

PASAA

Volume 55

January - June 2018

**The Effectiveness of Online and Paper-Based
Formative Assessment in the Learning of English
as a Second language**

Paul Joyce

Faculty of Law, Kindai University, Japan

Email: pauljoyce@hotmail.com

Abstract

This article investigates the relative effectiveness of online and paper-based formative assessment in the learning of English as a second language. For this study, 74 Japanese university second language (L2) students took online quizzes, while 71 were given paper-based quizzes to complete. The quiz content was drawn from learning material found in the class textbook. Both groups were assigned the exercises as homework. At the end of the 15-week study period, both groups were administered a summative assessment that also covered learning points from the textbook. After controlling for the learners' L2 proficiency through their TOEIC test scores, the performance of the two groups on the formative quiz was compared through correlation and regression analysis. While there was found to be a statistically significant relationship between the online formative assessment and summative exam

scores, this was not the case for the paper-based homework. When compared through an analysis of covariance, there was found to be a significant interaction between group and homework on exam scores. The online group was more affected by their homework score than the pencil and paper group. The findings from this study have implications for teachers, learners and curriculum developers.

Keywords: online learning, Interactive learning, Formative assessment, Summative assessment

Introduction

Educators are constantly aspiring to maximize student achievement. One means of seeking improved learner outcomes is through formative assessment. Whereas the purpose of summative assessment is to reveal whether students have reached the learning objectives of the course, formative assessment provides information to teachers and students on their progress in accomplishing these learning goals. Formative assessment has many learning benefits. Firstly, it has been shown that frequent formative testing results in greater continuous study throughout a course. As a result, summative scores increase (Fitch, Drucker & Norton, 1951). Secondly, by receiving feedback on their quiz answers, formative assessment enables teachers and learners to direct their effort towards material that has yet to be mastered (Black & Wiliam, 1998).

However, the benefits of formative assessment are argued to go beyond simply encouraging greater or better directed study. Indeed, laboratory studies have shown the likelihood of retrieving information from memory is enhanced by taking a test on learned information relative to simply restudying it again (e.g. McDaniel & Masson, 1985; Roediger & Karpicke, 2006). This is known as the testing effect. The testing effect is hypothesized to be related to memory retrieval. It is argued that effortful retrieval of material enhances the permanence with which it is stored in memory (Bjork

& Bjork, 1992). However, outside of highly controlled laboratory experiments, the evidence for the benefits of formative assessment in subsequent summative evaluation is less clear cut. Nevertheless, while there are some studies that do not show an association between formative and summative test performance (e.g. Haberyan, 2003; Kluger & DeNisi, 1997) most do (e.g. Barbarick, 1998; Hagen, 2000). A factor that is regarded as of importance in predicting the value of formative to summative assessment is the degree of overlap in the two types of tests. This includes how much similarity there is in the content of the material tested (e.g. Martin & Srikameswaran, 1974) and the format of the tests used (Dobson, 2008). The greater the similarity in the test format (e.g. multiple choice, short answer etc.) between the formative and summative tests, the more likely the completion of the former will improve performance on the latter.

Traditionally, formative assessment has been conducted through pencil and paper style tests. However, in recent years, owing to advances in technology, there has been tremendous growth in the use of online assessment tools (Angus & Watson, 2009). Owing to the ease with which students can access these materials outside of the classroom, such online quizzes have often been set as homework at the expense of pencil and paper style assessment methods (Johnson, 2006). Computer assisted learning offers a number of advantages over traditional approaches. These include the advantage to teachers of reduced administration from no longer having to collect and grade homework, input test scores or redistribute homework. Furthermore, there are benefits to students in receiving immediate feedback on their performance, and from the opportunity to repeat an assessment. In order to systematically evaluate the educational value of online formative assessment, there have been a number of research studies on the subject.

In some cases, online formative assessment has not been found to be associated with improved learning outcomes. In the context of a mathematics course for adult learners, a group of 22 students was provided with supplementary online materials that

included interactive exercises (Li & Edmonds, 2005). The control group was taught in a traditional style class. When the pre-test and post-test scores of the treatment group were compared with that of the 16 members of the control group, those students who had access to the online materials were not shown to have significantly benefitted from their study medium. While the small sample size of this study raises doubts over the findings, it is not the only study to reach such conclusions. Indeed, in the case of a developmental psychology course, access to computerized formative assessment in preparation for summative assessment was found to result in poorer exam performance (Brothen & Wambach, 2001). In the case of this study, an explanation for the findings could relate to the course textbook closely corresponding with the online questions. Due to this overlap, it is possible that the participants heavily relied upon the course materials to answer the online quiz questions. While this strategy may have maximised their online performance, the students' reduced cognitive engagement with the quiz questions could have detracted from the benefits usually attained from formative assessment.

In contrast, there have been a number of studies that have suggested online formative assessment holds educational value. Orr and Foster (2013) examined the effects of frequent online quizzing on student test performance in a university biology class. Participants were assigned 10 online quizzes to complete over the semester. Each quiz contained 10 test items. It was found that those students who attempted each of the formative assessments performed better than those who did not try any of them. Similarly, Kibble (2007) provided the 350 learners in his study with two online formative quizzes prior to summative examinations. Each of these quizzes was comprised of 20-30 multiple choice items. The researcher wished to learn whether manipulating the incentives for completing the quizzes influenced student participation. The results showed quiz scores to be significantly correlated with summative test performance and that by increasing incentives for involvement from zero percent per quiz to two percent, student participation in the formative assessments increased. In a further pertinent study,

Johnson (2006) also evaluated the relationship between online quiz usage and achievement. This research was conducted with 112 undergraduate educational psychology students who had the opportunity to attempt 14 true-false and 14 short-answer quizzes. As with the previous two studies, it was discovered that student use of online formative quizzes was related to increased academic achievement in the summative final exam. However, while the results from the above studies suggest the value of online formative learning towards summative assessment, the role of learner proficiency upon test scores was not taken into account. Therefore, it is unclear whether participation in the formative tests and later enhanced achievement on the summative tests was simply due to relative underlying proficiency.

To address the role of proficiency in formative assessment and its subsequent influence on summative assessment, Angus and Watson (2009) conducted a study with a first-year applied mathematics class. 1500 students based in Australia were provided with four online formative quizzes to complete over a semester. After controlling for proficiency, there was found to be a significant relationship between completion of the formative quizzes and performance on the semester ending summative test. However, while this well-designed study provides the most compelling case for the value of online formative assessment in improved learning outcomes, it is unclear whether the findings can be generalised to learners with different cultural backgrounds who are studying different academic subjects and using a different style of learning materials.

The above studies provide some guidance on the value of online formative assessment. However, as teachers have typically set pencil and paper based homework, it is also important to compare the effectiveness of computerized quizzes with the traditional approach to formative assessment. In a study involving 122 undergraduate biology students, the relative merits of paper-based lecture notes and practice quizzes were compared with their digital equivalents (Macedo-Rouet, Ney, Charles, & Lallich-Boidin, 2009). The paper-based group was found to fare significantly better

than the online learning group on the identical practice quizzes. However, the researchers surmised that the difference was due to the relative ease with which students could search the paper-based materials for the answers to quiz questions, rather than the medium in which the test items was delivered. To better understand learning differences solely from the medium in which quizzes are administered, Bonham, Deardorff and Beichner (2003) conducted a study with several hundred students who were enrolled in a physics class. The homework assignments reflected the practical differences in the implementation of online and paper-based homework. While the paper-based group was asked to provide detailed drawings and supporting calculations, due to limitations in computer technology, the online group was just required to provide their answers. Also, regarding the deadline for homework, in order to allow time for their responses to be graded by hand, the paper-based group had to submit their homework earlier than the online participants. Nevertheless, despite the difference in the two types of homework, the researchers did not find a significant difference in the summative test results for the two groups.

From reviewing the literature, there is some evidence to support the educational value of formative assessment. However, as research findings have conflicted, there is by no means a clear consensus on the benefits of computerized formative quizzes towards summative evaluation. In addition, research comparing the benefits of online formative assessment with traditional pencil and paper learning is both sparse and divergent in its findings. Furthermore, since previous studies have generally focused on science classes in western countries, there is considerable doubt whether the findings from these works generalize to other contexts. With this in mind, the following research questions were pursued:

1. In the university EFL context in Japan, is formative online quiz performance significantly related to summative learning?
2. In the university EFL context in Japan, is formative pencil

and paper quiz performance significantly related to summative learning?

3. In the university EFL context in Japan, is formative online quiz performance significantly more effective than paper-based homework towards summative learning?

Methodology

Participants

The study took place at a university in Japan. The participants were native Japanese L1 speakers who were enrolled at the university as non-English majors. All of the participants were second year students taking a weekly EFL speaking/listening class. The students had been routinely divided into six class groups based on their proficiency. Half of the classes used online homework and half paper-based homework. Aside from this, the syllabi and course materials were the same for all of the participants. The researcher was the instructor for all of the classes. There were 77 students in the online homework group and 76 in the paper-based homework group. However, as the list-wise approach to missing data was used, only the data from the learners that completed the final summative exam and the TOEIC test was included. This left 74 participants in the online group and 71 in the paper-based group. In terms of proficiency, the participants could broadly be described as being from a false beginner to an intermediate level.

Materials

The content for the formative and summative assessment was drawn from an English language textbook (Helgeson, Brown & Wiltshier, 2010). During the 15-week study period, the participants covered five units of the textbook. The final summative exam was a multiple-choice test that drew on material from each of these five units. It was comprised of 60 equally-weighted test items. Of these, there were 15 listening questions, 12 grammar, 20 vocabulary and 13 conversation items requiring the learners to order sentences to form a natural exchange. The final exam was worth 20 percent of

the students' grade on the course and was administered in the last week of the 15-week course.

The online homework group received their formative assessment through the open-source Moodle learning platform (Version 2.8.10, 2016). The students accessed 31 multiple-choice online quizzes through a username and password. For each of the five units of the textbook that was covered during the semester, there were five online quizzes to complete. In addition, there were a further six quizzes that drew from material from all five units. Of these, there were 13 vocabulary, six grammar, and one conversation quiz. They each contained ten original test items focused on learning points from the textbook. There were also 11 listening quizzes which were comprised of between four and eight questions. These items were identical to those found in the textbook and the textbook audio was embedded into them. There were typically four possible answer choices for each quiz item. Students could attempt each quiz as many times as they wished. Once a quiz was completed, students were provided with automated feedback on which items they had correctly and incorrectly answered. Each of the quizzes was weighted equally, and the participants were awarded the highest score they attained on each one. The online system also enabled the participants to track their overall progress. To reward the students for their work, the online materials were worth ten percent of the students' overall course grade.

For the purposes of their formative assessment, the paper-based homework group completed a series of cloze activities. These cloze tasks were printed in the back of the students' textbooks and were drawn from the classroom listening activities. For each of the five textbook units that were covered, the participants completed between one and four cloze conversation tasks. The number of items within each cloze task ranged between six and 24 questions. In total, the paper-based group was tasked with completing 127 items. By undertaking the cloze activities, the participants reviewed vocabulary and grammar points that were directly evaluated in the final summative assessment. Furthermore, while completing their homework, the students were exposed to a large number of

conversations. The content of these conversations was not directly evaluated in the final summative test. However, since completing this work involved reviewing the semantic, syntactic and pragmatic properties of natural conversation, it was expected that this would aid the participants in the conversation items found in the final summative test. As was the case with the online homework, the paper-based formative assessment was worth ten percent of the students' final grade.

As a proxy for L2 proficiency, the students' TOEIC test scores were used. All of the participants were administered the pencil and paper version of the standardised test which consists of two timed sections of 100 listening and reading questions.

Procedure

The data collection occurred over an 18-month period. In the autumn semester of their first year, the data was collected for the paper-based group. Exactly a year later, data was collected for the online group. Aside from the nature of the formative assessment, both classes had the same teacher and followed the same syllabus. The summative assessment for both groups was identical and was administered in the last week of the 15-week course. Both groups were asked to prepare for the final exam. All of the participants were encouraged to review the relevant chapters of the textbook, particularly the list of keywords and target grammar provided. However, the online group was also reminded that they could also complete the computerized quizzes. The students took the TOEIC test as part of their institutional requirements. This was taken during week 8 of the course. The reliability of the TOEIC test has been found to be .95-.96 (Liao, Hatrak, & Yu, 2010). The excellent test consistency value is strong evidence for the comparability of TOEIC test scores across test administrations.

The online homework group was informed about their formative assessment in the first week of the 15-week course. They could undertake the work at any point during the semester. Throughout the semester, the learners were reminded about their homework and encouraged to complete it. Aside from when they

were first introduced to the online materials in the first week of the course, the learners accessed the quizzes outside of class. The online quizzes could be completed on a wide range of internet connected devices including smart phones, PCs, tablets and Apple computers. However, since all of the online group accessed the quizzes using their own smart phone when they were introduced in class, it is assumed that this was the main means of usage. The deadline for completing the online homework was the start of the last class of the 15-week semester. The students were provided with such a late deadline because the feedback on their answers and the provision of their overall cumulative total score did not require any teacher intervention. Furthermore, minimal administrative effort was required to collect their final scores.

The paper-based homework group was also notified of their homework requirement in the first week of class. They submitted their formative assessment in two parts. In week 10 of the class, they brought the homework pertaining to the first three units of the course that had been covered. And in week 13, they were required to bring the work relating to the last two units of the course. As the quizzes were low-stakes tests, they were marked in class by a fellow student. This also provided the learners the chance to ask questions. After the marking was complete, the teacher collected the scores from the students. Unlike the online formative assessment, the homework was submitted prior to the very end of the semester. This was done for two reasons. Firstly, it allowed students who were absent from class or who had forgotten their textbook to submit their homework at a later date. And secondly, the interval between the submission date and the end of the semester provided time for the teacher to enter the student scores into the grade book. In the last two weeks of the semester, both the online and pencil and paper based group were reminded of the summative test and encouraged to review the textbook material.

Results

Research question one: In the university EFL context in Japan, is formative online quiz performance significantly related to summative learning?

As previously mentioned, there were 31 different online quizzes and students were able to complete each as many times as they wished. Of the 74 participants who were in the online group, 73 completed at least one quiz and the highest number of quizzes completed by one person was 80. The students undertook a mean average of 45.30 quizzes ($SD = 22.10$). Of these, 20.47 quizzes ($SD = 14.84$) on average were being completed for a second time or more. As can be seen in Table 1, using the highest score that they accrued on each test that they took, the students achieved an average score of 70.26% ($SD = 33.36$) overall on the formative quizzes. Their average score on the final summative exam was 83.06% ($SD = 7.95$). The online group scored between 210 and 535 on the TOEIC test with an average score of 410.07 ($SD = 65.69$).

Table 1: Descriptive results for the online formative quiz group

Online formative		Summative exam		TOEIC	
Mean (%)	SD	Mean (%)	SD	Mean	SD
70.26	33.46	83.06	7.95	410.07	65.69

To evaluate the strength of the relationship between scores on the online formative homework and the summative exam, while holding the effect of L2 proficiency constant, a partial correlation analysis was performed. The online quiz scores were found to be significantly related to how well students performed on the summative exam, ($r = .25, p < .05$). The results suggest a positive relationship of moderate strength (Cohen, 1988) between performance on the summative and formative tests.

The data was also subjected to an analysis of covariance. The covariate, TOEIC scores, was significantly related to summative exam scores, $F(1, 37) = 31.57, p < .001$, partial $\eta^2 = .46$. There was also a significant effect of online formative quiz scores on summative exam scores after controlling for the effect of TOEIC

scores, $F(35, 37) = 2.00, p < .05, \text{partial } \eta^2 = .66$.

Research question two: In the university EFL context in Japan, is formative pencil and paper quiz performance significantly related to summative learning?

There were 71 students who were tasked with completing paper-based quiz questions as part of their course requirements. Of these, 68 completed at least some of this work. The highest score on this assignment was 100%. As shown in Table 2, the mean score on the paper-based quiz homework was 72.08 ($SD = 28.31$). And the average score on the summative class exam was 81.83 ($SD = 7.74$). The TOEIC scores of the pencil and paper group ranged between 255 and 650 with an average score of 406.41 ($SD = 78.80$).

Table 2: Descriptive results for the pencil and paper formative quiz group

Pencil and paper formative		Summative exam		TOEIC	
Mean (%)	SD	Mean (%)	SD	Mean (%)	SD
72.08	28.31	81.83	7.74	406.41	78.80

When the effect of L2 proficiency was controlled, there was not found to be a statistically significant relationship between the pencil and paper formative assessment and the summative exam, ($r = -.07, p = \text{n.s.}$).

When an analysis of covariance was performed, the covariate (TOEIC scores) was significantly related to the summative exam scores $F(1, 60) = 5.08, p < .05, \text{partial } \eta^2 = .08$. However, as expected, there was not found to be a significant relationship between the summative exam and pencil and paper formative scores after controlling for English proficiency through the TOEIC scores, $F(9, 60) = .49, p = \text{n. s.}$

Research question three: In the university EFL context in Japan, is formative online quiz performance significantly more effective than paper-based homework towards summative learning?

To determine whether there is was a difference between the

online homework and the paper-based homework groups, an analysis of covariance (ANCOVA) was conducted. The summative exam scores were used as the dependent variable, TOEIC scores and group affiliation as the independent variables, and TOEIC scores as the covariate. As expected, TOEIC scores were significantly related to exam scores, $F(1, 97) = 17.18, p < .001$, partial $\eta^2 = .16$. Neither homework nor group was found to be significantly related to the dependent variable. However, there was a significant interaction between group and homework on exam scores, $F(5, 97) = 2.68, p < .001$, partial $\eta^2 = .12$. This indicates that the two groups were affected differently by the homework they were given. From examining a graph of the interaction between group and homework on exam scores, it can be seen that the summative exam scores of the online quizzes group were significantly more affected by their homework score than the pencil and paper group. This result suggests that completion of the online homework was more beneficial to test performance than undertaking the traditional style formative materials.

Discussion

The results from this study shed light on the relationship between the medium of formative tests and summative test results. The response to the first research question showed that, after controlling for L2 proficiency, performance on the formative online quizzes was related to summative test scores. In regard to the second research question, after removing the influence of L2 proficiency, the pencil and paper formative assessment used in this study was not found to form a relationship with summative test results. Finally, after removing the influence of proficiency, the formative quiz score performance of the online group was found to be significantly more related to summative test performance than that of the pencil and paper group. In other words, completion of the online homework was more beneficial to test performance than undertaking the traditional style formative homework.

The results from this study both accord with many of the previous studies in the field and stand in contrast with others. To

explore the reasons for this, it is worth restating that the research took place outside of the confines of a controlled laboratory experiment. As such, the medium in which the formative assessment occurred was not manipulated in isolation. Instead, the inherent differences in the two mediums used had a number of subsequent effects. As will be explained, these dissimilarities likely explain why completion of the online formative quizzes was found to be related to summative test performance and why the pencil and paper homework was not.

Firstly, a difference between online and traditional quizzes is ease of grading. The students' online scores could be accessed online at anytime. And since each student's quizzes were graded and their cumulative score tallied automatically, minimal administrative effort was required to transfer the learners' scores into the teacher's grade book. On the other hand, checking the paper-based homework was dependent upon the student coming to class and bringing their textbook. As this did not always happen, some time needed to be allowed for the submission of late work. Also, time was needed for data entry and to transform the raw scores into a final percentage. Due to these differences, while the deadline for the online quizzes could be set as the last week of class, the pencil and paper homework was due a few weeks earlier. As there was a longer gap between the pencil and paper group's deadline and the summative test, there was more chance for the memory trace to fade due to decay, interference or consolidation (Carpenter, 2012). As discussed in the Introduction, Macedo-Rouet, Ney, Charles and Lallich-Boidin (2009) found their paper-based group performed better than their online group in the summative assessment. One reason for the difference in the findings from that study and this one is that they did not provide their online formative assessment group any additional time.

A second difference in how the medium of the formative quizzes could influence learning relates to feedback. While the online quizzes gave immediate information to the students on how they had performed, the feedback on the traditional style homework occurred collectively during class time. Since early feedback is more

effective than delayed feedback (Bangert-Drowns, Kulik, Kulik & Morgan, 1991), this may also help account for the results from this study. This point may also help explain the difference in the findings between those studies reporting benefits to online learning and those that did not. In the majority of studies that found formative online quizzes beneficial to summative assessment (e.g. Angus & Watson, 2009; Kibble, 2007; Johnson, 2006), feedback to participants on their performance in the formative quizzes was provided immediately. On the other hand, for the studies that did not find computerised formative assessment to be superior to pencil and paper style formative exercises (e.g. Bonham, Deardorff & Beichner, 2003; Macedo-Rouet, Ney, Charles, & Lallich-Boidin, 2009), feedback was not provided to either group prior to grading.

Thirdly, the online and traditional formative testing mediums differ in the opportunity and incentives to review. In the case of the computerised questions, the participants could attempt them afresh as many times as they wished and thereby potentially increase their score. As was discussed, many of the participants took this chance. On the other hand, once the pencil and paper group's homework had been graded, it could not be reattempted anew. Since effortful retrieval and repeated quizzing are associated with learning (Karpicke & Roediger, 2007), this difference helps explain the results from this study. In addition, the opportunity and incentive to meaningfully review past quizzes also helps explain the results from previous studies. The research that found online work beneficial to summative learning tended to allow the unlimited submission of formative quizzes (e.g. Kibble, 2007; Johnson, 2006), while the studies that were equally or more supportive of paper-based quizzes did not (e.g. Bonham, Deardorff & Beichner, 2003; Macedo-Rouet, Ney, Charles, & Lallich-Boidin, 2009).

Lastly, in this study, the two mediums differed in their functionality. As previously discussed, a number of listening quizzes were made available to the online group. These materials contained embedded audio material that the students could easily access while completing the quiz questions. On the other hand, due to the nature of the materials they were using, the pencil and paper

group did not have the same opportunity. Conversely, as was discussed in the Introduction, due to limitations in computer technology, Bonham, Deardorff and Beichner (2003) disadvantaged their online group by requiring less detailed and comprehensive formative homework from them. Thus, medium functionality helps explain the results from this research, and the findings from a diverging study.

The findings from this study have implications for classroom teaching practice. The results show that online formative assessment is an effective means of improving student performance on summative tests in the field of L2 acquisition. Therefore, with the increasing availability of free online study tools such as Moodle (2018), Blackboard (2018), and Quizlet (2018) and the spread of smart phones and internet connected computers, there is an opportunity for instructors to enhance their students' learning, save in-class instructional time, and reduce the time spent on grading and administration. However, through comparing this study with previous research in the field, it is clear that to gain the educational benefits from using online formative quizzes, the differences between computerised quizzes and pencil and paper materials need to be exploited. Firstly, as has been discussed, due to the minimal administrative burden placed on teachers by online quizzes, the deadline for their completion can be much closer to the final summative test than that of traditional style quizzes. By enabling students to incorporate the online quizzes into their preparation for summative tests, teachers can provide a powerful incentive for learners to complete them. Secondly, in light of the educational benefits of immediate feedback to learners, teachers should ensure that online quizzes offer such support. For instance, when designing online materials, teachers should consider allowing students know which questions they have answered correctly and potentially even provide hints to questions they answer incorrectly. Thirdly, teachers should provide the opportunity and incentive for learners to resubmit online quizzes. As was the case in this study, by students being awarded the highest score they achieved on a quiz, there was an inducement for students to repeat quizzes and

many learners took this opportunity. Additional study from repeating quizzes results in greater learning. Lastly, when designing online quizzes, instructors should seek to take advantage of the functionality of the medium. For instance, since online quizzes lend themselves to the use of audio and video, there is the opportunity for instructors to include multi-media resources in their materials. This provides opportunities for students to improve their listening. In short, when incorporated wisely into a program of study, there are clear benefits to both students and instructors of online out-of-class formative assessment.

As discussed, the use of online formative assessment allows instructors to realise educational benefits through adjustments and improvements in such areas as the grading deadline, the immediacy of feedback, the opportunity and incentive to resubmit quizzes, and quiz functionality. And decisions regarding these points help explain the similarities and differences in the results from this study and previous research. However, the manner in which these points were incorporated into the research was subjective. As such, the lack of a principled, research-based approach to their inclusion in this study constitutes a limitation. A further limitation to this research relates to the lack of clarity over the individual and collective influence of the points above. That is, it not possible to differentiate how much of the benefit derived from the formative online materials was due to the later deadline for their completion, the availability of feedback, the opportunity to resubmit work, or quiz functionality etc. Nevertheless, since the context for this study was the classroom rather than the laboratory, the research design needed to realistically reflect how online and paper-based formative quizzes are employed by teachers. Therefore, rather than conducting a highly controlled experiment, the goal was ecological validity. On this point, it should also be noted that there were more online quiz materials than paper-based materials available for students. Although this is a further contaminating factor, it also reflects the authentic nature of the study. While there is no additional cost in adding online materials to be completed out of class, printing large quantities of additional paper-based materials

that are checked in class, is expensive and reduces the time for instruction. A final limitation pertains to the use of different versions of the TOEIC test as the proxy for L2 English proficiency. As discussed, the TOEIC test is a highly reliable measure of proficiency. Nevertheless, the use of differing versions introduced unwanted, albeit limited, error variance into the results.

In conclusion, this study provides support for using online formative quizzes for summative L2 English proficiency assessment. As such, the research provides evidence for the generalizability of laboratory testing effects in the classroom setting. However, by comparing the results from this study with those of previous works, the importance of the way in which quizzes are incorporated into the course has been highlighted. Therefore, while online formative quizzes can be a significant tool for educators to use in helping their students retain course content, their effectiveness is mediated by such factors as the immediacy of feedback, the opportunity and incentive to resubmit quizzes, and quiz functionality. The findings from this study have implications for teachers, learners and curriculum designers. Nevertheless, questions remain regarding student affect, the effectiveness of different item formats for online language acquisition, and how to encourage learner participation.

The Author

Paul Joyce is an Associate Professor at Kindai University in Osaka, Japan. He is from Britain. He undertook his M.A. (TESL/TEFL) at Birmingham University (England) and conducted his Ph.D. in Applied Linguistics at Surrey University (England). His research interests include vocabulary acquisition, language testing, and second language listening.

References

- 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(2), 255-272.

- Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*(2), 213-238.
- Barbarick, K. A. (1998). Exam frequency comparison in introductory soil science. *Journal of Natural Resources and Life Science Education, 27*, 55-58.
- Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes*. (Vol. 2, pp. 35-67). Hillsdale, NJ: Erlbaum.
- Black, P. & Wiliam, D. (1998). *Inside the Black Box: Raising standards through classroom assessment*. King's College, London, United Kingdom.
- Blackboard [Computer software]. (2018). Retrieved from <http://www.blackboard.com>
- Bonham, S. W., Deardorff, D. L., & Beichner, R. J. (2003). Comparison of student performance using web and paper-based homework in college-level physics. *Journal of Research in Science Teaching, 40*(10), 1050-1071.
- Brothen, T. & Wambach, C. (2001). Effective Student Use of Computerized Quizzes. *Teaching of Psychology, 28*(4), 292-294.
- Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science, 21*(5), 279-283.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Dobson, J. L. (2008). The use of formative online quizzes to enhance class preparation and scores on summative exams. *Advances in Physiology Education, 32*, 297-302.
- Fitch, M., Drucker, A. J., & Norton, J. A. Jr. (1951). Frequent Testing as a Motivating Factor in Large Lecture Classes. *Journal of Educational Psychology, 42*, 1-20.

- Haberyan, K. A. (2003). Do weekly quizzes improve student performance on General Biology Exams? *The American Biology Teacher*, 65, 110-114.
- Hagen, J. P. (2000). Cooperative learning in organic II: Increase in retention on a commuter campus. *Journal of Chemical Education*, 77(11), 1441-1444.
- Helgesen, M., Brown, S., & Wiltshier, J. (2010). *English firsthand 2 student's book* (4th ed.). Hong Kong: Pearson Longman.
- Johnson, G. M. (2006). Optional online quizzes: College student use and relationship to achievement. *Canadian Journal of Learning and Technology*, 32, 105-118.
- Karpicke, J. D. & Roediger, H. L. III. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, 57, 151-162.
- Kluger, A. N., & DeNisi, A. (1997). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254-284.
- Kibble, J. (2007). Use of unsupervised online quizzes as formative assessment in medical physiology course: Effects of incentives on student participation and performance. *Advances in Physiology Education*, 31(3), 253-260.
- Li, Q. & Edmonds, K. (2005). Mathematics and at-risk adult learners: Would technology help? *Journal of Research on Technology in Education*, 38(2), 143-166.
- Liao, C.W., Hatrak, N., & Feng, Y. (2010). *Comparison of Content, Item Statistics and Test-Taker Performance for the Redesigned and Classic TOEIC Listening and Reading Tests*. ETS.
- Macedo-Rouet, M., Ney, M., Charles, S., & Lallich-Boidin, G. (2009). Students' performance and satisfaction with web vs. paper-based practice quizzes and lecture notes. *Computers and Education*, 53, 375-384.
- Martin, R. R., & Srikameswaran, K. (1974). Correlation between frequent testing and student performance. *Journal of Chemical Education*, 51(7), 485-486.
- McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory

representations through retrieval. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 11, 371-385.

Moodle [Computer software]. (2018). Retrieved from <https://moodle.org/>

Orr, R., & Foster, S. (2013). Increasing student success using online quizzing in introductory (majors) biology. *CBE-Life Sciences Education*, 12, 509-514.

Quizlet [Computer software]. (2018). Retrieved from <https://quizlet.com>

Roediger, H. L. III & Karpicke J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249-255.