A COMPARISON OF TWO STUDENT COHORTS UTILIZING BLACKBOARD CAA WITH DIFFERENT ASSESSMENT CONTENT: A LESSON TO BE LEARNT.

Graham Farrell and Ying K Leung

A Comparison of Two Student Cohorts Utilizing Blackboard CAA with Different Assessment Content: A Lesson to be Learnt.

Graham Farrell
Swinburne University Australia
gfarrell@swin.edu.au

Ying K Leung
Hong Kong I.V.E.

Abstract

This study looks at two cohorts of students sitting a mid semester test using the Blackboard Multiple Choice Computer Aided Assessment (CAA) package, but with vastly different content to be assessed. One test was the simple Multiple Choice Question (MCQ) format of a stem followed by four simple text options whist the other utilized the MCQ format to compare complex SQL scripts to meet the required outcome stated in the stem. In total 465 students completed a questionnaire as part of the standard subject evaluation for the two subjects. The questions were designed to evaluate their opinion of the testing procedure and highlight their concerns. The results were statistically analyzed using the Chi-squared test for significant difference between the two student cohorts, producing some interesting results. We conclude that a CAA package should be well matched with the content to be assessed. In this study it is apparent that there was a serious problem with the mismatch between the content being tested and the CAA type chosen to do the job. The study also highlights that previous exposure by the students to the CAA is intrinsic to the success of the exercise. In addition we also observe that the students generally demonstrate acceptance of CAAs as a reliable, time efficient and trustworthy assessment mechanism, encouraging the use of CAAs. We also observed some interesting response regarding gender, but have not attempted to draw any conclusions in the area at this stage.

1   Introduction

The use of online testing is fast becoming a significant component of the educational strategies. The increasing usage can be attributed to the supporting technology that permits the distribution of assessment tools via the internet. Most of the existing computer based learning packages and management systems being utilized have CAA facilities to create, distribute and automatically grade assessment tasks.

Some educators fear that the technology is artificially driving the usage of CAA. Instructors are lured towards online assessment option, with the promises of faster turn around, greater feedback and automated student assessment recording.

The question that we ask is "In some instances is the use of online assessment invalid?"

There is no doubt that the benefits from the use of a good CAA package are significant and should be enthusiastically pursued. In many cases the use of CAA is not only valid, but preferred, especially when test material complies with the requirements of the CAA system of choice. The concern that we would like to raise is that when the content does not comply with the requirements of the CAA. In many cases CAA packages are being utilized as a summative assessment test where the material to be assessed does not fit into the functionality of the package. In these situations the students often voice dissatisfaction and complain bitterly about the testing procedure.

This report examines in detail two occurrences, one where the material suited the Blackboard CAA and another example where it apparently did not. It analyzes the collected data making some conclusions and recommendations.

2 Literature Review

Assessment plays a critical role in education, both as a means of summative and formative assessment. Black and William (1998) refer to "Assessment" as being a group of activities undertaken by both teachers and students in assessing themselves, generating feedback and grades. The Principles and Standards for School Mathematics (2000) state that assessment tasks should be fully integrated into the classroom as a routine activity and not a standalone event or interruption.

Instructors today have numerous assessment options available to them, including multiple choice questions, short answers, essay type answers, case study reports and many other options equally effective. It is accepted wisdom that instructors have an appreciation of the different choices of assessment and often choose the correct type for a particular purpose (Assessment Tools, 2003). There has been much discussion about the different assessment strategies employed to assess various levels of knowledge. For instance, it is very difficult to test the higher levels of Bloom's (1956) taxonomy of educational objectives utilizing the multiple choice question (MCQ) format of testing (Bacon 2003). It is generally accepted that MCQs can assess the lower levels of Bloom's (1956) taxonomy, being Knowledge and Comprehension, and is extensively utilized world wide for this purpose. Bloom's (1956) work does in fact demonstrate the ability of using MCQs as a means of testing all of the educational levels, including application, analysis, synthesis and evaluation, however, these types of questions are difficult to construct. Lambert, Schuwirth and Van Der Vleuten (1996) emphasize the growing dissatisfaction with the MCQ format as they rely on recognition of the

correct answers, while some see MCQs as only demonstrating knowledge of isolated facts (Wilson and Case 1993). Wilson and Case (1993) also state that they fear this "undue emphasis on recall" will "stimulate students to learn in a like mode". Similarly, Lambert, Schuwirth and Van Der Vleuten (2003) have strongly argued that the question format is not as crucial as the construction of the question, criticizing the validity of the constructed questions, asking if the questions actually test what they purported to test.

Multiple Choice Questions (MCQs) continue to be a popular assessment option for instructors today. Bacon's (2003) research demonstrates that MCQs when constructed properly have the ability to test a broad range of fields at various levels effectively and efficiently. MCQs relished in the hypermedia environment permitting the quick adoption of the MCQ format to the internet. Consequently there are numerous online MCQ testing packages available to instructors. Some of the appeal to instructors is the intrinsic nature of these MCQ packages to randomly generate questions from a predefined test bank, automatically give answers, suggest direction for the learner, offer adaptive learning strategies, automatically record grades and monitor the students' progress; all contributing to the students' management of their learning path.

Bacon(2003) considers one of the advantages of using MCQs is the elimination of "Subjective" marking, where "lack of reliability" can lead to inconsistency, whilst Ashburn(1938) demonstrates examples of where this has been apparent during the remarking of assessment tasks.

3 Method

The two subjects that this report focuses on, Database 1 (DB1) and Data Communications (DC), requested the students to complete a questionnaire on various aspects of the subject as part of the standard subject review process. Included was a series of questions that focused specifically on the mid semester Blackboard CAA test that they completed. The total number of students surveyed was 465, comprising of 404 DB1 and 61 DC students. The data was collected and analyzed.

3.1 Demographics

3.1.1 Database 1

Of the 404 Database students who responded to the questionnaire there were 267 males and 137 females, consisting of 166 undergraduate and 238 postgraduate students. The average age of the postgraduate students was 28 years with 98% of them being older than 20 whilst average age for the undergraduate students was 22 with 62% of them being 20 years of age or above.

3.1.2 Data Communication Students

Of the 61 DC students surveyed there were 48 male and 13 female. The majority of the students were undergraduates, but a firm figure of the distribution was not available. It is known that 74% of the students were

between 20 and 29 years of age. The DC students had extensive experience with the Blackboard student management system as it is a second year subject.

3.2 Tests for DB1 and DC

Both cohorts of students were required to complete an online test as part of their assessment. The tests for the two subjects were delivered during the tutorials under supervision utilizing Blackboard. Importantly, the content of the two tests were significantly different.

The DB1 test required the student to identify the correct Structured Query Language (SQL) script to complete a given task from a choice of 4 options, whereas the DC test was designed to evaluate the student's knowledge of networking protocols and the application of programs to support the construction of networks in the industry environment.

3.3 Surveys Structure

The subject evaluation survey contained 13 questions specifically designed to evaluate the testing process, implementation and demographics.

The response rate was above 75% for the DB1 students and 52% for the DC students.

4 Results and Preliminary Analysis

Non parametric statistical analysis was applied to the data where appropriate, utilizing the Chi-squared test for significant difference between the two cohorts.

For aid the reader the summaries have been grouped together under each of the survey questions below.

4.1 Q 1. Have you previously participated in an Online MCQ test before, not associated with this test?

49% of the DB1 and 84% of DC students registered having previous exposure to an online MCQ test. This shows a significant difference ($P\aleph^2(3)\approx 24.544$: $P< .001$).

4.2 Q 2. Did you use the sample test on Blackboard before sitting the test?

Only 58% of the DB1 students did the practice test whilst 85% of the DC students did.

On further investigation it was discovered that the practice test for DB1 was made available for the students 2 weeks before the sitting of the test, while the practice DC test was available to the students at the beginning of the semester.

4.3 Q 3. If you used the sample tests, did they assist you in doing the real test?

Of the students who used the sample test 66% of the DB1 students and 73% of the DC students considered the sample tests to be helpful.

4.4 Q 4. What testing method do you prefer to use to be able to read the entire question on one page/screen? Paper, Online or Either?

74% of DB1 students registered a preference to a traditional paper based test whilst only 36% of the DC students preferred a paper based test. This shows a significant difference ($P\aleph^2$ (7)≈ 41.465 : P< .001).

The main reason stated by the DB1 students was that the test required them to scroll down the screen moving the initial lines off the screen, making script comparison exceedingly difficult.

The DB1 students also stated that the ability to highlight or mark a paper version when comparing scripts is highly advantages and was not available in the online test.

4.5 Q 5. What testing method do you prefer to use to be able to concentrate on one question at a time? Paper, Online or Either?

58% of the DB1 students prefer the paper based test for this issue again referring extensively to the need for scrolling in their tests compared to 36% of the DC students, being significantly different ($P\aleph^2$ (5)≈ 16.99 : P< .001)

A number of DB1 students stated that resulting eye strain from concentrating on the screen whilst scrolling continually had a detrimental effect on their concentration and consequential scores.

4.6 Q 6. What testing method do you prefer to use to be able to move from question to question with ease? Paper, Online or Either?

A significant difference of 65% of DB1 students demonstrated a preference for using the paper based tests over the 43% of DC students ($P\aleph^2$ (5)≈ 39 : P< .001). In the supportive comments the students felt that the familiarity with paper permitted a quick flick through. Interestingly, it was stated by those who supported the online option that the ability to navigate through the questions is faster and very simple once you know how.

4.7 Q 7. What testing method do you prefer to use to be able to check answers before submitting? Paper, Online or Either?

Again, 64% of the DB1 students prefer the paper based test significantly greater than the 36% of the DC students ($P\aleph^2$ (5)≈ 30.776: P< .001).

4.8 Q 8. What testing method do you prefer to use to minimize the time it takes to process tests and receive result back to students.? Paper, Online or Either?

The greater majority of all of the students from both cohorts (above 90%) appreciate the speed from the automation that this technology supplies.

4.9 Q 9. What testing method do you prefer if the percentage of the final grade were large? Paper, Online or Either?

58% of the DB1 students and 33% of the DC students preferred paper based tests.

This again is a significant difference between the DC and the DB1 students' preference to the paper based test. ($P\chi^2$ (5)≈ 15.3678: P< .001).

The DB1 students test had a 30% weighting towards their final grade while the DC students weighting was 15%. The DB1 students would be mindful of the large contribution this test has to the final grade increasing their anxiety during and after the test.

4.10 Q 10. What testing method do you prefer if you wish to verify the mark you were awarded.? Paper, Online or Either?

42% of DB1 students prefer the paper option compared to 15% DC students, which is a significant difference. ($P\chi^2$ (5)≈ 20.711: P< .001).

The DB1 students are not as vigorously opposed to the online option as previously demonstrated. They consider the accuracy of the online recording to be reliable and show confidence on the CAA to authentically reproduce their recorded result.

4.11 Q 11. What testing method do you prefer if you wish to score a better grade.? Paper, Online or Either?

Again we see a significant difference in the preference for paper based testing for the DB1 students (60%) compared to the DC students (20%) ($P\chi^2$ (5)≈ 44.23: P< .001).

The comments from the DB1 students in support of paper based tests mentioned familiarity, ease of use of paper and the ability to highlight script components. The comments by the DC students in support of the online option were confidence in the scoring accuracy and familiarity due to previous exposure.

4.12 Q 12. Do you think you would have scored a better mark if you had of been exposed to a number of self evaluation practice tests for revision purposes during the semester as part of the course before sitting the final test?

It is apparent that most students consider the exposure to online self assessment beneficial in their preparation to final assessment tasks, with 89% DC students and 84% DB1 acknowledging the advantages of frequent pre-exposure.

4.13 Q 13. What do you consider being important in the design of an online assessment package?

The students consider the following as being important in the design and implementation of an online assessment package:

- "1 question per screen" (Dominant response)
- Clear layout, minimize eye strain and cognitive load
- Stability of system (Eg Test would automatically terminate if the screen was resized)
- Index to questions for quick reference, being able to move in a non linear fashion
- Practice tests should be available early in the semester.
- Usability issues. (Icon affordance, font size)

4.14   Gender Bias

There was not any strong evidence that indicated that there was any real difference between the responses from the females to the males. However, there were some observations of interest in their supportive comments. It was observed that 25% of females responded that they needed more time to complete the test compared to no registered comments from the male cohort. Also, more than 50% of the males registered a concern regarding presentation issues compared to none from the females. No conclusions can be drawn from these observations but the authors consider the observations noteworthy.

5   Conclusion and Discussion

The results of this study was both encouraging and concerning. The extensive use of CAAs is encouraged and in many cases proved to be a reliable and efficient means of grading the students.

The discussion that follows summarizes the main points and suggests some strategies to ensure the success of running a CAA task.

It was observed that a significant number of DB1 students registered no prior experience with online testing, including the practice test. In actual fact DB1 students were only given a 2 week window of opportunity to practice with the online testing system. In contrast the majority of DC students utilized the online practice test from the beginning of the semester. The authors strongly suggest that prior exposure to the CAA be encouraged as early in the subject as possible, as the response from the both cohorts clearly support it.

It was observed that the mature aged postgraduate students utilize the practice tests more than the younger, undergraduate students. For those with experience in teaching both undergraduates and postgraduates this result should not be of any surprise. The general approach to education from postgraduates is in most case more mature and responsible demonstrating a willingness to utilize most available resources to assist them along their learning path.

The DB1 students demonstrated a significant difference to the DC students in preference to paper based questions for the following issues:

- Maneuverability between questions.
- Need to check all answers before submission.
- Ability to concentrate on a single question at a time.

In addition one of the most recurring comments from the DB1 students was

- The need to be able to highlight areas of script when comparing them, which is difficult to simulate with CAA, unless customized.

The issue of the need to scroll the page down when comparing SQL scripts was by far the most concerning for the DB1 students. More than 80% of them registered scrolling as an issue. In contrast, this was not an issue for the DC students, as their test questions all sat comfortably on a single screen. This highlights the need for a match between the content being tested and the CAA chosen. In this case it was apparent that the DC content fitted well into the constraints of the CAA whilst the DB1 test did not.

It was encouraging to see that the large majority of the students appreciated the automation of the CAA and demonstrated a general confidence towards the verification of marks when it came to challenging their grades.

The weighting of this assessment task of the overall grade was significantly higher for the DB1 students (30% of the final) than for the DC students (15% of the final), raising the stakes for them, contributing to the majority of DB1 students preferring paper based tests for testing this type of material.

As a result of this study, the DB1 mid-semester SQL assessment task has returned to the traditional paper based test with a positive outcome, while DC instructor continues to successfully utilize the Blackboard CAA test.

In conclusion, it is apparent from this study that while CAAs offer great opportunities, but the content of the tests should be well matched with the CAA of choice. This is evident for the DB1 script comparison exercise, where the students needed to view the alternative scripts at the one time and would have benefited from being able to highlight components for closer scrutiny. In addition, the exposure to the CAA to be used should occur early, perhaps as a formative assessment task, as identified by Farrell and Leung (2002). The authors feel that CAA should be pursued with vigour as the benefits are in favour of both the instructor and the students. Its true potential will be maximized with good planning and implementation.

References

Ashburn, Robert (1938). An experiment in essay-type question. *Journal of Experimental Education 7 (1): 1-3*

Assessment tools for Assessment, Evaluation and Curriculum Redesign workshop: *month 7*
http://www.thirteen.org/edonline/concept2class/month7/index_sub2.html *(Last accessed Aug 2003)*

Bacon, Donald R (2003): Assessing Learning Outcomes: A Comparison of Multiple-Choice and Short-Answer Questions in a Marketing Context: Journal of Marketing Education. Vol 24. No 22. Sage publications [not yet published]

Black. P and William D (1998): Inside the Black Box: Raising Standards Through Classroom Assessment. Phi Delta Kappan October 1998· Volume 80· Number 2 P 139-149 http://www.pdkintl.org/kappan/kbla9810.htm

Black. P and William D (March 1998): Assessment and Classroom. Learning *Assessment in Education, Vol 5 March* P. 7-74

Bloom, Benjamin S (1956): Taxonomy of educational objectives, hand book 1: Cognitive domain. *New York: Longman Green.*

Farrell, G and Leung (2002), Designing an Online Self-Assessment Tool Utilizing Confidence Measurement. *Conference Proceedings IFIP 8.4 WG (2002)*

Lambert W.T. Schuwirth and C.P.M. van der Vlueten (1996): Quality Control: Assessment and Examinations: *http://www.oeghd.or.at/zeitschrift/1996h1-2/06_art.html (Last accessed Aug 2003)*

Principle and Standards for School Mathematics (2000): *National Council of Teachers of Mathematics - Standards 2000 Project Chpt 2*

Wilson, R. B. and Case, S. M.: Extended Matching Questions: An Alternative to Multiple-choice or Free-response Questions: *Journal of Veterinary Medical Education. Volume 20:3.*
http://www.utpjournals.com/jour.ihtml?lp=jvme/jvme203/ExtendedMatchingQuestions.html *(Last accessed Aug 2003)*