

The complementary use of audience response systems and online tests to implement repeat testing: A case study

Abstract:

Although learning theories suggest that repeat testing can be highly beneficial for students' retention and understanding of material, there is, so far, little guidance on how to implement repeat testing in higher education. This article introduces one method for implementing a three-stage model of repeat-testing via computer-aided formative assessment by employing audience response systems (ARS) and online tests complementarily. The first stage utilises ARS for immediate testing throughout lectures, the second stage facilitates delayed testing using online tests between lectures, and the third stage employs ARS to aid in-class revision of the previously studied material at the beginning of subsequent lectures. Using the example of a Business Economics course taught to MBA students at a UK University, the study investigates how two cohorts of students ($n_1 = 46$, $n_2 = 48$) perceived repeat testing to affect their understanding of the subject as well as their learning motivation and behaviour.

The exploratory research indicates that most students perceived all three test stages as helpful to develop their understanding of the subject. However, students who favoured a deep approach to learning rated testing more positively than students who preferred a surface approach. Surprisingly, students who favoured a deep approach to learning also reported a greater influence of the tests on their learning motivation and behaviour than students who preferred a surface approach. These findings mitigate concerns that experienced, effective learners might perceive repeated, multiple-choice based testing as unhelpful or disruptive to their learning.

Practitioner notes

What is already known about this topic

- Experiments suggest that repeat testing can enhance students' ability to recall and understand information.
- Ideally, tests on the same material should be spaced to take account of the advantages and disadvantages of immediate and delayed testing.
- Audience Response Systems (ARS) and online tests can positively affect students' attitude towards a course as well as their learning.

What this paper adds

- An explanation of the implementation of repeat testing using ARS and online tests to complement each other.
- The research indicates that repeat testing using multiple choice type questions can be helpful to students' understanding as well as their learning motivation and behaviour.
- Students who favour a deep approach to learning tend to perceive the impact of repeat testing on their understanding more positively than students who prefer a surface approach. Repeat testing also appears to impact more on the learning motivation and behaviours of students who favour a deep approach to learning.
- The consistent use of repeat testing does not necessarily lead to "testing fatigue", even for mature students who favour a deep approach to learning.

Implications for practice

- ARS and online tests can be used to complement each other to facilitate repeat testing.
- Requiring students to share ARS clickers can help facilitate collaborative learning and therefore mitigate key shortcomings of tests which utilise multiple choice questions.

Introduction

Using questions to promote generative learning, ie. the active processing and practice of relevant material and the integration of new information into pre-existing knowledge, has long been acknowledged as a powerful teaching and learning tool (Crooks, 1988; Mayer et al., 2009; Roediger and Karpicke, 2006). In the context of blended learning, computer aided assessment has noticeably improved the ability of lecturers to facilitate student participation in question based learning, in particular in large classes.

In classroom situations audience response systems (ARS), which give students the possibility to log answers to questions posed by the lecturer via hand-held devices (commonly known as "clickers"), permit large numbers of students to participate simultaneously in answering questions (Kay and LeSage, 2009). Prior research suggests that the use of ARS can help improve students' lecture attendance, their attention during lectures and their interest in a subject (MacGeorge et al., 2008; Kay and LeSage, 2009).

Moreover, ARS also improve the ability of insecure or introvert students to participate actively in class (Trees and Jackson, 2007).

Outside of the classroom, online tests have improved lecturers' ability to set, mark and provide timely feedback on homework questions, as online systems can noticeably reduce the workload involved (Angus and Watson, 2009; Lee, Courtney and Balassi, 2010).

While there is a large amount of literature which describes and analyses the integration of either ARS or online tests in courses (Trees and Jackson, 2007; MacGeorge et al., 2008; Kay and LeSage, 2009; Angus and Watson, 2009; Lee et al., 2010), we were unable to find any literature which considers the complementary use of ARS and online tests to enhance student learning. This is particularly surprising since cognitive theory suggests that, while individual tests can be useful in improving students' understanding and recall of material, repeat testing might be considerably more beneficial as it allows to space tests more effectively (Roediger and Karpicke, 2006). Using both ARS and online tests allows testing students repeatedly on the same material both in the short and longer term. Moreover, since both types of tests influence students' learning differently, they might be able to accommodate each others' didactic shortcomings.

However, alternatively, students might perceive the use of ARS and online tests as substitutive and therefore redundant. As mature students, in particular those who have a deep approach to learning, already tend to employ successful learning strategies, they might perceive highly structured learning activities such as repeat testing as an unnecessary imposition. This could lead to "testing fatigue" and affect students' motivation towards a course negatively.

This paper discusses the complementary use of ARS and online tests to facilitate repeat testing. We begin by exploring the rationale for introducing repeat testing using ARS and online tests and then outline a repeat testing model employed in a Business Economics course for post-experience MBA students. Based on the results of a survey of two cohorts of the course we explore students' perceptions of the impact of the three stages of repeat testing employed on their understanding of economics and their learning motivation and behaviour. Moreover, we analyse how students' approach to learning affected their perception of repeat testing.

Literature review

Prior research provides ample evidence that testing learned material affects students' ability to process and recall information. By providing diagnostic feedback on tests, lecturers can further help students identify and remedy misconceptions and gaps in their knowledge (Crooks 1988; Mayer et al., 2009). Diagnostic feedback can also enhance the metacognitive skills of students and encourage them to adopt more efficient learning strategies (Thomas and McDaniel, 2007; Mayer et al., 2009).

However, even if no feedback is provided, testing students' knowledge can have a more positive effect on students' ability to recall information than restudying the material (Roediger and Karpicke, 2006; Mayer et al., 2009). One frequently cited reason for this observation is that retrieving information from memory leads to an elaboration of memory traces and the creation of additional retrieval routes, which improves the likelihood that the information can be recalled correctly in the future (Roediger and Karpicke, 2006; Roediger and Butler, 2011).

As memory traces tend to be stronger, the more demanding or effortful the retrieval is (Kang, McDermott and Roediger, 2007), the type of questions asked can influence how effective tests are in developing memory traces. Requiring students to work out answers to questions and apply knowledge in different contexts involves much greater mental effort and active participation than merely remembering facts or definitions. Moreover, questions which focus on the active practice and application of recently acquired knowledge help students to develop their understanding of the material (Crooks 1988; Mayer et al., 2009; Roediger and Karpicke, 2006).

In addition to the type of questions asked, the timing of testing can also affect students' learning. While testing material shortly after it has been initially learned increases the likelihood that students correctly recall information and consequently establish a memory trace to correct rather than incorrect information (Butler and Roediger, 2007), testing material after longer intervals requires greater effort and therefore enhances students' ability to remember material for longer (Roediger and Karpicke, 2006; Roediger and Butler, 2011).

“Repeat testing” allows lecturers to take advantage of the benefits of both immediate and interval tests. Moreover, repeat testing should also encourage “distributed practice”, ie. the regular, spaced study of material. Cognitive theories suggest that spaced practice

and repeated testing tend to increase encoding variability as material is considered in different psychological contexts. This is expected to further strengthen the persistence of memory traces (Mozer, Pashler, Cepeda, Lindsey and Vul, 2009).

In higher education teachers tend to employ either immediate testing using ARS or delayed testing using online tests (Trees and Jackson, 2007; MacGeorge et al., 2008; Kay and LeSage, 2009; Angus and Watson, 2009; Lee et al., 2010). However, the literature on repeat testing and the analysis of the advantages and disadvantages of testing using ARS and online tests suggests that both forms of assessment could be used in conjunction to facilitate repeat testing and compensate for the short-comings of each method.

Asking students questions immediately after material has been discussed using ARS enables students and lecturers to gain instant feedback on students' understanding. In contrast to the traditional method of asking individual students to answer questions, the use of ARS enables all students to actively participate in working out and giving the answer (Trees and Jackson, 2007; Kay and LeSage, 2009). Moreover, the ability of ARS to instantly display the aggregate results of all students' chosen answers provides a much more realistic picture of the class' learning progress (Kay and LeSage, 2009). ARS can therefore not only help students to immediately identify misconceptions and gaps in their knowledge (Crooks 1988; Mayer et al., 2009), they can also help lecturers ascertain material which they need to discuss in more detail or more clearly (Edmonds and Edmonds, 2008).

One disadvantage of ARS is that questions tend to be presented in a multiple choice format, which does not require students to verbalise answers themselves and, indeed, allows students to guess answers. This problem can be avoided or at least reduced by asking students to share ARS clickers. In this case, students have to discuss the questions and agree on the answers. In particular when questions require students to apply their knowledge, rather than simply identify correct definitions or facts, the discussions can lead to peer-to-peer teaching as students explain to each other how they worked out the answer (Edmonds and Edmonds, 2008).

While ARS allow the immediate testing of material throughout teaching sessions, they can also be used for interval testing, eg. as part of the revision of material in subsequent teaching sessions. However, prior research suggests that the use of ARS does not impact

noticeably on students' preparation between teaching sessions (MacGeorge et al., 2008). Since repeat testing is supposed to encourage distributed practice, such as the revision and practice of the new material prior to subsequent tests (Crooks 1988), this is another possible limitation of ARS.

By contrast, setting students graded homework appears to be more effective in engaging students in regular revision and practice of new material (Palocsay and Stevens, 2008; Geide-Stevenson, 2009). While the initial development of high quality questions, distracters and feedback takes considerable time and effort (Sim, Holifield and Brown 2004), automated homework systems require less time for marking, particularly in large classes, and therefore allow lecturers to set homework more regularly (Angus and Watson, 2009, Lee et al., 2010).

Another benefit of using online tests to set homework is the timeliness of the feedback. As “feedback has to be paid attention to in order to enhance performance” (Duijnhouwer, Prins and Stokking, 2012, p. 172), the automatic release of feedback immediately after the test has been completed should enhance the chance that students use it constructively.

Finally, by asking students to take online tests a few days prior to the subsequent teaching sessions, lecturers are able to assess students' performance and identify topic areas where many students show weaknesses. This allows for a targeted use of the – usually quite limited – time for revision in class. Moreover, by using ARS to facilitate the revision, a further round of repeat testing can be implemented which can also facilitate collaborative learning and further students' ability to develop explicit arguments.

This can be helpful as, compared to the use of ARS, online tests also have a number of limitations. In particular there is usually little to no opportunity to benefit from collaborative learning. Moreover, although online tests can include a wider variety of question types than ARS, the dominance of multiple choice and multiple answer questions means that students are often not required to actively verbalise arguments and might indeed simply guess answers.

Method

Design of the repeat testing model

The course chosen for the implementation of repeat testing in line with the rationale discussed above is a Business Economics course which is part of the core curriculum of a full-time MBA programme at a UK University.

Due to the hierarchical nature of the subject, it is essential that students have a good grasp of the material discussed in previous teaching sessions in order to be able to follow subsequent sessions. Therefore, economics courses seem particularly likely to benefit from repeat testing. In addition, MBA students are notoriously wary of economics due to the subject's reputation as difficult and irrelevant (Gregorowicz and Hegji 1998; Polutnik, 2010). It was hoped that repeat testing would help build students' confidence and improve their perception of the subject, and thereby enhance their learning motivation and behaviour.

The taught component of the course consisted of eleven four-hour teaching sessions taught at weekly intervals. Teaching sessions consisted of a mix of traditional lectures, class discussions, organised group work and simulations.

In line with the discussion in the previous chapter, repeat testing was implemented in the course as follows:

To facilitate immediate testing in the classroom (see figure 1), students were asked to answer multiple-choice questions using ARS at various times throughout the lecture components of each teaching session (Immediate Test). Individual ARS questions were asked after the explanation of each new economic concept or theory. To focus students' attention on reasoning rather than recall, questions required students to apply newly learned economic concepts and theories and not merely to recall facts or definitions. In order to encourage collaborative learning and the verbalisation of arguments, students were asked to share ARS clickers with a neighbour with whom they had to discuss the question and agree on a joint answer. This was followed by a review of the results and the correct answer. If necessary, this led to a more detailed discussion of the rationale for the correct answer and how students could avoid arriving at wrong conclusions.

[Figure 1 about here]

As identified in figure 1, the implementation of delayed testing was operationalised in two stages. The first stage required students to participate in weekly online tests within five days of the preceding teaching session (Delayed Test 1). The online tests were hosted on the University's internet platform, so that students were able to log into the tests using their individual University access codes and passwords from any internet connected computer. Students were instructed to revise the preceding week's lecture material, using their own notes as well as the course lecture notes, the recommended textbook and other articles, before taking the test. The questions set in the online tests covered topics discussed in the lectures as well as in the recommended reading. Feedback on the tests, which consisted of multiple choice and multiple answer questions, was provided immediately after submission, and included information on the student's score, which questions they answered correctly or incorrectly, as well as diagnostic feedback which explained how to work out the answers or why alternatives were incorrect.

Students were advised that if they scored less than 70% in an online test, they should revise those topics they struggled with again and retake the test (Delayed Test 1b). To discourage the memorisation of correct answers and reduce the risk that students might copy their peers' results, the online system did not permit students to print or save their feedback and the tests were set up so that the sequence of questions and answer options was randomly allocated for each test attempt.

Requiring students to take the online tests at least two days prior to the subsequent teaching session ensured that the lecturer had time to review the class' performance, identify topic areas for review in the next teaching session and consider alternative ways to explain the relevant material. Moreover, it also introduced a minimum time interval between the first and second stage of the delayed testing.

The second stage of the delayed testing (Delayed Test 2) took place at the beginning of the subsequent teaching session. Using students' performance in the online tests as a guide, ARS were used to ask questions about those topic areas students struggled with most. The attempt statistics of the online tests provided data on the percentage of test attempts which answered each question correctly as well as the distribution of answers per answering option (see figure 2). While the first feature of the attempt statistics allowed the identification of questions where the class performed comparatively poorly,

eg. the wrong answer was chosen in more than 20% of attempts, the second feature provided information on students' misconceptions or where they might have gone wrong in their reasoning to work out the answer.

[Figure 2 about here]

While during the immediate testing stage individual questions were asked focussed on material covered immediately previously, in this second stage of delayed testing students were usually asked five to six questions covering different topics. Whereas during the first stage of delayed testing, feedback was only provided after students had submitted all of their answers to the online test, in this second stage of delayed testing, feedback was provided after each question was answered.

After this initial revision session at the beginning of each teaching session the repeat testing sequence started all over again, as the subsequent discussion of new material was accompanied by the use of ARS for immediate testing.

The use of ARS was an integral part of the design of each teaching session, so all attending students participated in the immediate tests and the second stage of the delayed testing. The need for students to discuss and agree their answer with a partner and the display of the number of clicker responses on the question slides in real time helped to facilitate participation rates in excess of 85% for each question.

By contrast, although online-tests were classified as formative assessment and therefore theoretically obligatory, effectively there was no hard sanction for non-participation as the online-tests did not contribute to course credits or the summative course mark. However, students who failed to take a test by the deadline were reminded by e-mail to engage in revision and take the test before the subsequent teaching session. The average participation rate in the weekly online tests for the first cohort was 92% (ranging from 98% to 76%), whereas for the second cohort it was 82% (ranging from 100% to 68%). Regarding students' engagement in the tests, 69% (69%) of students of the first (second) cohort took all online tests on time prior to the next lecture session, while 2% (9%) took less than half of the online tests on time.

The didactic rationale for using both ARS and online testing and for regular revision was explained to the students at the beginning of the course. Moreover, to strengthen

the link between the formative assessment and the summative assessment, students were advised that 30% of the marks of the final summative exam would be allocated to a multiple choice test. Although the content of the exam questions would differ from the questions used in ARS or online tests, it was indicated that both would provide a good practice for this part of the exam.

From the lecturer's perspective, while the initial design of questions, distractors and feedback for the ARS and, in particular, the online tests did take considerable time and effort, the workload involved reduced noticeably for the second student cohort, as the majority of questions developed for the first two stages of the repeat testing model were reusable and only required slight improvements or updating. For the second cohort the workload concerning the implementation of the repeat testing model therefore related mainly to monitoring participation in the online tests and developing questions for the second stage of the delayed testing.

Sample

The research was conducted on two cohorts of full-time MBA students at a UK University. In order to qualify to join the MBA programme applicants need to have at least three years management experience and usually either an undergraduate degree with at least a 2.1 classification (which is comparable to a 3.3 US GPA) or a professional qualification of similar standard. The two cohorts consisted of 46 students from 26 countries in the first year and 48 students from 18 countries in the second year. Students' age ranged from 26 to 45 years, with an average age of 31 in both years. 34% of students in the first cohort and 48% in the second cohort were female.

After the course had been completed, students were asked to participate in an anonymous online survey about their perceptions of the didactic features employed. The survey covered questions regarding a wide range of features of the course (including eg. the use of lecture notes, case studies, group work), in addition to those which were specifically geared towards eliciting their perception of the different elements of repeat testing.

The response rate for the survey of the first (second) cohort was 65.2% (67.3%). Respondents came from 19 (17) different countries, their average age was 31.8 (31.9) years; 40% (45%) of respondents were female. This suggests that they are a reasonable

representation of both cohorts. Both in terms of the descriptive statistics of the survey results and the subsequent analysis, the findings from both cohorts were consistent with each other, ie. there were no significant cohort effects.

As students' attitudes towards revision, testing and feedback are affected by their approaches to learning, the survey included the revised 20 item two-factor Study Process Questionnaire developed by Biggs, Kember and Leung (2001). 'Student approaches to learning' (SAL) theory suggests that students' learning strategy is affected by factors such as "students values' and motives, their perception of task demands, teaching and assessment methods, classroom climate, and so on" (Biggs et al., 2001, p. 134). Since many of the variables which determine students' learning strategies are unobservable as they depend on students' personal characteristics and prior experiences as well as often subtle characteristics of the learning environment (Ballantine, Duff and McCourt Larres, 2008), educational researchers have developed SAL inventories to identify differences in students' approaches to learning. Based mainly on cognitive processing theory, SAL inventories tend to rate students' preferences for a 'deep' approach to learning, which focuses on "looking for meaning in the matter being studied and relating it to other experiences and ideas with a critical approach", and for a 'surface' approach, which relies on "rote learning and memorization in isolation to other ideas" (Duff and McKinstry, 2007, p. 184).

Students who prefer a deep approach to learning (DA) are expected to perceive opportunities to engage actively in developing their understanding and to work steadily on revision more positively than students who display a surface approach to learning (SA). However, as students who favour a deep approach to learning have a greater intrinsic interest in learning and understanding, they might view testing using multiple choices type questions as disruptive to their own learning strategies. This suggests that they could perceive the later stages of repeat testing negatively.

To identify the degree to which students' engage in deep or surface learning we conducted a principal component analysis across the Study Process Questionnaire items. As expected, correlations between individual survey items indicated that students with a greater proclivity towards deep learning tend to be less inclined towards surface learning. We therefore applied oblimin factor rotation with Kaiser normalization. As

indicated by the component plot (figure 3), the items load reasonably well on the two factors in line with the predictions developed by Biggs et al. (2001).

[Figure 3 about here]

Moreover, the Cronbach's alpha values (table 1) for the items which make up the two scales are comfortably above 0.8 which suggests that the scales are internally reliable.

[Table 1 about here]

Using course grades or scores in standardised pre- and post-tests to measure the impact of repeat testing on students' learning progress would not have yielded reliable estimates, due to the diversity of students in terms of prior knowledge, academic aptitude, cultural background, language skills, as well as the range of didactic features which impact on student learning and which could not be controlled for (Sosin, Blecha, Agarwal, Bartlett and Daniel, 2004; Akyol and Garrison, 2011; Brasfield, McCoy and Milkman, 2013; Green, 2014). In line with prior literature (eg. DeBourgh 2008, MacGeorge et al., 2008, Akyol and Garrison, 2011), we therefore relied on a survey of students' own perceptions of how the different elements of repeat testing affected their understanding of the subject as well as their learning motivation and behaviour.

Analysis and discussion

Students' perceptions of repeat testing

The mean scores for the survey questions considering the usefulness of testing were all comfortably above 5, with standard deviations of below 0.8 for the use of ARS (2a, 2b, 2c, 2d, 2i) and below 0.95 for the use of online tests (3a, 3b, 3c, 3d). Given the 6-point Likert scale, this indicates that the overwhelming majority of students perceived the use of ARS and online tests as helpful to further their understanding of business economics. This suggests that the tests did not merely serve to assess students' knowledge and understanding but that the assessment was a constructive part of students' learning process.

[Table 2 about here]

[Table 3 about here]

The survey results also indicate that both online tests (3e Mean 5.1, SD 0.89, 3f Mean 4.8, SD 1.07, 3g Mean 4.4, SD 1.02) and the use of ARS for revision in subsequent teaching sessions (2j Mean 4.7, SD 1.10) incentivised many students to engage more diligently in revision. Repeat testing therefore appears to have been successful in encouraging students to increase their exposure to the relevant material in line with the repeat testing model (Figure 1).

The finding that the use of ARS to review material from previous teaching sessions (2j) encouraged students to engage in revision is inconsistent with prior research by MacGeorge et al. (2008). However, it is not clear whether MacGeorge et al. used ARS only for immediate testing or also specifically to revise the material of previous lectures. Moreover, a comparison of the responses using a Wilcoxon Signed Rank Test ($Z = -2.671$, $p < .01$) for questions 3e and 2j (see table 4) indicates that students found the online quizzes a greater incentive to participate in revision than the use of ARS.

[Table 4 about here]

It is not possible to discern how students' revision behaviour would have been affected if online tests would have been omitted. However, the results suggest that the use of ARS for revision and online tests are not perfect substitutes. This contention is further supported by the results regarding question 3j "Participating in online tests took up too much time" (Mean 3.9, SD 1.45). It indicates that students felt that the online tests made them spend more time on revision than they would have preferred.

Interestingly, students who tended to participate regularly in the weekly online tests and therefore presumably spent more time taking them, appear to have perceived them to be less onerous, as suggested by the significant negative Kendal tau correlation between questions 3k "I participated regularly in the weekly online tests" and 3j "Participating in online tests took up too much time" ($r = -.270$, $p < .05$).

The survey results also suggest that students found the revision at the beginning of each teaching session helpful to their understanding of economics (2i Mean 5.2, SD 0.79),

even though it involved testing students on key material for a third (or fourth) time. The high mean and low standard deviation of this item suggest that most students maintained a positive perception of the usefulness of the revision, which indicates that “testing fatigue” was not a problem. Comparing students’ perceptions of the helpfulness of online quizzes and the revision of material at the start of lectures to their understanding of economics, a Wilcoxon Signed Rank Test ($Z = -1.703$, $p < .1$) for questions 3a and 2i (see table 5) indicates that 60.3% of the sample rated both the same, 12.7% thought the revision and 27% thought the online tests were more helpful. This suggests that, although the Wilcoxon Signed Rank Test is statistically significant in favour of the online tests, 73% of the sample rated the revision as at least as helpful as the online tests.

[Table 5 about here]

This could be related to the fact that, although questions repeatedly covered the application of the same economic theories and concepts, the context and examples as well as the feedback tended to vary. Prior research into students’ perception of feedback highlights the need to rephrase feedback given repeatedly, even if it factually covers the same content, in order to maintain students’ engagement (Mavrikis, Gutierrez-Santos, Geraniou and Noss 2013).

Finally, the survey results show that most students felt the requirement to discuss answers to ARS questions with their neighbour was helpful to their understanding (2h Mean 5.0, SD 0.88). This suggests that many students benefitted from the opportunities for collaborative learning and peer-to-peer teaching ARS offered, which are not available in online tests.

Given the variation in the participation in the online tests between students, it would have been helpful if students’ participation records could have been linked to their survey responses. However, due to anonymity requirements this was not feasible. The fact that the distribution of the responses to a survey question on “I participated regularly in the weekly online tests” (3k Mean 5.5, SD 0.64) was limited to scores between 4 (agree) and 6 (strongly agree), suggests that students with a poor participation record in the online tests were not represented in the sample. As discussed earlier, 2% of the first and 9% of the second cohort engaged in less than half of the

online tests in time prior to the subsequent lectures. Therefore, unfortunately, these students' perceptions are probably not reflected in the survey.

The impact of students' learning orientation on their perception and use of repeat testing

As previously discussed, prior research suggests that students' approach to learning is likely to affect their perception of didactic methods. Students who tend to favour a surface approach to learning are expected to perceive didactic methods which require them to actively develop and employ their knowledge to solve problems more negatively than students who favour a deep approach to learning. However, students with a strong preference for a deep approach to learning, while being comfortable with question based learning per se, might perceive repeated multiple-choice based testing as unhelpful due to their limited complexity.

The Kendall's tau correlations of the measures for deep (DA) and surface learning approaches (SA) with the Likert scores in tables 2 and 3 indicate that students who prefer a deep approach to learning perceived the testing as comparatively more beneficial to their understanding of economics (2a $r = .174$, $p < .1$; 2b $r = .195$, $p < .1$; 2c $r = .241$, $p < .05$; 2f $r = .261$, $p < .01$; 2i $r = .243$, $p < .05$; 3b $r = .211$, $p < .05$) than students who favour a surface learning approach (2b $r = -.254$, $p < .05$; 2f $r = -.203$, $p < .05$; 2i $r = -.177$, $p < .1$; 3b $r = -.109$, $p < .1$).

Moreover, somewhat surprisingly, the Kendall's tau correlations also indicate that students with a stronger orientation towards a deep approach to learning rated the impact of the three testing stages on their learning behaviour and motivation comparatively highly (2d $r = .295$, $p < .01$; 2j $r = .197$, $p < .05$; 3e $r = .165$, $p < .1$; 3f $r = .212$, $p < .05$; 3g $r = .272$, $p < .01$; 3i $r = .245$, $p < .01$). These results not only alleviate concerns that students who favour a deep approach to learning might perceive repeated testing using multiple choice type questions as unhelpful. They also indicate that even comparatively highly self-motivated students, who have a general preference to develop a coherent understanding of subjects, appreciate support to encourage and structure their learning.

By contrast, students' preference for a surface approach to learning was significantly negatively related to their perception of the impact of testing on their learning

motivation and behaviour (2d $r = -.172$, $p < .1$; 3d $r = -.218$, $p < .05$; 3e $r = -.256$, $p < .01$; 3f $r = -.198$, $p < .05$; 3g $r = -.219$, $p < .05$; 3h $r = -.209$, $p < .05$). In line with these findings students with a preference for a surface approach to learning also reported a lower participation in the weekly online tests (3k $r = -.254$, $p < .05$). While these results do not indicate that repeat testing is detrimental to or of no use for students who favour a surface approach to learning, they do suggest that their positive impact on these students' learning motivation and behaviour is more limited.

Conclusion

This paper explores how computer aided assessment in form of ARS and online tests can be used complementarily to implement repeat testing in a real educational setting. The model applies ARS to immediately test students' understanding during lectures, online tests between lectures to facilitate an initial stage of delayed testing and ARS for revision at the beginning of subsequent lectures as a second stage of delayed testing.

An exploratory survey of post-experience MBA students participating in a Business Economics course applying repeat testing indicates that, while students' learning approaches affect their perception of repeat testing, overall repeat testing can be successfully employed not only to help students develop their understanding but also improve their learning motivation and behaviour. Surprisingly, students who favoured a deep approach to learning did not only perceive the use of ARS and online tests to be more helpful to develop their understanding of economics than students who preferred a surface approach, the findings also indicate that repeat testing had a greater impact on their learning motivation and behaviour. This suggests that even students who have an interest in developing a comprehensive, critical understanding of a subject might not necessarily be able to develop suitable study behaviours on their own and might therefore benefit from organised learning activities which encourage revision and help identify misunderstandings or gaps in their knowledge.

Moreover, results which indicate that students valued all three stages of the repeat testing model suggest that "testing fatigue" does not necessarily need to be a problem for repeat testing. However, as indicated in prior research discussed above, students' perception of repeat testing is likely to be affected by the types of questions set and the feedback provided. Academics' ability to ask analytical rather than factual multiple

choice questions, to vary the context of questions and to adjust their explanations is likely to influence students' perceptions of and engagement with repeat testing. Moreover, subjects which don't lend themselves to asking analytical multiple choice questions and non-hierarchical subjects are less likely to benefit from repeat testing.

Finally, the survey results indicate that students perceived the need to share ARS clickers with other students as helpful to their learning. This might be because sharing clickers requires students to explicitly verbalise arguments, which reduces the temptation for students to simply guess an answer and might lead to peer-to-peer teaching as students explain their reasoning to each other.

In summary, the complementary use of audience response systems and online tests to implement repeat testing can positively affect students' learning behaviour and learning progress.

References

- Angus, S.D. & Watson, J. (2009). Does regular online testing enhance student learning in the numerical sciences? Robust evidence from a large dataset. *British Journal of Educational Technology*, 40, 255–272.
- Akyol, Z. & Garrison, D.R. (2011). Understanding cognitive presence in an online and blended community of inquiry: Assessing outcomes and processes for deep approaches to learning. *British Journal of Educational Technology*, 42, 233-250.
- Ballantine, J., Duff, A. & McCourt Larres, P. (2008). Accounting and business students' approaches to learning: A longitudinal study. *Journal of Accounting Education*, 26, 188-201.
- Biggs, J., Kember, D. & Leung, D.Y.P. (2001). The revised two-factor study process questionnaire: R-SPQ-2F. *British Journal of Educational Psychology*, 71, 133-149.
- Brasfield, D., McCoy, J. & Milkman, M. (2013). Gender, measurement choice and student achievement in introductory economics. *Journal of Economics & Economic Education Research*, 14, 73-88
- Butler, A.C. & Roediger, H.L. (2007). Testing improves longterm retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, 19, 514-527.

Crooks, T.J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58, 438-481.

DeBourgh, G. (2008). Use of classroom “clickers” to promote acquisition of advanced reasoning skills. *Nurse Education in Practice*, 8, 76-87.

Duff, A. & McKinstry, S. (2007). Students’ approaches to learning. *Issues in Accounting Education*, 22, 183-214.

Duijnhouwer, H., Prins, F.J. & Stokking, K.M. (2012). Feedback providing improvement strategies and reflection on feedback use. *Learning and Instruction*, 22, 171-184.

Edmonds, C.T. & Edmonds, T.P. (2008). An empirical investigation of the effects of SRS technology on introductory accounting students. *Issues in Accounting Education*, 23, 421-434.

Geide-Stevenson, D. (2009). Does collecting and grading homework assignments impact student achievement in an introductory economics course? *Journal of Economics and Economic Education Research*, 10, 3-14.

Green, A. (2014). The case for the traditional classroom. *International Review of Economics Education*, 16, 87-99.

Gregorowicz, P. & Hegji, C.E. (1998). Economics in the MBA curriculum: Some preliminary survey results. *Journal of Economic Education*, 29, 81-87.

Kang, S.H.K., McDermott, K.B. & Roediger, H.L. (2007). Test format and corrective feedback modulate the effect of testing on memory retention. *European Journal of Cognitive Psychology*, 19, 528-558.

Kay, R.H. & LeSage, A. (2009). Examining the benefits and challenges of using audience response systems. *Computers & Education*, 53, 819-827.

Lee, W., Courtney, R.H. & Balassi, S.J. (2010). Do online homework tools improve student results in principles of microeconomics courses? *American Economic Review: Papers & Proceedings*, 100, 283-286.

MacGeorge, E.L., Homan, S.R., Dunning, J.B., Elmore, D., Bodie, G.D., Evans, E., Khichadia, S., Lichti, S.M., Feng, B. & Geddes, B. (2008). Student evaluation of

audience response technology in large lecture classes. *Education Technology Research and Development*, 56, 125-145.

Mavrikis, M., Gutierrez-Santos, S., Geraniou, E. & Noss, R. (2013). Design requirements, student perception indicators and validation metrics for intelligent exploratory learning environments. *Personal and Ubiquitous Computing*, 17, 1605-1620.

Mayer, R.E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., Bulger, M., Campbell, J., Knight, A. & Zhang, H. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology*, 34, 51–57.

Mozer, M.C., Pashler, H., Cepeda, N., Lindsey, R. & Vul, E. (2009). Predicting the optimal spacing of study. In Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C.K.I. & Culotta, A. (Eds.), *Advances in Neural Information Processing Systems 22* (pp. 1321-1329). La Jolla, CA: NIPS Foundation.

Palocsay, S.W. & Stevens, S.P. (2008). A study of the effectiveness of webbased homework in teaching undergraduate business statistics. *Decision Sciences Journal of Innovative Education*, 6, 213-232.

Polutnik, L. (2010). The case for economic reasoning in MBA education revisited. *The American Journal of Economics and Sociology*, 69, 78-84.

Roediger, H.L., & Butler, A.C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15, 20-27.

Roediger, H.L. & Karpicke, J.D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1, 181-210.

Sim, G., Holifield, P. & Brown, M. (2004). Implementation of computer assisted assessment: lessons from the literature. *Alt-J Research in Learning Technology*, 12, 215-229.

Sosin, K., Blecha, B., Agarwal, R., Bartlett, R. & Daniel, J. (2004). Efficiency in the Use of Technology in Economic Education: Some Preliminary Results. *The American Economic Review*, 94, 53-258.

Thomas, A.K. & McDaniel, M.A. (2007). The negative cascade of incongruent generative study-test processing in memory and metacomprehension. *Memory & Cognition*, 35, 668-678.

Trees, A.R. & Jackson, M.H. (2007). The learning environment in ARS classrooms. *Learning, Media and Technology*, 32, 21-40.

Tables and Figures

Figure 1: Repeat testing model

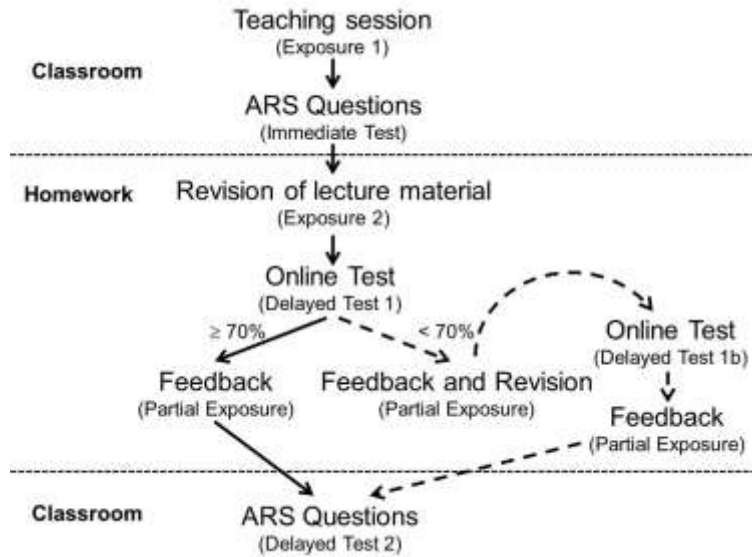


Figure 2: Example of a multiple choice question and its feedback in an online quiz

| Question and lures | <i>% answered</i> |
|--|-------------------|
| If the US government would impose an additional 40% tariff on imported steel, steel prices in the USA would | |
| - increase while global steel prices would remain the same; | 20% |
| - increase while global steel prices would fall; | 40% |
| - increase while global steel prices would increase; | 33.33% |
| - remain the same while global steel prices would fall. | 6.67% |
| <i>Unanswered</i> | 0% |
| <p>Automatic feedback</p> <p><i>Correct answer:</i> increase while global steel prices would fall.</p> <p>Since the USA is a very large economy compared to the global steel market, if the USA would impose a tariff, it would affect global demand and supply sufficiently to have an impact on global steel prices. As the tariff would make foreign steel imports into the USA more expensive, steel prices in the USA would increase, even though domestic steel producers would probably be able to expand their domestic supply to some degree.</p> <p>As foreign producers can export less to the United States, the available supply of steel on world markets is higher, which leads to a reduction of global steel prices. The larger the tariff imposing economy compared to the world market, the more it is able to affect world market prices. This means that the prices which consumers in the tariff-imposing country have to pay are less than the old price plus the tariff, since some of the tariff costs are effectively borne via falling prices on international markets by the foreign steel producers.</p> <p>PS: The reduction of the prices on the global steel market would not only affect steel producers in steel exporting countries such as China, India and Australia, but also in steel importing countries such as the UK and Germany. As world market prices for steel would fall, German and UK producers who struggle in terms of price competitiveness against cheaper Chinese, Indian and Australian suppliers might now become too expensive and be at risk of having to leave the market. They might therefore lobby their own governments either to intervene against the USA's tariffs or to impose tariffs as well.</p> | |

Figure 3: Principal component analysis – component plots

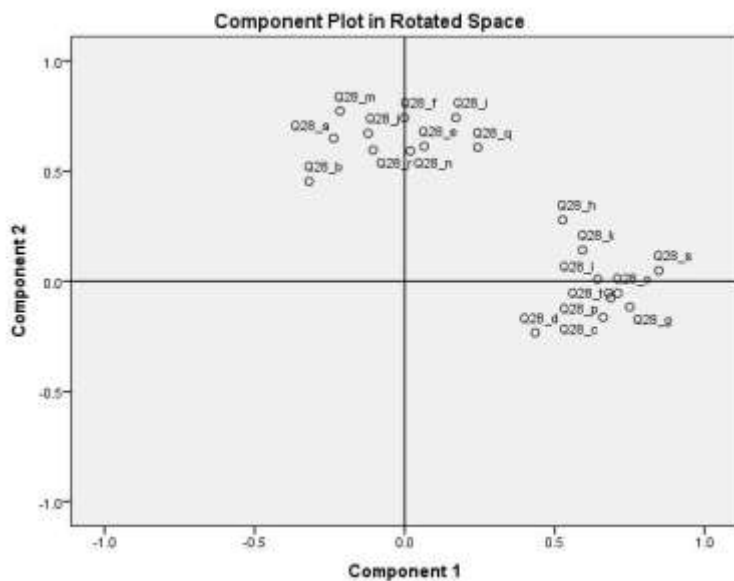


Table 1: Deep and Surface approaches to learning

| | Min | Max | Mean | Median | SD | Cronbach α |
|----|--------|-------|-------|---------|-------|-------------------|
| DA | -2.102 | 2.554 | 0.000 | -0.1035 | 1.000 | 0.854 |
| SA | -1.558 | 2.697 | 0.000 | -0.1253 | 1.000 | 0.858 |

*Table 2: Students' perceptions of the use of ARS - Kendall's tau correlations
(N=63; 6-point Likert scale)*

| | Mean (SD) | DA | SA |
|--|---------------|-----------------|-----------------|
| 2a) Thinking about how to answer ARS questions in lectures helped my understanding of economics. | 5.3 (0.67) | .174* .083 | -.165 .101 |
| 2b) The explanation of the answers to ARS questions helped my understanding of economics. | 5.4 (0.66) | .195* .054 | -.254** .012 |
| 2c) The ARS questions helped me to identify which concepts I understood well and which I didn't | 5.3 (0.66) | .241** .017 | -.078 .437 |
| 2d) Participating in ARS questions helped me identify issues I needed to learn more about. | 5.3 (0.69) | .295** .003 | -.172* .086 |
| 2e) I think that the result of the ARS questions encouraged the lecturer to explain concepts and their applications more clearly the second time around. | 5.3 (0.65) | .295*** .003 | -.226** .025 |
| 2f) The use of ARS helped my attention in lectures. | 5.1 (0.78) | .261*** .009 | -.203** .040 |
| 2g) The use of ARS helped me to participate more actively in this class. | 5.0 (0.84) | .186* .061 | -.031 .757 |
| 2h) I think I learned more using ARS because I had to discuss the answer with my neighbour before voting. | 5.0 (0.88) | .159 .109 | -.175* .076 |
| 2i) The revision of the material from the previous week at the beginning of lectures helped my understanding of business economics. | 5.2 (0.79) | .243** .014 | -.177* .074 |
| 2j) The revision of the material from the previous week at the beginning of lectures increased my motivation to revise between lectures. | 4.7 (1.10) | .197** .040 | -.143 .138 |
| 2k) There were too many ARS questions in lectures. | 2.4 (1.23) | -.174* .068 | .337*** .000 |

* Correlation is significant at the 0.10 level (2-tailed); ** at the 0.05 level (2-tailed); *** at the 0.01 level (2-tailed).

The Likert scale ranged from 1 = strongly disagree to 6 = strongly agree.

*Table 3: Students' perceptions of online tests - Kendall's tau correlations
(N=63; 6-point Likert scale)*

| | Mean (SD) | DA | SA |
|---|---------------|-----------------|------------------|
| 3a) Online tests were helpful to develop my understanding of business economics. | 5.4 (0.68) | .046 .645 | -.156 .120 |
| 3b) The explanation of the answers to the questions in the online tests helped my understanding of economics. | 5.2 (0.87) | .211** .034 | -.189* .058 |
| 3c) Participating in online tests helped me identify concepts which I understood well and which I didn't. | 5.3 (0.71) | .127 .206 | -.150 .137 |
| 3d) Participating in online tests helped me identify issues I needed to learn more about. | 5.3 (0.93) | .144 .148 | -.218** .028 |
| 3e) My incentive to revise the lecture material between lectures increased because I was required to participate in online tests. | 5.1 (0.89) | .165* .093 | -.256*** .009 |
| 3f) Because of the online tests I tended to revise the lecture material more carefully between lectures than I otherwise would have done. | 4.8 (1.07) | .212** .030 | -.198** .042 |
| 3g) If I scored less than 70% I did some more revision of the lecture material before I retook the test. | 4.4 (1.02) | .272*** .006 | -.219** .027 |
| 3h) For the online tests I tended to consider carefully the feedback on questions I answered wrong. | 5.3 (0.85) | .091 .361 | -.209** .042 |
| 3i) For the online tests I tended to consider carefully the feedback on questions I answered correctly. | 4.3 (1.34) | .245*** .009 | -.035 .713 |
| 3j) Participating in online tests took up too much time. | 3.4 (1.46) | -.009 .923 | .247*** .008 |
| 3k) I participated regularly in the weekly online tests. | 5.5 (0.64) | .157 .130 | -.254** .014 |

* Correlation is significant at the 0.10 level (2-tailed); ** at the 0.05 level (2-tailed); *** at the 0.01 level (2-tailed). The Likert scale ranged from 1 = strongly disagree to 6 = strongly agree.

Table 4: Wilcoxon signed rank test of delayed testing methods' impact on motivation to engage in revision

| | N | Mean Rank | Sum of Ranks |
|-----------------------------------|----|-----------|--------------|
| 3e_2j Negative Ranks ^a | 11 | 15.27 | 168.0 |
| Positive Ranks ^b | 25 | 19.92 | 498.0 |
| Ties ^c | 27 | | |
| Total | 63 | | |

Notes: a. 3e < 2j. b. 3e > 2j. c. 3e = 2j. Z = -2.617 (based on negative ranks). Sig. (two-tailed) 0.008. 3e) "My incentive to revise the lecture material between lectures increased because I was required to participate in online tests". 2j) "The revision of the material from the previous week at the beginning of lectures increased my motivation to revise between lectures".

Table 5: Wilcoxon signed rank test of delayed testing methods' impact on understanding

| | N | Mean Rank | Sum of Ranks | % of sample |
|-----------------------------------|----|-----------|--------------|-------------|
| 3a_2i Negative Ranks ^a | 8 | 13.06 | 104.50 | 12.7% |
| Positive Ranks ^b | 17 | 12.97 | 220.50 | 27.0% |
| Ties ^c | 38 | | | 60.3% |
| Total | 63 | | | 100.0% |

Notes: a. $3a < 2i$. b. $3a > 2i$. c. $3a = 2i$. $Z = -1.703$ (based on negative ranks). Sig. (two-tailed) 0.088. 3a) “Online tests were helpful to develop my understanding of business economics”. 2i) “The revision of the material from the previous week at the beginning of lectures helped my understanding of business economics”.