AJET | ASCILITE

The effect on student behaviour and achievement of removing incentives to complete online formative assessments

Stephen Agnew
University of Canterbury, Christchurch, New Zealand

Jane Kerr
Clinical and Pharmaceutical Research Trust, Christchurch, New Zealand

Richard Watt
University of Canterbury, Christchurch, New Zealand

This research explored the effect of incentives to complete online quizzes during a course. When a 1% weighting incentive per quiz was removed, student engagement dropped dramatically. There is also evidence that students who continued to complete quizzes did so with less vigour, spending less time on each quiz, starting them later in the week and having fewer attempts per quiz on average. The average mark per quiz was also lower once an incentive to complete them was removed. There was no significant difference in examination performance based on how many quizzes a student continued to attempt after incentives were removed. However, a comparison to a control group of students who sat the same invigilated assessments showed that, relative to term test performance, final exam mean score was 7% lower for the cohort who had the incentive to complete online quizzes removed. This differed from the control group, who showed no difference between term test and final examination mean scores when quiz incentives were maintained for the entire course. Building on previous research, this study demonstrates that a binary variable representing engagement in online quizzes did not capture the quality of that online engagement.

Implications for practice or policy:
- Completion of online formative assessments by students is reduced if course leaders remove small-stakes incentives.
- The removal of small-stakes incentives by course leaders harms student motivation and achievement. Specifically, students who complete formative online assessment without incentives have fewer attempts, start them later in the availability window, spend less time completing them and record a lower mean score than those with incentives.
- Average final examination achievement is lower when incentives to complete online formative assessments are removed by course leaders.

Keywords: online assessment, quizzes, incentives, formative assessment, quantitative

Introduction

As technology has developed, so has the ability to deliver continuous assessment quickly and efficiently to large groups of students. This has been particularly impactful in the tertiary sector, where previously large physical spaces were required to assess the student body. Now many courses contain online quizzes as part of their assessment, with a small mark allocation to encourage students to complete them. The need to incentivise good work habits has been recognised (Scheyvens et al., 2008; Williams, 1997), with Cook and Babon (2017), stating "while students typically recognize the value of undertaking prescribed preparatory work prior to lectures, this is generally insufficient motivation for them to regularly complete preparatory readings" (p. 25). Research has also discussed the benefits of continuous assessment in terms of improving student learning (Rezaei, 2015) and as a way of encouraging students to remain engaged in a course (Holmes, 2015). In their 2018 study, Day et al. suggested that continuous assessment is a useful tool for lecturers, to encourage consistent work effort throughout a course. Quantifying the benefits of continuous online assessment, however, can be challenging. As discussed in the Literature review section, studies have measured the effect of continuous online assessment on student perceptions of their learning, rather than assessing the impacts of learning itself. Other research has found that students who engage in continuous online assessment are more likely to achieve higher marks on summative, high-stakes assessment such as examinations. These findings do, however, ignore the possibility that there is no causal element to this

relationship. In other words, completion of small-stakes online continuous assessment identifies students who are doing better academically because they are engaged, motivated students rather than identifying any inherent benefit from actually completing the online assessment tasks. This is an important consideration. Time spent setting up, delivering and monitoring continuous online assessment has an opportunity cost. If online quizzes identify higher achieving students, rather than improving achievement, they may not be the best use of resources. This study fills a gap in the literature by quantifying any change in student behaviour when removing an incentive to complete online quizzes. Secondly, a more nuanced approach measured the quality as well as quantity of engagement in online quizzes. Lastly, a between-group comparison measured the effect of continuous online assessment on achievement in high-stakes examinations.

Literature review

While undertaking a study examining testing of middle-grade science students, McDaniel et al. (2011) theorised that quiz completion should promote learning through processing material, with subsequent retrieval and retention. Others have also promoted the benefit to learning of feedback provided by quizzes to a broader age range of students (Butler & Roediger, 2008; Carpenter & DeLosh, 2006; McDaniel & Masson, 1985; McDaniel et al., 2011; McDaniel et al. 2012; Pashler et al., 2005; Roediger & Karpicke, 2006). In a study of 458 undergraduate geography students at the University of Melbourne, Cook and Babon (2017) evaluated the use of weekly online quizzes (worth 20% of the course grade) as an incentive for students to complete preparatory reading, with the goal of encouraging active learning. They used student course evaluations to examine student perceptions. They found that quizzes had a positive effect in encouraging preparatory reading completion; there was also a high level of engagement with the quizzes, with average completion rates of 94%. They did not attempt to measure the impact of quiz completion on overall course grades.

There exists a body of research examining the testing effect – improvements in performance on final examinations due to practice tests on the same material (Adesope et al., 2017). In a 2017 meta-analysis, Adescope et al. concluded there is a positive impact on learning (and hence achievement on high-stakes final examinations) of having prior testing on the same material. They found that practice multiple-choice tests have the largest effect size; however, there is a question mark over the longevity of the effect on long-term retention. In a recent 2018 study of university students, however, Greving and Richter found no testing effect from multiple-choice testing when retention of learning content was tested 1, 12 and 23 weeks after the last lecture.

Heterogeneity between variables adds complexity to proving a link between online engagement in a university course and achievement. Studies have analysed correlations between levels of engagement and academic achievement. In her 2020 study, Nieuwoudt compared the achievement of groups of students with varying levels of online engagement. Whilst factors such as whether classes were synchronous or asynchronous, influenced levels of engagement, Nieuwoudt found a positive relationship between levels of engagement and final grades. When higher examination results are correlated with greater levels of online quiz engagement, it is difficult to establish causality. Are higher performing students more likely to complete online quizzes, without any substantial improvement in their learning? Or is it a direct result of quiz completion that a student obtains higher levels of achievement? (Thomas et al., 2017). If it is the former, the time taken to set up and monitor online quizzes may be better spent in other areas.

Hoskins and Van Hooff (2005) attempted to answer two research questions: Firstly, which students voluntarily utilise web-based learning? And, also, does this use influence their academic achievement? To this end, Hoskins and Van Hooff identified students who engaged in web-based learning in a second-year psychology class and recorded the number and length of accesses and use of the bulletin board. They found that older students were more likely to engage online, along with higher achieving students. In addition, they deduced that students who were more engaged in the web-based environment supporting the class were significantly more likely to achieve better academic results. However, they did not establish a causal relationship between web-based engagement and higher academic results.

The conclusion that more academically able students are more likely to engage in bulletin board use is equally plausible.

There is a body of literature examining the importance of self-directed learning (SDL) and self-regulated learning (SRL). These terms contain similarities, predominantly an underlying personality perspective, and have been used synonymously in the past (Onah et al., 2020). There are nuanced differences, however, with SDL considered to be a more macro concept, with learners potentially deciding and defining their learning tasks, while with the more micro concept of SRL, students may regulate their learning under a third party's direction (Onah et al., 2020). The course in which participants were drawn for this study had learning outcomes defined by the course lecturer. The ability of students to self-regulate their learning could therefore be an influential factor in determining student success. Studies have confirmed that students with more developed SRL skills academically outperform students with less developed SRL skills (Barnard-Brak et al., 2010; Chen & Huang, 2014; Schunk, 2005; Yen et al., 2018). When using a between-group experimental design, Uz and Uzen (2018) found that those who received a blended instruction approach developed SRL skills better than those in the group exposed to traditional instruction. Students in the treatment group reported that the blended learning environment increased motivation (Uz & Uzen, 2018). However, without the ability to establish causality, online metrics such as bulletin board use could still be interpreted as an identifying characteristic of a more academically able student rather than as a direct contributing factor that improves their academic performance. This sentiment is echoed in the work of Johnson (2006), where few of the 112 students studied made use of optional online quizzes designed to prepare students for examinations. Johnson found completion of the online quizzes to be correlated with improved attainment on an invigilated examination. She noted that quiz use may have caused higher achievement or that higher performing students may have been more likely to complete the quizzes. Similar findings resulted from J. D. Kibble et al. (2011), who reported a positive correlation between online quiz participation and performance on summative assessment. When comparing quiz and final performance for students, Gholami and Zhang (2018) indicated that "some behavioural patterns have strong implications for a student's final learning outcome in the course, with more active study habits being correlated with stronger final performance" (p. 1).

While acknowledging a number of "confounding factors in student performance", Angus and Watson (2009, p. 255) examined behavioural patterns. They used a regression analysis that incorporated variables for prior student aptitude, gender and in-course mastery, including the number of quiz attempts, rather than quiz mark, as a proxy for effort. They also identified attendance at a voluntary support scheme as a proxy for student effort. The main finding was that higher exposure to online quizzes was correlated with increased scores on the final exam, but the exposure was not causative.

A subset of the literature has examined how incentivising quiz completion through allocating a small portion of the final course grade to online quiz completion can change student behaviour. In a study examining the behaviour of medical sciences students, J. Kibble (2007) applied a range of different settings between zero weighting and 2% for each online quiz, across different cohorts of students. The different conditions when sitting the quizzes included an unlimited number of attempts, credit for attempting a quiz, credit based on the mark achieved on the quiz, a limited number of attempts within a 1-week window and unlimited attempts. Two of J. Kibble's models were similar to settings compared in this study. When offered one attempt within a 1-week window, 52% of students attempted the quizzes. But, when students were offered a 1% credit for the better of two allowable quiz attempts in a 1-week window, the proportion of students who participated increased to 97%. In a follow-up survey, more than 90% of students indicated that they had completed quizzes to earn credit (J. Kibble, 2007).

While attempting to identify a relationship between self-assessment quiz taking behaviour and final exam scores, Ozarslan and Ozan (2016) also found the correlation that learners who regularly completed voluntary self-assessment quizzes also achieved better results on the final exam. A secondary finding was the lack of a correlation between the number of quiz attempts and the final exam score, when multiple quiz attempts were allowed (Ozarsland & Ozan, 2016). However, similarly to other studies (Jensen et al., 2014; McDaniel et al., 2012), it could be that higher ability students are more likely to attempt and do better, on both formative and summative assessments.

Other researchers have cautioned against using small-stakes assessment as motivation to encourage learning, rather than an intrinsic motivation, warning that students may see the assessments as a means of *mark harvesting* rather than as learning opportunities (Cook & Babon, 2017; Williams, 1997). The marks weighting given to continuous, small-stakes assessments also needs to be considered as this is "often the main signal in determining the amount of effort a student will put into a task" (Cook & Babon, 2017). An advantage of continuous online assessment is that learners are given an opportunity to self-assess their own progress. McMillan and Hearn (2008, as cited in Ozarslan & Ozan, 2016, p. 15) defined this process as occurring "when students judge their own work to improve performance as they identify discrepancies between current and desired performance".

Aims of the research

A constraint of much of the previous research evaluating this aspect of education is the difficulty of conducting randomised, between-group experiments where there is the potential for one of the groups to perform better academically as a result of the experiment. The difficulty outlined above in differentiating between causality and correlation when assessing the impact of student engagement on student achievement is magnified when using cross-sectional data gathered from convenience samples of university students.

The current study attempted to circumvent these challenges by using an integrated approach to answer the following research questions:

(1) How does changing the incentives to complete low-stakes online quizzes impact on student behaviour in terms of quiz completion?
(2) Can a causal relationship be established between examination marks and student engagement in online quizzes? And if so, what are the findings?

Rather than comparing students' differing levels of online engagement across the duration of a course, this project incorporated the dimension of adjusting the incentives for students to complete online quizzes within a course. If reducing the incentive to complete online quizzes changes the behaviour of both higher and lower performing students, it may be possible to tease out causal links by comparing academic achievement before and after changes in incentives.

Method

In 2018, a first-year introductory economics course of 337 students at a mid-range university was selected as the cohort for this study, as it contained a mixture of students studying towards commerce, science, arts and engineering degrees. This allows findings to be considered generalisable across different student cohorts. The course consisted of one semester, broken into two terms of 7 and 5 weeks respectively, with a 2-week mid-semester break in the middle. For each of the last 5 weeks of Term 1, students completed a weekly online quiz worth 1% each. For these 5 weeks, a quiz opened at 0900hrs each Wednesday and closed at 23.55hrs that Friday. Within that time period, each student was allowed a maximum of two attempts on that week's quiz, with their highest mark being recorded. Each quiz attempt was given up to 60 minutes in duration for a 10-question multiple-choice quiz. The time allowed for each attempt was deliberately generous so the students could self-assess on their first attempt, then make use of their notes to improve performance in their second attempt. The goal was for each weekly quiz to be treated as a learning experience, with students being told whether their answers were correct or not, thus highlighting areas of content requiring more reinforcement. On the last day of term, the students sat a higher-stake term test with a 25% weighting. In order to examine any potential change in student motivation, data from a range of behavioural variables were collected for each student, namely:

- the day a particular quiz was first attempted
- the duration in minutes and seconds of each student's first quiz attempt
- the number of attempts a student made at each quiz
- the number of quizzes a student had at least one attempt at.

The quizzes were intended to give students an opportunity to self-assess, judge and improve their own learning. Consequently, the earlier a student attempted a quiz in the Wednesday to Friday window, the more time they had to complete this process. When the methodology for this study was first devised, the amount of time a student spent completing a quiz was perceived to represent how much effort a student was putting into the self-assessment process. A quiz of short duration could represent a glib attempt at the quiz or a quick attempt at mark harvesting. Similarly, the number of attempts for each quiz could be representative of effort: one attempt to get a satisfactory mark or two attempts when actively engaging in the self-assessing and subsequent improvement process. However, for a very able student, a short single attempt could achieve full marks, indicating an existing good grasp of the content. Lastly, we deemed the number of quizzes a student attempted at least once to be a variable to measure overall effort in the course and engagement in self-assessment and improvement. Each student's term test percentage was recorded, for use as a dependant variable in Term 1 and a proxy for ability in the Term 2 analysis.

The students had been informed at the start of the course through lectures, the course website and the course outline that the 5 weeks of Term 2 would each contain a quiz following the same parameters as Term 1, with a final examination at the end of the course. The final examination had a 55% weighting, with 25% of the marks coming from material covered in Term 1 and 75% of the marks allocated to material covered in Term 2. During the semester break and the first week of the second term, students were informed of an assessment regime change, as follows:

> We are trialling a different assessment regime this semester:
> You all need to go to the "5% for quizzes 6 to 10" tab under the course menu on the Learn web page. When there, you will see some instructions, and a link where you can choose one of two options.
> Option A:
>   By clicking "Yes" you will receive 5% from quizzes 6 to 10 whether or not you complete them. Note that you will still be able to complete the tutorial quizzes for learning purposes if you want to, you will just receive the maximum of 5% available regardless of the marks you score on each tutorial quiz. If you click yes, you will receive the 5% available from quizzes 6 to 10 even if you do not attempt them.
> Option B:
>   By clicking the "No" box you will receive the mark you earn out of 5% from completing quizzes 6 to 10.

The value to learning of continuing to complete the quizzes was reinforced to students on numerous occasions.

Of the 331 students, 277 chose option A, 7 chose option B, with 47 not responding. The 47 students who did not respond were considered to be disengaged from the course and were removed from the sample. In addition, the 7 students who chose option B were also removed. They demonstrated no incentive to change their behaviour, as they continued on the same assessment regime as Term 1. This left a sample of 277 students. The same data was then collected for the duration of Term 2 as was collected in Term 1, with term test percentage replaced by final examination percentage.

To allow for ordinary least squares (OLS) regressions to be carried out, several of the variables were recoded:

(a) The day a quiz was first attempted variable was recoded as 1 for Wednesday, 2 for Thursday and 3 for Friday. An average was then calculated for each student, for each term (*average day quiz commenced*).
(b) An average of the duration of each student's first quiz attempt was calculated for each term (*average quiz duration*).
(c) An average of the number of attempts a student took for each quiz was calculated for each term (*average quiz attempts*).
(d) A fourth variable was also derived showing the number of quizzes each student had at least one attempt at for each term (*quizzes completed*).

As mentioned earlier, these eight variables (four variables repeated across two terms) could be considered behavioural variables representing particular approaches students took to the completion of their quizzes in each term. A larger number in variable (a) could be thought of as a student being more likely to leave their first quiz attempt until the last minute and treating the quizzes less as a learning and improving opportunity, and more as mark accumulation. Variables (b) and (c) could be considered effort variables, with a smaller average duration and lower number of quiz attempts representing students who are rushing through their quizzes or treating them more glibly. Equally, these variables could potentially be ability variables, where good students required less time and fewer attempts to gain maximum marks for each quiz. Variable (d) can confidently be interpreted as an effort variable, with fewer quizzes attempted indicative of less motivated students.

An OLS regression was run to establish if the eight behavioural variables were correlated with the dependant variable of term test percentage. In line with previous research outlined in the literature review above, final examination percentage was also included as an independent variable as a proxy for student ability. The model (Model 1) therefore took the following form:

$$\text{Term Test \%} = \alpha + \beta_1 \text{ Count2} + \beta_2 \text{ Day2} + \beta_3 \text{ Duration2} + \beta_4 \text{ Attempts2} + \beta_5 \text{ Count1} + \beta_6 \text{ Day1} + \beta_7 \text{ Duration1} + \beta_8 \text{ Attempts1} + \beta_9 \text{ Final Examination \%}$$

Where:

Term Test % = grade received in the term test at the end of Term 1
A = intercept
Count2 = the number of the five quizzes in term 2 that were attempted at least once
Day2 = the average day of starting the five quizzes in Term 2
Duration2 = the average length of the first quiz attempt for the 5 quizzes in Term 2
Attempts2 = the average number of attempts for each of the 5 quizzes in Term 2
Count1 = the number of the five quizzes in term 1 that were attempted at least once
Day1 = the average day of starting the five quizzes in Term 1
Duration1 = the average length of the first quiz attempt for the 5 quizzes in Term 1
Attempts1 = the average number of attempts for each of the 5 quizzes in Term 1
Final Examination % = the % earned on the final examination.

A similar regression was run for Term 2, with a dependant variable of final examination percentage and term test percentage included as a proxy for student ability, giving the model (Model 2):

$$\text{Final Examination \%} = \alpha + \beta_1 \text{ Count2} + \beta_2 \text{ Day2} + \beta_3 \text{ Duration2} + \beta_4 \text{ Attempts2} + \beta_5 \text{ Count1} + \beta_6 \text{ Day1} + \beta_7 \text{ Duration1} + \beta_8 \text{ Attempts1} + \beta_9 \text{ Term Test \%}$$

All eight behavioural variables were included in each regression in an attempt to capture any changes in behaviour between Term 1 and Term 2.

The same term test and final examination used in the course had been used in a previous offering of the same course and therefore could be used to establish a baseline performance on the final examination relative to the term test. This essentially made the previous cohort of students, who had no change in incentive throughout the course, into a control group. A difference in differences approach could then be used to see if examination performance relative to term test performance for the group with changing incentives was substantively different from the previous control cohort.

The assumptions of OLS regressions were all satisfactorily met. Initial concerns that multi-collinearity may have been an issue between some of the variables were unfounded as the variable inflation factor was less than 2.5 for all variables apart from *quizzes completed: Term 2* (7.2) *and average quiz attempts: Term 2* (6.2), which were both still less than 10.

To ensure that no student was unfairly disadvantaged, all 331 students in the course, including those who did not respond, were given 5% for quiz completion in Term 2. This research was conducted within the ethical guidelines pertinent to our institution.

Results and discussion

The first research question sought to establish whether and how student behaviour would change with the removal of incentives to complete the quizzes. As can be seen in Table 1, there is a significant change in student behaviour. In Term 1, 55% of the sample completed all five Term 1 quizzes; however, only 6% completed all five Term 2 quizzes with the incentive of 1% for each quiz removed. At the other extreme, where only 1% completed no quizzes in Term 1, 51% completed no quizzes in Term 2. For those students who continued to attempt the quizzes in Term 2, their quiz score on average is lower in Term 2 than in Term 1, suggesting less effort was forthcoming in Term 2 quiz attempts.

Table 1
Percentage of students completing quizzes

Quizzes completed	Term 1	Term 2	Term 1 average quiz score	Term 2 average quiz score
5 quizzes	55%	6%	0.90	0.81
4 quizzes	26%	9%	0.83	0.76
3 quizzes	12%	9%	0.71	0.75
2 quizzes	4%	10%	0.77	0.69
1 quiz	1%	14%	0.63	0.59
0 quizzes	1%	51%	0.00	0.00

This finding is substantiated when between-term comparisons are made for the other behavioural variables, as shown in Table 2:

Table 2
Student behaviour when completing quizzes

Student behaviour	Term 1	Term 2
Average time commenced (Wed = 1, Thurs = 2, Fri = 3)	2.24	2.33
Average duration	14.95	11.63
Average # of attempts (1 or 2)	1.65	1.50

On average, those students who continued to complete the quizzes in Term 2 started the quizzes later, spent less time completing them and had fewer attempts at each quiz when compared to Term 1.

To answer the first research question, this study therefore demonstrates that the removal of the 1% per quiz incentive resulted in a large behaviour change, where over half of the sample did not attempt a single quiz in Term 2, a large change from the 1% in Term 1. There is also evidence that those students who continued to complete the quizzes in Term 2 did so with less enthusiasm.

The results of the OLS regressions run to answer the second research question of whether there is a casual impact on examination percentage of behaviour changes in quiz completion are shown in Table 3.

Table 3
OLS regression coefficients examining quiz behaviour and summative assessment scores

Variables	Term test %	Examination %
Quiz count Term 2	0.878 (0.692)	-0.740 (0.704)
Average day quiz commenced Term 2	-2.645 (1.573)	2.208 (1.649)
Average quiz duration Term 2	0.042 (0.110)	-0.030 (0.115)
Average number of attempts per quiz Term 2	2.526 (2.186)	-1.584 (2.291)
Quiz count Term 1	1.606** (0.799)	1.497 (0.838)
Average day quiz commenced Term 1	0.391 (1.524)	-5.064*** (1.564)
Average quiz duration Term 1	-0.268** (0.106)	0.120 (0.112)
Average number of attempts per quiz Term 1	2.156 (2.850)	-4.179 (2.975)
Term test %	--------	0.757*** (0.045)
Final examination %	0.691*** (0.041)	--------
Adjusted R^2	0.595	0.593

** and *** denote statistical significance at the 5 and 1% levels respectively.

For Model 1, none of the Term 2 behaviour variables correlate with term test percentage. This is predictable, as none of the Term 2 content was covered in the term test, with the term test being sat before the incentives for quiz completion were removed. There are, however, two Term 1 behaviour variables that correlate with term test percentage at the 95% confidence level (CI). For every extra quiz a student completed in Term 1, there was a 1.6% increase in term test percentage. This suggests that doing quizzes is beneficial for student learning as measured by the term test results. It is probable that high achieving students are more likely to complete more quizzes, meaning the quiz count Term 1 variable could be a proxy for student ability. However, the fact that quiz count Term 1 is still significant even with the inclusion of final examination percentage as an independent variable in the regression, suggests the increase in term test percentage is being driven by the benefit of completing more quizzes.

The second variable that significantly correlates with term test percentage is the average quiz duration in Term 1 variable. Somewhat surprisingly, the negative coefficient shows that for every minute longer spent on average on the quizzes in Term 1 there is approximately 0.25% decrease in term test percentage. This suggests that rather than being a behavioural variable, average quiz duration may in fact be another proxy for ability, with more-able students completing the quizzes in less time. As expected, final examination percentage is closely correlated with term test percentage at the 99% CI, with every 1% increase in term test percentage correlated with a 0.69% increase in final examination percentage. The model explains 59.5% of the variation in term test mark.

Given that two of the quiz-related variables from Term 1 correlate with term test percentage, an expectation could arise that some Term 2 quiz variables might correlate with final examination percentage. However, as Table 3 shows, this is not the case. Term test percentage again correlates with final examination percentage, with an effect size of 0.76% increase in term test percentage for every 1% increase in final examination percentage.

The only other variable that correlates significantly with the final examination result is the average day quiz commenced variable from Term 1. The effect size is large, with a 5% increase in the final examination percentage for every day later a quiz is first started on average across the term. Given that 25% of the examination was derived from the material covered in Term 1, it is perhaps unsurprising that a Term 1 quiz behaviour variable correlates with final exam percentage. The quiz count variable from Term 1 that is significant in the first regression is also significant in the second regression, but only at the 90% CI level. What is surprising is that no Term 2 variables are correlated with final examination percentage. This finding strengthens a case that the quizzes in Term 2 were undertaken with less effort than in Term 1, and therefore were less beneficial. Certainly, the results shown in Tables 1 and 2 do lend support to this hypothesis.

Reflecting on the second research objective of examining whether behaviour changes in quiz completion have a casual impact on examination percentage, we can say that Term 2 quiz behaviour does not correlate with final examination performance, while Term 1 quiz behaviour does. Although in Term 1 completing more quizzes correlates positively with term test percentage, completing more quizzes in Term 2 does not correlate with final examination percentage. One possible explanation for this finding is that students did not put as much effort into the Term 2 quizzes, for which there is some evidence. This means they did not receive the benefit from doing the quizzes properly in Term 2, as they did in Term 1, because in Term 1 the quizzes counted towards a student's final grade. Marks were allocated directly on how well each student did in the quizzes. To further explore this hypothesis, additional analysis was done on the term test percentage relative to the final examination percentage.

The final examination mean score of 50% was much lower than the term test mean of 57%. This provides evidence that students performed less well in the final examination because they had not completed the quizzes as rigorously in Term 2 as they had in Term 1. To strengthen this finding, exam results from a historical class are compared with the student cohort from this study. The 2018 term test and final examination were the same as a previous year's term test and examination where possible. The prior year's assessments were chosen as they predated the historical set of copies of term tests and final examinations held on file at the university library for student use. The final examinations were identical between the two years. The two term tests were the same apart from 10 multiple-choice questions which were different due to minor content tweaking over time. To make the term test marks comparable, the 10 marks which the tests did not have in common were removed, and student percentages recalculated based on the remaining 90% of the tests.

Direct comparisons of term test and final examination percentages are not particularly illuminating, due to likely differing academic abilities between the two student cohorts. Instead, a comparison is made of final examination performance relative to term test performance for each year between the years. In the historical year, the assessment regime followed the traditional format, with no difference in the treatment of the online quizzes between Term 1 and Term 2, thus providing the baseline for comparison. As Table 4 shows, with the consistent quiz assessment regime across both terms, the term test mean is the same as the final examination mean. However, in the treatment year, the mean of the final examination is 7% lower than the term test mean.

Table 4
Summative assessment mean scores of control and treatment groups

Years	Term test mean %	Exam mean %
Historical year	52%	52%
2018 year	57%	50%

For the group of students who continued to do at least one quiz in Term 2, the gap is slightly larger with the same final examination mean, and a term test mean of 60%. This is further evidence to demonstrate that those who continued to complete the online quizzes in Term 2 were not putting in the same amount of effort in terms of treating them as a learning activity as they did in Term 1, when the quizzes had marks allocated to them based on achievement. This was true for students across all levels of ability. Term test and examination means were compared for both the control and treatment years, with the comparison stratified by quartile based on term test percentage (Table 5). In the baseline year, there was a small improvement in achievement from the term test to the examination for the bottom two quartiles, with a small decrease for the top two quartiles. In the treatment year, there was a larger decrease in achievement on the examination relative to the term test for all four quartiles, with the decrease in achievement of the top quartile almost four times larger in the treatment year than the baseline year. The results show that even the most academic students had a much bigger decrease in their achievement on the final examination relative to their term test performance when compared to the control year. For every quartile, there was a statistically significant difference ($p < 0.001$) when the differences in the means between the test and examination in the treatment year were compared to those of the control year.

Table 5
Comparison between control and treatment groups of term test and examination means by quartile

	Quartile 1		Quartile 2		Quartile 3		Quartile 4	
Years	Test	Exam	Test	Exam	Test	Exam	Test	Exam
Historical year	28	31	45	48	58	57	74	71
2018 year	33	28	54	45	67	52	81	70

The findings of this research confirm similar results mentioned in the literature review, that giving credit for completing coursework encourages participation (J. Kibble 2007; Holmes, 2015; Day et al., 2018; Cook & Babon, 2017). Where Kibble found that by allocating 1% credit per quiz was sufficient to increase participation from 52% to 97%, this study found removing 1% credit per quiz saw participation fall from 99% of students attempting at least one quiz to 49%. There is a general consensus that continuous assessment is beneficial to short-term learning (Cook & Babon 2017; Day et al., 2018; McDaniel et al., 2011; Rezaei 2015) with positive relationships between completion of continuous assessment such as online quizzes and student achievement. There is a dissenting finding from Greving and Richter (2018) that there was no testing effect from multiple-choice testing, especially in the longer term. Whatever the findings, it has been difficult to isolate out the fact that higher achieving students are more likely to engage in continuous assessment. This leaves some doubt over any causal relationship between participation in continuous assessment and academic achievement. Where Ozarslan and Ozan (2016) found a correlation between voluntary self-assessment quizzes and final examination performance, Gholami and Zhang (2018) also found a correlation between study habits and final performance. McDaniel et al. (2011) presented an intuitively appealing argument that "quizzing requires active processing of the target material and more specifically requires retrieval, a process that improves retention" (p. 400). However, the control and treatment group approach in our research allows for a stronger assertion regarding the effect of online quiz engagement on examination attainment.

A pertinent addition to the literature from our research is the finding that mere completion of online quizzes may not correlate with overall attainment. This is supported by the lack of a correlation between the number of quiz attempts and final exam score when multiple quiz attempts were allowed in the report from Ozarslan and Ozan (2016). An explanation for this lack of correlation could be that when faced with multiple attempts at the same quiz, students may not treat each quiz as a learning activity, but instead as an opportunity to harvest marks. This study also found a similar lack of correlation between examination performance and the number of quizzes attempted when incentives were removed. However, the evidence shows that those who continue to complete the quizzes do so with less rigour, thus reducing the benefit of continuing to complete quizzes relative to those who stopped doing quizzes altogether. When overall performance of the cohort as a whole is compared to a control group, the effect of removing incentives to complete quizzes is substantial and statistically significant.

Conclusion

The unique contribution this paper makes in explaining the benefit gained from allowing continuous self-assessments is twofold. By using an experimental approach with a control group and a treatment group, two important effects are separated out. First, the behaviour of students completing continuous assessment changes as a result of changes in incentives, and second, there is an effect on academic achievement in high-stakes assessments due to changing behaviour in completing continuous assessment activities. There was a significant increase in the number of students who completely disengaged from attempting online quizzes when a 1% incentive per quiz was removed (1% to 51% of the cohort). Of students who remained engaged, it was to a lesser degree, by completing fewer quizzes. There is also evidence that those who continued to complete quizzes did so with less vigour, by spending less time on each quiz, starting it later in the week and having fewer attempts per quiz on average. The average mark per quiz was also lower once the incentive to complete them was removed.

In addressing the first research aim, to examine how changing incentives to complete low-stakes online quizzes impacts student behaviour in terms of quiz completion, this research found that there are nuanced changes in student behaviour. A large number of students stopped doing the quizzes when the incentive to complete them was removed. Even those who continued to complete them, as a group, completed them to a lower standard. On average, students attempted them later in the week, spent less time completing them and had fewer attempts at each quiz. This is despite the students being regularly informed when the quizzes were open and the benefit to their learning of completing the quizzes. We assume a large group of students who stopped completing quizzes, did not see the value of self-assessment or use the quizzes as learning opportunities. So, for the majority of students, having a small incentive was sufficient to at least motivate them to attempt the quizzes.

This has implications for the second research objective, to establish whether there is a casual impact on examination percentage marks as a result of behaviour changes in quiz completion. Running OLS regressions found no significant differences in examination performance related to number of quizzes a student continued to attempt after incentives were removed. We propose that using incentives to encourage students to complete quizzes may be a critical factor in value being added to the grade received in high-stakes exams.

Even students who continued to do some or all of the online quizzes completed them later in the week with fewer attempts and spent less time on them. Term 1 quiz completion (when the quizzes were being completed more rigorously) was correlated with a higher term test and final exam percentage. This suggests that when quizzes are done as self-assessment and learning opportunities, they do correlate with improved grades in high-stakes assessments. Therefore, by changing the incentive to do online quizzes (as we did in this experiment), students' final exam outcomes are impacted, as quizzes are completed more haphazardly, even by those who continue to do them. To strengthen this apparent causal relationship, a comparison with the historical control group of students showed that, relative to term test performance, final exam mean was 7% lower for the cohort who had the incentive to complete online quizzes removed. The control group showed no difference between term test and final examination mean scores when quiz incentives were maintained for the entire course.

An implication of this research for future studies is that a binary variable for quiz completion does not capture the nuanced effect of the quality of each quiz attempt. Binary variables representing engagement in an online activity do not capture the full quality of said engagement. Although the variable of quiz mark rather than quiz completion may be more likely to capture the quality of each quiz attempt, quiz mark may also be influenced by student ability. Often, in previous research projects, students were grouped into categories such as attempted quizzes or did not attempt any. A similar grouping occurs with respect to contribution to online discussion boards. Even among students who are participating in a course through online contribution, the quality of this online participation is an important contributory factor. When quality of engagement is considered, the control-intervention approach used in this research finds a substantive, causal relationship between incentivised online quiz completion and final examination attainment.

Limitations and future research

One of the strengths of this research – having a comparison baseline group who also sat the same assessments – creates a limitation. More research is needed to support the generalisability of the results. Although the subjects were from a range of academic backgrounds, a similar research approach in different subject areas other than economics would help affirm the findings.

References

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, *87*, 659–701. https://doi.org/10.3102/0034654316689306

Angus, S., & Watson, J. (2009). Does regular online testing enhance student learning in the numerical sciences? Robust evidence from a large data set. British Journal of Educational Technology, 40, 255–272. https://doi.org/10.1111/j.1467-8535.2008.00916.x

Barnark-Brak, L., Lan, W., & Paton, V. (2010). Profiles in self-regulated learning in the online learning environment. *The International Review of Research in Open and Distance Learning*, *11*(1), https://doi.org/10.19173/irrodl.v11i1.769

Butler, A., & Roediger, H. L., III. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, *36*, 604–616. https://doi.org/10.3758/MC.36.3.604

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, *34*, 268–276. https://doi.org/10.3758/BF03193405

Chen, C. M., & Huang, S. H. (2014). Web-based reading annotation system with an attention-based self-regulated learning mechanism for promoting reading performance. *British Journal of Educational Technology*, *45*(5), 959–980. https://doi.org/10.1111/bjet.12119

Cook, B., & Babon, A. (2017). Active learning through online quizzes: better learning and less (busy) work. *Journal of Geography in Higher Education*, *41*(1), 24–38. https://doi.org/10.1080/03098265.2016.1185772

Day, I., Van Blankenstein, F., Westenberg, P., & Admiraal, W. (2018). Explaining individual student success using continuous assessment types and student characteristics. *Higher Education Research & Development*, *37*(5), 937–951. https://doi.org/10.1080/07294360.2018.1466868

Gholami, A., & Zhang, L. (2018). Student behaviour in unsupervised online quizzes: A closer look. In *Proceedings of the 23rd Western Canadian Conference on Computing* (pp. 1–6). Association for Computing Machinery. https://doi.org/10.1145/3209635.3209650

Greving, S., & Richter, T. (2018). Examining the testing effect in university teaching: Retrievability and question format matter. *Frontiers in Psychology*, 9, 1–10. https://doi.org/10.3389/fpsyg.2018.02412

Holmes, N. (2015). Student perceptions of their learning and engagement in response to the use of a continuous e-assessment in an undergraduate module. *Assessment & Evaluation in Higher Education*, *40*(1), 1–14. https://doi.org/10.1080/02602938.2014.881978

Hoskins, S., & Van Hooff, J. (2005). Motivation and ability: which students use online learning and what influence does it have on their achievement? *British Journal of Educational Technology*, *36*(2), 177–192. https://doi.org/10.1111/j.1467-8535.2005.00451.x

Jensen, J., McDaniel, M., Woodard, S., & Kummer, T. (2014). Teaching to the test…or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. *Educational Psychology Review*, *26*(2), 307–329. https://doi.org/10.1007/s10648-013-9248-9

Johnson, G. (2006). Optional online quizzes: College student use and relationship to achievement. *Canadian Journal of Learning and Technology*, *32*(1). https://www.learntechlib.org/p/42799/

Kibble, J. (2007). Use of unsupervised online quizzes as formative assessment in a medical physiology course: Effects of incentives on student participation and performance. *Advances in Physiology Education*, *31*(3) 253–260. https://doi.org/10.1152/advan.00027.2007

Kibble, J. D., Johnson, T., Khalil, M., Nelson, L., Riggs, G., Borrero, J., & Payer, F. (2011). Insights gained from the analysis of performance and participation in online formative assessment. *Teaching and Learning in Medicine*, *23*(2), 125–129. https://doi.org/10.1080/10401334.2011.561687

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L., III. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, *103*(2), 399–414. https://doi.org/10.1037/a0021782

McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(2), 371–385. https://doi.org/10.1037/0278-7393.11.2.371

McDaniel, M. A., Wildman, M., & Anderson, J. (2012). Using quizzes to enhance summative-assessment performance in a web based class: An experimental study. *Journal of Applied Research in Memory and Cognition*, *1*(1), 18–26. https://doi.org/10.1016/j.jarmac.2011.10.001

Nieuwoudt, J. E. (2020). Investigating synchronous and asynchronous class attendance as predictors of academic success in online education. *Australasian Journal of Educational Technology*, *36*(3), 15–25. https://doi.org/10.14742/ajet.5137

Onah, D. F., Pang, E. L. L., & Sinclair, J. E. (2020). Cognitive optimism of distinctive initiatives to foster self-directed and self-regulated learning skills: A comparative analysis of conventional and blended-learning in undergraduate studies. *Education and Information Technologies*, *25*, 4365–4380. https://doi.org/10.1007/s10639-020-10172-w

Ozarslan, Y., & Ozan, O. (2016). Self-assessment quiz taking behavior analysis in an online course. *European Journal of Open, Distance and E-learning*, *19*(2), 15–31. https://doi.org/10.1515/eurodl-2016-0005

Pashler, H., Cepeda, N., Wixted, J., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 3–8. https://doi.org/10.1037/0278-7393.31.1.3

Rezaei, A. (2015). Frequent collaborative quiz taking and conceptual learning. *Active Learning in Higher Education*, *16*(3), 187–196. https://doi.org/10.1177/1469787415589627

Roediger, H. L., III, & Karpicke, J. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Scheyvens, R., Griffin, A., Jocoy, C., Liu, Y., & Bradford, M. (2008). Experimenting with active learning in geography: Dispelling the myths that perpetuate resistance. *Journal of Geography in Higher Education*, *32*(1), 51–69. https://doi.org/10.1080/03098260701731496

Schunk, D. H. (2005). Self-regulated learning: The educational legacy of Paul R. Pintrich. *Educational Psychologist*, *40*(2), 85–94. https://doi.org/10.1207/s15326985ep4002_3

Thomas, J., Wadsworth, D., Jin, Y., Clarke, J., Page, R., & Thunders, M. (2017). Engagement with online self-tests as a predictor of student success. *Higher Education Research & Development*, *36*(5), 1061–1071. https://doi.org/10.1080/07294360.2016.1263827

Uz, R., & Uzun, A. (2018). The influence of blended learning environment on self-regulated and self-directed learning skills of learners. *European Journal of Educational Research*, *7*(4), 877–886. https://doi.org/10.12973/eu-jer.7.4.877

Williams, A. (1997). Making the most of assigned readings: Some alternative strategies. *Journal of Geography in Higher Education*, *21*(3), 363–371. https://doi.org/10.1080/03098269708725442

Yen, C., Bozkurt, A., Tu, C., Sujo-Montes, L., Rodas, C., Harati, H., & Lockwood, A. (2018). A predictive study of students' self-regulated learning skills and their roles in the social network interaction of online discussion board. *Journal of Educational Technology Development and Exchange*, *11*(1). https://doi.org/10.18785/jetde.0901.03

Corresponding author: Stephen Agnew, steve.agnew@canterbury.ac.nz