

Can Using Online Formative Assessment Boost the Academic Performance of Business Students? An Empirical Study

S. W. Oellermann¹ and A. van der Merwe²

¹Durban University of Technology, Riverside campus, Pietermaritzburg, Box 13262 Cascades, 3202, Republic of South Africa

²Durban University of Technology, Riverside campus, Pietermaritzburg, Box 101112 Scottsville, 3201, Republic of South Africa

E-mail: ¹<sueo@saint.co.za>, ²<alexvdm@dut.ac.za>

²Telephone: +27338458868, ²Fax: +27338458831

Can Using Online Formative Assessment Boost the Academic Performance of Business Students? An Empirical Study

S. W. Oellermann¹ and A. van der Merwe²

¹*Durban University of Technology, Riverside campus, Pietermaritzburg, Box 13262 Cascades, 3202, Republic of South Africa*

²*Durban University of Technology, Riverside campus, Pietermaritzburg, Box 101112 Scottsville, 3201, Republic of South Africa*

E-mail: ¹<sueo@saint.co.za>, ²<alexvdm@dut.ac.za>, ²Telephone: +27338458868, ²Fax: +27338458831

KEYWORDS Online learning. summative assessment. multiple choice questions. self-test. gender.

ABSTRACT The literature indicates that, where online learning has been blended with traditional teaching methods, more effective results have been achieved than by using only face-to-face or online instruction. To strengthen the findings of two of the researcher's previous studies on the impact of using the self-test tool of an online classroom on summative assessment results, two student cohorts used an online classroom created in Blackboard, a learning management system (LMS). The time spent by each student in the online classroom and their scores in subsequent multiple-choice question (MCQ) tests were measured. The survey data were analysed using bivariate correlations and tests on standardised regression coefficients showed self-test time as a significant predictor of summative MCQ performance while controlling for ability. Exam MCQ performance was found to be associated, positively and significantly, with annual self-test time at the 5 percent level and a significant relationship was found between

MCQ marks and year marks. It was concluded that student use of the self-test tool in formative assessments has a significant bearing on students' year marks. The negative nature of the standardised beta coefficient for gender indicates that, when year marks and annual self-test time are considered, males appear to have performed slightly better than females.

I. INTRODUCTION

Hodgson (2011), an expert in Applied Linguistics with 17 years' experience in teaching communication in English, commented in an interview with the researcher that most of the students at the Riverside campus of the Durban University of Technology (DUT) appear to lack sufficient proficiency in English to cope with tertiary studies. Creative teaching interventions are therefore necessary. There appears to be growing interest in the possibility that online learning can improve many aspects of the learning process as well as produce technologically literate graduates who are able to apply their skills effectively in an increasingly digital workplace. Clearly, the use of online learning is gaining ground as technology advances but its effectiveness is still in question and there have been calls for more empirical evidence of its benefits in tertiary level education.

Following the successful learning outcomes achieved from a preliminary study in 2007 (Oellermann 2009), a similar online classroom was constructed in Blackboard, a learning management system (LMS) for the 2011 Management 1 class to gauge the usefulness of online learning as part of a blended learning approach in teaching a larger class (Oellermann 2011). Findings of both studies were similar, and were consistent with those of other researchers (Osguthorpe and Graham 2003).

All students agreed that repeating the self-test helped them engage more easily with the learning material, improved their knowledge of the subject and helped them answer summative multiple-choice questions (MCQs). Oellermann (2009, 2011) acknowledged that the weaknesses of both studies include the small size of the samples and the failure to take into account intervening variables such as gender that could have skewed the results and cautioned that more robust research should be undertaken to provide more solid evidence and strengthen research findings. The sample of the current study is therefore larger and its objectives are threefold:

- To test for consistency of findings between this study and the results of the researcher's studies conducted in 2007 and 2011 as well as the studies recorded in the literature generally.
- To investigate a possible correlation between students' use of the online classroom and their subsequent performance in a MCQ test.
- To determine the effect of gender on students' performance in multiple-choice items following use of the self-test tool.

Subsequent sections include a brief review of the literature, a brief look at blended learning including the use of the self-test tool in formative assessment, discussion of the research design, presentation and analysis of the data, limitations of the study and the conclusions made.

II. THE LITERATURE REVIEW

1. Blended learning

Blended learning is gaining ground in tertiary level education around the world because it makes use of many of the advantages presented by new technology through using an online classroom, while retaining the benefit of personal interaction that comes with traditional face-to-face teaching. Promoters of blended learning view it as an opportunity to redesign the development, scheduling and delivery of higher education courses through a combination of physical and virtual instruction (Vaughn 2007). Garrison and Kanuka (2004:1) warned that internet information and communication technologies are likely to be the “defining transformative innovation for higher education in the 21st century.” Kanuka (2006) later suggested that the higher education institutions of the world could lose their monopoly as providers of education if they do not use technology to offer more innovative modes of delivery that are suited to the needs of their students. With new forms of learning such as Webinars, blogs and podcasts becoming more common, the experts believe that the majority of formal and informal learning experiences in the future are likely to be blended ones (So and Bonk 2010).

An online classroom provides an interactive learning space that fully engages the attention of the learner. All course materials, exercises and interactive revision tests can be made available online and can be accessed any time by students, which forces them to take responsibility for the

pace and timing of their learning. Both summative and formative assessments can be conducted online. A personal presence is maintained with students during a reduced number of face-to-face contact periods where students can iron out any difficulties they may have in studying online.

2. Some advantages of blended learning

2.1. Flexibility

Badenhorst and De Beer (2004) believe that wider student access can be more comfortably achieved by having more than one method of delivery. Vaughn (2007) observed a similar perspective held by academics at the University of Calgary who claim the advantage of a reduction in classroom time as a result of their online learning interventions.

An investigation of blended learning in tertiary institutions, conducted by Hernes (2006), noted that a number of universities throughout the world have sought to meet the challenges of increasing programme flexibility, improving quality, increasing student access, and reducing costs with new technology. He contends that the new methods provided by blended learning may reduce duplication of effort through access to open source courseware and software, and that “these changes promote a learner-centred pedagogy,” where students are more involved in and take responsibility for their learning (Hernes 2006). Stacey and Gerbic (2008) observed that the integration of physical and virtual learning enhances flexibility and enables both teachers and their students to become learners.

2.2. Active learning

Active learning generally implies student engagement in the learning process through performing meaningful learning activities while focusing and reflecting on what they are doing (Bonwell and Eison 1991). Online learning demands active student involvement, and Yoder and Hochevar (2005) found that their students performed better when tested on material where they were actively involved in the learning process, rather than on material where other techniques were used. Their investigation was thorough: it covered three classes over three successive years, all covering the same material with the same instructor, and they found that student performances were better where more active learning had taken place both within and between classes. Prince (2004), in a study of various types of active learning in the literature,

acknowledges the extensive evidence supporting the benefits of student engagement in the learning process.

2.3. Motivation

Martens et al. (2010) refer to the fairly extensive evidence indicating that intrinsic student motivation arises when students' need for autonomy (where students have choices), competence (where students experience success) and social relatedness (working in a socially positive or accepting environment) are addressed. Blended learning can address all of these issues; students have a choice in the way that they learn, the self-test tool with feedback gives them a feeling of competence in that their learning improves with every repetition of the test, and the facilitator creates a friendly, accepting and encouraging environment in face-to-face contact sessions. Brown et al. (1996) maintain that the feedback given to students on their performance can be a vital incentive to them and assist them in taking more control and responsibility for their learning.

3. Online learning and the self-test tool

Following a previous intervention using the self-test tool in formative assessment, Rienties and Woltjer (2004) conducted an interesting experiment in 2003 at the University of Maastricht, The Netherlands. Economics and business students were forced to do two online formative self-tests with feedback, each consisting of 20 multiple-choice questions, in order to qualify to join a group discussion. They hoped that the first self-test would give the students an opportunity to gauge whether they had studied enough to pass the second test, which was the prerequisite for joining the post-discussion group. The intention was to force students to read the textbook and come prepared for the ensuing group discussions (for which most students had been ill-prepared in the past). There were 30 groups, and 346 students completed a Likert-type survey. Tutors confirmed that all students had read the literature before attending the groups and, as a result, the discussions went to a deeper level; and therefore, were considered useful. A subsequent student survey revealed that all students had worked harder, believed that they had learnt more, and were more interested in the course.

The researchers cautioned that their findings should be viewed in the light of other innovations introduced at the time, and changes in the course over the previous years; and therefore, warrant further investigation. A caution here is that offered by Lowry (2004) who tried to prevent rote learning by supplying feedback that showed why the answer was incorrect, without giving the correct answer. The intention was to prompt students to use their understanding and not just memorise. Rientjies and Woltjer (2004) had a similar reason for limiting the number of times students were permitted to do self-tests to only two. St Clair (2009) found that his purely online economics students were often more comfortable online than in traditional classes and that generally, their grades were higher than he would have expected in a traditional class. However, this was a personal opinion, as he does not present any scientific research to corroborate this finding.

Such new teaching and learning strategies, that are evolving with new technology, demand reflection on how existing learning theories can be combined with them in a way that will maximise learning that is appropriate, meaningful and successful (Ally 2004; Cook n. d.). The theory that most informs this study is the behaviourist theory of learning, and Ebert (n. d.) asserts that it is still prominent and relevant in current education. Behaviourist theory, as espoused by Skinner (1958), showed behaviour as being shaped and maintained by “reinforcing consequences.” His teaching machine encouraged students to be actively involved in their learning by pressing a button to record their responses in a MCQ test. Feedback given for incorrect answers would prompt the student to make another choice and the machine would only continue to the next question once the student had chosen the correct one. Correct choices made were thus reinforced by being allowed to proceed to the next question.

According to Ally (2004), online testing of learner achievement should be integrated into the learning process, and suitable feedback should be given to learners so that they can monitor their progress and take any necessary corrective action. The self-test tool of the online classroom with immediate, detailed feedback provides students with the formative assessment that enables them to do this. As students repeatedly attempt forced-choice questions, they are rewarded with higher test marks, having learnt from immediate feedback that explains why their original responses were incorrect. Lowry (2005) acknowledges the role of feedback in enhancing learning in the self-test tool and its role in formative assessment.

Ebert (n. d.) notes that many behaviourist tactics are relevant in current teaching because much of the curricula in education focus on content that requires memorisation of factual information before higher-level, problem-based learning can take place. This is especially so at first year level where a foundation needs to be laid that enables students to develop an understanding of basic course concepts. Richardson (2002) claims that through responding to a multitude of questions, that cover a large amount of course material, students can demonstrate an understanding of simple logic in a fairly short amount of time.

Although some educationists may categorise true and false quizzes at the lowest level of learning according to Bloom's taxonomy of cognitive levels, Richardson (2002) suggests that they can be used at higher levels. For example, students can be provided with a set of information that is new to them – a set of data or written work or a case study – then asked various forced-choice questions related to the content or the presence/absence of certain characteristics in the work (Oellermann 2009) or to solving a problem.

According to Richardson (2002) in a suitable set of true and false questions each sentence should be concise and straightforward and should ideally comprise only one concept. Landsberger (2007) recommends that negatives or double negatives should best be avoided. Second-language students in particular find them tricky. To avoid answers being too obvious, true statements should be about the same length as false statements. Vague or absolute words or statements (like “all,” “always,” “never,” “none” and “only”) should be avoided, as they tend to be too obviously false (Richardson 2002:5).

According to Test tips (Test tips. True and false n. d.), a useful guide in setting true and false tests, a set of true statements that cover the material to be tested should first be written. Then, half of these statements should be converted to false though not negative statements. Obviously, it would be necessary to randomise the sequence of true/false responses to avoid a discernible pattern. It has been noted that “true” is marked more often in guessing and that assessing “false” statements tends to be more challenging for the students (Landsberger 2007:1).

2.4. Previous studies in blended learning: 2007 and 2011

In the researcher's two previous studies in using an online classroom as part of a blended mode of instruction, an extensive question bank of true and false self-tests covering topics in the whole syllabus was loaded into the self-test tool for the following reasons:

- The immediate feedback given on incorrect answers assisted students in engaging with the material and in preparing for assessments and being detailed, it helped them to understand more fully why their answers were incorrect.
- Grades were for student information only and not for formal grading. This gave students a safe place where they could make mistakes without being embarrassed.
- Students could realise a sense of competence as their results improved with every repetition of the tests which should be motivating for students.
- With only two choices, the student is likely to guess correctly on half the questions, even if he/she has done little studying (Oellermann 2009) which is encouraging for second language students.

Such formative testing is an example of how Skinner's behaviouristic teaching methods can be adapted to an online learning environment (Skinner 1958). Self-tests with suitable, and detailed feedback were used for formative assessments and in revision for tests and exams. Student grades were not used for course marks. Students could learn from repeating self-tests and receiving immediate written feedback on why answers were incorrect. In view of the literature, their scores were expected to improve every time they repeated the test.

Lowry (2005) found that a group of students who voluntarily made use of computer-aided formative MCQ assessments with feedback performed significantly better in end of module summative assessments than did a control group who used none of the learning material. A short formative assessment of five MCQs followed each lecture. Students received immediate feedback, which explained why the answer was incorrect but did not give them the correct answer. In this way, they were prompted to think and work out the answers themselves and not just memorise. Lowry (2005) concedes that the different achievements of the two groups in the summative assessment could be explained by the fact that the more competent students elected to use the self-assessments, while the others did not.

For this reason, a second test on similar material was conducted on the same two groups and no option for formative assessment was given. The differences in their scores were minimal, which confirmed that the differences between the two groups were not significant. Factors

affecting the validity of this research are that different lecturers presented the material, and student demographics and prior computer competence levels were not taken into account. Interesting informal feedback was that the self-test tool increased students' confidence, which is likely to have improved their perception of the subject.

MCQ assessment accounts for up to 40 percent of term tests and 25 percent of the end of year exam, both summative assessments, so the self-test tool, if well utilized, can make an important contribution to student success. Other authors whose findings in experiments with the self-test tool support this are Ibabe and Jauregizar (2009). They found that students who used interactive online self-assessment with feedback performed better academically in summative assessments. Such experiments constitute a clear acknowledgment of the importance of the self-test tool as a learning tool.

2.5. MCQ testing and gender

In some cases, the role of gender has been found to influence student scores in MCQs. Ng and Chan (2009:240) used a variety of MCQ test methods in an engineering course and found that, while females scored better than males in psychology theory, males had the advantage in “numerical, mechanical, three-dimensional spatial perception, general knowledge, verbal reasoning, mathematics and spatial items”. They noted that females generally found the difficult questions more difficult and the more simple, straightforward questions easier, and acknowledged that such gender-related discrepancies in achievement might be due to factors such as the test area, instruction/scoring conditions and item difficulty.

Hassmen and Hunt (1994) acknowledge the existing research findings that indicate that MCQ testing favours males over females. They conducted a study that compared gender-related scores achieved by undergraduates in a traditional MCQ test, with those achieved by them in a MCQ test where they were permitted to check if their answers were correct. They found less disparity between the scores of the two genders in the MCQ test where students were permitted to check their answers than in scores in the traditional MCQ test, and suggest that the different cognitive styles of males and females could explain this finding.

Conversely, Deepak et al. (2011) found that, of the 200 final year medical students at the University of Dammam in Saudi Arabia who sat the MCQ tests, the female students showed an overall better performance in all theory assessments than did the male students, mainly because

they obtained high scores in surgical disciplines. They attribute female success to the following psychological traits related to learning that they claim are borne out by the literature:

- Vigilance in environmental scanning may give females better recall of details learnt.
- Superior reading skills, perceptual speed and memory are attributed to females.

They acknowledge that the provision of extra test time (to allow for language differences) might have minimised the female tendency to experience exam nerves and postulate that the imaginative perception and associative memory that females possess, and that is required in the surgical disciplines, may have enabled females to outperform males in this discipline. Deepak et al. (2011) acknowledge that learning approach, (females are said to have a superficial approach) diligence and socio-cultural factors could also favour females.

Parker (2006) compared the performance of students in microeconomics at five South African universities and found that males generally outperformed females. As her study controlled for student ability and effort, she suggests that this result could support previous research findings that show that males generally outperform females in MCQ tests. The course grade, on average, at the five universities was based on more than 60 percent MCQ assessment and this could lend more weight to her suggestion.

III. METHODOLOGY

This analysis is based on a study of two 2012 cohorts of management students at DUT Riverside campus comprising a total population of 173. Following a comprehensive data cleansing process, whereby sixteen students who had dropped out during the year, 13 who had not used the online classroom and cases with missing relevant variables were excluded, the sample was reduced to 130. The units of analysis are fairly homogeneous in that they were all black and were all students in their first year of study in a management subject on the Riverside campus of DUT. The sample is reasonably representative of the population. The sample gender proportions are: sample males = 49.2% and population males = 46.8% while sample females = 50.8% and population females = 53.2%. The average age of the sample is 20.7 years while that of the population is 21 years.

The time spent by each student in the online classroom during the year and in using the self-test tool was measured in minutes and harvested from the Blackboard system. This

data as well as individual student scores in MCQ tests, student year marks and gender were uploaded in the Statistical Package for Social Sciences (SPSS), version 21. MCQ test scores were correlated with usage rates of the online classroom and compared for gender. The Cronbach's alpha test score of .713 confirmed that the survey instrument has a high degree of internal consistency and this lays a sound foundation for establishing the credibility of the findings of the study. Furthermore, Little's missing completely at random (MCAR) test confirmed that there are no distinct patterns of missingness in the data. The data were subsequently analysed, described and illustrated using basic descriptive statistical techniques. Following the bivariate correlation analyses, multiple regression modelling was employed in an effort to control for possible intervening variables not accounted for by zero-order correlations. Comparisons were made with the findings from the researcher's previous studies and the findings of other researchers, as noted in the literature review, in order to check for consistency.

Students each wrote one MCQ test (Test 3) consisting of 50 questions (each with at least four alternative answers) based on topics covered in the learning modules section of the online classroom and were required to indicate their gender on their answer scripts. The true and false revision self-tests with feedback were made available for student use in the assessments tool of the online classroom two weeks prior to the test. The same procedure was used to prepare students for answering 25 MCQs in the final examinations.

IV. RESULTS AND DISCUSSION

1. Student use of the online classroom and their assessment results

1.1. Test 3 mark and prior self-test time

Students' marks obtained in Test 3 (consisting of 50 MCQs) were correlated with their time spent using the self-test tool in the online classroom in the two weeks prior to the test and Table 1 presents the results. Pearson's correlation of these two variables is flagged as significant at $r = .316$ ($p < 0.01$), which indicates a positive and medium effect size of prior self-test time on Test 3 mark. This positive relationship between prior use of a self-test tool and student performance in a MCQ assessment is consistent with the findings of Ibabe and Jauregizar (2009), Lowry (2005), and Rienties and Woltjer (2004). This is also consistent with both of the researcher's previous

studies (Oellermann 2009, 2011) where all students agreed that the formative self-tests helped them answer MCQs in summative assessments.

1.2. Exam MCQ mark and prior self-test time

The amount of time students spent using the self-test tool in the two weeks prior to the examination was correlated with their subsequent achievement in the examination MCQ test and this is reflected in Table 2. Pearson's correlation figure of $r = .100$ (small effect size) in Table 2 is not flagged as significant. This result is surprising because significantly positive correlations were noticed in previous years. A possible explanation might be related to the increasing academic weakness of students. Because they require a lot of time to prepare for assessments, tests in the various subjects are not conducted in the same week, and this fact will account for the positive correlation between time spent in the self-test tool prior to Test 3 and the test marks they achieved. Such leeway is not afforded students at exam time when they might have to write two exams in one day. In such a case, students would prioritise their studying. As MCQs account for only 25 percent of the exam mark, they would focus more on studying subject material that would enable them to answer written questions. For Test 3 the focus would be on learning in the self-test tool, as they know that this prepares them for MCQs and 100 percent of the Test 3 questions were MCQs.

1.3. Exam MCQ mark and annual self-test time

The total amount of time students spent using the self-test tool throughout the year was correlated with the marks they achieved in the MCQ section of the exam and is illustrated in Table 3. This correlation ($r = 0.238$), though having a small effect size, is nevertheless flagged as significant and one could therefore deduce that student use of the self-test tool over the whole academic year prepared them to some extent to successfully answer exam MCQs. Again, this is consistent with the findings of Ibabe and Jauregizar (2009), Lowry (2005), and Rienties and Woltjer (2004), as well as with both of the researcher's previous studies (Oellermann 2009, 2011).

A moderating factor that could lower the rate of student usage of the online classroom is the poor state of the student computer laboratories, a fact that is acknowledged by computer technicians on the DUT, Riverside campus; computers are old and have little memory. A lot of time was spent assisting students in accessing the online classrooms. On a number of occasions, access

was gained using the student's credentials from the researcher's personal computer when it was not possible to do so using computers in the computer laboratory.

2. Year mark and Test 3 mark

Using descriptive statistics, MCQ scores in Test 3 (which consisted of 100 percent MCQs) were compared with the year mark (which consisted of approximately 40 percent MCQs). Table 4 shows the results thereof. The mean and maximum scores for the Test 3 marks were found to be significantly higher than the mean and maximum scores for the year marks. This result is confirmed by a Pearson's correlation of the same two variables which yielded a positive, significant coefficient of large effect size $r = .508$. These results, shown in Table 4, indicate that students who did well in Test 3 generally achieved a higher year mark and that the students in the sample did better in MCQ tests than in other forms of assessments such as assignments and assessments where written answers were required. This could imply that students' superior MCQ marks may be attributed to their use of the self-test tool. If this is the case, this result serves to strengthen the results shown in Table 1, where the relationship of Test 3 marks, and prior use of the self-test tool, showed a positive medium correlation.

Intervening variables such as students' writing ability and the difficulty of the different types of assessment would need to be considered. In this respect, it should be mentioned that, in assessing the written work of DUT students, allowance is made for second language students who have varying degrees of difficulty in expressing themselves. This finding is perhaps an indication of the extent to which MCQs (and student prior use of the self-test tool in Test 3) helped students achieve a higher year mark.

3. MCQ marks and gender

The scores for Test 3 (50 MCQs) and the exam MCQs were correlated with student gender and this is shown in Tables 5 and 6. In uploading data into SPSS, the following values were allocated in terms of gender: Male = 1; Female = 2. A Spearman's correlation was deemed the most suitable correlation test to use in correlating gender (a categorical variable) and test scores (a scale variable). The negative nature of Spearman's correlation coefficient in Table

5 of .04 indicates that, as one variable increased, the other variable decreased. This means that male MCQ scores were better than female scores, but only marginally, and the coefficient is not flagged as significant. This is a similar finding to that of Hassmen and Hunt (1994), who attributed their finding to a possible difference in cognitive style between the two genders and Parker (2006), who found that males generally outperformed females in microeconomics at five South African universities where assessments consisted of more than 60 percent MCQs.

Spearman's correlation coefficient of $-.129$ shown in table 6 reflects better performance by males than females in the exam MCQs. Although the males might have performed better than females in the exam MCQs than in Test 3, the figures are still not flagged as being significant.

V. MULTIPLE REGRESSION ANALYSIS AND DISCUSSION

1. The bivariate analyses

The results, shown in Section IV, were arrived at by using bivariate analyses. The problem with such zero order correlation coefficients is that one has no idea how the possible intervening variables may influence them. A way of helping to control for such potentially moderating factors is in the use of multiple linear regression on the basis that test scores (the dependant variable) are generally accepted to be normally distributed. Subsequent tests for normality in the SPSS, however, suggested only crude approximations of bell shaped distributions of year marks, Test 3 marks and exam MCQ marks. It is likely that these slightly skewed distributions would normalise given a larger sample size and so validity of the regression results is claimed for this sample only.

2. Multiple regression analysis

A test was performed using the multiple regression statistical technique where the Test 3 marks were used as the dependent variable. Multiple independent variables in the test included gender, self-test time prior to Test 3, self-test time for year and year mark. Such multiple explanatory variables (or predictors) are useful in estimating the score of the dependant or

criterion variable, which in this case is the Test 3 mark, and thus hold the promise of confirming the findings made in the bivariate analyses. By measuring the naturally occurring scores on each predictor variable, one may attempt to establish which set of variables has the most effect or can best predict Test 3 marks (SPSS for Psychologists, nd.).

Using the enter method, a significant model emerged [$F(3,126) = 22.74$; $p < 0.0005$]. Adjusted R square = .336. Significant variables are shown below:

Predictor variable	Beta	<i>p</i>
Selftesttimepretest	.294	$p < 0.0005$
Gender	-.070	$p = 0.332$
Yearmark	.498	$p < 0.0005$

The regression model is significant at the 1 percent level and the adjusted R square of .336 indicates that the model accounts for 33.6 percent of the variation in Test 3 marks.

The results in Table 7 suggest that year mark, with a high *t* value of 6.932 and a small sig (*p*) value of .000, has the biggest impact on Test 3 marks. This result confirms the finding, reported on in Table 4, of a high Pearson's correlation coefficient of .508 for Test 3 mark and year mark, and suggests that a high Test 3 mark is associated with a high year mark. As shown earlier in Table 1, the Test 3 mark and prior use of the self-test tool yielded a medium yet significant Pearson's correlation coefficient of .316, so it can reasonably be assumed that use of the self-test tool enhanced student learning as measured by students' year marks.

Table 7 also shows self-test time (as recorded from Blackboard LMS) as a statistically significant predictor of performance at the 1 percent level. This is important because it shows that the year mark is not the only predictor of student performance and that, regardless of the student's academic ability as reflected in the year mark, student time spent in the self-test tool of the online classroom, as part of the researcher's teaching blend, is a significant predictor of how well students perform in summative assessments.

Those students who did not achieve a duly performed (DP) year mark of 40 percent were excluded from writing the final examination, including the MCQ portion thereof. Sample selection bias could skew the results of the study if the more academically capable students were using the online classroom, and their better student performance in assessments is due to their academic superiority rather than time spent in the self-test tool. The introduction of the year mark control variable which is not exclusive, as a proxy for student ability, serves to neutralise to

some extent the risk of sample selection bias. This measure enhances the validity and reliability of the study's findings.

The negative value of the standardised beta coefficient for gender in Table 7, prompted a further investigation. A regression analysis was done with the exam MCQ mark as the dependant variable (or performance indicator) and year mark, self-test time for year (recorded by Blackboard LMS), and gender as independent variables or predictors. Table 8 illustrates this.

Using the enter method, a significant model emerged [$F(3,125)=14.226$; $p < 0.0005$].

Adjusted R square = .237. Significant variables are shown below:

Predictor variable	Beta	<i>p</i>
Yearmark	.425	$p < 0.0005$
Gender	-.159	$p = .044$
Seltesttimeyear	.188	$p = .017$

The regression model is significant at the 1 percent level and an adjusted R square of .237 means that the model accounts for 23.7 percent of the variation in the student exam MCQ performance. Although having only modest explanatory power, this model shows that the exam MCQ performance is positively, and significantly, associated with the amount of time that students spend doing self-tests throughout the year at the 5 percent level.

Gender is also flagged as a significant predictor of the exam MCQ performance at the 5 percent level. The negative nature of the standardised beta coefficient for gender indicates that when year marks and self-test time for the year are taken into account, males appear to have performed slightly better than females. This is consistent with the findings of Hassmen and Hunt (1994), and Parker (2006). More importantly for this study, self-test time for the year is significantly and positively associated with the exam MCQ performance.

3. Limitations of the study

Obvious limitations of the study include the following:

- The relatively small size of the sample
- The fact that other institutions are not included in the study
- The narrow definition of student achievement, which is gauged by only two MCQ test marks

- The failure to accurately and objectively account for intervening variables such as learner motivation, educational background, previous computer experience, academic ability and language proficiency.

VI. CONCLUSIONS

This study produces empirical evidence that the self-test tool with feedback in the online classroom, if used as a revision tool in formative assessment, leads to improved MCQ marks in summative assessments, which in turn should lead to an improved year mark. The extent of this benefit is reflected in the findings shown in Tables 1 and 4 and to a significant but lesser degree by the finding shown in Table 3.

Year marks were captured in an attempt to neutralise self-selection bias to some extent; those students who did not achieve a duly performed (DP) year mark of 40 percent were excluded from writing the final examination, including the MCQ portion thereof. Correlating year marks, which are not exclusive, with time spent online, should reduce bias and strengthen the reliability and validity of the study's findings. Furthermore multiple regression results, controlling for ability (proxied by year mark), confirm that time spent online is correlated positively with performance/achievement.

The finding that the amount of time that students spent in the self-test tool of the online classroom is a significant predictor of how well they perform in summative assessments is consistent with similar findings in the literature as reported by Ibabe and Jauregizar (2009), Lowry (2005), and Rienties and Woltjer (2004), and in both of the researcher's previous two studies (Oellermann, 2009; 2011).

Negative, weak and insignificant Spearman's correlations showed that male MCQ scores were very slightly better than female scores. Although much weaker, this is a similar finding to that of Hassmen and Hunt (1994) and Parker (2006). Consistency of findings over the three studies conducted by the researcher, backed up by similar findings in the literature, point to a modest degree of dependability, credibility and stability reliability (Neuman 2000) of the research.

REFERENCES

ALLY M 2004. Foundations of educational theory for learning. In Anderson T and Elloumi F. eds. *Theory and practice of online learning*. Athabasca, Ca: Athabasca University. From <http://cde.athabascau.ca/online_book/pdf/TPOL_book.pdf> (Retrieved November 15, 2013)

BADENHORST JJC, DE BEER KJ 2004. *Blended learning at the Central University of Technology, Free State*. From <http://www.google.co.za/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=2&ved=0CC8QFjAB&url=http%3A%2F%2Femerge2004.net%2Fconnect%2Fsite%2FUploadWSC%2Femerge2004%2Ffile20%2Femerge2004article.doc&ei=EJQkUr_XGc2Uswb0xIDIBg&usg=AFQjCNElzFH0_woiIUSwpeg8TLy7zeO0Lw> (Retrieved September 9, 2013)

BONWELL CC, EISEN JA 1991. Active learning: Creating excitement in the classroom. *Eric Digest*. From <<http://www.ntlf.com/html/lib/bib/91-9dig.htm>> (Retrieved October 10, 2011)

BROWN S, RACE P, SMITH B 1996. *500 tips on assessment*. London: Kogan Page.

COOK J n. d. *A constructivist approach to online course design to enhance interaction and learner motivation in K-12*. EDTECH. Boise State University. From <<https://sites.google.com/a/boisestate.edu/edtechtheories/a-constructivist-approach-to-online-course-design-to-enhance-interaction-and-learner-motivation-in-k-12>> (Retrieved November 8, 2013)

DEEPAK KK, AL-UMRAN KU, AL-SHEIKH MH, AL-RUBAISH A 2011. The influence of gender on undergraduate performance in multiple choice testing in clinical disciplines at the University of Dammam, Saudi Arabia. *Al Ameen Journal of Medical Sciences*, 4(2):123-130. From <<http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=ea1730d9-25c6-4da9-a701bb7988925e86%40sessionmgr10&vid=1&hid=15>> (Retrieved June 14, 2012)

EBERT AK n. d. *Behaviourism vs. constructivism in the technological secondary education classroom*. EDTECH. Boise State University. From
<<https://sites.google.com/a/boisestate.edu/edtechtheories/behaviorism-vs-constructivism-in-the-technological-secondary-education-classroom-1>> (Retrieved November 11, 2013)

GARRISON DR, KANUKA H 2004. Blended learning: uncovering its transformative potential in higher education. *The Internet and Higher Education*, 7(2):95-105, 2nd quarter. From
<http://ac.els-cdn.com/S1096751604000156/1-s2.0-S1096751604000156-main.pdf?_tid=5419b86c-76bc-11e3-9e3a-00000aacb362&acdnat=1389003878_360117a9a98e72844c8789c206dbfcf4> (Retrieved October 21, 2011)

HASSMEN P, HUNT D P 1994. Human self-assessment in multiple choice testing. *Journal of Educational Measurement*, 31(2):149-160. Summer. From
<<http://www.jstor.org/stable/10.2307/1435174?origin=api>> (Retrieved June 14, 2012)

HERNES G 2006. The university: current challenges and opportunities. *The Virtual University and e-learning*. UNESCO. From <<http://www.unesco.org/iiep/virtualuniversity/home.php>>
(Retrieved July 3, 2013)

HODGSON L 2011. Interview with Mrs L. Hodgson, Lecturer, Durban University of Technology, 24 October.

IBABE I JAUREGIZAR J 2010. Online self-assessment with feedback and metacognitive knowledge. *Higher Education*, 59(2):243-258. From
<<http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=6c7871e2-365a-437e-99bb-b5456bd1e83e%40sessionmgr12&vid=13&hid=24>> (Retrieved September 14, 2011)

KANUKA H 2006. Has e-learning revolutionized education? *Embassy newspaper*, 9 Sept. From
<File://E:\Has eLearning Revolutionized Education - Embassy -Newspaper Online.htm>
(Retrieved November 2007)

LANDSBERGER J 1996. Constructing true and false tests. *Study guides and strategies*. From <<http://www.studygs.net/teaching/tsttak2a.htm>> (Retrieved September 10, 2007)

LOWRY R 2005. Computer aided self-assessment - an effective tool. *Chemistry Education Research and Practice*, 6(4):198-203. From <http://www.rsc.org/images/Lowrypaper_tcm18-41072.pdf> (Retrieved October 26, 2011)

MARTENS R, DE BRABANDER C, ROZENDAAL J, BOEKAERTS M, VAN DER LEEDEN R 2010. Inducing mind sets in self-regulated learning with motivational information. *Routledge Educational Studies*, 36(3):322-326. Jul. From <<http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=6c7871e2-365a-437e-99bb-b5456bd1e83e%40sessionmgr12&vid=8&hid=24>> (Retrieved September 14, 2011)

NEUMAN W L 2000. *Social research methods - qualitative and quantitative approaches*. 4th ed. Needham Heights, Mass: Allyn and Bacon.

NG AWY, CHAN AHS 2009. The testing methods and gender differences in multiple-choice assessment. *AIP Conference Proceedings on 1 October 2009*, 1177(1):236-243. From <<http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=4e10c0e3-4ca2-444b-991d-1a310c29ed9b%40sessionmgr110&vid=1&hid=126#>> (Retrieved June 14, 2012)

OELLERMANN SW 2009. Does using the self-test tool in the online classroom lead to inspired learning? *The Journal of Independent Teaching and Learning*, 4:5-15.

OELLERMANN SW 2011. Gauging the effectiveness of using an online classroom as part of a blended learning approach in teaching Management 1 students. Unpublished paper. Durban University of Technology.

OSGUTHORPE R T, GRAHAM CR 2003. Blended learning environments: definitions and directions. *Quarterly Review of Distance Education*, 4(3):227-233. From

<http://www.mendeley.com/research/blended-learning-systems-definitions-directions/#page-1>>
(Retrieved October 26, 2011)

PARKER K 2006. The effect of student characteristics on achievement in introductory microeconomics in South Africa. *South African Journal of Economics*, 74(1):137. From <<http://onlinelibrary.wiley.com/doi/10.1111/j.1813-6982.2006.00054.x/pdf>> (Retrieved September 2, 2013)

PRINCE M 2004. Does active learning work? A review of the research. *Journal of Engineering Education*, From <http://clt.jhsph.edu/resources/views/content/files/150/Does_Active_Learning_Work.pdf> (Retrieved August 2, 2013)

RICHARDSON E 2000-2001. Alabama professional development modules. *Alabama Department of Education*. From <http://www.alabamapepe.com/profdevmodule/t_false/truefals.pdf> (Retrieved August 13, 2013)

RIENTIES B, WOLTJER G 2004. Regular online assessment, motivation and learning. From <<http://arno.unimaas.nl/show.cgi?fid=887>> (Retrieved June 14, 2012)

SKINNER BF 1958. Teaching machines. *Science*, 128(3330):969-977. From <<http://apps.fischlerschool.nova.edu/toolbox/instructionalproducts/edd8124/fall11/1958-Skinner-TeachingMachines.pdf>> (Retrieved November 12, 2013)

SO H-J, BONK CJ 2010. Examining the roles of blended learning approaches in computer-supported collaborative learning (CSCL) environments: a Delphi study. *Educational Technology & Society*, 13(3):189-200. From <<http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=c0b99231-81a9-49b4-a468-759e4ffd7e16%40sessionmgr111&vid=4&hid=121>> (Retrieved July 24, 2013)

SPSS FOR PSYCHOLOGISTS n. d. *Multiple regression*. From
<<http://www.palgrave.com/pdfs/0333734718.pdf>> (Retrieved December 1, 2014)

STACEY E, GERBIC P 2008. *Success factors for blended learning*. From
<<http://www.ascilite.org.au/conferences/melbourne08/procs/stacey.pdf>> (Retrieved October 15, 2011)

ST CLAIR DJ 2009. My experience with teaching online: confessions and observations of a survivor. *Merlot Journal of Online Learning and Teaching*, 5(1). from
<http://jolt.merlot.org/vol5no1/stclair_0309.htm> (Retrieved October 15, 2011)

TEST TIPS. *True or false* n. d. From
<http://imagecache2.allposters.com/images/pic/CAMB/27512_b~Test-Tips-True-or-False-Posters.jpg> (Retrieved November 15, 2007)

VAUGHN N 2007. Perspectives on blended learning in higher education. *International Journal of e- Learning*, 6(1):81-94. From <<http://dutlib.dut.ac.za:2057/docview/210364938>>
(Retrieved August 10, 2013)

YODER J D, HOCHVAR CM 2005. Encouraging active learning can improve students' performance on examinations. *Teaching of Psychology*, (32)2:91-95. From
<<http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=4&hid=13&sid=906d94ff-145e-48e4-a5e0-11831df5f550%40sessionmgr13>> (Retrieved October 24, 2011)

6. TABLES

Table 1: Test 3 mark and prior self-test time

		Test3mark	selftesttimepretest
Test3mark	Pearson Correlation	1	.316**
	Sig. (2-tailed)		.000
	N	130	130
Selftesttimepretest	Pearson Correlation	.316**	1
	Sig. (2-tailed)	.000	
	N	130	130
**. Correlation is significant at the 0.01 level (2-tailed).			

Table 2: Exam MCQ mark and prior self-test time

		ExamMCQ	Selftesttimepreexam
ExamMCQ	Pearson Correlation	1	.100
	Sig. (2-tailed)		.259
	N	129	129
selftesttimepreexam	Pearson Correlation	.100	1
	Sig. (2-tailed)	.259	
	N	129	130

Table 3: Exam MCQ mark and annual self-test time

		Selfsttimeyear	ExamMCQ
Selfsttimeyear r	Pearson Correlation	1	.238**
	Sig. (2-tailed)		.007
	N	130	129
ExamMCQ	Pearson Correlation	.238**	1
	Sig. (2-tailed)	.007	

	N	129	129
**. Correlation is significant at the 0.01 level (2-tailed).			

Table 4: Year mark and Test 3 mark

		Yearmark	Test3mark
Yearmark	Pearson correlation	1	.508**
	Sig. (2-tailed)		.000
	N	130	130
Test3mark	Pearson correlation	.508**	1
	Sig. (2-tailed)	.000	
	N	130	130
**. Correlation is significant at the 0.01 level (2-tailed).			

Table 5: Test 3 mark and gender

			Gender	Test3mark
Spearman's rho	Gender	Correlation coefficient	1.000	-.040
		Sig. (2-tailed)	.	.653
		N	130	130
	Test3mark	Correlation coefficient	-.040	1.000
		Sig. (2-tailed)	.653	.
		N	130	130

Table 6: Exam MCQ mark and gender

			ExamMCQ	Gender
Spearman's rho	ExamMCQ	Correlation coefficient	1.000	-.129
		Sig. (2-tailed)	.	.145

	Gender	N	129	129
		Correlation coefficient	-.129	1.000
		Sig. (2-tailed)	.145	.
		N	129	130

Table 7: Standardised beta coefficients

Model		Standardised coefficients	T	Sig.
		Beta		
1	(Constant)		3.725	.000
	Yearmark	.498	6.932	.000*
	selftesttimepretest	.294	4.096	.000*
	Gender	-.070	-.974	.332

*statistically significant at the 1 percent level.

Dependant variable = Test 3 mark (50 MCQs)

Table 8: Standardised beta coefficients

Model		Standardised coefficients	T	Sig.
		Beta		
1	(Constant)		7.777	.000
	Yearmark	.425	5.480	.000
	Gender	-.159	-2.038	.044*
	Selfsttimeyear	.188	2.415	.017*

*Statistically significant at the 5 percent level

Dependant variable = MCQ mark