rhul.ac.uk](mailto:arnaud.chevalier@rhul.ac.uk)

## 1. Introduction

Improving students' performance in education has been a long-standing goal of researchers and stakeholders alike (see reviews by Hanushek 2006 or Bishop 2006). In higher education, a large fraction of students fails to pass courses. This may reflect a lack of effort due to uncertainty about its return, high discounting of the future, subjective ability (mis)perceptions or mis-information about one's study progress due to lack of feedback. Since intrinsic motivation does not suffice to induce satisfactory average student performance, we examine whether incentives help increase students' effort and performance?<sup>1</sup>

Taking a lead from personnel economics (e.g. Lazear, 2000), many studies have focused on financial incentives. Leuven et al (2010) find that substantial rewards (up to €681) for passing all first year exams *increase* achievement among high ability, but *decrease* it among low ability students, potentially due to the crowding out of intrinsic motivation<sup>2</sup>. Angrist et al. (2010) conclude from a series of randomized trials - Angrist et al. (2009) and (2010), Cha and Patel (2010), MacDonald et al. (2009), and Barrow et al. (2010)- that there is an "emerging picture of mostly modest effects for cash award programs of this type at the post-secondary level" (p.1). These limited effects may result from crowding out between financial reward and intrinsic motivation (Frey and Jegen, 2002) or a mismatch between achievement targets and students' ability (Camerer and Hogarth, 1999). In addition, the incentives provided in these experiments (up to 5,000 US\$ per head) may be difficult to scale up considering limited higher education resources. We thus concentrate on (low cost) non-financial incentives<sup>3</sup>.

---

<sup>1</sup> This assumes that effort is an important input into the education production function. Metcalf et al. (2012) show a deterioration of pupils' achievement in years with a greater demand for leisure (i.e., a major sports event during exam time), especially for low ability boys.

<sup>2</sup> Garibaldi et al. (2012) and Scott-Clayton (2011) also show that financial incentives raise effort using regression discontinuity designs.

<sup>3</sup> Dolton et al. (1994) and Angrist et al. (2009) examine support services, and find small knowledge and performance effects. Beltz et al. (2012) find detrimental effects of belated reward and lax resitting constraints.

We conduct a controlled field study in two courses among first year undergraduate economics students at a large college of the University of London, consisting of variations in the incentives to participate in weekly online quizzes. These quizzes administered through a web platform are used as a measure of students' effort. The quasi-experimental setup allows us to pursue two lines of investigation: firstly, we compare quiz participation, our measure of student effort, within student across term weeks and courses, in which they face incentives to engage in continuous learning which vary across weeks: a) the provision of additional study material conditional on quiz participation, b) the awarding of a prize for the best quiz performance, c) declaring a quiz as compulsory coursework, and, in the second year, d) weighting quiz performance towards the course grade. This allow us to test the relative efficiency of different types of incentives. All incentives can be scaled-up easily: they are implementable in large classes and are low cost– in contrast to many financial incentives<sup>4</sup>.

Secondly, we change the assessment structure between two academic years; i.e. we focus particularly on assessment weighting, thus setting stronger incentives to participate in quizzes for the second cohort. By comparing the two cohorts we can assess whether and by how much quiz participation (effort) improves grade performance. Not only do we assess the impact of incentives on effort, but also on grade performance.

Our contribution to the literature is threefold. First, we provide empirical evidence on the effectiveness of several incentive mechanisms in inducing effort. Pozo and Stull (2006) investigate the combined effect of additional math courses and assessment incentives for first year economics students<sup>5</sup> and find an average 2% achievement gain among students in the treatment group with larger gains among the weakest students. However, the interpretation of these achievement effects is unclear: they are either due to the effect of additional math

---

<sup>4</sup> Ball et al (2006) study a costlier teaching tool and finds performance effects of similar magnitude.

<sup>5</sup> The treatment group's overall course grade depends on the best result from a pre-university math exam and a second exam following a Maths course. The control group can choose to attend the same Maths course, which may enhance their performance in the economics courses, but which will not affect their course grade directly.

training on performance (in economics courses) or due to the assessment incentive. Grove and Wasserman (2006) evaluate the effect of assessment weighting (worth 15% in the treatment group and zero in the control group) only and report higher participation rates in completing these problem sets in the treatment group. They find a grade increase among freshmen of 2.4 to 4.6 percentage points. Our quasi-experimental setup allows the identification of students' behavioural response to various incentives. We test their effect within the same student population and include low and high stakes incentives, thus covering a broad set of non-financial incentives. Only assessment weighting incentives and course requirements have large effects – increasing weekly quiz participation between 30 and 55 percentage points. We also find heterogeneous effects across ability types and gender, in a way that is consistent with a trade-off between intrinsic and extrinsic motivation (Frey and Jegen, 2002) and gender differences in risk behaviour. Overall quiz participation increased over the full duration of the course and grades increase by 0.25 of a standard deviation on average, and up to 0.4 of a standard deviation for students of below median ability. These performance increases are in the order of magnitude of the results of large financial stakes as incentives (see Angrist et al, 2010, and Leuven et al. 2010).

Our second contribution is the opening of the 'blackbox' of the relationship between assessment incentives and student performance. We show in a simple theoretical model that assessment incentives increase the benefit of participating in continuous learning through a direct (grade) reward in addition to the indirect reward of helping to prepare for the exam. However, the performance effect of these incentives will depend on two factors: (i) students' susceptibility to incentives, and (ii) the effectiveness of the incentivised assessment tool – online quizzes in our case – in preparing students for exams, and transmitting and testing knowledge. As well as testing the impact of various incentives on effort, we assess the direct effect of effort (quiz participation) on grades and its heterogeneity with respect to gender,

ability level and student origin and also include usually unobserved characteristics such as risk aversion and discount rates which are measured in a separate questionnaire.

Third, we extend the literature on the effect of feedback in higher education. Our quiz technology has the added benefit of reducing students' uncertainty about their ability and the return to effort by providing regular feedback to students. Bandiera et al. (2012) report that feedback increases test score by about 1%.<sup>6</sup> Moreover, the probability of obtaining a first class grade on the post-feedback essay increases by 4 percentage points (from a base line of 17%), highlighting the larger performance impact of feedback provision on more able students<sup>7</sup>. While our study does not focus on feedback effects *per se*, we provide feedback, conditional on having completed the quiz. We find that in the absence of incentives, many students – especially males and low ability students- are not willing to exert effort to obtain feedback. Hence, the feedback effects found in the literature may be conditional on participation. Reversely, combining assessment incentives with a feedback mechanism may result in even larger performance increases than the effects identified here.

In the remainder of the paper, we develop a simple model of students' effort choices conditional on their ability, which lends a structural interpretation to our reduced form results (Section 2). Section 3 describes the design of our quasi-experiment and the data. Section 4 presents empirical results. Section 5 concludes.

---

<sup>6</sup> Rather than giving feedback on performance, Evensky et al. (2007) study the effect of weekly feedback on (self-reported) cumulative study effort in absolute terms and relative to the class average. They find a small grade increase of 0.18 percentage points.

<sup>7</sup> Azmat and Iriberry (2011) conduct their experiment in schools rather than in higher education. They find that the provision of relative feedback induced a 5% increase in students' grades, and no heterogeneity in the treatment.

## 2. A Simple Economic Model

We propose a simple inter-temporal model of study behaviour in which students wish to maximise their utility from leisure and course performance, i.e. an exam at the end of term or final overall course grade. Our model is similar to the one used by Bandiera et al (2012), but focuses on changes in assessment rules rather than on feedback.

Our model has two periods. In period 1, i.e. term time, lecturing and other study activities, including the online quizzes, take place. Period 2 is the pre-exam time during which students can study for the exam. Students derive utility from their overall course grade, which is revealed at the end of period 2. Students choose their effort level maximising the following utility function:

$$U = u_1(l_1) + \beta * u_2(g(e_1, e_2, a), l_2) \quad (1)$$

where utility in period 1 depends only on their choice of leisure time  $l_1$  and utility in period 2 depends on their course grade  $g$  and  $l_2$ .  $\beta < 1$  is the discount factor. In each period  $t$ , students choose an effort level  $e_t \in [0,1]$ , measured in time units. If they choose an effort of zero, their entire time endowment ( $T=1$ ) is spent enjoying leisure. At the maximum effort level of 1, no time is left for other activities. In both periods, students face the same time constraint  $l_t + e_t = 1$ .

The overall grade production function  $g$  is a weighted average of grades in both periods, and depends on period specific effort and time-invariant ability  $a$ :

$$g = (1 - c) * g_2(a, e_1, e_2) + c * g_1(a, e_1, e_2) \quad (2)$$

where  $c$  is the assessment weight in period 1. Exam performance  $g_t$  is a monotonously increasing function with decreasing marginal returns in effort  $e$ . Grades depend on academic ability,  $a$ , which is heterogeneous across students. For simplicity, we

assume that the grade function is linear in ability. For the moment, we also assume that at the beginning of period 1, students are endowed with ability  $a \in [0,1]$ , which is drawn from a uniform distribution between 0 and 1, and that students know their ability (This assumption shuts down any feedback effect of period-one assessments. We will discuss feedback effects later in this section). Final grades are the given by:

$$g = (1 - c) * (a * \ln(\gamma e_1 + e_2) + \varepsilon_2) + c * (a * \ln(e_1) + \varepsilon_1). \quad (3)$$

Grades are also affected by idiosyncratic iid shocks ( $\varepsilon_2$ ) with mean zero and standard deviation 1. These shocks reflect any circumstantial factors that may hinder or propel a student's performance.

In our study,  $g_1$  represents performance in the online quizzes and  $g_2$  exam performance. Our cross-cohort approach varies the assessment weighting  $c$  in period 1 from zero (control) to a positive weighting  $c > 0$  (treatment). Exerting effort (and participating in the assessments) in period 1 is costly as it crowds out leisure time. However, regardless of assessment weighting, effort  $e_1$  directly results in better grades in period 2 since exam performance  $g_2$  depends on effort in both periods, and we assume:  $\frac{\partial g_2}{\partial e_1} > 0$ . The effectiveness of period 1 effort on period 2 grades depends on the depreciation of knowledge,  $\gamma < 1$ , and is lower than that of effort exerted in period 2:

$$\frac{\partial g_2}{\partial e_1} = \frac{a\gamma}{\gamma e_1 + e_2} < \frac{\partial g_2}{\partial e_2} = \frac{a}{\gamma e_1 + e_2}$$

Students may prefer to exert effort in period 2 for two reasons: first, knowledge acquired at an earlier date depreciates at a rate  $\gamma$ . Secondly, students are impatient and value leisure in period 1 higher ( $\beta < 1$ ).

Thus, at the beginning of the year the maximisation problem of each student is:



$$U = u_1(l_1) + \beta * u_2(g(e_1, e_2, a), l_2) \quad \text{s.t.} \quad l_1 + e_1 = 1 \text{ and } l_2 + e_2 = 1 \quad (4)$$

With the introduction of positive assessment weights  $c$ , we expect students to shift effort from period 2 to period 1 for the following reasons:

1. The marginal grade return to period 1 effort can be written as:

$$\frac{\partial g}{\partial e_1} = \frac{(1-c)*a*\gamma}{(\gamma e_1 + e_2)} + \frac{c*a}{e_1} > 0$$

If ability  $a$  is greater than zero and students exert positive effort in period 1<sup>8</sup>, and the assessment weight  $c$  is positive, the following holds:

$$\frac{c * a * e_2}{e_1(\gamma e_1 + e_2)} > 0$$

and the marginal return to effort exerted in period 1 will be greater in a course with positive assessment weight  $c$  relative to the same course with zero assessment weighting.

2. At the same time, the marginal benefit of effort in period 2 decreases due to the lower weight of the final exam in the overall course grade  $(1-c) < 1$ :  $\frac{\partial g}{\partial e_2} = \frac{(1-c)*a}{\gamma e_1 + e_2} < \frac{a}{\gamma e_1 + e_2}$ .

3. In the control group, the overall grade depends only on the realisation of the final exam shock  $\varepsilon_2$  with mean zero and variance one. In the treatment, the overall grade depends on the exam shocks in the continuous online assessments and in the final exam. If the two shocks are independent and have a variance of 1, the variance of the overall shock

$$\text{Var}(c * \varepsilon_1 + (1 - c) * \varepsilon_2) = c^2 * \text{Var}(\varepsilon_1) + (1 - c)^2 * \text{Var}(\varepsilon_2) = 2c(c - 1) + 1 < 1$$

is lower than one, the variance of the one-time shock with weighting 1 in year 1.

In our study,  $g_1$  consists of multiple assessments during the term. Hence, the variance of the iid shocks is lower in period 1 than in period 2 (where only one exam takes place). By participating in all assessments, students can thus substantially reduce the variance of shocks

---

<sup>8</sup> We also plausibly assume that marginal (grade) returns to effort are decreasing.

affecting the overall grade:  $\text{Var}(\hat{\varepsilon}_1 + \hat{\varepsilon}_2) < \text{Var}(\hat{\varepsilon}_1) < \text{Var}(\hat{\varepsilon}_2)$ . Risk averse students should thus increase their effort in period 1 when we introduce positive assessment weights.

An additional benefit of period 1 assessments, that we have ruled out so far by assuming that students know their ability  $a$ , is feedback. Participation in assessments in period 1 yields feedback, which helps students to learn about their unknown ability  $a$  and to better determine how much effort is needed to reach their optimal grade  $g^*$ . This is especially relevant in situations in which students enter a new environment with unknown performance standards. It would also matter when students take new subject matters and are thus uncertain about their subject-specific ability. Bandiera et al. (2012) provide a detailed model of feedback effects – and vary feedback exploiting different assessment rules across academic departments. Since we do not vary feedback – just the incentive to exert effort to obtain it, we only sketch the role of feedback provided in the period 1 assessments: Let us assume that students form a prior belief  $\hat{a}$  about their ability<sup>9</sup>. Since it is a noisy signal they are unsure how much effort is needed to reach their optimal grade  $g^*$ . In both years of our field study, students can buy a signal of their ability by exerting effort  $\hat{e}_1$  (in quizzes) which allows them to observe their resulting grade  $\hat{g}_1 = a * \ln(\hat{e}_1) + \hat{\varepsilon}_1$ . Due to the unobserved shock  $\hat{\varepsilon}_1$ , some uncertainty about ability and mapping of effort to grades remains. Repeated participating in period 1 assessments reduces the noisiness of the signal. If students perceive the unobserved exam shock to be zero, their ability posterior is:  $\tilde{a} = \frac{\hat{g}_1}{\ln(\hat{e}_1)}$ . Due to Bayesian updating, there is convergence to true ability,  $\tilde{a} \rightarrow a$ . If students decide not to participate in the assessment, they do not receive a signal and the best estimate of their true ability  $a$  remains their prior belief  $\hat{a}$ . However, feedback is costly as students need to exert effort in first period

---

<sup>9</sup> The first year undergraduate students in our experiment come from a wide variety of countries and educational systems. Since they are just starting university, it seems reasonable to assume that they take their school performance as a signal of their ability but that this is a noisy signal of their relative ability in the new environment (and the strongly selected peer group).

assessments to learn about their ability. Assessment incentives increase the benefit of such effort by providing additional grade returns which should increase the number of students deciding to obtain such feedback.<sup>10</sup>

For these reasons, we expect students to exert more effort in the first period, i.e. participate in quizzes, when positive weighting of the quizzes is introduced. Whether students exert an overall higher effort or simply shift their effort between period 2 and 1, depends on (i) the weighting of the course grade,  $c$ , (ii) the adequacy of the updating of their priors on their own ability,  $a$ , (iii) their discount factor  $\beta$ , and (iv) the presence and size of random shocks, which can be interpreted as luck in exams or as exogenous shocks which may affect the student, like illness. As such, the overall effect of assessment weighting on grades is ambiguous and an empirical matter.

The conceptual framework we have set out above is simplistic in two aspects. Firstly, we may underestimate the incentives to exert effort in the first period in our theoretical model. Effort shifting may be even more beneficial than shown if study time in period 1 is *more* rather than less productive in the final exam than period 2 effort. We assume study times in both periods to be substitutes in the production of the final grade and depreciation of knowledge over time,  $\gamma < 1$ . However, if learning benefits from repetition and each week's course material builds on last weeks' contents, then effort in both periods might be complements and  $\gamma \geq 1$ .

Secondly, we assume that students maximize their first year exam score. As first year grades do not count towards their overall degree classification and a simple pass requirement

---

<sup>10</sup> If the quality of the signal increases in the effort students exert (at the extensive or intensive margin), then the role of feedback will be different between the years due to the effort incentive described above.

is needed to progress to the second year<sup>11</sup>, this assumption need not hold empirically. If this motive is absent, we do not expect high ability types to exert (much) more effort under assessment weighting with a pass requirement, since their high ability makes failure unlikely. However, grades play an important role as signals of qualifications to employers and admissions offices, which use these grades to assess applicants for internships, student jobs and admissions to Master programmes. Types with a medium or low ability for whom the pass requirement may become binding should exert more effort to meet the threshold with this grade discontinuity, unless their ability is too low to achieve the requirement.

### **3. Study design and Data**

We conducted the study among incoming first year economic students at a large college of the University of London.<sup>12</sup> Excluding repeating students, the first cohort consisted of 206 students, the second of 242 students. We test the role of incentivised assessment on the continuous study effort of students and overall student performance. The focus is on first year undergraduate students who enter a new teaching and grading environment, may lack self-discipline to study, or heavily discount the graduation deadline. As such, they face considerable uncertainty at the beginning of their university studies and have greater need for regular feedback on their work.

The study was carried out in two courses that run over two terms each: Principles of Economics and the statistics part of Quantitative Methods for Economics<sup>13</sup>. Both courses are core requirement of the first year; i.e. students need to pass them in order to progress. As such, they are high stakes modules<sup>14</sup>. Four different faculty members teach these four one-term modules of the two courses. Both courses run in parallel and are mandatory. The material,

---

<sup>11</sup> This is not specific to the studied institution. The first year of a degree does not count towards the final degree class at most UK institutions.

<sup>12</sup> The experiment is designed to treat all students within a cohort equally, so that no ethical concerns apply.

<sup>13</sup> The math part of this module was not subject to quizzes nor to the change of assessment weighting.

<sup>14</sup> Note that, the estimated effects from our experiments may still represent a lower bound since first year in this college – like in most other British institutions – does not count towards the final degree classification.

including the quiz question database, remained identical between the two years. Each course includes weekly lectures (2 or 3 hours), a compulsory tutorial, and -in the case of statistics- a compulsory computer lab session. Apart from these two core courses, economics majors are required to complete three additional courses. These courses contained no continuous online assessment and assessment rules remained the same across both years.

In the first year, we introduced on-line quizzes for both courses, testing the concepts taught in the previous week. In both years, we informed students about quizzes in the same way and encouraged participation. To minimise cheating, quizzes had to be completed within a predefined window of one to three days following the last weekly lecture. We also randomised the ordering of response items to each question across students and students had to complete the quiz within 60 minutes once they had started it. After the closing of the time window, students were informed about their overall score, the correct and their stated answer for each question and – in some cases – detailed explanations.

We then incentivise quizzes in some weeks, and not in others. We further vary the type of incentive across weeks to assess how different incentives affect quiz participation, our proxy for effort. This ‘within’ student design allows us to account for students’ unobserved characteristics. Table 1 lists these five incentive types and Table 2 shows their timing across courses and weeks. The first treatment is a simple participation incentive that gives access to seminar exercise solutions conditional on quiz participation in the corresponding week (“Solution”).<sup>15</sup> The second incentive is a performance incentive in the form of a £20 book voucher for the best quiz performance in a week (“Voucher”). It represents a small monetary incentive, which should increase quiz participation and performance in the respective week. However, the incentive is likely to affect only students who believe to have a chance to win it.

---

<sup>15</sup> In one of the modules (macroeconomics), access to the material is conditional on achieving a quiz grade above 30% .

Our third incentive (“Compulsory”) makes one quiz per term mandatory, thus introducing a participation incentive and an (albeit limited) performance incentive.<sup>16</sup>

In Year 2, we repeat this incentive structure and add assessment as a stronger incentive tool: three to four out of the total number of weekly online quizzes contribute 2.5% (“Assessed 2.5%”) or 5% (“Assessed 5%”) each towards the final course grade, for a total of 20% (Principles in Economics) and 10% (Quantitative Methods)<sup>17</sup>. If this small assessment weight is salient enough, then we expect an increase in student participation and performance in weeks in which assessment weights apply.

In both years, we collect information on students’ weekly quiz participation. The data also includes administrative data on seminar attendance and students’ characteristics. Furthermore, we have information on students’ performance in other first-year modules and conduct a survey in the first week of the Autumn term.

Our key variables of interest are the students’ ability, effort and performance. We approximate ability using university entry scores that are computed from students’ school completion grades<sup>18</sup>. Effort is largely unobserved and difficult to measure. In this study, we observe three different measures of effort: quiz participation, quiz grades and the time students require to complete an online quiz, which are observed weekly. Furthermore, we ask students about their self-study time and lecture attendance. While this gives us a quite

---

<sup>16</sup> In the Statistics module, the compulsory quiz was conducted in-class as a paper-and-pencil test in week 11. In the Principles of Economics course, we informed students that admission to final exam was conditional on having completed at least three out of four formative assessments: these were two essays and the two “compulsory” on-line quizzes (one per term). However, this threat could not be strictly enforced by college rule.

<sup>17</sup> The lower overall weighting in Quantitative Methods is due to the experimental design: we use quizzes in the statistics part, but not in the mathematics part.

<sup>18</sup> For international students, we use a combination of the academic equivalencies scales published by the University of Brighton ([www.brighton.ac.uk/international/equivalencies](http://www.brighton.ac.uk/international/equivalencies)) and the scales used by the admissions office of the College in which we conduct the experiment. We deviate in the valuation of the international baccalaureat as the equivalence scales seem to conservative given the high quality of this school degree programme.

As a robustness check, we alternatively used standardised scores on other first year courses that are not part of our experiment – with very similar results. We are presenting results based on pre-university scores as they are not subject to substitution of effort within the first year and will not be affected by our experiment.

complete picture of effort spent in different course-related activities, only quiz participation is available for all students, every week, all other measures are conditional on quiz participation.<sup>19</sup> Hence, we rely on this measure as our proxy for effort.<sup>20</sup>

Measurement of course grades, i.e. student performance, is key in the second part of the analysis, in which we want to identify the effect of assessment incentives on student performance. Since this identification strategy relies on cross-cohort differences, we needed to develop an examination tool that is identical across cohorts<sup>21</sup> *and* complies with student assessment requirements in higher education institutions. The latter prohibit the exact replication of exams across academic years. If these two conditions in measurement are not met, a grade change across years may reflect variation in the difficulty of the exam or the marking process rather than the impact of incentives. Hence, we rely on marks from term-time exams which take the form of multiple-choice exam with questions that are randomly drawn from a large database of test questions which remains unchanged across years. Across students, the exam is thus in expectation identical across years. Since this exam type was used only in Principles of Economics, we restrict our discussion of performance effects to this course. The exam is a compulsory (online) exam with a participation rate of over 90%. For each student, we observe two such interim grades, one per term. Marking is automated and thus not subject to any bias (Hawthorne effect). To minimise cheating, we restricted (i) the test duration to 60 minutes once started, (ii) the overall time window for the test to 24 hours, (iii) questions are drawn randomly for each student and (iv) correct answers are only

---

<sup>19</sup> We observe increases of all of these measures of effort in incentivised weeks. We do not report them due to the composition bias resulting from the participation constraint. If only high ability students participate when no incentive is in place, the experimental evidence will under-estimate the effect of incentives on duration of quiz participation and performance. Alternatively, if quiz participants tend to be weaker students with a high demand for feedback, then the experiment would give us upwards biased estimates of incentives on effort.

<sup>20</sup> Another measure that is observed for all students in all weeks is seminar attendance which is surveyed in administrative records. However, participation is mandatory and enforced through official warning letters if students fail to show up, so this measure is unsuitable for our purposes.

<sup>21</sup> Even identical exams across years may be flawed as students in the second year may be able to get hold of the previous exam.

provided when all quizzes have been completed. If cheating was taking place, we would expect exam outcomes to be positively correlated across students who start the online quiz at a similar time. We do not find any evidence that a student's quiz grade increases in the number of students starting the quiz around the same time (see Table 3 in which we look at different time ranges: +30 sec, +1 min, +2 mins), and neither do we find any significant correlation between the starting time and one's grade<sup>22</sup>.

With the exception of introducing assessment weights, we keep all course parameters – content, timing, delivery etc.- constant between the two cohorts. Since our identification relies on differences across the two years in the second part of the analysis, we test whether the two cohorts are similar in terms of observable characteristics. Table 4 shows that the second entry cohort is about 3 months younger, has a larger fraction of males (62 relative to 52%), more economics majors, is of lower academic ability – as measured by entry grades – and contains more British students. These composition shifts were in part related to the impending reform in education financing, which substantially increased tuition fees for British and EU students the following year. To avoid the fees hike, British students rushed to enter higher education under the old system. To rebalance our sample, we use propensity score matching based on age, gender, citizenship, degree major and ability, and only keep individuals who matched<sup>23</sup>. The control individuals are then reweighted according to the matching process, and these weights are used throughout the analysis. Column 3 of Table 4 shows that the matched cohorts are well-balanced on observable characteristics and are statistically indistinguishable. We also rely on additional survey evidence<sup>24</sup> eliciting students' time preferences, risk aversion and their self-confidence to show that the samples match well

---

<sup>22</sup> These results are available from the authors upon request.

<sup>23</sup> The matched sample is obtained using an Epanechnikov kernel with bandwidth 0.01 using a program developed by Leuven and Sianesi. 16 individuals from the second cohort are not matched and are dropped from the analysis. 9 individuals from the first cohort are never used as controls are also dropped.

<sup>24</sup> We conduct these surveys in the first lecture of Principles of Economics in both years.



on those dimensions that were not part of the matching procedure (Table 4) This suggests that our matched sample is balanced not only on observables but also on (usually) unobservable characteristics and preference parameters.

Table 5 summarizes our measures of effort and performance for the matched sample across the two years by course<sup>25</sup>. Students become significantly more active in continuous learning via quizzes during the second year, as increased quiz participation and quiz grades show. The number of attempted quizzes increases by 25% (Principles) to 72% (Quantitative Methods) and quiz grades increase by 10 to 18%. Even the time spent on each quiz increases significantly in our sample in both courses. Mid –term grades increase by 3 marks.

#### **4. Results**

##### **a. The Relationship between Incentives and Student Effort**

First, we investigate the effectiveness of different incentives in inducing students to engage in continuous learning. We administer seven to ten quizzes per term and course and confront students with the five different incentive mechanisms described in the last section. We vary incentives across quiz weeks, terms and years as shown in Table 2. Figures 1A reports the weekly quiz participation rate for the two years Principles of Economics . Vertical lines mark weeks with incentivised quizzes. Over the duration of the term, we see a large amount of attrition in quiz participation: In year 1, for example, quiz participation starts at about 60% for the first three quizzes and drops to around 20% for the last weeks of the term for both courses. In the second term, participation is even lower. The particularly low quiz participation in weeks without incentives suggests that students' demand for feedback is low when obtaining such feedback requires effort. Figure 1B shows similar results for the Quantitative Methods course.

---

<sup>25</sup> \*\* in the third column depicts statistically significant differences across years at the 10% significance level.

Figures 1A and B also show that soft incentives – book voucher and provision of seminar solutions, marked by dotted vertical lines - do not appear to have much impact on participation. High-stake incentives on the contrary, such as making the quiz a compulsory part of the coursework, considerably increase participation to above 90%. In year 2, participation in weeks without external incentives is almost the same in both years, confirming that in the absence of incentives our cohorts make a very similar allocation of effort. In weeks with assessed quizzes (marked by solid and dashed vertical lines), participation is always greater than 80% and substantially higher than participation in the same weeks in year 1. These figures suggest that students react strongly to incentives in the form of assessment rules or compulsion.

To estimate the effect of incentives on weekly quiz participation ( $q$ ), we first estimate the following model using ordinary least squares:

$$q_{ict} = \alpha + \sum_z \beta_z Incentive_{zct} + \gamma a_i + \sum_k \delta_k x_{ick} + \varphi T_t + \mu_c + \varepsilon_{ict} \quad (5)$$

where the subscripts stand for individual  $i$  at time  $t$  (measured in weeks) in course  $c$ . Each individual has a set of  $k$  individual characteristics  $x$ , is endowed with ability  $a$ , and can be exposed to a set of  $z$  incentives. Additionally, we allow for time-specific fixed effects  $T$  in the form of a year and term dummy and a linear trend in week, and a course-specific fixed effect  $\mu_c$ . Individual characteristics are age in months, sex, dummies for Chinese, British or other nationality, degree subject<sup>26</sup> and ability. We further control for weeks with additional academic burdens, i.e. assignment deadlines in other courses. All standard errors (in Tables 6 to 10) are estimated using the robust Huber-White-Sandwich estimator and are clustered by student identifier. Additionally we estimate a model with individual fixed effects  $\mu_i$  to

---

<sup>26</sup> Both courses draw students from different degrees, e.g. Economics, Management, Economics and Mathematics and other combinations.

account for unobserved heterogeneity with virtually identical results.<sup>27</sup> Panel A in Table 6 reports the OLS estimates of the effect of our set of different incentives on quiz participation for both courses pooled (column 1), and separately (columns 2 to 3). All incentive effects are measured relative to student effort in non-incentivised weeks.

First, we find that participation in continuous learning varies with ability regardless of the incentives provided. The differences are large: students in the 1<sup>st</sup> quartile of the ability distribution are 8.5 percentage points less likely to participate in the quizzes than students in the 4<sup>th</sup> ability quartile.

We find that strong incentives such as compulsion or assessment weights have a positive effect on quiz participation. Attaching assessment weights of 2.5% to the quizzes boosts quiz participation by 38 percentage points. The effect is large given the relatively low weight of 2.5% per assessed quiz. Interestingly, doubling the incentive weight to 5% increases quiz participation to about 56%, i.e. only about 1.5 times, instead of doubling it. This is consistent with the idea that (i) students interpret weighting as a signal that the quiz contents cover important course material, and expect higher returns to knowledge about the covered topics relative to non-assessed contents; or (ii) some low ability (or low motivation) students will never participate in quizzes, e.g. because they think they will fail the course regardless of their effort. Declaring an online quiz as compulsory coursework has a similarly strong effect on participation (increase by 66 percentage points). Since the enforcement of such a rule is difficult and costly in practice, we expect it to work best in the first year when students do not have experience with university practices yet. In subsequent years, its impact may weaken. Note that compulsion has a weaker effect in Quantitative Methods. However,

---

<sup>27</sup> The results are available from the authors upon request. We report the OLS results here, as the fixed effects estimator does not allow us to look at some parameters of interest, such as that for different ability levels, as they are time-invariant.

this test in a class and was not computer based; so the differences in effect may be due to these differences in set-up.

The “Solution” incentive, which gives access to problem-set solutions conditional on quiz participation, has a negative effect on participation. The split of our results into the course-specific result in column 3 and 4 gives an indication why: the incentive has the expected positive effect in the Principles course but a strongly negative one in Quantitative Methods. For this course, the assessment period occurred during the mid-term break. It is likely that the “holiday” effect overcompensates the incentive effect for this course. The small positive effect of the incentive in the Principles course – a 3 percentage point increase in quiz participation - is not surprising as problem-set solutions can be shared among fellow students, so it is a low stake incentive.

Only one incentive has a consistently negative impact on quiz participation - the promise of a prize reward in the form of a £20 book voucher for the best quiz performance. It reduces participation by about 8.2 to 10 percentage points. Since only high ability students are likely to win the voucher with a salient probability, a prize is unlikely to increase mean participation and may crowd out intrinsic motivation.

Neither controlling for ability nor accounting for unobserved heterogeneity alters the effect of the incentives on quiz participation. We also do not find evidence of cohort effects, confirming that cohorts are balanced and do not systematically differ in their intrinsic motivation to participate in quizzes.

So far, we have shown that there are strong and statistically significant differences in quiz participation, our measure of effort, between incentivised and non-incentivised weeks. However, when faced with a mixed schedule of incentivised and non-incentivised quizzes,

students may simply shift effort between weeks rather than increasing effort overall.<sup>28</sup> In consequence, there may be displacement effects which lead us to overestimate the impact of incentives on student effort. We investigate displacement effects in Panel B of Table 6, by estimating the effect of incentives on participation relative to participation in the first week of term in which students were not informed about the nature of future quizzes yet. As Table 2 shows, quizzes started a week earlier in the Principles course than in Quantitative Methods. In the following week, students were informed about the quiz schedule and information on the varying nature of quizzes in Principles. Given the similarities in the delivery of the two courses, students may have anticipated the same to apply in Quantitative Methods. Hence, while we can use the first quiz as a control for Principles<sup>29</sup>, this measure may be flawed for Quantitative Methods due to the 1-week time lag in the quizzes. In the following, we therefore report results only for the Principles course.

The estimates on the effect of incentives on quiz participation in Panel B are very similar to those in Panel A, and participation in non-incentivised weeks does not significantly differ between cohorts.<sup>30</sup> We conclude that there is no evidence of displacement effects of the incentives on participation in non-incentivised weeks. This is confirmed in Table 7: the total number of completed quizzes increases significantly when assessment weighting is introduced (column1).

We now relax the assumption of a common treatment effect and allow for heterogeneous impacts of incentives across students with respect to gender, ability and

---

<sup>28</sup> Even a shift in effort from later to earlier weeks may have beneficial effects on overall performance as these courses have a modular structure in which topics covered in later weeks cover contents introduced previously. Shifting study effort towards the first half of the term may enable students to follow lectures and seminar better throughout the course.

<sup>29</sup> Indeed, participation in the first week quiz does not differ between the two cohorts for this course.

<sup>30</sup> As further robustness check of our identification strategy, we investigate, but find no evidence of any cross-cohort differences in quiz participation in the first week.

nationality<sup>31</sup>. We add interaction effects of the incentives with student characteristics and estimate the following model:

$$y_{ict} = \alpha + \sum_z \beta_z Incentive_{zct} + \sum_{zk} \beta_{zk} Incentive_{zct} * x_{ick} + \varphi T_t + \mu_i + \varepsilon_{ict} \quad (6)$$

Our previous results have shown that there is an 8.5 percentage point gap in quiz participation between students in the bottom and top quartiles. We find evidence of a stronger effort impact of all incentives for students in the lowest ability quartile but no heterogeneity in the impact of incentives for the other ability quartiles (Table 8). These are strongest (and of virtually equal size) for the two assessment incentives and compulsion. All of these incentives increase participation by an additional 25 to 26 percentage points for students in the bottom relative to students in the top ability quartile. Assessment weighting is the only incentive type which has a stronger effect on students in the second ability quartile. Hence, students with below median ability are more receptive to assessment weighting than other students. The salience of the weight however has little impact on participation. In column 2 of Table 7, we also allowed for heterogeneous incentive effects by ability and find that total effort in quizzes increases most strongly in the lowest ability group. Strong incentives thus narrow the participation gap between students of different ability.

Male students are more receptive to assessment weighting than female students and somewhat more receptive to the tournament situation created by the book voucher (see Croson and Gneezy (2009) for similar findings of gender behaviour in tournament situations). Finally, non-Chinese foreign students react much stronger to assessment weighting than British and Chinese students. Their participation increases by 74 percentage points compared to between about 40 percentage points among Chinese students and 60 percentage points among British students. Finally, we investigate heterogeneous treatment effects with respect

---

<sup>31</sup> We find no heterogeneity in the reaction to incentives between economics and non-economics majors, and thus do not report them.

to the discount rate and risk aversion: we find no evidence that incentive have a differential effect by discount rate, while we do find a stronger reaction to the assessment incentives and to compulsion among less risk-averse students.<sup>32</sup>

All results show that low-cost incentives, particularly those based on assessment rules, strongly affect continuous learning effort by students, as measured by quiz participation. We further show that there are no or only small displacement effects, so that overall effort in quizzes increases. Figure 3A illustrates the shift in the distribution of the overall number of completed quizzes to the right. Our findings are consistent with our model predictions: the introduction of assessment incentives will result in an increase in period 1 (i.e. term time) effort. These effects are particularly strong among students of below median ability whose participation is significantly lower in the absence of quizzes. Figure 3B illustrates this and also shows the higher responsiveness of lower ability students to incentivised quizzes.

### **b. The Impact of Assessment Rules on Student Performance**

If students exert additional effort by increasing quiz participation, when given external incentives in the form of assessment weights, does student performance increase? If our incentives only lead to intertemporal substitution of effort, there may be no improvement in performance. To investigate performance changes, we switch to an identification strategy which exploits variation in assessment incentives between cohort 1 and 2.

As discussed in Section 2, we measure performance using students' grades in the mid-term exam. It is identical in expectation between the two years, as a set of exam questions is randomly drawing from the same large question database for each student. Since this measure

---

<sup>32</sup> Results available from the authors upon request. Risk attitudes are obtained from questions on personal medical and travel insurance ownership, smoking, interest charges on your credit cards, playing the lottery and gambling, saving account ownership, jaywalking, dangerous sports activities. Discount rates are elucidated from 5 questions regarding the present values of hypothetical prices in one year time

is only available for Principles in Economics, we focus our analysis on this course in the following and drop all information relating to Quantitative Methods. For each student, we observe two interim grades, one per term. Participation on these exams does not substantially differ across waves, so there is no evidence that our results are driven by selection into the exam. Our model specification changes to:

$$S_{it} = \alpha + \beta T_t + \gamma a_i + \sum_k \delta_k x_{ikt} + \varepsilon_{it}$$

where  $S$  denotes the exam grade,  $T$  defines the treatment, i.e. exposure to assessment incentives, which is equal to 1 in year 2.

Figure 2 compares the distribution of grades, our performance measure, in Principles of Economics, across year 1 and 2 by term. In both terms, the grade distribution shifts to the right in year 2. Thus, our descriptive results point towards a grade increase during the second year of the study. The effect seems particularly pronounced among low ability students.

The first Panel of Table 9 shows OLS estimates of the incentive effect of the quizzes on student performance controlling for the same characteristics as above, including student's ability. We normalise grades in each year using z scores, i.e. computing the difference of each grade from the average grade and dividing by the standard deviation of grades. We find that assessed quizzes increase the average normalised grade by 0.25 of a standard deviation, i.e. 3.8 marks or about 5.5% (column 1 of Table 9). The effect is statistically significant at the 5% level. Next, we look at incentive effects along the grade distribution (columns 2 to 6). We find evidence of grade shifts everywhere along the grade distribution but at the very top – the 90<sup>th</sup> grade percentiles (column 6). They are concentrated particularly at the lower end (about 0.3 of a standard deviation in the first decile) and in the middle of the grade distribution where grade effects are around double those at the top of the distribution. These findings suggest that incentives help students from the lower part of the distribution and thus have the



potential to reduce failure rates. Towards the top of the grade distribution, incentive effects fade. These incentive effects are consistent with larger intrinsic motivation at the top.

Next, we look into the mechanism behind these performance effects of assessment incentives. There are two reasons why performance increases particularly in the lower half of the grade distribution: a) we have shown in Tables 7 and 8 that lower ability students increase their total effort in quizzes by more than higher ability students. As such, incentives help level the playing field by balancing effort across the ability scale; b) it may be that continuous learning via online quizzes is more effective for lower ability students.

In the second panel of Table 9, we show IV estimates of the returns to effort, measured by the number of attempted quizzes, on performance. Since we have shown in Table 8 that quiz effort does not only depend on incentives, but varies endogenously with ability and other (observed and unobserved) student characteristics, we need to use an IV estimator to pin down the grade return of participating in an additional quiz. We use the exogenous assessment weights as our instrument as it affects effort in quizzes but has no additional impact on grades. The first column of Table 9 panel B shows the validity of the instrument: once we control for the number of quizzes, there are no additional differences in performance across cohorts. In the second column, we show the first stage, confirming that the instrument correlates positively with the number of quizzes. The IV estimates in panel B show that each additional quiz increases grades by 0.16 of a standard deviation – and column 2 to 6 confirms that the effect is more pronounced in the first and second grade quartile. Quiz participation has thus a causal effect on exam performance. Panel C contains reduced form results including an interaction term of the number of quizzes and a cohort dummy. The estimates neither provide empirical evidence that a given number of quizzes are more effective with assessment incentives than without, nor do they show evidence that assessment weights coerce students into participating in quizzes but decrease the quality of effort.

In summary, we find a significant and quantitatively large effect of moderate assessment incentives on grades. Their effect is concentrated at the lower to middle part of the grade distribution, hence the provision of incentivised continuous learning tools does not only increase grades but also reduces the *variation* in grades between students, and can help raise pass rates.

## **5. Conclusions**

This paper presents the results of a field study in which we manipulate the incentives for students to participate in on-line quizzes. We find that non-financial incentives are highly effective in inducing additional effort among students. Student effort improves particularly through assessment incentives, which raise online quiz participation by up to 55 percentage points. Assessment weights need not be very high – in our study, students react strongly to incentives which count for only 2.5% of the overall course grade. Doubling this weight to 5% only increases participation by an additional 15 percentage points. We also find that the effect of assessment weights is heterogeneous as they are particularly effective among low ability students. In contrast, rewarding top performers with prizes can lead to reduced participation, consistent with inadequate targeting of such incentives to heterogeneous student ability.

Inducing students to participate in quizzes is an effective means of improving their performance: we find that each additional quiz attempted improves grades by 2.7 grade points on average (4%) (or around 0.15 of a standard deviation) with the effect again being stronger for pupils below the median. These effects of such incentives are of a comparable magnitude to studies that implemented large (and costly) financial incentives (see Angrist et al, 2010, and Leuven et al. 2010). They are also comparable to the effects of relative and absolute feedback found in Bandiera et al. (2012).

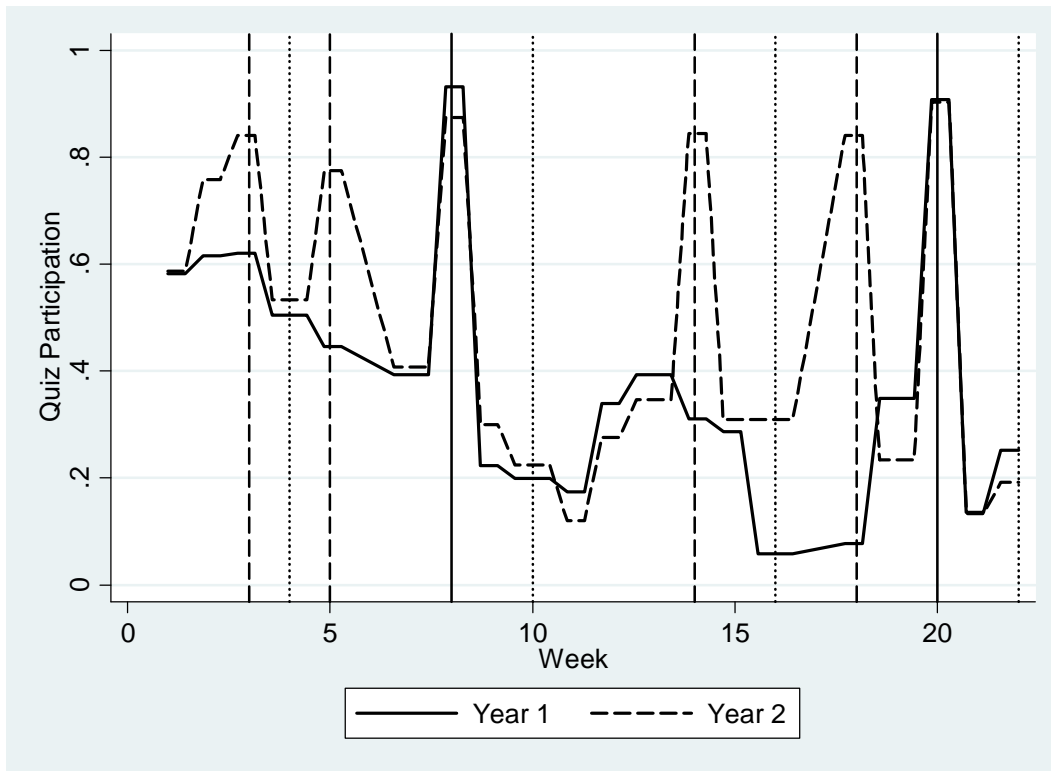
## References

- Angrist, J., Lang, D. and Oreopoulos, P. (2009) Incentives and Services for College Achievement: Evidence from a Randomized Trial, *American Economic Journal: Applied Economics* 1(1), 136-63.
- Angrist, J. and Lavy, V. (2009) The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial, *American Economic Review* 99(4), 1384-1414.
- Azmat, G. and Iriberry, N. (2010) The importance of relative feedback information: Evidence from a natural experiment using high school students, *Journal of Public Economics* 94, 435-52.
- Ball, S.B. Ekel, C. and Rojas, C. (2006) Technology Improves Learning in large principle of economics classes: Using our WITS. *American Economic Review* 96(2), 442-6.
- Bandiera, O. Larcinese, V and Rasul, I. (2012) Blissful ignorance? Evidence from a natural experiment on the effect of individual feedback on Performance. LSE mimeo.
- Beltz, P., Link, S. and Ostermaier, A. (2012) Incentives for Students: Evidence from Two Natural Experiments, Ifo Working Paper #133.
- Bishop, J. (2006) Drinking from the fountain of knowledge: Student incentive to study and learn-externalities, information problems and peer pressure. CAHRS Working Paper #04-15.
- Camerer, C. F. and Hogarth, R. M. (1999) The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework, *Journal of Risk and Uncertainty* 19(1-3), 7-42.
- Croson, R. and Gneezy, U. (2009) Gender Differences in Preferences. *Journal of Economic Literature* 47(2), 448-474
- Dolton, P.J., Klein, J.D. and Weir, I. (1994) The economic evaluation of peer counselling in facilitating computer use in higher education. *Education Economics* 2(3), 313-26.
- Evensky, J., Grove, W.A., Hu, Y. and Wasserman, T. (2007) Closing the Loop: Enhancing Collegiate Performance by Self-Assessment, mimeo.
- Frey, B and Jegen, R. (2002) Motivation crowding theory, *Journal of Economic Surveys* 15(5), 589-623.
- Grove, W. A. and Wasserman, T. (2006) Incentives and student learning: A natural experiment with economics problems sets. *American Economic Review* 96(2), 437-41.

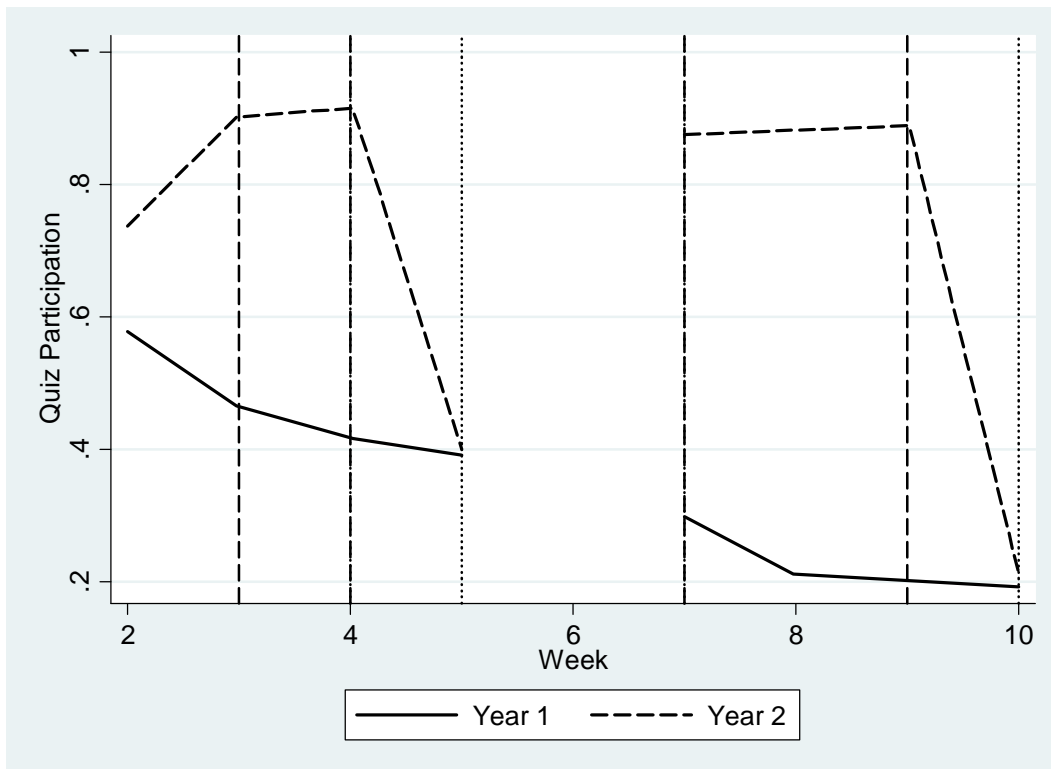
- Hanushek, E. 2006 School Resources in Handbook of the Economics of Education, (E. Hanushek and F. Welsh), Volume 2, Elsevier
- Lazear, E. P. 2000. Performance Pay and Productivity, *American Economic Review* 90(5), 1346-1361.
- Leuven, E Osterbeek, H and van der Klaauw, B. (2010) The effect of financial rewards on student's achievement: Evidence from a randomized experiment. *Journal of the European Economic Association* 8(6), 1243-65.
- Metcalfé, R., Burgess, S. and Proud, S. (2012) Using the England football team to identify the education production function: Student effort, educational attainment and the World Cup. University of Oxford, mimeo.
- Pozo, S. and Stull, C.A. (2006) Requiring a Math skill unit: Results of a randomised experiment. *American Economic Review* 96(2), 437-441.
- Stock, J.H. and Yogo, M. 2005. Testing for Weak Instruments in Linear IV Regression. In D.W.K. Andrews and J.H. Stock, eds. *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge: Cambridge University Press, 2005, pp. 80–108.

**Tables and Figures**

**Figure 1A: Weekly quiz participation – Principles of Economics**



**Figure 1B: Weekly quiz participation – Quantitative Methods**



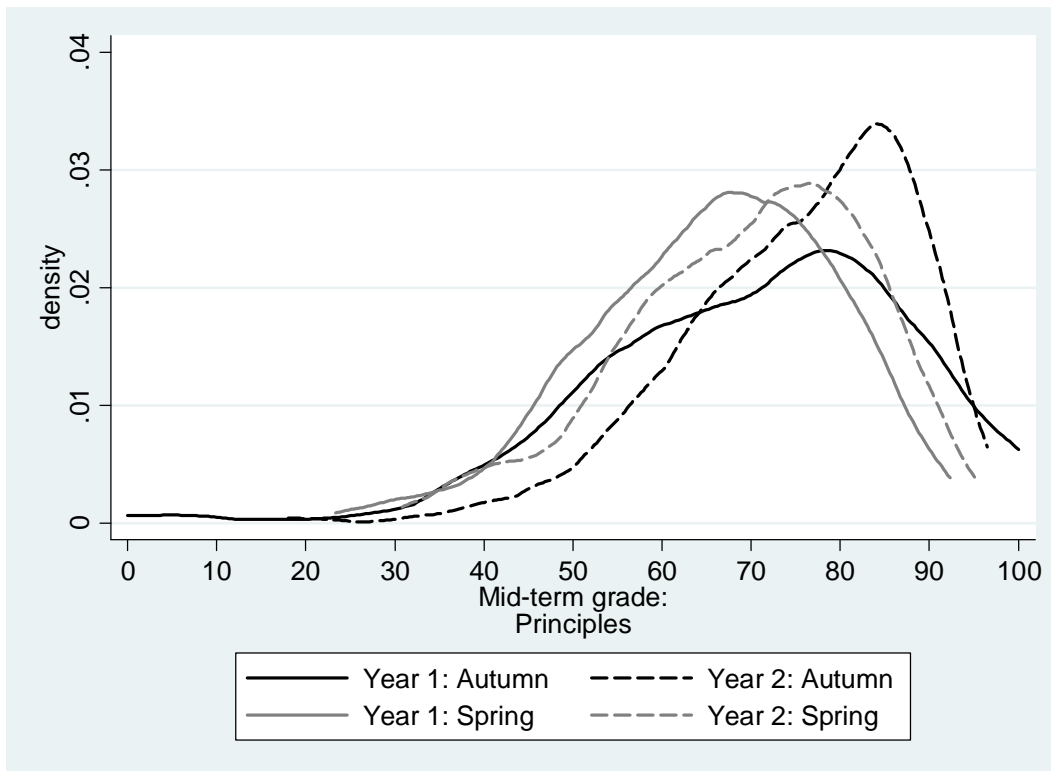
Source: Quiz – Total number of students: 206 in year 1 and 240 in the second cohort

Note: dash vertical lines refer to second cohort incentives only – Assess 2.5%

Solid vertical lines refer to year 1, compulsory incentive and second cohort – Assess 5%

Dotted vertical lines refer to soft incentives: book voucher and solution provision

**Figure 2: Distribution of Mid-Term Mark in Principles of Economics.**



**Table 1: Incentives**

Label	Incentive type	Incentive design
o	No incentive	Quiz with formative feedback
Solution	Participation only	Access to exercise sheet solutions conditional on quiz participation
Voucher	Participation & performance: monetary	£20 book voucher - prize for best quiz performance
Compulsory	Participation & (weak) performance	Quiz mandatory part of coursework
Assessed 2.5%	Participation & performance: assessment	Assessed quiz, counting 2.5% towards the overall course grade
Assessed 5%	Participation & performance: assessment	Assessed quiz, counting 5% towards the overall course grade

**Table 2: Incentive structure in online quizzes by year and course**

Week	Principles of Economics		Quantitative Methods	
	Year 1	Year 2	Year 1	Year 2
<b>Autumn</b>				
1	o	o	-	-
2	o	o	o	o
3	o	Assessed 2.5%	o	Assessed 2.5%
4	Solution	Solution	Voucher	Voucher +Assessed 2.5%
5	o	Assessed 2.5%	Solution	Solution
6	Term break		Term break	
7	o	o	Voucher	Assessed 2.5%
8	Compulsory	Assessed 5%	o	-
9	o	o	-	Assessed 2.5%
10	Voucher	Voucher	o	Voucher
11	o	o	Compulsory*	Compulsory*
<b>Spring</b>				
1	o	o	-	-
2	o	o	-	-
3	o	Assessed 2.5%	-	-
4	o	o	-	-
5	Voucher	Voucher	-	-
6	Term break		Term break	
7	o	Assessed 2.5%	-	-
8	o	o	-	-
9	Compulsory	Assessed 5%	-	-
10	o	o	-	-
11	Solution*	Solution*	-	-

Note: - designates weeks without an online quiz. A star indicates that access to solution was conditional on getting a mark of 30 or above.

**Table 3: Evidence on Cheating: Fraction of students starting exam and performance**

	Start time +/- 1mn		Start time +/- 30s		Start time +/- 2mn	
	Norm. grade	Dif in grade	Norm. grade	Dif in grade	Norm. grade	Dif in grade
% started	3.084 (3.426)	37.990 (57.62)	7.559 (5.983)	13.633 (87.182)	4.040 (2.873)	18.344 (41.302)
Cohort	0.332 (0.102)**	-1.738 (1.775)	0.316 (0.103)**	-1.703 (1.749)	0.330 (0.101)**	-1.701 (1.787)
R <sup>2</sup>	0.14	0.17	0.14	0.17	0.14	0.17

**Table 4: Descriptive statistics – Population characteristics and outcomes**

	Full sample		Ability Sample		Matched Sample	
	Year 1	Year 2	Year 1	Year 2	Year 1	Year 2
<i>Observed individual characteristics</i>						
Age (in months)	233.6 (1.152)	230.9 (0.955)*	231.2 (1.110)	228.8 (0.786)*	227.2 (0.771)	229.1 (0.842)*
Male	0.534 (0.035)	0.616 (0.031)**	0.500 (0.042)	0.610 (0.034)**	0.637 (0.035)	0.614 (0.036)
UK citizen	0.291 (0.032)	0.525 (0.033)**	0.404 (0.041)	0.600 (0.034)**	0.560 (0.036)	0.582 (0.036)
Chinese citizen	0.126 (0.023)	0.079 (0.017)*	0.116 (0.027)	0.054 (0.016)**	0.067 (0.018)	0.058 (0.017)
Econ. Major	0.481 (0.035)	0.698 (0.030)**	0.500 (0.042)	0.717 (0.032)**	0.731 (0.033)	0.698 (0.033)
Ability (University entry tariff score)			329.80 (6.970)	303.66 (4.621)**	311.99 (5.078)	306.08 (4.472)
No. Observations	206	242	146	205	137	189
<i>(Unobserved) preference parameters</i>						
Risk <sup>a</sup>	6.303 (0.122)	6.404 (0.108)	6.269 (0.143)	6.444 (0.113)	6.262 (0.121)	6.438 (0.115)
Confidence <sup>b</sup>	13.007 (0.221)	12.671 (0.211)	12.900 (0.255)	12.744 (0.232)	12.998 (0.240)	12.732 (0.240)
Lottery <sup>c</sup>	194.9 (20.4)	138.9 (16.0)**	195.0 (24.1)	131.9 (17.0)**	182.5 (19.1)	132.4 (17.1)*
Discount rate <sup>d</sup>	4.322 (0.440)	4.439 (0.431)	4.056 (0.512)	4.200 (0.464)	4.247 (0.486)	4.008 (0.476)
No. Observations	149	155	110	133	103	128

Note: \* and \*\* indicate significant mean differences between waves at the 10% and 5% level respectively. Standard errors reported in parentheses. Econ Major denotes enrolment in Economics (or Economics & Finance) programs. Matched samples obtained from kernel matching (Epanechnikov) with bandwidth (0.01).

a: Risk attitudes and discount rate are obtained from a class-room survey conducted in week 1, term 1. Risk is the sum of scores obtained from the following questions: Do you have a personal medical insurance, do you smoke, do you take travel insurance, have you incurred interest charges on your credit cards, do you play lotteries, do you have a saving account, did you play slot machines last week, do you cross at pedestrian crossing, do you do any dangerous sport. b: Confidence is the sum of scores obtained from the following questions: I feel comfortable speaking to a bank manager about loans, I enjoy challenging situations, I'm not scared of being in debt, I handle uncertainty well. c: Lottery: is the value expressed to play a lottery with 50% chance of winning a price of £1,000. d: Discount rate is elucidated from 5 questions regarding the present values of hypothetical prices in one year time



**Table 5: Changes in various margins of student effort between years, matched sample**

	Year 1	Year 2
<i>Principles in Economics</i>		
No. of quizzes attempted	8.002 (0.365)	9.904 (0.391)**
Average quiz grade (%)	59.52 (0.693)	65.20 (0.679)**
Average quiz time (min)	32.45 (0.677)	35.11 (0.738)**
Mid-term grade	70.863 (0.831)	73.37 (0.785)**
<i>Quantitative Methods</i>		
No. of quizzes attempted	2.947 (0.163)	5.058 (0.111)**
Average quiz grade (%)	57.27 (0.820)	67.08 (0.645)**
Average quiz time (min)	13.99 (0.720)	22.90 (0.669)**
Normalised average grade in other first year courses	-0.051 (0.055)	0.096 (0.051)**
No. Obs.:	137	189

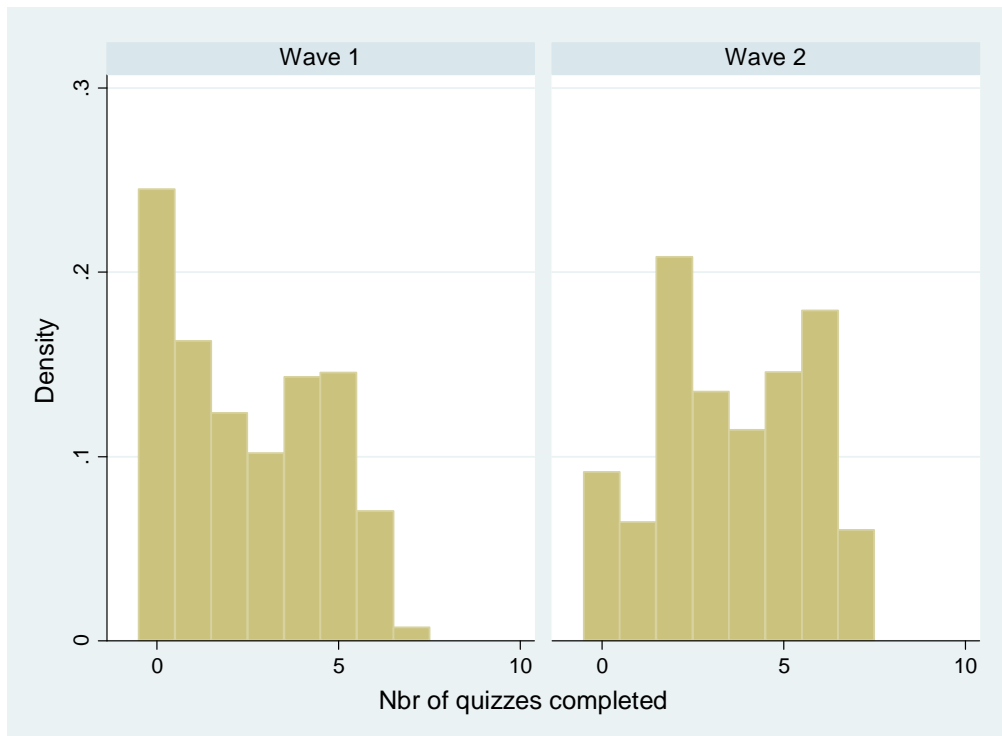
Note: \* and \*\* indicate significant mean differences between waves at the 10% and 5% level respectively. Standard error in parentheses.

**Table 6: Incentives and quiz participation, matched sample**

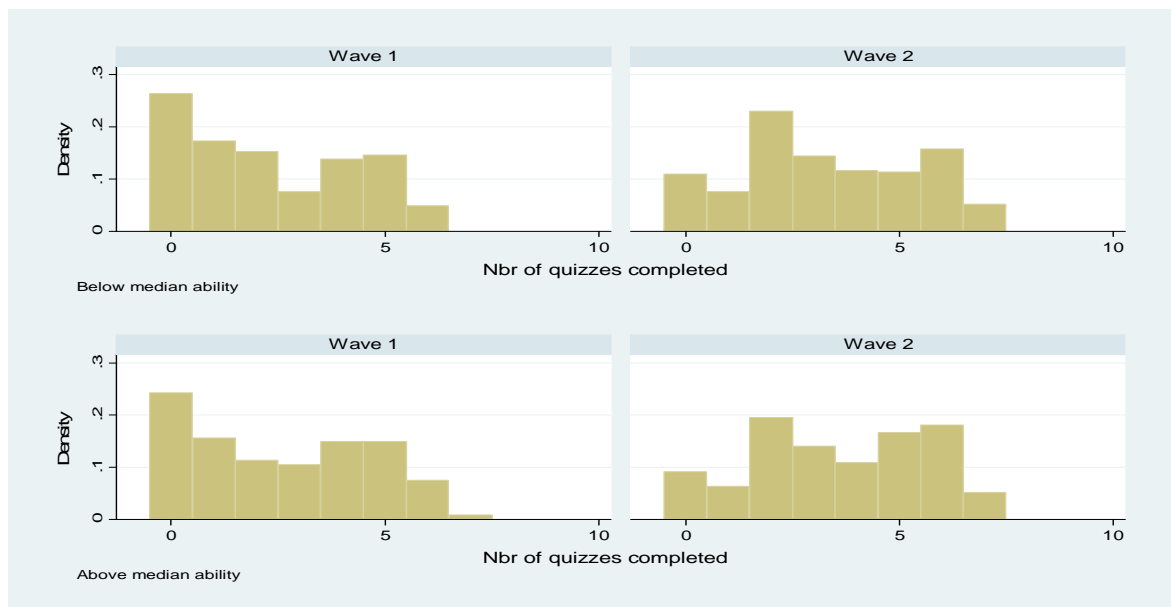
	A: The impact of incentives relative to non-incentivised weeks			B: Displacement effects
	Both courses	Quantitative Methods	Principles of Economics	Principles of Economics
	OLS	OLS	OLS	OLS
No incentive				0.022 (0.035)
Solution	-0.052 (0.015)**	-0.527 (0.050)**	0.017 (0.017)	0.037 (0.037)
Book voucher	-0.088 (0.013)**	-0.078 (0.024)**	-0.091 (0.016)**	-0.072 (0.040)*
Compulsory	0.663 (0.043)**	0.124 (0.074)*	0.720 (0.037)**	0.741 (0.060)**
Assess 2.5%	0.381 (0.019)**	0.313 (0.029)**	0.418 (0.023)**	0.435 (0.035)**
Assess 5%	0.603 (0.026)**		0.616 (0.025)**	0.636 (0.043)**
Cohort 2	0.061 (0.047)	0.115 (0.070)	0.023 (0.041)	0.025 (0.042)
Quant. Method	0.009 (0.044)			
Ability Q1	-0.084 (0.043)**	-0.083 (0.059)	-0.085 (0.044)*	-0.085 (0.044)*
Ability Q2	-0.056 (0.057)	-0.028 (0.079)	-0.067 (0.053)	-0.067 (0.053)
Ability Q3	0.018 (0.045)	0.004 (0.056)	0.023 (0.047)	0.023 (0.047)
Observations	9,108	2608	6500	6500
[individuals]	[327]	[326]	[325]	[325]
R2	0.28	0.35	0.28	0.29

Note: Other independent variables are: an indicator of term, a linear in week, term specific trend, course specific trend, dummies for gender, Chinese Nationals, other non-UK nationals, subject of degree and term and a linear in age (in month), tests in that week in other modules, essay in that week in other modules. Robust standard errors clustered at individual level.

**Figure 3A: Number of quizzes completed before in-term exam – Principles of Economics**



**Figure 3B: Number of quizzes completed before in-term exam – Principles of Economics**



**Table 7: Incentives and overall quiz participation, Principles:**

	Homogenous	Heterogenous
wave	1.225 (0.300)**	
Ability Q1	5.852 (3.183)*	5.372 (2.886)*
Ability Q2	5.995 (3.263)*	7.578 (3.461)**
Ability Q 3	6.802 (3.082)**	7.398 (2.936)**
Ability Q 4	6.472 (3.083)**	7.530 (2.874)**
Wave * Ability Q1		1.988 (0.375)**
Wave * Ability Q2		0.672 (0.656)
Wave * Ability Q3		1.313 (0.487)**
Wave * Ability Q4		0.983 (0.613)
R <sup>2</sup>	0.733	0.738
Obs	650	650

Note: standard errors clustered at the individual level. Model estimated without a constant. Other independent variables are: an indicator of term, a term specific trend, course specific trend, dummies for gender, Chinese Nationals, other non-UK nationals, subject of degree and a linear in age (in month), tests in that week in other modules, essay in that week in other modules. The total number of completed quizzes refers to completed quizzes before the in-term exam.

**Table 8: Incentives and quiz participation, Principles - Heterogenous Effects**

	Heterogeneous incentive effects by characteristic X					
	Ability			Gender	Nationality	
	Q1	Q2	Q3	Male	UK	China
No incentive		0.011 (0.042)		0.019 (0.032)	0.076 (0.027)**	
No incentive * X	0.097 (0.048)**	-0.021 (0.066)	-0.046 (0.081)	0.004 (0.049)	-0.079 (0.048)	-0.128 (0.081)
Solution		0.007 (0.058)		0.036 (0.037)	0.079 (0.033)**	
Solution * X	0.109 (0.062)*	0.013 (0.076)	-0.027 (0.093)	0.003 (0.051)	-0.069 (0.052)	-0.016 (0.077)
Book voucher		-0.137 (0.048)		-0.131 (0.039)**	-0.012 (0.037)	
Book voucher * X	0.145 (0.056)**	0.071 (0.062)	-0.013 (0.082)	0.093 (0.048)*	-0.080 (0.049)	-0.223 (0.086)**
Compulsory		0.642 (0.053)**		0.685 (0.045)**	0.825 (0.058)**	
Compulsory * X	0.261 (0.078)**	0.074 (0.124)	0.007 (0.088)	0.086 (0.085)	-0.119 (0.094)	-0.248 (0.095)**
Assess 2.5%		0.271 (0.075)**		0.399 (0.045)**	0.523 (0.041)**	
Assess 2.5% * X	0.253 (0.084)**	0.189 (0.092)**	0.066 (0.111)	0.057 (0.062)	-0.128 (0.061)**	-0.248 (0.095)**
Assess 5%		0.452 (0.081)**		0.603 (0.047)**	0.741 (0.044)**	
Assess 5% * X	0.255 (0.091)**	0.204 (0.095)**	0.136 (0.110)	0.053 (0.061)	-0.144 (0.062)**	-0.332 (0.102)**
R <sup>2</sup>		0.33		0.32	0.32	
F-test		2.09**		3.06**	2.53**	

Note: F-test: F test on the heterogenous components. Sample size: Matched sample: n=6500 Number of individuals, N:=325)

Other independent variables are: an indicator of term, a linear in week, an interaction between week and term Controls include: dummies for gender, Chinese Nationals, other non-UK nationals, subject of degree and term and a linear in age (in month), tests in that week in other modules, essays due in that week in other modules. Robust standard errors clustered at the individual level.

**Table 9: Incentives and Quiz Participation Effects on Normalised Grades: Principles of Economics**

<i>A: Reduced form estimates</i>							
	All	Q10	Q25	Q50	Q75	Q90	
Year 2	0.250 (0.124)**	0.305 (0.090)**	0.400 (0.143)**	0.366 (0.140)**	0.145 (0.069)**	-0.062 (0.065)	
(pseudo) R <sup>2</sup>	0.09	0.04	0.06	0.09	0.08	0.07	
<i>B: Instrumental variable estimators</i>							
	All	All	IVQ10	IVQ25	IVQ50	IVQ75	IVQ90
Year2	-0.027 (0.114)						
No. quizzes (SE)/[ 95% CI]	0.171 (0.032)**	0.155 (0.061)**	0.061 [-0.12,0.20]	0.136 [-0.09,0.30]	0.171 [0.06,0.30]	0.145 [-0.10,0.33]	0.056 [-0.13,0.27]
1 <sup>st</sup> stage		1.615 (0.0298)**					
F-test		29.34					
(pseudo) R <sup>2</sup>	0.21						
<i>C: Cohort specific effects</i>							
	All	Q10	Q25	Q50	Q75	Q90	
No. quizzes	0.191 (0.047)**	0.197 (0.028)**	0.234 (0.064)**	0.232 (0.009)**	0.187 (0.058)**	0.101 (0.031)**	
Year 2	0.127 (0.176)	-0.371 (0.138)**	-0.096 (0.282)	0.246 (0.041)**	0.351 (0.199)*	0.156 (0.146)	
Nbr Quizzes * Year 2	-0.046 (0.046)	0.028 (0.035)	-0.008 (0.074)	-0.083 (0.010)**	-0.086 (0.054)	-0.063 (0.038)*	
(pseudo) R <sup>2</sup>	0.22	0.13	0.15	0.19	0.13	0.10	

Note: Matched Sample - Sample size is 584 for mid term (292 students). Standard errors are clustered at the individual level in OLS and probit regressions. Quantile regressions are estimated separately for all quantiles. IV quantiles are estimated simultaneously, the first stage is estimated by OLS. Confidence interval are constructed from bootstraps (200 replications)

Controls include: dummies for gender, Chinese Nationals, other non-UK nationals, subject of degree and term and a linear in age (in month) and ability quartiles.