INCLUSIVE CLASSROOMS IN ECONOMICS: UNDERSTANDING STUDENT ENGAGEMENT USING MENTIMETER

Arpita Ghosh[1]

Abstract

In this paper, I examine whether different types of non-game-based audience polls affect students' in-class performance and participation. I administered two types of Mentimeter polls in small group classroom setting for two first-year undergraduate Economics courses, one timed quiz and the other an un-timed quiz. Using ordinary least squares methods, my analysis indicates that students seemed to perform 16-17% worse in timed quizzes. Interestingly, contrary to the existing literature, female students did not perform worse in these quizzes. Students were quite keen to participate in both types of quizzes with an overall average of approximately 80% completion. Finally, students' gender does not affect their participation or performance conditional on the gender of the tutorial leaders.

Keywords: Mentimeter, Student Response Systems, Student engagement, inclusive classroom, formative exercises

JEL Codes: A22

Introduction

Student engagement and interaction is a widely studied aspect in higher education research. Among others, Student Response Systems (SRS), like Mentimeter, are widely used to facilitate active teaching and increase student engagement (Skoyles & Bloxsidge, 2017). Especially during the pandemic, educators increasingly used SRSs to allow students to participate when teaching online/hybrid, as they helped them to measure students' participation and understanding of the core class materials (Pichardo et al., 2021). Given the adaptiveness and anonymity features of many SRS, they can be used to track students' understanding/performance without 'answer anxiety' (Skoyles & Bloxsidge, 2017). Moreover, these platforms can provide a much-needed voice to everyone in a classroom and not just a select few (Hill & Fielden, 2017).

However, a question remains if these platforms are over-used in certain cases - i.e., if students show less engagement with them over time and how to effectively use them in settings where they can benefit students more. Moreover, given that existing research shows that mobile technology or SRS can complement and not substitute for traditional teaching (Barreiro-Gen, 2020), the impact of such platforms in within classroom formative exercises has not been explored explicitly to the best of my knowledge.

[1] Assistant Professor in Economics (Lecturer, Education and Scholarship), Department of Economics, University of Exeter. Postal address: Department of Economics, Streatham Court, University of Exeter Business School, Rennes Drive, Exeter, UK, EX44PU. Email: arpitaghoshecon@gmail.com, A.Ghosh2@exeter.ac.uk. The author thanks the anonymous referee, the Editor of the Journal, Sadia Malik, the participants at the TeachECONference 2024 (CTaLE, UCL), and the American Economic Association Thirteenth Annual AEA Conference on Teaching and Research in Economic Education (CTREE) 2024 for important comments and suggestions. They additionally thank participants at the Teaching Day 2023 from the Department of Economics, University of Exeter for insightful comments on an early version of this work. Finally, the author also acknowledges the Economics Scholar programme at Exeter and the excellent research assistance from Camron Manrique.

In this paper, I investigate a few simple questions, starting with how students perform in seminars when they engage with multiple choice questions (MCQs) through Menti quizzes. I then explore if these students participate in these quizzes consistently in each seminar and examine if there are any gender specific patterns to their performance and participation.[2] I created two types of Mentimeter polls in seminars for two first-year undergraduate Economics modules: some timed and others un-timed. I administered these two types of formative quizzes in the small group seminars of both the modules. Analysis of 1,238 observations using OLS over two academic years indicate that in line with existing literature, students seem to perform worse in timed quizzes. The underlying reason could be that students need more time to reach the correct answer. Equally, they may feel less stressed in a non-competitive environment (Walstad & Robson, 1997). However, in contrast to existing literature, I find that female students do not perform worse than their male cohort mates in timed quizzes. I also define an additional outcome variable to measure participation rate, which is the proportion of the number of questions attempted over the total number of questions asked in each quiz. Students seem to be participating consistently in both types of quizzes with an overall average of 80%. Moreover, students perform slightly better in seminars with female seminar leads; however, in these classes female students do not tend to show differential performance or participation rate. Additional results show students perform worse in seminars with a higher number of quantitative questions and demonstrate slightly more participation in seminars which occur very early or very late in the day.

This study has connections with three strands of literature which are directly relevant. Research has shown that time limited tests are less reliable, less inclusive/equitable, and less valid as a true reflection of students' knowledge (Gernsbacher et al., 2020). I also find similar results in this paper. Moreover, it is widely discussed in the literature that females perform worse than males in timed MCQs testing including tests like standardized assessment tests (Bolger & Kellaghan, 1990; Gneezy et al., 2003; Graetz & Karimi, 2022; Griselda, 2020; Walstad & Robson, 1997) and females tend to skip more MCQs in comparison to their male counterparts (Conde-Ruiz et al., 2020; Riener & Wagner, 2017; von Schrader & Ansley, 2006). Contrary to this, I provide evidence that this conclusion may not be as generalizable in different contexts. In this paper, I find female students do not perform differentially to their male counterparts in formative quizzes. These classes were also largely comprised of non-economics major students studying Economics subjects. Consequently, in these types of settings, the gender difference in performance may not be applicable. Finally, non-game-based audience response platforms like Mentimeter make students comfortable to take active part in the classroom (Gokbulut, 2020; Mohin et al., 2022). The findings of high level of participation on average and more participation in seminars which took place in very early or very late in the day show that students in general prefer these anonymous modes of in-class participation. However, a crucial consideration in interpreting the results of this study is the lack of other demographic controls like the background of students or past exam performances. This paper uses anonymous responses from Mentimeter quizzes and hence cannot link the data with such controls.

Overall, this study contributes to broadening the understanding of anonymous in-class student participation and indicates ways of making classrooms more inclusive. The following section briefly details the key literature in the area, followed by the background and data of this study, and concludes with results, limitation, and discussion.

[2] In this paper, I use the term gender to denote student's self-declared identity in response to the question "Which gender you associate yourself with?" where the options were 'Male', 'Female', 'Non-Binary', and 'Prefer not to say'. Hence there are no clear distinctions between the social construct of gender and biological identity in this paper.

Literature

This section provides a structured overview of the related literature and details the ways this study relates and contributes to it. The first strand of literature focuses on understanding student responses to clickers or SRS, and it is quite extensive. The secondary strand of literature is the gender gap in performances in MCQs in timed settings and quantitative questions.

In the first strand, there are three key themes under which the related interdisciplinary and economics pedagogy literature in SRS use can be summarized. The first is student engagement. The common idea under this umbrella states that SRS like Mentimeter (or similar) improves student engagement, and students enjoy using it (Barreiro-Gen, 2020; Elliott, 2003; Gokbulut, 2020; Graham et al., 2007; Mayhew et al., 2020). Among these papers, many provide also quantitative evidence that clicker usage increases student interactivity (Heaslip et al., 2014; Laxman, 2011).

The second theme in this broad category involves student performance, which demonstrates mixed results. Here the key idea is that SRS like Mentimeter usage may improve academic attainment or perceived learning (Mayhew et al., 2020), allowing staff to gauge students' level of understanding (Elliott, 2003). The existing evidence generally suggests no positive effect of SRS use on academic attainment. Furthermore, there are no significant differences in attainment between clicker and nonclicker (traditional pen and paper, for example) sessions (Hayter & Rochelle, 2013; Johnson et al., 2008; Mohr, 2013). A similar pattern has been shown for Mentimeter specifically (Madiseh et al., 2022). Another software like Mentimeter, called TurningPoint, also had no positive impact on performance (Barreiro-Gen, 2020), while others like VotApedia is shown to have positive educational outcomes for students (Dunn et al., 2013)

The final aspect of this literature is the Equality, Diversity, and Inclusion (EDI) consideration. SRSs provide platforms for students who would not normally contribute (whether because of gender, confidence, or other reasons) and play an active role in class (Graham et al., 2007; Madiseh et al., 2023; Mayhew et al., 2020; Pichardo et al., 2021). Hence, anonymity of these SRS platforms is appreciated by students (Heaslip et al., 2014; Wood, 2020). Mentimeter and similar systems can provide a safer classroom environment (Guarascio et al., 2017).

I detail some of the key papers from all the three themes above to set up the context of my study. In one of the early papers, Mohr (2013) shows that while students using the iclicker[3] appear to have higher average exam scores, the significant effect disappears once demographic and academic variables are controlled for in a business and economic statistics course. There are many such studies which concentrate on SRS in classrooms, often with no conclusive evidence of performance differences between traditional and SRS use (Hayter & Rochelle, 2013; Imazeki, 2014; Johnson et al., 2008; Stowell, 2015). This literature provides many potential reasons for this lack of evidence, including bandwidth, class size, and potential polling fatigue with extensive polling (more than 15 questions), etc. In this context, the risk of distraction in SRS use is also important. Mohr (2013) shows that in their study, 38% of students reported they were distracted "sometimes" and 4% "most of the time" while using mobile phones. In contrast, Heaslip et al. (2014) showed that iclicker use resulted in positive feedback from students in a post-test questionnaire, such as its use "was fun," provided a "better understanding of material," and "[got the] class involved." Overall, this paper showed positive impact on interactivity, while anonymity was also important, with students not feeling embarrassed to get questions wrong. Imazeki (2014) demonstrates some of the advantages of 'Bring Your Own Device' (BYOD) over traditional clickers. This area has been well-

[3] iClicker is a handset device that allows students to respond to multiple choice, true/false and yes/no questions displayed in the lecture slides. This is relevant, as Mentimeter has similar functions on the students' devices, along with further functions.

researched in the literature. Benefits of BYOD includes convenience for students, open-ended questions, low commitment (no need to buy clickers), low cost (vast majority of students already have mobile phones, and the software (including Menti) is often free and can be upgraded for a small premium). BYOD also has several downsides which included distraction risk, need for signal/bandwidth, excluding students without devices, integration issues with university systems (although premium Mentimeter allows educator to easily download excel sheets with polling results). Guarascio et al. (2017) specifically compares Socrative (mobile technology like Mentimeter) and traditional SRSs or clickers. Socrative was found to create a 'safer' classroom environment where students were more confident to ask questions through the technology.

From the EDI perspective, in a relatively contemporary study, Wood (2020) examined Mentimeter usage within the geography department at Newcastle University, UK. In this work, students generally responded positively to Mentimeter usage, as it made them feel more involved and they enjoyed its anonymity. Mentimeter is found to increase student enjoyment and engagement and provide a platform for less confident students to give their thoughts (student voice) (Mayhew et al., 2020). This also increased inclusivity for students who struggled to participate due to culture, gender, disability, or other means. Additionally, 68% of respondents said Mentimeter use had increased or significantly increased learning, with most others responding the same level of learning, pointing to a perception that Mentimeter increases student attainment. This work builds upon Mayhew (2019) which looked at the use of Mentimeter in political science. In another interdisciplinary study in Spain using Mentimeter, Pichardo et al., (2021) show the anonymity allows less confident students to wilfully participate, as well as leaving a 'right to remain silent,' i.e. nobody is forced to answer. Mentimeter usage includes the opinions of all class members. Madiseh et al. (2023) studied a large group of students taking a foundation course at a public university in Oman and showed generally positive perceptions from students regarding its use in class.

As mentioned above, the secondary broad research field related to this paper is on the gender gap differences in academic performances especially when the evaluation methods are MCQs versus constructed responses. Literature in this area has often established that female students perform worse than their male counterparts in MCQs (Bolger & Kellaghan, 1990; Ellison & Swanson, 2023; Gneezy et al., 2003; Guez et al., 2020; Lumsden & Scott, 1987; Markowsky & Beblo, 2022). Moreover, some papers have found that there is often a gender gap in quantitative performance (Kahn & Ginther, 2017; Khasawneh et al., 2021). Part of this literature also finds that women tend to leave quantitative questions blank more often than males (Conde-Ruiz et al., 2020; Saygin & Atwater, 2021), and they may perform worse in competitive environments (Cahl'iková et al., 2020). However, this gender gap may not always exist (Chan & Kennedy, 2002). These gender gaps may also disappear in repeated competitions (Cotton et al., 2013) and may be reduced by abolishing negative marking in exams (Coffman & Klinowski, 2020).

In the context of these two broad strands of literature, my paper adds to the understanding of student engagement techniques used in classrooms from both inclusivity and academic outcomes perspectives. Specifically, I study how different types of polls might affect students differentially, both in their in-class formative quiz performance, as well as in their participation, conditional on their stated gender. In this context, this paper relates to that of Mu & Paparas (2015) who also study mobile technology in classrooms for non-economists. One of the modules (for both academic years) I examine is a module of undergraduate introductory economics for purely non-economist students, and the other (available only for the first academic year) is an introductory finance course for both economics and non-economics major students. Moreover, most of the existing papers study students' and educators' perceptions and student performances but, to my knowledge, they don't explore student's completion rates (or

participation rate) in formative classroom exercises. Finally, with calls for de-emphasis of MCQs in grade compositions to reduce gender gap in student performances in Economics (Engelhardt et al., 2024), this paper highlights the need of deep examinations of trends in student performances before restructuring assessments.

Background

In this section, I describe the background of the underlying data used for analysis. For the empirical analysis, I use anonymous participant level data from two different first-year courses,[4] one course (module 1 hereafter) for both academic years of 2022-23 and 2023-24 and one course (module 2 hereafter) for 2022-23.[5] These two first year modules (one introductory economics for non-economists and one introductory finance for economists and non-economists) had no pre-requisites; however, they had two key differences. Module 2 was more quantitative in nature in comparison to module 1, and the seminars in module 2 were bi-weekly whereas the seminars in module 1 were weekly.[6]

I designed the adoption of Mentimeter quizzes in both these modules to create an inclusive environment where all students feel able to participate without feeling peer pressure to get the answers correct. In my university, attendance is not recorded, and as a result, seminar and lecture attendance tend to suffer. I intended to promote student participation using low-stakes formative quizzes. These ensured that students felt that attending the seminars was worthwhile. and they were preparing them actively for the exams. Moreover, as the exams for both these modules were time-limited and included some MCQs, I aimed to provide timed quizzes to give the students opportunities to practice answering questions in a time-bound environment. However, I did not want to disadvantage a section of the students in these timed formative quizzes. This experiment provided me the opportunity to assess whether the seminars of these two modules were inclusive to students from different genders. All the data collected during the observation period in this study used the Mentimeter timed and non-timed quiz approach.

In both modules, I implemented two types of in-class seminar quizzes: timed and un-timed MCQs or quizzes. The seminars were run by nine instructors, two lecturers (assistant professors in Economics), including the module lead (the convenor of the module, with all administrative responsibilities), and seven first-time tutors. Tutors of a module (in this case, one lecturer and all first-time tutors) were solely responsible for delivering the tutorials, and they were provided with all materials and instructions beforehand. Both lecturers, as well as 3 out of 7 tutors were female. The timed MCQs were provided with 30 seconds of countdown. For the un-timed MCQs, tutors were instructed to wait about 60-70 seconds or to move on to the explanation/next question once all participants answered (whichever is sooner). The allocation of timed vs un-timed quizzes was alternative per instructor. Specifically, if the first seminar (in a week) of instructor 1 received an untimed quiz, the second seminar received the timed quiz. The following week, for the same instructor, the first seminar would receive the timed quiz, and the second would take the untimed one. Students can only see the seminar information of the ones they are allocated to, and they cannot choose to attend a different seminar due to space constraints in general, based on the seminar lead. They were also not informed beforehand which seminar would receive which type of quiz; however, they were not barred from attending another tutorial in case of any unforeseen schedule clashes. These circumstances, where a student informed the module (course) leader to ask for a different seminar timing to attend, were only a handful in these two years. However, as attendance is not

[4] The terms ‘course’ and ‘module’ are used interchangeably in this paper. In the UK courses are often referred to as modules.

[5] Module 2 was led by me only for the 2022-2023 academic year due to teaching allocation constraints.

[6] Seminars are small group classes with approximately 20-30 students in each session.

recorded in general, students could have attended a different seminar than their original allocation, and data on these cases are unavailable. As my analysis is at the aggregate level, this does not pose any significant concern for the results.

I was the module leader for both modules. For the first academic year, module 1 seminars were delivered by two Economics Scholars and the module lead (2 seminars each). Module 2 seminars were delivered by a new lecturer (2 seminars) and the module lead (4 seminars). For the second academic year (module 1 only), all the seminars were delivered by five Economics Scholars (2 seminars each), who were different from the first academic year's Economics Scholars. The seminar timings were scheduled (by the central timetabling team) across all days of the week with start timings ranging from 8:30 a.m. to 5:30 p.m. For approximately half the cases, seminar times were scheduled at 12:30 p.m., 1:30 p.m., or 3:30 p.m. Overall, across the two modules, 5 out of the total 22 seminars were scheduled at 8:30 a.m. or 5:30 p.m. Anecdotally, seminars at these two times are the least popular for students, as they find it difficult to attend these sessions and/or concentrate on the study materials. Tuesdays and Wednesdays had the fewest sessions, as approximately less than 1 in 5 seminars were on these two days (Wednesday afternoons are generally preserved for sports events in most universities in the UK).

The sets of questions to be covered in the seminars were uploaded in the module virtual environment one week in advance for students to have the time to study and attempt beforehand. This practice embraced the evidence in the literature that providing students with materials to prepare ahead of the face-to-face learning sessions (flipped classroom techniques) and testing their knowledge in Just-in-Time learning with BYOD can consolidate their learning better (Santamarina-Campos et al., 2021; Åhman et al., 2021). All students in each session were treated equally, seeing the same number of timed and un-timed MCQs/quizzes over the semester. The questions used in the Menti were all MCQs (short), focused mainly on the concepts and at times on solving small mathematical problems.[7] All the seminars involved two sections of questions, MCQs and short answer quantitative or discursive questions. Tutors were instructed to first concentrate on the short answer questions and then on the MCQs in each session.

In these sessions, I asked only one optional personal question on gender identity. Figure A1 shows an example of the Mentimeter slide with the demographic question. Figure A2 shows the participant information shown before the start of the quizzes.[8] These quizzes did not contribute to summative assessments; they were part of the formative assessments which were only used in seminars. This information was also conveyed to the students clearly in each seminar sessions. According to university policy, attendance is not recorded in large or small group teaching sessions. However, information on expected attendance in each session is available, and generally the seminars are comprised of 19-26 students. Students do not always attend the seminars despite frequent encouragements from the module lead and tutors. However, as an approximate example, in week 3 of the academic year 2022-2023 (the first week of seminar) of module 1, 73% students participated in the Menti polls. More students may have attended the session and not participated in the quiz, but this cannot be verified.

Data and Method

In this section, I describe the data, key variables, sample size, and other details followed by the empirical specifications. The data used in the analysis come from the Mentimeter quizzes administered in the small group seminars. The seminar leaders would use the pre-assembled quizzes in the classroom and participant level data was recorded for each seminar.

[7] Examples of four such questions (from each module) are provided in Appendix B.
[8] Figures A3 and A4 show the participant numbers by week for both the modules.

These data from each session were combined and then analyzed to answer the research questions. Importantly, the data only records participant-level information in each session (timed and untimed). There is no way to link participant information from one session to another as these quizzes were administered anonymously. As a result, the unit of analysis in this paper is the participant in each session, which may have the same student in each session or different students altogether. Finally, the data used in this paper are anonymous and solely come from student responses to the Mentimeter sessions. As a result, I cannot link other demographic characteristics of these students, such as whether they are international students, or their past exam performances to control for ability.

Outcomes

I use two different metrics of student performance and engagement. The first one is a measure of student performance which is defined by the number of correct answers provided by a participant over the total number of questions. The other metric is a measure of participation or completion which has been used in the literature in different context of student engagement (Braun & Sellers, 2012; Kibble, 2011; Orr & Foster, 2013; Salemi, 2009). This variable is defined by the number of questions attempted by each student over the total number of questions. Notably, these two outcomes are highly correlated, with the Pearson correlation coefficient of 0.798. This denotes that generally students who consistently complete the quizzes also perform well in them. These two metrics are used as the outcomes of interest in the empirical analysis and are discussed in the results section.

Independent variables

The key independent variables in this paper are the type of quizzes faced by the respondents (timed or untimed) and their self-declared gender identity. The other variables I control for include whether the module lead is also the seminar lead and the number of quantitative questions in each quiz. These variables are constructed for the sessions. The identities of the tutors were known to me, and quiz questions were prepared by me for all seminars. I also have information on the timing of each seminar, i.e., whether these sessions are early in the morning (8.30 am) or quite late (for the UK) in the evening (5.30 pm). Anecdotally, students do not like to attend these early or late sessions, and, as a result, the timing for the seminars is included as a control for the analysis.

Sample size

Each observation in this study is per response by a student in a quiz. Crucially, these observations may come from a unique student being present in each seminar or different students in different seminars. Table 1 shows the total observations (responses by students) collected from the Mentimeter sessions. It also depicts the reasons behind dropping some observations and the final observation count. For example, I dropped observations where participants did not prefer to disclose their gender identity (24 such observations). I also excluded 25 observations where the participants reported their gender to be non-binary, as such observations were too few to analyze patterns. The final count of observation for analysis is 1238 over two academic years and both modules.

Table 1: Waterfall of observations in the analysis

Total Observations	1451
Blank in Gender Question	164
"Prefer not to say"	24
Non-Binary	25
Final Observations	1238

Importantly, as mentioned before, I only have data for module 2 for the 1st academic year. Moreover, because all data are anonymous, I cannot link participants to their final summative results for these modules or any other demographic characteristics. The discussion section describes these aspects in more detail to underscore the possible concerns in generalizing the results of this paper.

Summary statistics

Table 2 below provides summary statistics for the key variables in the empirical analysis. Across all the observations from respondents, 54% of the time the responses were from students who self-identified as females. Largely, the observations are from module 1 (as described before, module 2 data is only available for the second academic year, and it also had bi-weekly seminars). Overall, 56% of the time the gender of the seminar lead was female. Seminar times were either early or late in the day for 22% of the response cases. The maximum number of questions in quizzes was 15 with an average of approximately 9 questions.

Table 2: Summary statistics of key variables

Variables	Mean/Percent	Standard Deviation	Min	Max
Female students	54.20%			
Female tutors	56.87%			
Module 1 observations	83.93%			
Lecturer instructors	23.10%			
Timed MCQ	50.48%			
Seminar times too early/too late in day	22.05%			
Number of questions	8.78	3.85	4	15
Number of questions attempted	7.19	3.93	0	15
Number of quantitative questions	0.63	1.42	0	6
Total observations	1238			

Absentees: robustness

I also define an additional variable called "absentee". An absentee is defined as a participant who did not start the quiz at all or stopped towards the end of the quiz (i.e., stopped attempting two or more consecutive questions from the end). Tables A1 and A2 in the appendix

show the distribution of such responses by participants.[9] This variable provides a unique perspective on the student behavior in classroom engagement by gender. Moreover, this variable is useful to understand if there are any underlying trends in 'polling fatigue' of students.[10] It is important to reiterate here, that the data includes module 1 for both academic years and module 2 only for one academic year. Table A1 shows the number of absentees by gender and module for the first academic year for both modules. In module 1, the number of absentees is almost the same by gender, and in module 2, the difference is negligible (11 females and 14 males). Table A2 shows for module 1, the number of absentees by gender over the two academic years of 2022-2023 and 2023-2024. There is an increased number of absentees who identify as female in the second year in comparison to the first (59 in year 2 versus 11 in year 1). Even though the number of absentees increase for participants who identify as male, this is only about 150% in comparison to over 400% for female absentees.

Empirical Specifications

The key point of interest is to estimate whether average performance or completion rates differ by type of quizzes administered, controlling for the key characteristic of student gender. This can be estimated using the ordinary least squares (OLS) model with dummy variables for the type of quiz (timed vs un-timed) and student gender (male vs female). Moreover, the interaction between the type of quiz and the student gender would capture the differential effects on the outcomes for female students in a timed quiz setting. The OLS model delivers unbiased estimates with the least variance subject to some crucial assumptions. As we cannot control for other student characteristics or connect students' outcomes across sessions, there can be omitted variable bias in this case.

In this regard, one issue could be that high performing students choose to attend untimed quizzes, leading to the results discussed in the next section. There are two reasons that this is not a potential concern. One, students were not aware of the oscillating nature of the timed vs un-timed quizzes, that is each seminar sessions saw both timed and un-timed quizzes. Given these are formative in class exercises, it is unlikely any student would have changed their behavior in a systematic way to attend un-timed quiz sessions. Secondly, as the timetable is centrally managed, students, in general, do not have the flexibility to choose which session to attend. However, as attendance is not monitored and past performance of students cannot be linked with the data, this concern cannot be addressed systematically within the empirical results. Another potential concern could be that seminars which were early (8:30 a.m.) or late (5:30 p.m.) in the day could have had only un-timed quizzes resulting in better performance of students. As each seminar would have timed quiz one week and un-timed the next, this should not affect the results. Given these considerations above, the key estimating regression equation is defined as follows:

$$(1)\ Outcome_i = \alpha + \beta TypeQuiz_i + \delta GenderStudent_i + \eta TypeQuiz_i * GenderStudent_i + \gamma X_i + \epsilon_i$$

where $〚Outcome〛\_i$ is either performance or participation by participant i. Our main parameter of interest is η which denotes the differential outcomes by gender of the student in

[9] Absentees denoted as "No" in tables A1 and A2 are students who show erratic participation behaviour not according to the definition above. An example would be a student who participated in all questions apart from a couple from the middle of the quiz.

[10] I use this variable to exclude these responses (both 'yes' and 'no' in the tables A1 and A2) for the robustness regressions in the appendix. These numbers are quite small and do not give us any qualitatively different results.

different types of quizzes (timed or un-timed). β is the type of quiz, δ is the pure effect of student gender, X_i is a series of student response specific controls such as seminar lead (whether the module lead is also the seminar lead), timing of the seminar/tutorial, number of quantitative questions, etc. Finally, ϵ is the error term. Ex-ante, following the literature, one would expect that all students perform worse in timed quizzes and females will perform worse in timed quizzes in comparison to males. Moreover, from tables A1 and A2, we may expect that female students, on average, may show less participation than their male cohort mates.

In addition to the above, I examine whether having a female seminar lead has a differential effect on student performance or tenacity conditional on student gender. In this case, a triple interaction term with the gender of the tutor, student gender, and type of quiz will demonstrate the differential effect on the performance of the female students in a time quiz setting with a female tutor. Hence, the second key estimating equation of interest is as follows:

$$
\begin{aligned}
(2)\ Outcome\_i &= \alpha + \phi\ [\![GenderTutor\_i + \kappa\ [\![TypeQuiz]\!]\_i \\
&+ \theta\ [\![GenderStudent]\!]\_i + \rho\ [\![TypeQuiz]\!]\_i * [\![GenderStudent]\!]\_i \\
&* [\![GenderTutor]\!]\_i + \chi X\_i + \iota\_i
\end{aligned}
$$

where the key parameter of interest is ρ which captures the interaction of type of quiz with tutor and student gender. If there are any role model effects on the outcomes of female students, we may expect them to perform better and demonstrate more completion rate in seminars led by female tutors (Griffith, 2014; Porter & Serra, 2020); especially in Economics where there exists a lack of such role models.

Results

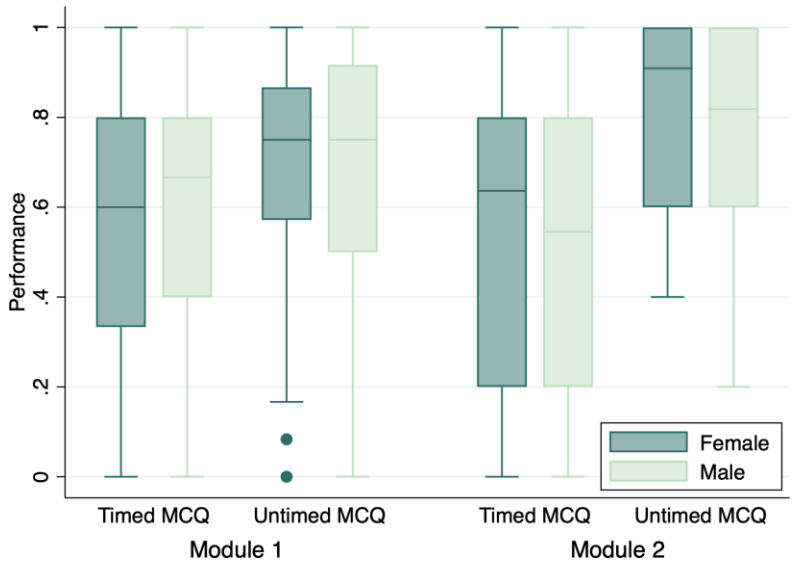
In this section, I present first some graphical evidence for both the key outcomes and then I present a set of regression results.

Graphical evidence

Figure 1 below shows that females do not perform significantly differently than males across modules and type of quizzes. This figure is the first level evidence for the key finding of this paper. Both males and females perform slightly worse in timed MCQs in module 2. In comparison to this, they perform quite well in the un-timed quizzes for this module. It is important to reiterate that the module 1 data in these graphs include both academic years, and module 2 is only present for academic year 2022-2023.[11]

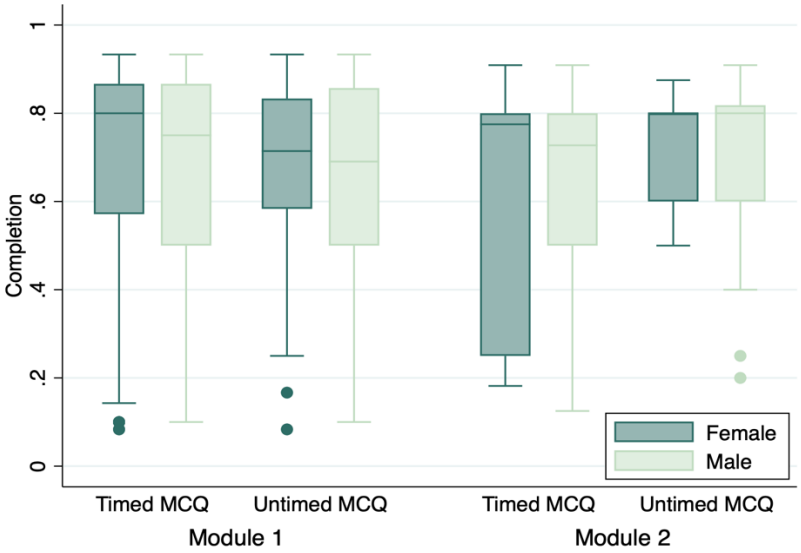
[11] Figure A5 in appendix show the broad distribution of performance of students in timed and untimed quizzes. The average performance of students in the untimed MCQ is slightly higher than the timed quizzes. Average completion/participation is slightly less in un-timed quizzes in aggregate (appendix figure A6). Figure A7 shows that on average, in timed quizzes students performed slightly worse in comparison to un-timed quizzes in both modules. Students demonstrate slightly lower completion in timed MCQs in module 2, which is more quantitative in nature (figure A8). This is perhaps in line with the lack of potential preparation by the students prior to the seminars leading to them not being able to answer the Mentimeter quizzes. Finally figure A9 shows that female and male students do not exhibit widely different completion behaviour overall.

Figure 1: Do females 'perform' differently in timed vs untimed quizzes compared to males?



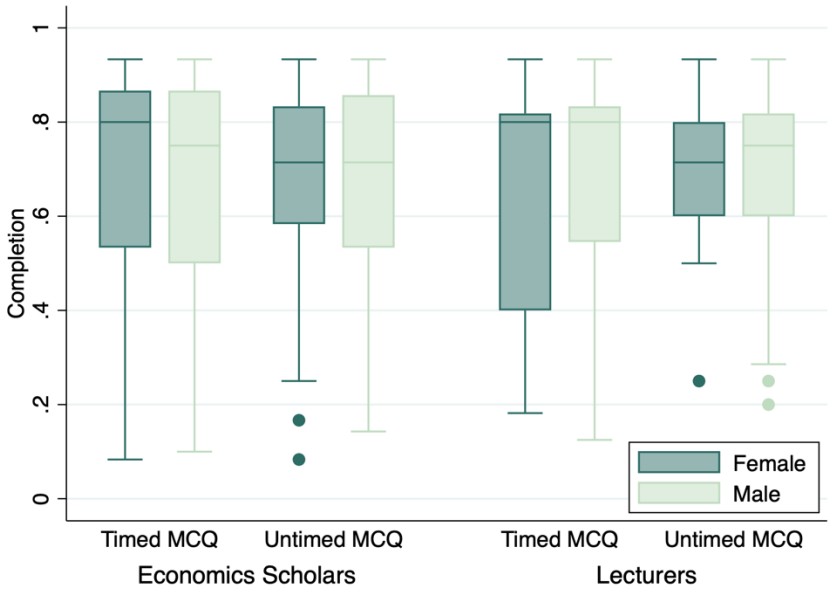
Notes: In module 1 timed MCQs- females: lower hinge 0.33, upper hinge 0.8 and males: lower hinge 0.4, upper hinge 0.8; in module 1 untimed MCQs – females: lower hinge 0.57, upper hinge 0.87 and males: lower hinge 0.5, upper hinge 0.92; In module 2 timed MCQs- females: lower hinge 0.2, upper hinge 0.8 and males: lower hinge 0.2, upper hinge 0.8; in module 2 untimed MCQs – females: lower hinge 0.6, upper hinge 1 and males: lower hinge 0.6, upper hinge 1. Whisker cutoffs [0-1].

Figure 2: Do female students show different 'completion rate' vs males in timed vs un-timed quizzes?



Notes: In module 1 timed MCQs- females: lower hinge 0.57, upper hinge 0.87 and males: lower hinge 0.5, upper hinge 0.87; in module 1 untimed MCQs – females: lower hinge 0.58, upper hinge 0.83 and males: lower hinge 0.5, upper hinge 0.86; In module 2 timed MCQs- females: lower hinge 0.25, upper hinge 0.8 and males: lower hinge 0.5, upper hinge 0.8; in module 2 untimed MCQs – females: lower hinge 0.6, upper hinge 0.8 and males: lower hinge 0.6, upper hinge 0.82. Whisker cutoffs (0-1).

Figure 3: Do female students show different 'completion rate' in comparison to males in timed vs un-timed quizzes when tutors are closer to their age-group?



Notes: For Economics Scholars, timed MCQs- females: lower hinge 0.53, upper hinge 0.87 and males: lower hinge 0.5, upper hinge 0.87; Economics Scholars, untimed MCQs – females: lower hinge 0.58, upper hinge 0.83 and males: lower hinge 0.53, upper hinge 0.86; For lecturers, timed MCQs- females: lower hinge 0.4, upper hinge 0.82 and males: lower hinge 0.55, upper hinge 0.83; lecturers, untimed MCQs – females: lower hinge 0.6, upper hinge 0.8 and males: lower hinge 0.6, upper hinge 0.82. Whisker cutoffs (0-1).

As a reminder, completion (or participation) is defined as the number of questions attempted by each student over the total number of questions in each seminar. Across different types of quizzes both females and males demonstrate slightly lower average completion in untimed MCQs. This effect seems to come from module 1 largely for males (figure 2). It is important to note that in these cases the entire distributions are not significantly different from one another. The differences exist only on average. Finally, in timed MCQs students show slightly more average completion while working with lecturers as seminar leads (figure 3).

Regression results

Table 3 shows the results for the students' performance and completion rates.[12] This table shows the different specifications from our first empirical equation above (equation 1). Timed MCQs see worse average performances by the students across different specifications (1-4) in contrast to untimed MCQs. On average, this effect is about 16-17%, which can be interpreted as on average 46% of performance rate by students in comparison to the mean of the dependent variable. In other words, in timed quizzes, students would have gotten approximately 2 more questions incorrect out of 9 questions (on average) compared to un-timed quizzes. Students perform slightly worse when they encounter a relatively large number of quantitative MCQ questions. Both results are in line with the existing literature in this field, as discussed above. Interestingly, I do not observe any significant effects on the timed MCQ quizzes based on the gender of the students. Specifically, female students do not perform worse than their male counterparts in the timed MCQs.

[12] Appendix table A3 shows some other specifications for both performance and completion rates, without controls.

Table 3: Student performance and completion in different types of quizzes conditional on gender

	Outcome: Performance				Outcome: Completion			
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender	-0.0114	-0.0064	-0.0086	-0.0119	0.0001	0.0056	0.000	-0.0029
	(0.0206)	(0.0206)	(0.0205)	(0.0208)	(0.0187)	(0.0186)	(0.0185)	(0.0188)
Timed MCQ	-0.1687***	-0.1681***	-0.1689***	-0.1722***	-0.0778***	-0.0772***	-0.0793***	-0.0822***
	(0.0232)	(0.0233)	(0.0232)	(0.0231)	(0.0220)	(0.0221)	(0.0219)	(0.0218)
Gender * Timed MCQ	0.0120	0.0119	0.0112	0.0127	-0.0086	-0.0088	-0.0104	-0.0091
	(0.0318)	(0.0317)	(0.0317)	(0.0316)	(0.0303)	(0.0303)	(0.0300)	(0.0299)
Module Lead		0.0409*	0.0237	0.0677***		0.0456***	0.0020	0.0406**
		(0.0184)	(0.0205)	(0.0210)		(0.0170)	(0.0194)	(0.0189)
Timing of Seminar			0.0379***	0.0165			0.0963***	0.0776***
			(0.0192)	(0.0196)			(0.0178)	(0.0173)
No. Quant Questions				-0.0228***				-0.0200***
				(0.0068)				(0.0063)
Constant	0.7153***	0.7029***	0.7003***	0.7121***	0.8651***	0.8512***	0.8446***	0.8550***
	(0.0149)	(0.0158)	(0.0159)	(0.0164)	(0.0136)	(0.0143)	(0.0143)	(0.0150)
Observations	1238	1238	1238	1238	1238	1238	1238	1238
Adjusted R-squared	0.08	0.08	0.08	0.09	0.02	0.03	0.04	0.05
Mean Dependent Variable	0.63	0.63	0.63	0.63	0.82	0.82	0.82	0.82

Notes: This table shows the specifications for the equation 1 above. Dependent variables are Performance and Completion rates. Performance rate is defined as correct answers over total number of questions. Completion rate is defined as attempted questions over total number of questions. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Timed MCQs see lower completion rates by the students across specifications 5-8 in Table 3. On average this effect is about 7-8%, which can be interpreted as on average 72% completion by students in comparison to the mean of the dependent variable. Put differently, in timed quizzes, students would have attempted approximately 1-2 less questions out of 9 questions (on average) compared to un-timed quizzes. Students are slightly less tenacious when they encounter more quantitative MCQ questions. There are no gender effects on completion in timed MCQs. Seminars which had the module lead as the tutor seemingly observe higher completion rates and better performances compared to other sessions; however, this is not persistent across all specifications. As a reminder, the timing of seminars is denoted by seminars which occur very early in the day (8:30 a.m. UK time) or very late in the day (5:30 p.m. UK time). There is a slightly positive effect on the completion rates of students when seminar timings are extreme. This may denote that students who attend these seminars (which are least popular for students from anecdotes and observed attendance trends) tend to try to

attempt more questions and utilize their time. However, the students in the sample of such seminars could have been more disciplined on average.

Table 4: Student performance and completion in different types of quizzes with female seminar leads

	Outcome: Performance			Outcome: Completion		
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Female Tutor	0.0458**	0.0486***	0.0856***	0.0213	0.0227	0.0499**
	(0.0193)	(0.0185)	(0.0255)	(0.0179)	(0.0176)	(0.0242)
Weakly Seminars	-0.0434	-0.0516	-0.0451	0.0019	-0.0023	0.0026
	(0.0357)	(0.0327)	(0.0330)	(0.0321)	(0.0311)	(0.0314)
Module Lead	0.0281	0.0204	0.0193	0.0329	0.0280	0.0271
	(0.0304)	(0.0278)	(0.0278)	(0.0272)	(0.0261)	(0.0261)
Timing of Seminars	0.0176	0.0268	0.0243	0.0720***	0.0774***	0.0755***
	(0.0232)	(0.0213)	(0.0212)	(0.0194)	(0.0192)	(0.0191)
No. Quant Questions	-0.0262***	-0.0305***	-0.0304***	-0.0186**	-0.0209***	-0.0208***
	(0.0084)	(0.0075)	(0.0076)	(0.0072)	(0.0070)	(0.0071)
Female Student		-0.0136	0.0211		-0.0043	0.0212
		(0.0204)	(0.0279)		(0.0186)	(0.0269)
Timed MCQ		-0.1746***	-0.1761***		-0.0077	-0.0070
		(0.0228)	(0.0227)		(0.0299)	(0.0299)
Female Student * Timed MCQ		0.0151	0.0159		-0.0077	-0.0070
		(0.0315)	(0.0314)		(0.0299)	(0.0299)
Female Student* Female Tutor			-0.0627*			-0.0461
			(0.0320)			(0.0306)
Constant	0.6436***	0.7427***	0.7174***	0.7978***	0.8484***	0.8298***
	(0.0370)	(0.0356)	(0.0387)	(0.0336)	(0.0341)	(0.0375)
Observations	1238	1238	1238	1238	1238	1238
Adjusted R-squared	0.01	0.09	0.10	0.02	0.05	0.05
Mean Dependent Variable	0.63	0.63	0.63	0.82	0.82	0.82

Notes: This table shows the specifications for equation 2 above. Dependent variables are Performance and Completion rates. Performance rate is defined as correct answers over total number of questions. Completion rate is defined as attempted questions over total number of questions. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 4 shows the effects on students' performance and completion respectively when they have female seminar leads (following equation 2 above).[13] Female students do not show any differential completion rates when they have a female seminar lead (despite a slightly negative average performance rate). Overall, all students seem to perform better in seminars with female tutors, and they tend to show slightly higher completion in these classes as well. It is important to note that all male tutors were students, and some female tutorial leads were lecturers (including the module lead). As a result, these effects could be correlated with the experience and idiosyncrasies of the tutorial leads beyond their gender identities.

The negative effects of the timed MCQs in both performance and completion persist conditional on other controls across all the specifications in Table 4. Additionally, even though students do not perform significantly better in seminars which take place quite early or very late in the day, they tend to be more tenacious in these seminars, echoing the Table 3 results above. In line with the earlier results, there are no gender effects in performance or completion in timed MCQs, which is in contrast with existing literature such as Ellison & Swanson, (2023) and Griselda, (2020).

Robustness checks

Appendix tables A5 and A6 show the results without any of the absentee students (both "Yes and "No" students discussed in appendix tables A1 and A2) for student performances and completion/participation as robustness checks. This essentially means I rerun the analysis with 791 observations without any type of absentee students for the entire sample. Specifically, I only consider students who attempted all the quizzes in every session. The results are qualitatively the same as the main regression tables above, and there are no gender effects in either the student performance or participation/completion in timed MCQs.

Appendix table A7 shows the results for performance and completion for only module 1 for both years and excludes the module 2 sample as a robustness check. This is important as the cohort for module 1 in the academic year 2023-2024 was significant bigger (more than double) and there were 5 Economics Scholars as tutors in this year (in contrast to 2 Economics Scholars for the year before). Once again, there are no qualitative differences in the results, although for the second academic year there is a slight decrease in student performance and completion but not significant when controlling for other variables. Importantly there are no significant gender effects for timed MCQs in these robustness checks.

Finally, appendix Table A8 adds some additional variables as controls. Across all specifications, I add week numbers as controls, as there is a steady decline in student attendance as evidenced by figures A3 and A4. In columns 2 and 4, I also include tutor type (lecturers vs Economics scholars). Notably, I exclude the variable module lead in these 2 specifications, as module lead and tutor type (when lecturer) are highly correlated. This avoids a potential bad control problem. Columns 1 and 2 have performance and 3 and 4 have completion rates as outcomes. The key results of this paper are not altered in these cases, that is there are no gender effects in any of the outcome variables. The two variables, module lead (or lecturer as tutor) and the number of quantitative questions, are insignificant in these specifications.

Discussion and Conclusion

There are several limitations of this analysis. First, as mentioned above, I cannot control for students' academic ability or the final module performance as the Mentimeter session data are completely anonymous. Hence the results are solely representative of students' behavior in the seminar sessions. Furthermore, the absence of actual attendance data imposes additional

[13] Appendix table A4 shows some other specifications for both performance and completion rates, without controls.

challenges to the completeness of the results. Moreover, I cannot control for other demographic characteristics like racial identity of students, etc. The absence of these demographic controls, which are essential to determine student performances, may have generated significant omitted variable bias in these results. This is perhaps the most significant limitation of this paper. Finally, the Mentimemter data are collected for each session. As a result, these data cannot identify a participant across different seminar sessions. This, in essence, means that I cannot observe the trend of each student behavior in a longitudinal sense throughout the semester.

In conclusion, this paper used anonymous student performance data from Mentimeter quizzes from two modules to examine if the use of this SRS with timed setting is inclusive for female students. Results from 1,238 observations show that there are no gender effects in student performance controlling for the type of quiz they encountered. This result echoes some work which questions the existence of gender differences in performance (Chan & Kennedy, 2002). Specifically, in this context, although all students perform worse in the timed MCQs settings, female students do not perform worse than their male counterparts in these cases. This result could be in line with some recent evidence that high-performing female students show no significant differences in academic performances in MCQs with timed settings (Funk & Perrone, 2023); however, I cannot establish this in the context of this setting.

My finding that students perform worse in timed quizzes compared to un-timed quizzes comes with no surprise, given the vast literature in this area (Gneezy et al., 2003; Markowsky & Beblo, 2022; Walstad & Robson, 1997). Students have more time to ponder on the answers in un-timed MCQs, and they seemingly perform better. They may also feel less stressed in the un-timed quizzes. An important consideration here is that these polls are not part of the summative grades of students. Hence, even though the competitive aspect of timed quizzes is not tied to the grades students receive and can be looked at as formative exercises, the time countdown of 30 seconds may create a more stressful environment. Finally, some studies have found that providing more time does not necessarily improve student performance (Kennedy & Fiester, 2020; Portolese et al., 2016). In this regard, my findings show a contrast, at least for Economics courses, that un-timed quizzes improve student performance in formative quizzes.

The implications of SRS use in classrooms require further exploration. Future work may focus on the longitudinal effects of technology on student performances and classroom participation with demographically linked data. Moreover, in such analysis, it may be useful to combine objective measures, like final exam performances, with subjective outcomes, such as student perceptions of understanding of content and satisfaction.

Data declaration: All data (in pseudonymised form) and accompanying code are available upon request to the author.

References

Barreiro-Gen, M. (2020). Evaluating the effects of mobile applications on course assessment: A quasi-experiment on a macroeconomics course. *International Review of Economics Education*, *34*, 100184.

Bolger, N., & Kellaghan, T. (1990). Method of measurement and gender differences in scholastic achievement. *Journal of Educational Measurement*, *27*(2), 165–174.

Braun, K. W., & Sellers, R. D. (2012). Using a "Daily Motivational Quiz" to Increase Student Preparation, Attendance, and Participation. *Issues in Accounting Education*, *27*(1), 267–279. https://doi.org/10.2308/iace-10219

Cahl\'\ikov\'a, J., Cingl, L., & Levely, I. (2020). How stress affects performance and competitiveness across gender. *Management Science*, *66*(8), 3295–3310.

Chan, N., & Kennedy, P. E. (2002). Are multiple-choice exams easier for economics students? A comparison of multiple-choice and "equivalent" constructed-response exam questions. *Southern Economic Journal*, *68*(4), 957–971.

Coffman, K. B., & Klinowski, D. (2020). The impact of penalties for wrong answers on the gender gap in test scores. *Proceedings of the National Academy of Sciences*, *117*(16), 8794–8803.

Conde-Ruiz, J. I., Ganuza, J. J., & Garc\'\ia, M. (2020). Gender gap and multiple choice exams in public selection processes. *Hacienda Publica Espanola*, *235*, 11–28.

Cotton, C., McIntyre, F., & Price, J. (2013). Gender differences in repeated competition: Evidence from school math contests. *Journal of Economic Behavior & Organization*, *86*, 52–66.

Dunn, P. K., Richardson, A., Oprescu, F., & McDonald, C. (2013). Mobile-phone-based classroom response systems: Students' perceptions of engagement and learning in a large undergraduate course. *International Journal of Mathematical Education in Science and Technology*, *44*(8), 1160–1174.

Elliott, C. (2003). Using a personal response system in economics teaching. *International Review of Economics Education*, *1*(1), 80–86.

Ellison, G., & Swanson, A. (2023). Dynamics of the gender gap in high math achievement. *Journal of Human Resources*, *58*(5), 1679–1711.

Engelhardt, B., Johnson, M., & Siemers, S. (2024). Pedagogical practices and the gender gap in economics education. *Southern Economic Journal*.

Funk, P., & perrone, H. (2023). *Gender Differences in Academic Performance the Role of Exam Design in Multiple-Choice Tests*. https://doi.org/10.2139/ssrn.4401568

Gernsbacher, M. A., Soicher, R. N., & Becker-Blease, K. A. (2020). Four empirically based reasons not to administer time-limited tests. *Translational Issues in Psychological Science*, *6*(2), 175.

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, *118*(3), 1049–1074.

Gokbulut, B. (2020). The effect of Mentimeter and Kahoot applications on university students'e-learning. *World Journal on Educational Technology: Current Issues*, *12*(2), 107–116.

Graetz, G., & Karimi, A. (2022). Gender gap variation across assessment types: Explanations and implications. *Economics of Education Review*, *91*, 102313.

Graham, C. R., Tripp, T. R., Seawright, L., & Joeckel, G. (2007). Empowering or compelling reluctant participators using audience response systems. *Active Learning in Higher Education*, *8*(3), 233–258.

Griffith, A. L. (2014). Faculty gender in the college classroom: Does it matter for achievement and major choice? *Southern Economic Journal*, *81*(1), 211–231.

Griselda, S. (2020). Different Questions, Different Gender Gap: Can the Format of Questions Explain the Gender Gap in Mathematics? *Available in Ideas.Repec: Https://Ideas.Repec.Org/Jmp/2020/Pgr710.Pdf*.

Guarascio, A. J., Nemecek, B. D., & Zimmerman, D. E. (2017). Evaluation of students' perceptions of the Socrative application versus a traditional student response system and its impact on classroom engagement. *Currents in Pharmacy Teaching and Learning*, *9*(5), 808–812.

Guez, A., Peyre, H., & Ramus, F. (2020). Sex differences in academic achievement are modulated by evaluation type. *Learning and Individual Differences*, *83*, 101935.

Hayter, J., & Rochelle, C. F. (2013). Clickers: Performance and attitudes in principles of microeconomics. *Available at SSRN 2226401*.

Heaslip, G., Donovan, P., & Cullen, J. G. (2014). Student response systems and learner engagement in large classes. *Active Learning in Higher Education*, *15*(1), 11–24.

Hill, D. L., & Fielden, K. (2017). Using Mentimeter to promote student engagement and inclusion. *Unpublished Manuscript*.

Imazeki, J. (2014). Bring-your-own-device: Turning cell phones into forces for good. *The Journal of Economic Education*, *45*(3), 240–250.

Johnson, M., Robson, D., & others. (2008). Clickers, student engagement and performance in an introductory economics course: A cautionary tale. *CHEER Virtual Edition*, *20*, 4–12.

Kahn, S., & Ginther, D. (2017). Women and STEM. In *Available in: https://www.nber.org/papers/w23525*.

Kennedy, A. B., & Fiester, S. E. (2020). Reduced performance on examinations following untimed short assessments. *Studies in Educational Evaluation*, *65*, 100835. https://doi.org/10.1016/j.stueduc.2019.100835

Khasawneh, E., Gosling, C., & Williams, B. (2021). What impact does maths anxiety have on university students? *BMC Psychology*, *9*, 1–9.

Kibble, J. D. (2011). Voluntary participation in online formative quizzes is a sensitive predictor of student success. *Advances in Physiology Education*, *35*(1), 95–96. https://doi.org/10.1152/advan.00053.2010

Laxman, K. (2011). A study on the adoption of clickers in higher education. *Australasian Journal of Educational Technology*, *27*(8).

Lumsden, K. G., & Scott, A. (1987). The economics student reexamined: Male-female differences in comprehension. *The Journal of Economic Education*, *18*(4), 365–375.

Madiseh, F. R., Al Abri, A., & Sobhanifar, H. (2022). *Integration of mentimeter into the classroom: A scoping review*.

Madiseh, F. R., Al Abri, A., & Sobhanifar, H. (2023). Integrating Mentimeter to Boost Students' Motivation, Autonomy, and Achievement. *Computer-Assisted Language Learning Electronic Journal*, *24*(3), 232–251.

Markowsky, E., & Beblo, M. (2022). When do we observe a gender gap in competition entry? A meta-analysis of the experimental literature. *Journal of Economic Behavior & Organization*, *198*, 139–163.

Mayhew, E. (2019). No longer a silent partner: How Mentimeter can enhance teaching and learning within political science. *Journal of Political Science Education*, *15*(4).

Mayhew, E., Davies, M., Millmore, A., Thompson, L., & Pena, A. (2020). The impact of audience response platform Mentimeter on the student and staff learning experience. *Research in Learning Technology*, *28*.

Mohin, M., Kunzwa, L., & Patel, S. (2022). Using mentimeter to enhance learning and teaching in a large class. *International Journal of Educational Policy Research and Review*, *9*(2), 48.

Mohr, T. M. (2013). Iclickers and student performance. *International Review of Economics Education*, *14*, 16–23.

Mu, H., & Paparas, D. (2015). Incorporating the advantages of clickers and mobile devices to teach Economics to non-economists. *Cogent Economics & Finance*, *3*(1), 1099802.

Orr, R., & Foster, S. (2013). Increasing Student Success Using Online Quizzing in Introductory (Majors) Biology. *CBE—Life Sciences Education*, *12*(3), 509–514. https://doi.org/10.1187/cbe.12-10-0183

Pichardo, J. I., López-Medina, E. F., Mancha-Cáceres, O., González-Enr\'\iquez, I., Hernández-Melián, A., Blázquez-Rodr\'\iguez, M., Jiménez, V., Logares, M., Carabantes-Alarcon, D., Ramos-Toro, M., & others. (2021). Students and teachers using mentimeter: Technological innovation to face the challenges of the covid-19 pandemic and post-pandemic in higher education. *Education Sciences*, *11*(11), 667.

Porter, C., & Serra, D. (2020). Gender differences in the choice of major: The importance of female role models. *American Economic Journal: Applied Economics*, *12*(3), 226–254.

Portolese, L., Krause, J., & Bonner, J. (2016). Timed Online Tests: Do Students Perform Better With More Time? *American Journal of Distance Education*, *30*(4), 264–271. https://doi.org/10.1080/08923647.2016.1234301

Riener, G., & Wagner, V. (2017). Shying away from demanding tasks? Experimental evidence on gender differences in answering multiple-choice questions. *Economics of Education Review*, *59*, 43–62.

Salemi, M. K. (2009). Clickenomics: Using a Classroom Response System to Increase Student Engagement in a Large-Enrollment Principles of Economics Course. *The Journal of Economic Education*, *40*(4), 385–404. https://doi.org/10.1080/00220480903237950

Santamarina-Campos, V., Carabal-Montagud, M. A., Segarra-Oña, M., & Gasent-Blesa, J. L. (2021). Towards cctive and social learning through the combination of Flipped Classroom and Just-in-Time Learning. *ICERI2021 Proceedings*, 1695–1709. https://doi.org/10.21125/iceri.2021.0464

Saygin, P. O., & Atwater, A. (2021). Gender differences in leaving questions blank on high-stakes standardized tests. *Economics of Education Review*, *84*, 102162.

Skoyles, A., & Bloxsidge, E. (2017). Have you voted? Teaching OSCOLA with Mentimeter. *Legal Information Management*, *17*(4), 232–238.

Stowell, J. R. (2015). Use of clickers vs. mobile devices for classroom polling. *Computers & Education*, *82*, 329–334.

von Schrader, S., & Ansley, T. (2006). Sex differences in the tendency to omit items on multiple-choice tests: 1980–2000. *Applied Measurement in Education*, *19*(1), 41–65.

Walstad, W. B., & Robson, D. (1997). Differential item functioning and male-female differences on multiple-choice tests in economics. *The Journal of Economic Education*, *28*(2), 155–171.

Wood, A. (2020). Utilizing technology-enhanced learning in geography: testing student response systems in large lectures. *Journal of Geography in Higher Education*, *44*(1), 160–170.

Appendix: Additional Graphs and Tables

Figure A1: Demographic question asked in sessions

Figure A2: Information provided to participants in sessions

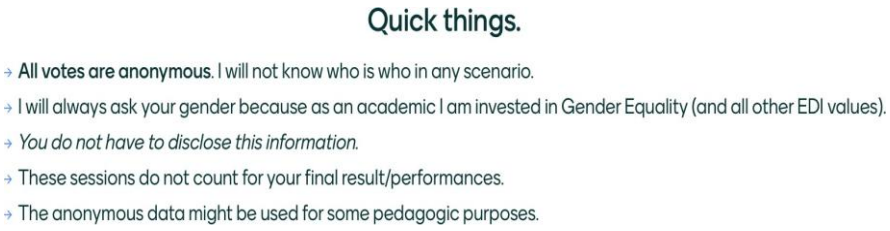


Figure A3: Number of participants by week for course 1

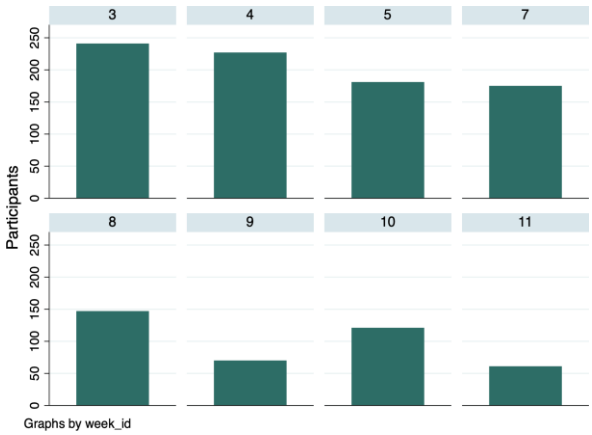


Figure A4: Number of participants by week for course 2

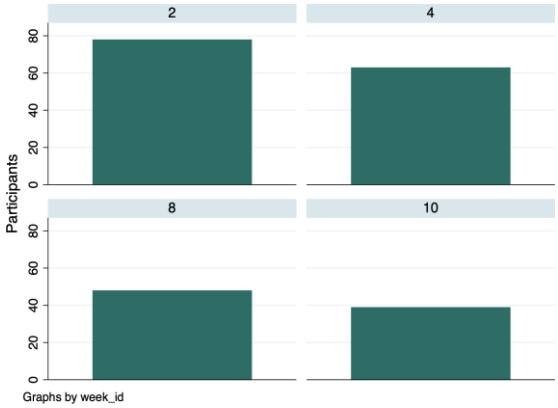


Table A1: Absentees by gender over two modules for year 1

Module	Absentees	Female	Male	Total
1	No	35	35	70
1	Yes	11	10	21
Total		**46**	**45**	**91**
2	No	23	47	70
2	Yes	11	14	25
Total		**34**	**61**	**95**

Table A2: Absentees by gender over module 1 for both years

Year	Absentees	Female	Male	Total
1	No	35	35	70
1	Yes	11	10	21
Total		**46**	**45**	**91**
2	No	96	56	152
2	Yes	59	25	84
Total		**155**	**81**	**236**

Figure A5: Do students 'perform' differently in timed vs untimed quizzes overall?

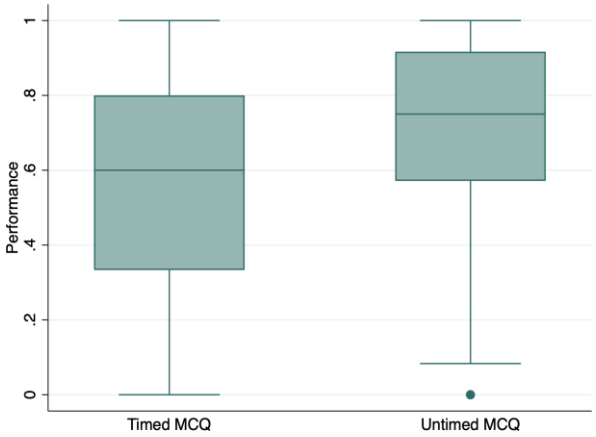


Figure A6: Do students show different 'completion' in timed vs untimed quizzes overall?

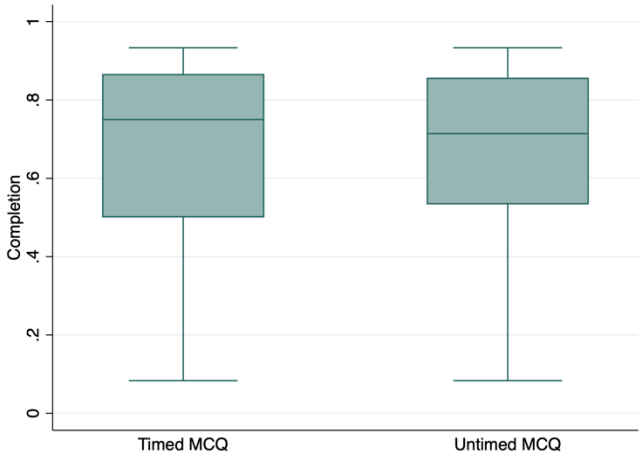


Figure A7: Do students 'perform' differently in timed vs un-timed quizzes, by module?

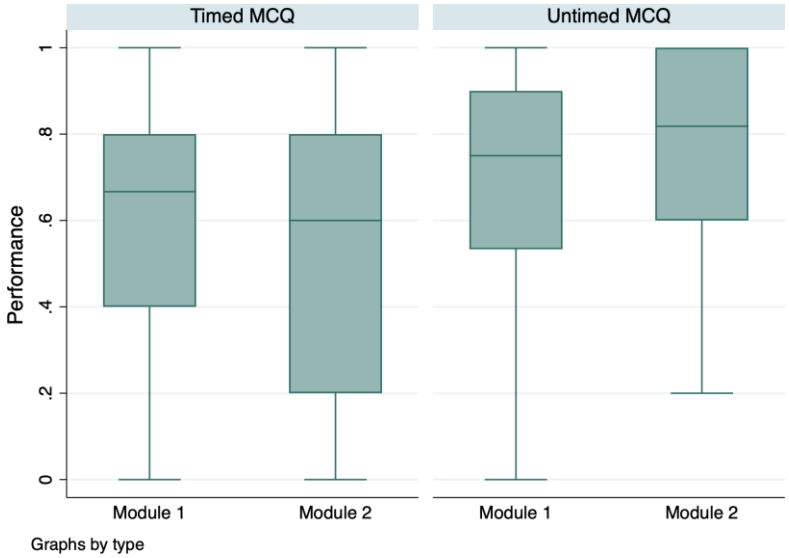


Figure A8: Do students show different 'completion' in timed vs untimed quizzes, by module?

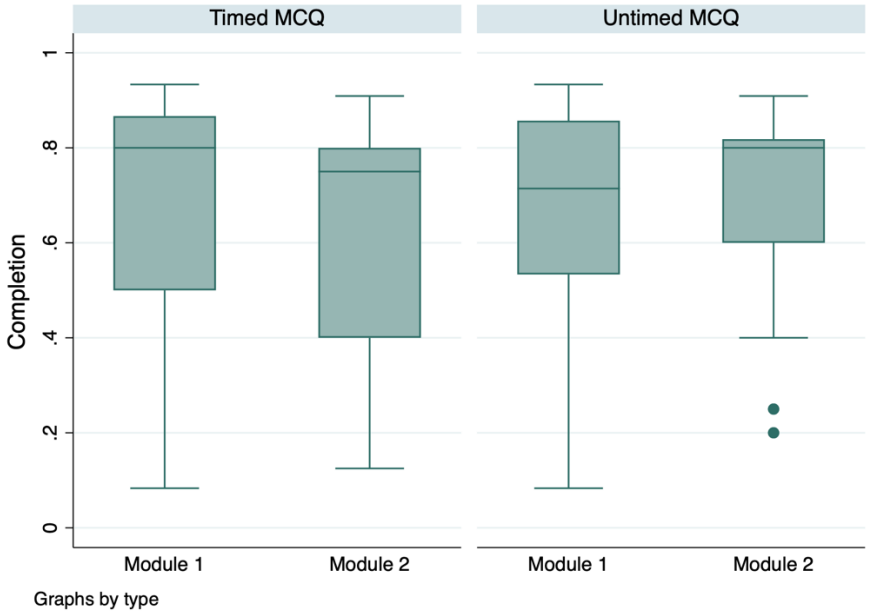


Figure A9: Do students show different 'completion' in timed vs untimed quizzes across gender identity?

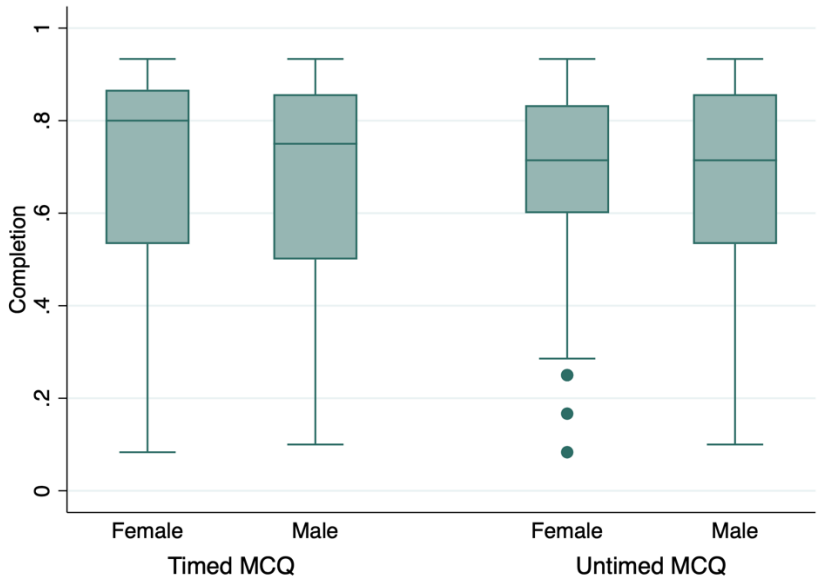


Table A3: Student performance and completion in different types of quizzes with female seminar leads

	Outcome: Performance			Outcome: Completion		
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Gender	-0.0087 (0.0166)		-0.0054 (0.0159)	-0.0060 (0.0153)		-0.0043 (0.0152)
Timed MCQ		-0.1623*** (0.0158)	-0.1621*** (0.0158)		-0.0826*** (0.0151)	-0.0825*** (0.0151)
Constant	0.6320*** (0.0121)	0.7092*** (0.0103)	0.7120*** (0.0132)	0.8266*** (0.0111)	0.8651*** (0.0093)	0.8674*** (0.0123)
Observations	1238	1238	1238	1238	1238	1238
Adjusted R-squared	-0.00	0.08	0.08	-0.00	0.02	0.02
Mean Dependent Variable	0.63	0.63	0.63	0.82	0.82	0.82

Notes: Dependent variables are Performance and Completion rates. Performance rate is defined as correct answers over total number of questions. Completion rate is defined as attempted questions over total number of questions. * p<0.10, ** p<0.05, *** p<0.01.

Table A4: Student performance and completion in different types of quizzes with female seminar leads

	Outcome: Performance			Outcome: Completion		
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Female Tutor	0.0524**	0.0567***	0.0478**	0.0329**	0.0459***	0.0261
	(0.0165)	(0.0177)	(0.0193)	(0.0155)	(0.0164)	(0.0180)
Weakly Seminars		0.0152	0.0359		0.0460**	0.0920***
		(0.0257)	(0.0292)		(0.0234)	(0.0263)
Module Lead			0.0363			0.0810***
			(0.0255)			(0.0224)
Constant	0.5975***	0.5823***	0.5616***	0.8047***	0.7587***	0.7127***
	(0.0121)	(0.0284)	(0.0316)	(0.0118)	(0.0262)	(0.0288)
Observations	1238	1238	1238	1238	1238	1238
Adjusted R-squared	0.01	0.01	0.01	0.00	0.01	0.01
Mean Dependent Variable	0.63	0.63	0.63	0.82	0.82	0.82

Notes: Dependent variables are Performance and Completion rates. Performance rate is defined as correct answers over total number of questions. Completion rate is defined as attempted questions over total number of questions. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table A5: Student performances in different types of quizzes - Robustness

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Gender	0.0023		0.0030	0.0027	0.0117	0.0126	0.0155
	(0.0164)		(0.0162)	(0.0206)	(0.0204)	(0.0204)	(0.0205)
Timed MCQ		-0.0721***	-0.0721***	-0.0725***	-0.0674***	-0.0673***	-0.0627**
		(0.0165)	(0.0165)	(0.0246)	(0.0244)	(0.0245)	(0.0249)
Gender * Timed MCQ				0.0007	-0.0052	-0.0045	-0.0073
				(0.0332)	(0.0330)	(0.0330)	(0.0332)
Module Lead					0.0763***	0.0866***	0.0697***
					(0.0169)	(0.0192)	(0.0214)
Timing of Seminar						-0.0189	-0.0099
						(0.0203)	(0.0210)
No. Quant Questions							0.0112*
							(0.0061)
Constant	0.7391***	0.7710***	0.7694***	0.7695***	0.7466***	0.7476***	0.7408***
	(0.0122)	(0.0102)	(0.0137)	(0.0154)	(0.0166)	(0.0168)	(0.0174)
Observations	791	791	791	791	791	791	791
Adjusted R-squared	-0.00	0.02	0.02	0.02	0.04	0.04	0.04
Mean Dependent Variable	0.74	0.74	0.74	0.74	0.74	0.74	0.74

Notes: Dependent variable: performance rate is defined as correct answers over total number of questions. We exclude all the absentee students from this analysis. * p<0.10, ** p<0.05, *** p<0.01

Table A6: Student completion in different types of quizzes - Robustness

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Gender	0.0044		0.0045	0.0060	0.0152	0.0131	0.0173
	(0.0107)		(0.0107)	(0.0136)	(0.0135)	(0.0134)	(0.0135)
Timed MCQ		-0.0061	-0.0061	-0.0042	0.0010	0.0009	0.0077
		(0.0108)	(0.0108)	(0.0167)	(0.0163)	(0.0162)	(0.0164)
Gender * Timed MCQ				-0.0035	-0.0095	-0.0111	-0.0152
				(0.0219)	(0.0214)	(0.0214)	(0.0213)
Module Lead					0.0779***	0.0536***	0.0286***
					(0.0070)	(0.0059)	(0.0055)
Timing of Seminar						0.0446***	0.0579***
						(0.0052)	(0.0065)
No. Quant Questions							0.0166***
							(0.0023)
Constant	0.9385***	0.9434***	0.9411***	0.9403***	0.9168***	0.9145***	0.9045***
	(0.0081)	(0.0067)	(0.0091)	(0.0104)	(0.0115)	(0.0116)	(0.0123)
Observations	791	791	791	791	791	791	791
Adjusted R-squared	-0.00	-0.00	-0.00	-0.00	0.04	0.05	0.07
Mean Dependent Variable	0.94	0.94	0.94	0.94	0.94	0.94	0.94

Notes: Dependent variable: completion rate is defined as total attempts over total number of questions. We exclude all the absentee students from this analysis. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table A7: Student outcomes for module 1 in both years - Robustness

	Outcome: Performance		Outcome: Completion	
Variables	(1)	(2)	(3)	(4)
Gender	-0.0053	-0.0086	0.0060	0.0004
	(0.0228)	(0.0227)	(0.0207)	(0.0204)
Timed MCQ	-0.1344***	-0.1370***	-0.0571**	-0.0609**
	(0.0261)	(0.0261)	(0.0245)	(0.0243)
Gender * Timed MCQ	-0.0129	-0.0121	-0.0173	-0.0139
	(0.0346)	(0.0345)	(0.0327)	(0.0323)
Year 2	-0.0324*	-0.0088	-0.0713***	-0.0299
	(0.0182)	(0.0233)	(0.0162)	(0.0201)
Module Lead		0.0042		-0.0118
		(0.0335)		(0.0291)
Timing of Seminar		0.0436*		0.0931***
		(0.0239)		(0.0222)
No. Quant Questions		-0.0135		0.0629***
		(0.0229)		(0.0199)
Constant	0.7237***	0.7019***	0.9085***	0.8518***
	(0.0211)	(0.0262)	(0.0187)	(0.0224)
Observations	1039	1039	1039	1039
Adjusted R-squared	0.06	0.06	0.03	0.05
Mean Dependent Variable	0.63	0.63	0.82	0.82

Notes: performance rate is defined as correct answers over total number of questions and completion is defined as attempted questions over total number of questions. We exclude module 2 observations from this sample. *p<0.10, **p<0.05, ***p<0.01

Table A8: Student performance and completion in different types of quizzes – extra controls

	Outcome: Performance		Outcome: Completion	
Variables	(1)	(2)	(3)	(4)
Gender	-0.0098	-0.0108	-0.0029	-0.0037
	(0.0200)	(0.0200)	(0.0187)	(0.0187)
Timed MCQ	-0.1616***	-0.1617***	-0.0765***	-0.0766***
	(0.0227)	(0.0227)	(0.0217)	(0.0216)
Gender * Timed MCQ	0.0089	0.0084	-0.0091	-0.0095
	(0.0308)	(0.0308)	(0.0296)	(0.0296)
Module Lead	0.0360		0.0236	
	(0.0261)		(0.0252)	
Lecturers		0.0070		-0.0018
		(0.0289)		(0.0257)
Timing of Seminar	0.0369*	0.0510**	0.0913***	0.1034***
	(0.0206)	(0.0208)	(0.0193)	(0.0190)
No. Quant Questions	-0.0032	0.0011	-0.0132	-0.0092
	(0.0084)	(0.0093)	(0.0083)	(0.0085)
Constant	0.8006***	0.8194***	0.9099***	0.9283***
	(0.0338)	(0.0374)	(0.0330)	(0.0334)
Observations	1238	1238	1238	1238
Week No. as controls	Yes	Yes	Yes	Yes
Adjusted R-squared	0.13	0.13	0.08	0.08
Mean Dependent Variable	0.63	0.63	0.82	0.82

Notes: Performance rate is defined as correct answers over total number of questions. Completion rate is defined as attempted questions over total number of questions. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Example quiz questions

Module 1

1. If MC=$5 and FC/Q=15, then what is AC?
   - $15
   - $5
   - $20
   - $10
2. Acquiring a firm that sells a substitute good would make the demand curve for youroriginal product
   - More inelastic
   - More elastic
   - Unchanged
   - None of the options

Module 2

1. At an annual interest rate of 7%, the future value of $5000 in five years is closest to:
    - $3565
    - $6750
    - $7013
    - $7035
    - $7050
2. When using the internal rate of return (IRR) investment rule, we compare
    - the average return on the investment opportunity to returns on all other investmentopportunities in the market.
    - the average return on the investment opportunity to returns on other alternatives inthe market with equivalent risk and maturity.
    - the NPV of the investment opportunity to the average return on the investmentopportunity.
    - the average return on the investment opportunity to the risk-free rate of return.
    - None