THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

Discussion Paper Series
UQ School of Economics

Repeated quizzing of building block mathematics concepts to improve grades in economics
2021-05-12
Jack Leggett, Bryan Morgan, Kam Ki Tang

Discussion Paper No.642

Repeated quizzing of building block mathematics concepts to improve grades in economics

Jack Leggett[a]
Bryan Morgan[b]
Kam Ki Tang[b,#]

a. School of Education, The University of Queensland
b. School of Economics, The University of Queensland
Corresponding author

Repeated quizzing of building block mathematics concepts to improve grades in economics

Abstract

Modern economics courses require a high level of mathematical proficiency, and a lack of such proficiency may reduce the performance of economics students or lead students to avoid economics courses altogether. In the experiment described in this study, we sought to improve economics students' mathematical skills, and in turn their grades, by repeatedly quizzing them on building block concepts. We analysed data from four semesters of a first-year mathematics course. In the earlier two semesters, under an old quizzing approach, students were quizzed five times, but no concept appeared on more than one quiz. In the last two semesters, under the new quizzing approach, important basic concepts appeared repeatedly across several quizzes. Results showed that performance with repeated concepts improved across quizzes and the new quizzing approach was associated with slightly higher final exam scores. Scores on the mid-semester exams, which occurred before most quizzes, were not consistently associated with the quizzing approach (and therefore the semesters), helping to rule out cohort effects and some other potential alternative explanations for the improvement in final exam scores.

Keywords: mathematics; quiz; building block

1. Introduction

Mathematisation and formalisation are common in modern economic studies in universities (Weintraub 2002; Milonakis 2017). This is the case for not only postgraduate courses[1] but also second- and third-year undergraduate courses. In Australia, the increasing role of mathematics in economic studies, however, often confronts the reality that a lot of university students are not proficient in even high-school mathematics. This lack of proficiency has consequences for university students' performance in their economic studies. Conversely, it could impact on the number of high school students choosing to study economics in the first place, either as a whole degree, or as a major or as electives within some other degrees (e.g. within a bachelor degree of business or arts).

An obvious - albeit ad hoc - remedy is to provide first-year undergraduate economics students basic mathematical training before they undertake more advanced economics courses. In fact, mathematics courses have become an ubiquitous feature of tertiary economics programs worldwide. It is well-known that mathematical proficiency requires repeated practising (Burns et al. 2014). For instance, instructors can give students recurring exercises on the same concepts. They can also test students on the same concepts repeatedly throughout a semester to alleviate the 'learn and forget' habit among students. In various subject areas, regular testing or quizzing often improves performance in other course assessments (see Sotola & Crede, 2020, for a recent meta-analysis). Taking a test can provide skill practice, but answering knowledge questions can also improve understanding and retention of concepts through so-called retrieval practice effects (Moreira, Pinto, Starling, & Jaeger, 2019; Pan & Rickard, 2018; Rowland, 2014).

The aforementioned approaches are, however, demanding in terms of exercise or assessment space and time. In particular, as the number of topics covered accumulates with coure progression, the trade-off between the space and time dedicated to different topics becomes more challenging. As such, the 'ideal' approach of repeating all or even most topics is not feasible in practice. There is also a conundrum that, because the (relatively) most advanced and challenging topics within a course are naturally covered towards the end of the semester, they are precisely the ones having the fewest opportunities for repetition.

These constraints mean that repetition of earlier, building block concepts is more feasible in practice. From the pedagogical viewpoint, it raises the following question:

> *Does the repetition of building block mathematics concepts help students improve their performance not only with those concepts but also with more advanced concepts, and thus improve their overall mathematical proficiency?*

At the University of Queensland (UQ) in Australia, we conducted an experiment to answer this question. Specifically, in a first-year mathematical course for economic studies,

[1] This is particularly the case for courses in Honours and PhD programs.

UQ students learn some building block mathematical tools like basic algebra and solving simple equations in the first few weeks of the semester before they dive into (relatively) more advanced topics such as unconstrained and constrained optimisation. For years, students have been tested regularly throughout the semester but earlier topics are *not tested a second time until the final exam*. However, in 2018 and 2019, we changed the testing scheme so that building block concepts were *tested repeatedly throughout the semester*. We then compared students' performance in quizzes and in the final exam between the students in semester 2 of 2018 and semester 1 of 2019 (the treatment group) and the students in previous semesters (the control group).

There is conflicting evidence on how useful basic mathematics courses are for improving student performance in economics. Hoag and Benedict (2010) found that college algebra courses, business calculus courses, and mathematics courses for students planning to become a mathematics teacher have helped improve students' performance in economic courses. However, using a natural experiment, Lagerlöf and Seltzer (2009) showed that basic mathematical courses have little effect on student performance in various compulsory economic courses. With the present study, we hope to provide more information about when and how the additional practice of building block mathematics concepts might improve performance in economics courses.

In addition to focusing on building block concepts, the new quizzing approach we implemented had two other major design features. Firstly, the concepts chosen for repeated practice were those that, according to results of quizzes and exams in previous semesters, a significant proportion of students found difficult. These building block concepts were required to answer more complex problems in the latter part of the course and in the final exam. As such, these concepts would seem to be those most likely to be limiting student performance. Secondly, within practical constraints, repetitions were scheduled so as to maximise the amount of time separating them. This second feature was intended to take advantage of the *spacing effect*—a highly reliable learning effect in which practice repetitions increase potency more, the more they are separated in time (Carpenter, Cepeda, Rohrer, Kang, & Pashler, 2012). In particular, a *spaced* practice typically improves long-term retention of skill and knowledge compared to the same amount of more closely scheduled (*massed*) practice. Roughly speaking, an implication of the spacing effect is that, once the total amount of practice to take place has been determined, the practice should be distributed as widely as possible within the time available—for example, throughout a semester (but see Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008; and Mettler, Massey, & Kellman, 2016, for more nuanced implications).

At the outset, there are several ways in which quizzing might lead to improved exam scores. For example, students may gain skill or knowledge directly from the processes involved in attempting to answer a quiz question, or from attending to feedback after their attempt, or even by reading the information given within the quiz itself. Less directly,

students may gain better awareness of their strengths and weaknesses by taking a quiz and may then arrange their study more effectively (Arnold & McDermott, 2013), or the presence of quizzes may simply motivate them to study course material more in general, with the specific content of the quizzes being unimportant. We will return briefly to these possibilities in the discussion section.

The rest of the present paper is organised as follows. Section 2 describes the experiment used to generate the data. Section 3 explains the methodology. Section 4 presents the results. The last section concludes with some discussion.

2. Data

At UQ, *ECON1050 Tools of Economic Analysis* is a compulsory first-year mathematics course for the Bachelor of Economics (BEcon) and Bachelor of Advanced Finance and Economics (BAFE) programs, as well for the economic majors of the Bachelor of Politics, Philosophy and Economics (BPPE) and Bachelor of Arts (BA) programs. ECON1050 is offered twice a year, in semester 1 (S1) and semester 2 (S2).

A series of five online quizzes were introduced to ECON1050 in 2016 as a formative assessment and to encourage students to keep up to date with the course material. Initially, each quiz tested topics from two preceding lectures. In S2 2018 and S1 2019, a number of additional, revision questions were added to quizzes two, four and five.

The online quizzes are created on a proprietary platform that provides useful analytics of student performance at the level of individual questions. Data for this paper were extracted from the program reports for each quiz, and also included scores on mid-semester and final exams.

3. Methodology

3.1 Participants

Assessment data were available for 891 unique cases (i.e. students) over four semesters. For students who had attempted the course more than once, only their first attempt was included. The coverage of the most recent four semesters before and after the change allowed a clean comparison involving both a first and second semester under both the old and new quizzing approach. Also, the use of data before 2000 avoided potential complications caused by the pandemic.

In S1, about 30-40% of students are from the BAFE program, which has a higher entry requirement than the other three programs. Hence, scores tended to be higher in S1. In S2, the cohort is more diverse and does not include BAFE students. Table 1 shows, for each semester, the numbers of participants whose data were analysed and the share of students enrolled in various programs.

Table 1. Distribution of students by semesters and programs

		S2 2017	S1 2018	S2 2018	S1 2019
		Control	Control	Treatment	Treatment
Number of enrolled students		244	261	246	308
Share of enrolled students by programs	BEcon	20%	11%	23%	12%
	BEcon/Dual	56%	20%	33%	21%
	BAFE	0%	43%	0%	32%
	BA (Econ major)	7%	7%	13%	7%
	BPPE	0%	6%	7%	6%
	BCom	11%	9%	13%	15%
	BBusMan	5%	2%	9%	5%
	Other	2%	2%	3%	2%
Share of students attempted quizzes (average)		85%	91%	85%	88%

Note: BAFE students only take the course in S1 because of program restrictions. BEcon = Bachelor of Economics; BAFE = Bachelor of Advanced Finance and Economics; BA = Bachelor of Arts; BPPE = Bachelor of Politics, Philosophy and Economics; BCom = Bachelor of Commerce; BBusMan = Bachelor of Business Management. BEcon/Dual are dual degree programs that include: BEcon/BArts, BCom/BEcon, BBusMan/BEcon, BEcon/BLaws, BEcon/BScience.

3.2 Design

The main dependent variables were scores on the mid-semester and final exams. The main independent variable was the quizzing approach. As mentioned, an old quizzing approach was used in the earlier two of the four semesters from S2 2017 to S1 2019, and the new quizzing approach was used in the latter two. An important secondary independent

variable was semester because there are known cohort differences between semesters in ECON1050. In particular, as mentioned, students enrolled in the academically demanding BAFE program take the course only in S1 of each year, typically increasing average performance in those semesters.

Because the first semester was S2 of 2017, and the last semester was S1 of 2019, to reduce confusion for the reader, and to allow clearer figures and tables, from here on S2 of each year will be referred to as *Semester A* and S1 as *Semester B*. Similarly, the period under the old quizzing approach (S2 of 2017 and S1 of 2018) will be referred to as *Year 1*, and the period under the new quizzing approach (S2 of 2018 and S1 of 2019) will be referred to as *Year 2*. Thus, the first of the four semesters is referred to as Year 1 Semester A, and the last as Year 2 Semester B. This terminology allows us to refer to specific individual semesters briefly and unambiguously.

3.3 Materials and procedure

The set of quizzed concepts was identical across all semesters. The names of quizzed concepts, and the quizzes on which they appeared, are given in Appendix A. In each semester, students are assessed in five quizzes, a mid-semester exam, and a final exam. In the semesters under the old quizzing approach, each concept was quizzed only once, while under the new quizzing approach some concepts in the first two quizzes appeared in additional questions both in those quizzes and in some later quizzes. Where possible, repetitions were distributed widely throughout the semester in order to maximise spacing effects.

The timing of the quizzes was similar in all four semesters. Apart from the revision questions, a quiz tested material from two previous lectures and was conducted after all students had the opportunity to attend tutorials for both lectures. For example, quiz 1 tested topics from lectures 1 and 2 and was open in week 4, quiz 2 tested lectures 3 and 4 and was open in week 5, and so on. Each quiz was open for 7 days and students were not required to complete a quiz in one sitting. However, students could not return to a question once it had been answered. Questions were individualised so as to have different numerical answers. A feature of the proprietary software used for the quizzes is that once a quiz was closed the student could see the correct working for questions they got wrong. Unfortunately, the program analytics did not indicate how many students took advantage of this feature.

Quizzes in Year 1 Semesters A and B did not contain revision questions. In Year 2 Semesters A and B, revision questions were contained in Quizzes 2, 4, and 5. Quiz 3 was a longer quiz and tested the first of the more complex topics, so it was decided not to burden students with additional revision questions in Quiz 3.

Quiz 2 contained four revision questions (out of 14 questions) from Quiz 1 covering three topics; namely set theory, interval notation and solving linear equations. In Quiz 4

there were four revision questions. The topics tested were interval notation and solving linear equations from Quiz 1 and quadratic functions and log functions from Quiz 2. Finally, Quiz 5 contained 5 revision questions; one question for each of the revision topics from Quizzes 1 and 2. Quizzes 4 and 5 each contained 15 questions in total.

In Year 1 Semesters A and B and Year 2 Semester B, the mid-semester exam was held at the end of week 7. In Year 2 Semester A the mid-semester exam was held at the end of week 6, so in this case, students were given extra time to complete Quiz 2. In all four semesters, the mid-semester exam tested material from lectures 1 to 4, in other words, the topics tested in Quizzes 1 and 2. Thus, most of the repetitions of concepts on quizzes occurred *after* the mid-semester exam.

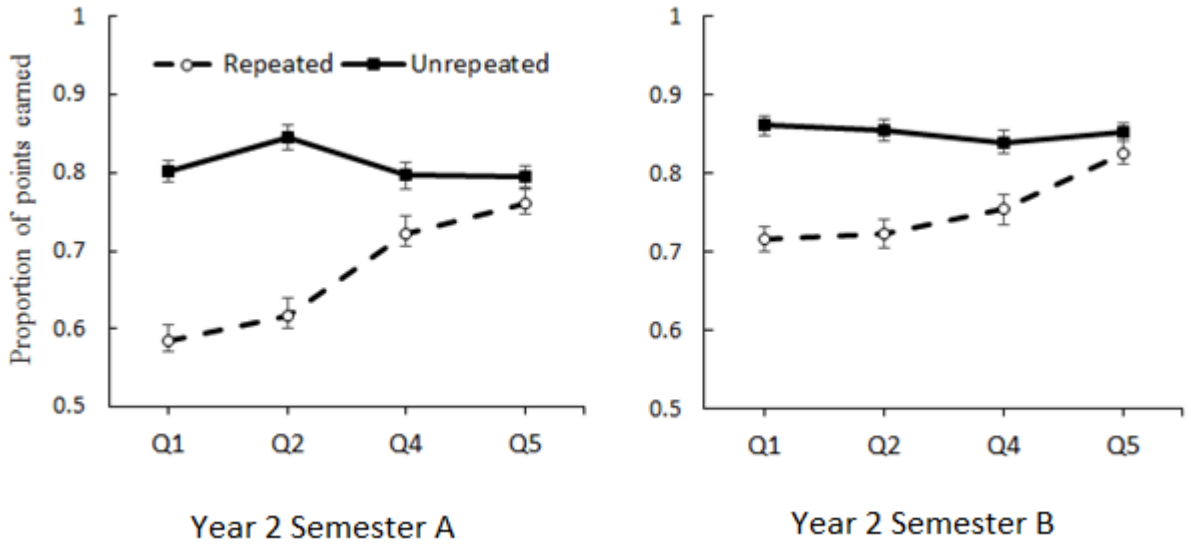
4. Results

4.1 Preliminary analyses: Improvement across quizzes

When repeated practice is effective, performance should typically improve across repetitions. Therefore, as an initial check, we conducted a three-way mixed ANOVA to compare performance across quizzes, under the new quizzing approach (i.e., in Year 2), with concepts that were repeated (appeared on more than one quiz) versus unrepeated (appeared on only one quiz). The dependent variable was the proportion of points earned of those available for the relevant concept type. For this analysis, data from Quiz 3 were excluded because Quiz 3 contained no repeated concepts. In addition, only participants who attempted Quizzes 1, 2, 4, and 5 were included, leaving 147 participants from Semester A and 218 participants from Semester B.

We first conducted a 4 (quiz number: 1, 2, 4, or 5) X 2 (concept repetition: repeated versus unrepeated) X 2 (semester: A versus B) mixed ANOVA, with semester as the between-participants factor. Mean scores are shown in Figure 1. Crucially, the two-way interaction between concept repetition and quiz number was significant ($F(3, 1089) = 44.34$, $p < .001$, $\eta_p^2 = .11$), reflecting greater improvement in scores for repeated than for unrepeated concepts. This ANOVA also revealed significant main effects of quiz number ($F(3, 1089) = 14.21$, $p < .001$, $\eta_p^2 = .04$), concept repetition ($F(1, 363) = 328.45$, $p < .001$, $\eta_p^2 = .48$), and semester ($F(1, 363) = 15.49$, $p < .001$, $\eta_p^2 = .04$). On average, scores were higher in later than in earlier quizzes within a semester, higher with unrepeated than with repeated concepts, and higher in Semester B than in Semester A. There was a marginally significant interaction between quiz number and semester ($F(3, 1089) = 2.47$, $p = .06$, $\eta_p^2 = .01$), and a significant interaction between concept repetition and semester ($F(1, 363) = 10.12$, $p = .002$, $\eta_p^2 = .03$), with both interactions reflecting more variable mean scores in Semester A.

Figure 1. The proportion of points earned for repeated and unrepeated concepts in Semester A and B of Year 2



Note: Quiz 3 was excluded because it contained no repeated concepts. Error bars represent within-participants standard errors.

The three-way interaction was also significant ($F(3, 1089) = 5.14$, $p = .002$, $\eta_p^2 = .01$), reflecting a weaker two-way interaction between concept repetition and quiz number in Semester B compared to Semester A. In other words, the difference between repeated and unrepeated concepts in improvement across quizzes was smaller in Semester B than in Semester A. However, follow-up two-way ANOVAs for each semester separately confirmed that the two-way interaction was significant in both Semester A ($F(3, 438) = 25.97$, $p < .001$, $\eta_p^2 = .15$), and Semester B ($F(3, 651) = 16.65$, $p < .001$, $\eta_p^2 = .07$).

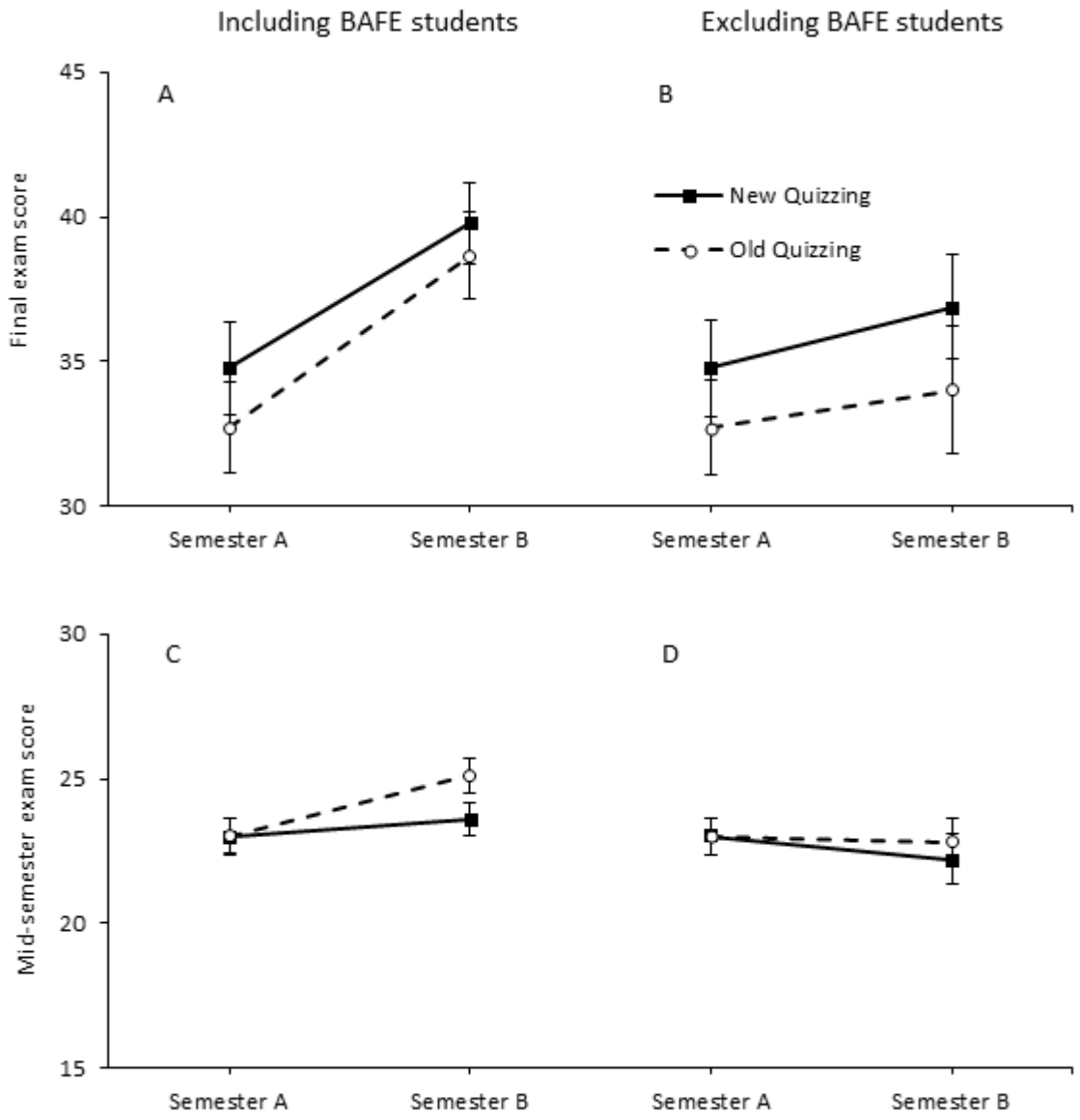
These preliminary results are not surprising and could have any of several causes. Most notably, because the unrepeated concepts were by definition tested only once in the quizzes, any improvement in individual unrepeated concepts over the semester could not be assessed. However, the lack of improvement in unrepeated concepts, as a whole, rules out two possibilities: (1) that the quiz content became easier over time, and (2) that total quiz performance improved during each semester because of changes in study behaviour. Meanwhile, in the case of repeated concepts, improvement was not inevitable and a lack of improvement would be clear evidence that the quizzes were ineffective. These initial results, therefore, provide some reassurance that the new quizzing approach was working as intended.

4.2 Main analyses: Differences in exam scores associated with the new quizzing approach

Average scores on each mid-semester and final exams are shown in Figure 2. A 2 X 2 between-participants ANOVA on the mid-semester exam scores revealed a significant main effect of semester ($F(1, 887) = 18.60$, $p < .001$, $\eta_p^2 = .021$), a significant main effect of quizzing approach ($F(1, 887) = 5.57$, $p = .019$, $\eta_p^2 = .006$), and a significant interaction effect ($F(1, 887) = 5.64$, $p = .018$, $\eta_p^2 = .006$). As can be seen in panel C of Figure 2, all three of these

results were due to a higher mean score in Semester B with the old quizzing approach than in the other semesters (across which scores were similar). Follow-up tests revealed a significant simple effect of quizzing approach on mid-semester exam scores in Semester B ($t(485) = -3.626$, $p < .001$, $d = .33$), but not in Semester A ($t(402) = .01$, $p = .992$, $d < .01$). Note that this significant difference is unlikely to be due to the new quizzing approach because it is in the wrong direction and most quizzes occurred after the mid-semester exam.

Figure 2. Mean exam scores by semesters and quizzing approaches



Note: Final exam scores are on the upper panel and mid-semester exam scores on the lower panel. Means shown in panels A and C were calculated from the full sample including BAFE students who tend to perform relatively well, while those shown in panels B and D were calculated after excluding BAFE students. Means in A semesters are unaffected by these

exclusions because BAFE students do not take ECON1050 in A semesters. The maximum possible score was 50 on each final exam and 30 on each mid-semester exam.

Concerning the final exam scores, a 2 X 2 between-participants ANOVA revealed a significant main effect of semester ($F(1, 887) = 51.89$, $p < .001$, $\eta_p^2 = .055$), and a significant main effect of quizzing approach ($F(1, 887) = 4.40$, $p = .036$, $\eta_p^2 = .005$), but no significant interaction ($F(1, 887) = .40$, $p = .526$, $\eta_p^2 < .001$). Scores were higher in B semesters than in A semesters (as would be expected given the presence of BAFE students in B semesters), and also higher in Year 2 under the new quizzing approach than in Year 1 under the old quizzing approach. In terms of raw scores, the new quizzing approach was associated with a small advantage in final exam scores of between one and two points out of 50, equivalent to between about two and four percentage points. Follow-up tests revealed that the simple effect of quizzing approach on final exam scores was marginally significant between A semesters ($t(402) = 1.725$, $p = .085$, $d = .17$), and not significant between B semesters ($t(485) = 1.16$, $p = .247$, $d = .11$).

The smaller (and non-significant) difference in final exam scores between B semesters may be due to higher overall student performance in Semester B of Year 1, under the old quizzing approach, most clearly shown in the relatively high mid-semester exam scores in that semester. The presence of BAFE students in the B semesters increases overall performance. Excluding BAFE students makes the cohorts in A semesters and B semesters more comparable, and this is reflected in the pattern of means shown in Panels B and D of Figure 2. Indeed, the exclusion of BAFE students abolishes all three significant effects on mid-semester exam scores: the main effect of semester ($F(1, 681) = 2.00$, $p = .157$, $\eta_p^2 = .003$), the main effect of quizzing approach ($F(1, 681) = .66$, $p = .418$, $\eta_p^2 < .01$), and the interaction ($F(1, 681) = .68$, $p = .411$, $\eta_p^2 < .01$). In addition, when BAFE students are excluded, the simple effect of quizzing approach on final exam scores between B Semesters becomes significant ($t(279) = 2.13$, $p = .034$, $d = .26$). Taken together, the results suggest that the new quizzing approach was associated with a small but reliable increase in final exam scores, but had no consistent effect on mid-semester exam scores.

4.3 Exploratory analysis: Moderation by student performance

Additional practice of building block mathematical concepts and skills should be of most benefit to weaker students. Therefore, we broke the sample into quintiles according to performance on the mid-semester exam, and then analysed the association of quizzing approach with final exam scores across these quintiles. According to a 2 X 5 between-participants ANOVA on final exam scores, there were significant main effects of quizzing approach ($F(1, 881) = 27.50$, $p < .001$, $\eta_p^2 = .03$), and mid-semester exam performance quintile ($F(4, 881) = 236.18$, $p < .001$, $\eta_p^2 = .52$), but the critical interaction between quizzing approach and quintile was not significant ($F(4, 881) = 1.81$, $p = .125$, $\eta_p^2 = .01$). However, the measured differences in mean final exam scores between the old and

new quizzing approaches were larger in the lower quintiles, as can be seen in Figure 3. In a 2 X 2 ANOVA restricted to only the second and fifth quintiles, the interaction between quintile and quizzing approach was significant ($F(1, 364) = 7.95$, $p = .005$, $\eta_p^2 = .02$). These results provide, in our view, weak evidence that the new quizzing approach was more beneficial for lower-performing students. This possibility is supported by the pattern of means and by a significant interaction, but the interaction was significant only in a restricted post-hoc analysis and in any case, could be due to differences in the room for improvement in the exam scores themselves rather than to differences in actual learning.

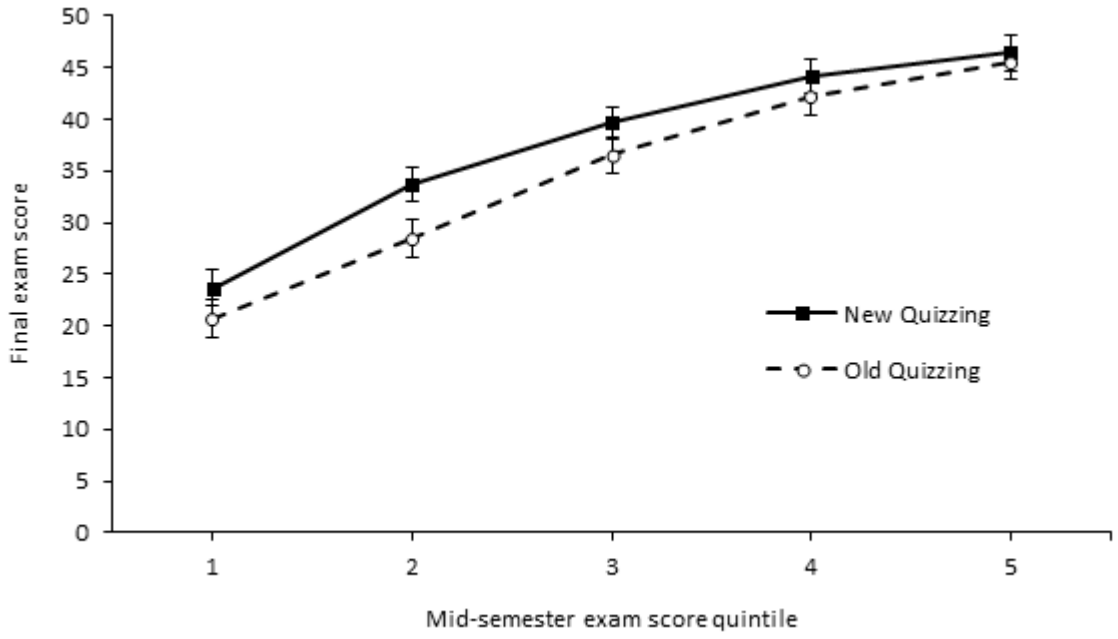


Figure 3. Average scores on the final exam according to relative performance on the mid-semester exam and to quizzing approach. Error bars represent 95% confidence intervals.

5. Discussion

We analysed data from four semesters of a first-year economics course to determine whether additional quizzing on building block mathematics concepts could improve overall grades. In the two semesters with the new quizzing approach, performance improved across quizzes with repeated but not with unrepeated concepts, consistent with the possibility that repeated quizzing was effective. According to a comparison with two previous semesters, the new quizzing approach was also consistently associated with slightly higher final exam scores but was not consistently associated with mid-semester exam scores. Finally, the new quizzing approach may have been most beneficial for lower-performing students, but the evidence on that issue was not very clear. The results are consistent with many studies that find benefits of quizzing (Sotola & Crede, 2020), and broadly consistent with the finding of

Hoag and Benedict (2010) that additional general mathematics training can improve grades in economics courses.

Our reliance on comparisons between semesters raises several potential alternative explanations for the findings that should be considered. Between semesters, there may be changes in average student ability or behaviour (i.e., cohort effects), or in teaching or marking practices. However, all of these potential explanations are rendered less plausible by the fact that average mid-semester exam scores were virtually identical between semesters. The exception was an unusually high average score on the mid-semester exam in a semester using the old quizzing approach, but this anomaly vanished when BAFE students were excluded. At the same time, a lack of variation in mid-semester exam scores is still consistent with the new quizzes being effective; the new quizzing approach would not be expected to have much effect on mid-semester exam scores, as the majority of repeated concepts were only repeated after the mid-semester exam.

We should also note that the apparent effect of quizzing on final exam scores was quite small. Small effects are more likely than large effects to occur by chance or indeed because of some unknown confound. On the other hand, a small effect is to be expected, given the low intensity of our intervention and the many other influences on student exam performance in any course. Overall, it seems that the full pattern of results is not easily accounted for by any single alternative explanation, and is consistent with the new quizzing approach having improved student performance.

As noted in the introduction, there are several ways in which quizzing may improve exam scores. Students might learn directly from attempting the quiz or attending to feedback, they might study more strategically after quizzes reveal their strengths and weaknesses with the course material, or quizzes might motivate more total study or more regular study over the semester. Our data cannot distinguish between these possibilities, although a change in study time or regularity seems unlikely since the new quizzing approach did not alter the number or timing of the quizzes themselves.

The findings of this paper have important implications. First of all, advanced topics tend to be more numerous, complex and heterogeneous. As such, it is more difficult to repeat testing of all advanced topics to the same extent, and the job of creating new test questions is also more demanding. In comparison, building block concepts are limited in number and their contents tend to be more standardised. As a result, the task of creating new questions for repeated quizzes may be more manageable. Second, advanced topics by nature have to be covered in the latter part of a course, leaving little space to quiz them repeatedly. This means that the repeated quizzing of building block concepts is more feasible in practice. Thirdly, building block mathematical concepts are embodied in most economics subjects, unlike more advanced mathematical concepts. Therefore, improving students' ability in handling those concepts is likely to generate benefits to students for their entire

economics program rather than just for a few subjects. In fact, arguably, being proficient in foundation mathematics would benefit students in any subject areas, in work and in life.

6. Conclusion and suggestions for further research

Quizzing is a simple and cheap intervention that can improve learning. Our findings suggest that additional spaced quizzing of basic mathematics concepts can improve performance in economics courses. We hope others might confirm and extend our findings. Future studies could manipulate the total amount of quizzing or the proportion of quizzing devoted to basic versus advanced concepts. In addition, when particular quizzed concepts can be related to particular exam questions, manipulations performed within-participant (between-concept) might be especially informative. For example, for each participant, some concepts could be repeated more often or on a different schedule compared to others, and then the effects could be assessed by comparing performance on corresponding exam questions. Such an approach was not possible in the present study (as most exam questions relied on multiple concepts), but has the advantages of ruling out many potential alternative explanations as well as allowing equitable treatment of all students. Measures of student attitudes and study behaviour would also be helpful in determining the extent to which these factors explain the benefits of quizzing in economics and other university courses.

Reference

Arnold, K. M., & McDermott, K. B. (2013) Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(3), 940-945. doi: 10.1037/a0029199

Burns, M. K., Ysseldyke, J., Nelson, P. M., & Kanive, R. (2015). Number of repetitions required to retain single-digit multiplication math facts for elementary students. *School Psychology Quarterly*, *30*(3), 398.

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review, 24*, 369-378. doi: 10.1007/s10648-012-9205-z

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science, 19*(11), 1095-1102.

Denny, E. (2014). Factors influencing the performance of non-economics majors in an introductory economics course. *International Review of Economics Education*, *17*, 1-16.

Hoag, J., & Benedict, M. E. (2010). What Influence Does Mathematics Preparation and Performance Have on Performance in First Economics Classes? *Journal of Economics and Economic Education Research*, *11*(1), 19.

Lagerlöf, J. N., & Seltzer, A. J. (2009). The effects of remedial mathematics on the learning of economics: Evidence from a natural experiment. *The Journal of Economic Education*, *40*(2), 115-137.

Mettler, E., Massey, C., & Kellman, P. J. (2016). A comparison of adaptive and fixed schedules of practice. *Journal of Experimental Psychology: General*, *145*(7), 897-917. doi: 10.1037/xge0000170

Milonakis, D. (2017). Formalising economics: social change, values, mechanics and mathematics in economic discourse. *Cambridge Journal of Economics*, *41*(5), 1367-1390.

Moreira, B. F. T., Pinto, T. S. S., Starling, D. S. V., & Antonio, J. (2019). Retrieval practice in classroom settings: A review of applied research. *Frontiers in Education, 4*, 1-16. doi: 10.3389/feduc.2019.00005

Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological Bulletin, 144*, 710-756. doi: 10.1037/bul0000151

Rowland, C. A. (2014). The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. Psychological Bulletin, 140(6), 1432-1463. doi: 10.1037/a0037559

Sotola, L. K., & Crede, M. (2020). Regarding class quizzes: A meta-analytic synthesis of studies on the relationship between frequent low-stakes testing and class performance. *Educational Psychology Review.* Advance online publication. doi: 10.1007/s10648-020-09563-9

Weintraub, E. R. 2002.*How Economics Became a Mathematical Science*. Durham: Duke University Press. https://www.dukeupress.edu/how-economics-became-a-mathematical-science.

Appendix A: The Distribution of Repeated and Unrepeated Concepts Across Quizzes

The following table shows the quiz or quizzes that a concept appeared. Under the old quizzing approach, concepts appeared only on the first quiz listed.

Concept	**Quizzes**
Integer Powers; Rules of Algebra; Fractions; Inequalities; Summation; Quadratic Equations; Fractional Powers	Q1
Essentials of Set Theory	Q1, Q2, Q5
Intervals and Absolute Values	Q1, Q2, Q4, Q5
Solving Equations	Q1, Q5
Two Linear Equations in Two Unknowns	Q1, Q2, Q4
Exponential Functions; Sums, Products and Quotients; Simple Rules for Differentiation	Q2
Quadratic Functions; Logarithmic Functions	Q2, Q4, Q5
Implicit Differentiation; Linear Approximation; Taylor's Formula; Polynomial Approximation; Extreme Points; The Extreme Value Theorem; Economic Examples; Further Economic Examples; Local Extreme Points	Q3
Functions of two Variables; Partial Derivatives with Two Variables; The Lagrange Multiplier Method; Sufficient Conditions; Why the Lagrange Method Works	Q4
Indefinite Integrals; Integration by Parts; Integration by Substitution; Matrices and Matrix Operations; Matrix Multiplication	Q5