AISHE Conference 2007
Teaching and Learning in the Changing World of Higher Education

30th & 31st August 2007
NUI Maynooth, Ireland

A Study of Blended Assessment Techniques in On-line Testing.

Eugene F.M. O'Loughlin,
School of Informatics,
National College of Ireland.
eoloughlin@ncirl.ie

Steven J. Osterlind,
University of Missouri-Columbia,
16 Hill Hall, Columbia,
MO 65211, USA.

Abstract

Over the past few years, blended learning has become more and more popular for educators and students alike. However, assessment is slow to follow this trend – blended assessment has not yet gained the same status as blended learning.

Traditional on-line testing using various types of multiple choice questions (MCQ) has some disadvantages compared to written assessments. Principle among these is that educator's cannot be certain if students have demonstrated knowledge levels appropriate to their marks – guessing and looking for patterns are obvious tactics used.

In this study, traditional methods of assessment are combined in an innovative way. Assessments used are primarily on-line MCQ-based, but for some key questions – written "follow-on" questions require written explanations on paper for choices made in the MCQ. For example, a student could be asked to identify the correct definition for a term from a list of possible answers given, and then asked to give an example in their own words of where the term is normally used. In this way, an educator can set an MCQ question and then ask for a further short explanation or description of an example that clearly illustrates student understanding.

Assessment results are gathered over four semesters in a two year experiment. Both undergraduate and post-graduate students were assessed in this method. Student performance in MCQ tests featuring “follow-on” questions is compared with traditional MCQ-only assessments for the same groups. Tests results are also examined to see if students benefit from the “follow-on” questions, where tests results including and excluding the “follow-on” questions are compared.

The key findings are that the method of blending assessment described is an effective way of combining MCQ-based questions with written questions in an assessment. A comparison of results in assessments that use MCQ tests featuring “follow-on” questions, versus traditional MCQ-only tests reveals that students’ benefited by getting higher marks in tests using “follow-on” questions.

Introduction

When we think of assessment, we generally think of it as the process of establishing, usually in evidence-based and measurable terms, knowledge, skills, attitudes, and beliefs. While there are many different methods of assessment available to educators, they often choose to rely on simple methods for testing (such as multiple-choice questions - MCQ) or by the traditional method of written examinations. Educators have for a long time debated the merits of multiple-choice type assessments compared with those of traditional "pencil-and-paper" assessments. The quandary for educators is often deciding which method is best suited to a particular situation. Too often, class size drives the assessment method chosen. When a large number of students are to be appraised a multiple-choice test is often used even though it takes a great deal of effort to design and author MCQs. Once written and used in an assessment, these tests require little time to assess and mark, especially if on-line assessment tools are used. When small classes are involved in the assessment, the more traditional approach of setting essay-type questions (which take relatively less time to author) works best for many educators, regardless of the fact that they require extended time for marking.

The quandary between item types is even more acute in the on-line environment. Students are rarely asked to provide long typed answers on a computer in a formal examination setting. Mostly, it just does not make sense to do so when there is a simpler and cheaper option available: hand-written exams in an exam hall. After all, computer labs or computer-based test centres rarely have more that a few dozen computers available. Consequently, most educators in the on-line environment rely on the tried-and-trusted, though limited, multiple-choice type of questioning. Typed answers are rarely more than simple fill-in-the-blank or short answer questions, making them relatively easy to mark.

Blended Learning

In recent years, blended learning has become more accepted as a way of learning and teaching. Driscoll (2002) describes blended learning as referring to four different concepts:

- To combine or mix modes of Web-based technology to accomplish an educational goal
- To combine various pedagogical approaches to produce an optimal learning outcome
- To combine any form of instructional technology with face-to-face instructor-led training
- To mix or combine instructional technology with actual job tasks

Bersin (2003) found that the key to blended learning seems to be selecting the right combination of media that will drive the highest impact for the lowest possible cost, and that programs with the highest impact blend a complex media with one or more simpler

tools. Bonk (2004) refers to the Perfect E-Storm, "where technology, the art of teaching, and the needs of students are converging." Bonk discusses thirty emerging technologies that are generating waves of new opportunities in online learning environments. Combined with the traditional classroom environment, teachers and students are now faced with powerful teaching and learning methods that make up the concept of "blended learning". Rosenberg (2006) describes "true" blended learning as a "combination of training (formal) and non-training (informal) approaches that support the smart enterprise (such as knowledge management, performance support, and coaching) in ways that improve the effectiveness and efficiency of learning".

So – we have blended learning, but what about blended assessment?

The "pencil-and-paper" vs. MCQ debate

First, let's compare "pencil-and-paper" assessment with MCQs. Multiple choice types of assessment have some disadvantages compared to written assessments. Principle among these is that educator's cannot be certain if students have demonstrated knowledge levels appropriate to their marks – guessing and looking for patterns are obvious tactics used. Knowing that the end-of-semester assessment will be composed of MCQ, which often tend to address superficial facts, may encourage learning of surface detail rather than a deeper understanding of the underlying concepts.

Even if questions are carefully worded, assessors cannot be sure that a student who answers correctly not only knows the correct answer, but also understands the subject being examined. Then, with MCQ guessing is the other obvious limitation. After all, a student who guesses the correct answer gets the same marks as one who fully understands the subject. Almost certainly this would not happen in pencil-and-paper tests. Also, students can select a correct answer for superficial reasons, such as when they vaguely remember reading something in a book about the subject, or by selecting the answer through a process of elimination.

While it is difficult for students to achieve high overall marks in an MCQ test with limited knowledge, they can get lucky and pass a test by guessing, looking for patterns, and reducing the number of possible correct answers by a process of elimination. Educators, therefore, cannot be fully satisfied with MCQ tests, regardless of whether they are paper-or computer-based. Throw in authentications and security issues in the on-line environment and there are many reasons why MCQ is not popular with some educators. Equally unpopular for essay exams is the effort of marking page after page of written scripts.

Blended Assessment

According to McCabe (2006) "blended assessment drives blended learning". McCabe found that blended learning and computer assisted assessment involves the tight coupling and interaction of learning components. When assessment and learning resources are blended together, students are encouraged to learn more effectively. Blended learning resources supported by blending several different types of computer assisted assessment are therefore extremely effective.

As students and educators are adopting new technologies for learning, we should at the same time be re-considering our traditional methods of assessment. Instead of taking sides in the “pencil-and-paper” vs. MCQ debate, perhaps a blend of both methods will work.

A blend of assessment methods

In this study, traditional methods of assessment are combined in an innovative way. In the experiments described below, the MCQ questions were authored and delivered in Moodle while students wrote their answers to the follow-on questions on paper which was handed up at the end of the assessment. Students don’t get their overall score until the follow-on questions are marked. A blend of assessment techniques in the on-line environment has been carried out at the National College of Ireland over four semesters in a two year experiment. Both undergraduate and post-graduate students were assessed in this method. Assessments used are primarily on-line MCQ-based – with between 10 and 20 questions per assessment. These assessments were delivered though the Moodle LMS in a computer laboratory under supervision. In effect, this part of the assessment is indistinguishable from a regular MCQ only on-line assessment.

The following is a typical MCQ question from the module Introduction to Java Programming:

> *Which of the following Java "for" statements contains a syntax error?*
>
> *A.* *for (i = 0; i < 10; i++)*
>
> *B.* *for (int i = 1; i <= numCalc. i++)*
>
> *C.* *for (int i = 0; i < 25, i++)*
>
> *D.* *for (int i = 0; i < sizeArray; i++)*

As most Java programmers should notice (and the diligent student in the course), there is a syntax error in option "B". The full-stop will generate an error in the Java compiler. Options A, C, and D will not generate a syntax error. As the question stands, a student selecting option "B" will get full marks regardless of whether they spot syntax errors or not. Perhaps option "B" simply looks wrong because it is a bit different from the others. No examiner can be certain that a student who selects option "B" above really understands *why* this code will generate an error. After all, the student was not required to write out a particular "for" statement which, if done correctly, would demonstrate a deeper understanding of syntax, as well as coding. It can also be said that if a student selects either options "A", "C", or "D" above that he or she may strongly feel that their selection does in fact contain an error, but they have no opportunity to explain why or justify their selection and consequently get no credit for their view.

Suppose instead the question was given in two parts – one part composed of an MCQ and the other as written. The first part would be exactly the same question as above and can be answered as normal. The second part could be what we call a "follow-on question" based on the first part. At its simplest, the follow-on question could be something like "Explain *why* you made your selection." A more complex version for the Java programming question above could be something like, "Identify the syntax error(s) in your selection and write a correct version". Short written answers on paper work best and are all that is required for the follow-on part of the question. Extra marks should be awarded for these explanations. These written responses will have to be marked separately if the MCQs are delivered on-line.

Suddenly, but usefully, the question becomes a harder one for students to get full marks. The more able student who can correctly identify the right option, explain fully what the syntax error is in this case, and suggest an acceptable correct alternative, will receive top marks. A student who selects the wrong answer and gives an invalid written answer will score lowest, perhaps even zero for the full question. A student who guesses the correct

answer, but does not know why the error occurs and cannot give a reasonable explanation, will still get full marks for selecting the correct option, but will score poorly in the follow-on question. What about the student who makes an incorrect selection, but writes an explanation as to what they believe the error is, and even writes an alternative, acceptable, version of the code? This student will still receive zero marks for the MCQ part, but perhaps their written answer should warrant some extra, even meritorious, marks.

Overall, this blended approach can encourage students to think a little deeper into the answer as they will have to provide an explanation for their selection. It is also encouraging to know that even if a student gets an MCQ wrong, there is still the possibility to get some marks in the follow-on question as has happened in this study. Students are not discouraged from guessing as negative marking does not apply. They may even be encouraged to guess and to attempt an explanation in the hope of gaining some marks.

The absolutism of MCQs being either correct or incorrect can disadvantage students. Supposing in the question on Java code the student is confident that he or she can eliminate two incorrect options out of the four choices available, but is undecided as to which of the remaining two options to select. The student did evidence that he or she can at least partially answer the question correctly by eliminating two options, but if they then selects the incorrect option from the remaining two, zero marks ensue, despite the fact that some knowledge was displayed. While a leaner may have demonstrated partial knowledge, they do not get any credit for this. If a follow-on question is used in this instance, the student has an opportunity to demonstrate their partial knowledge in their written response and perhaps gain some marks.

Blending MCQs and pencil and paper assessment in this way will give the assessor some insight into how students, and the class as a whole, are demonstrating true knowledge of the subject area. Assessors will also be able to give credit for partial knowledge in the cases where follow-on questions indicate this knowledge on behalf of the student.

In contrast, for pencil-and-paper assessment, students can write their explanations at length and will most likely gain at least some credit for even partial knowledge. Here, zero marks in written responses are less likely to be given, especially if the response illustrates some knowledge of the subject area. The examiner's judgement is now a factor in assessing a response.

Of course, using follow-on questions for each and every MCQ defeats the purpose of using MCQs in the first place. Such a strategy would be impractical and would also increase the assessor's workload. Instead, an appropriate balance for the number of follow-on questions compared to the total number of MCQs should be achieved. Our

practice is to set two to three follow-on questions per assessment. Even one such question in an MCQ test can be useful. This number is small enough to mark quickly, but at the same time get a clearer picture of students' levels of understanding.

It is important to note that greater care must be taken in phrasing both parts of this type of MCQ/follow-on combination. Questions must be meticulously worded so as to give students some opportunity to gain marks even if they make a wrong MCQ selection. If students are getting MCQs correct, but are unable to provide explanations for their selections – then guessing may be a factor. The assessor will have to re-visit such questions to ensure that they are carefully written so as not to hint at the correct option that is easy to guess. Writing MCQs just got harder.

Student Performance

So how do students perform under the blended assessment format described here? Our results show that students can benefit if blending is employed in assessments. In several experiments, students' overall scores for blended assessment (all featuring 10 to 20 MCQs and 2 or 3 follow-on questions) are compared with scores that exclude the marks for the follow-on questions. Figure 1 shows the results where overall class marks were poor for a class of 22 students:

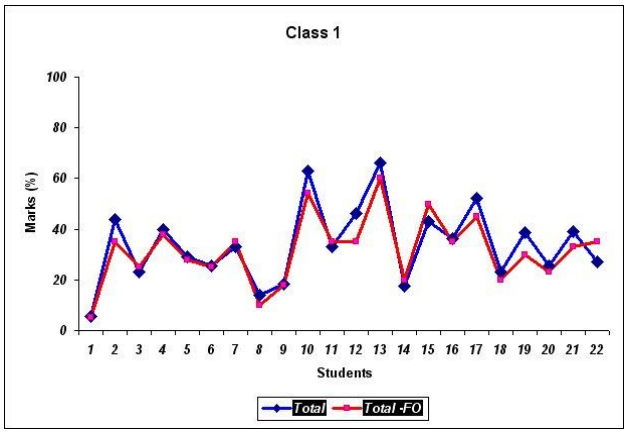

Figure 1 – a comparison of marks for full blended assessment (blue line) with marks excluding follow-on questions (red line).

In this assessment, 16 out of 22 students had their marks increased by between 0.4% and 11.2% when marks for the follow-on questions are included. The remaining 6 students show a decrease ranging from -1.9% to -8.1%.

Figure 2 shows the results where overall class marks were good for a small class of 8 students:

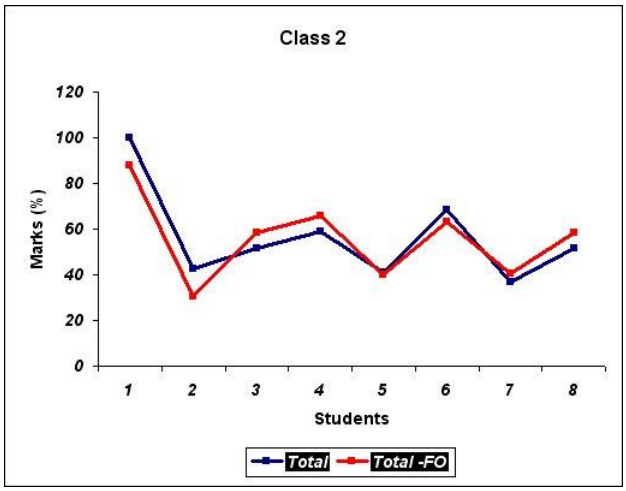

Figure 2 – a comparison of marks for full blended assessment (blue line) with marks excluding follow-on questions (red line).

In this assessment, 4 out of 8 students had their marks increased by between 3.8% and 7% when marks for the follow-on questions are included. The remaining 4 students show a decrease ranging from -0.6% to -12.1%.

Figure 3 shows the results where overall class marks were good for a small class of 8 students:

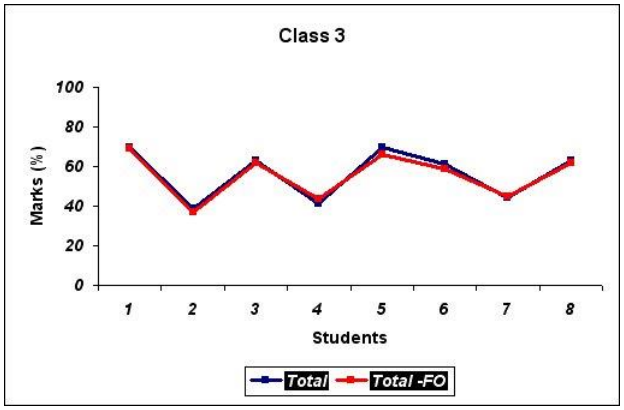

Figure 3 – a comparison of marks for full blended assessment (blue line) with marks excluding follow-on questions (red line).

Here, results are almost identical with 2 out of 8 students had their marks increased slightly by between 0.6% and 2.9% when marks for the follow-on questions are included. The remaining 5 students show a slight decrease ranging from -0.77% to -3.8%.

Figure 4 shows the results where overall class marks were good for a class of 19 students:

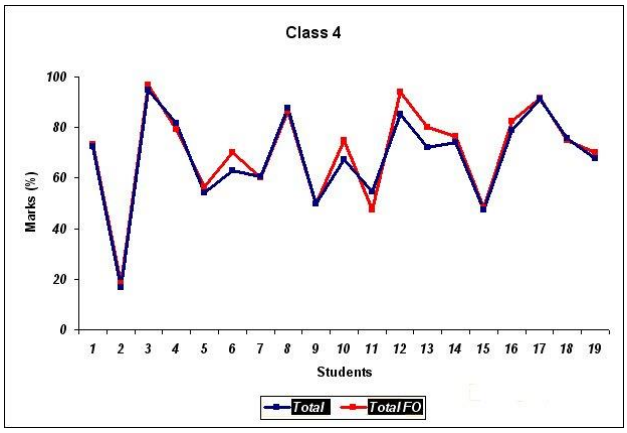

Figure 4 – a comparison of marks for full blended assessment (blue line) with marks excluding follow-on questions (red line).

In this assessment, 14 out of 19 students had their marks increased by between 0.5% and 9.06% when marks for the follow-on questions are included. The remaining 5 students show a decrease ranging from -0.7% to -7.1%.

Figure 5 shows the results where overall class marks were good for a class of 19 students:

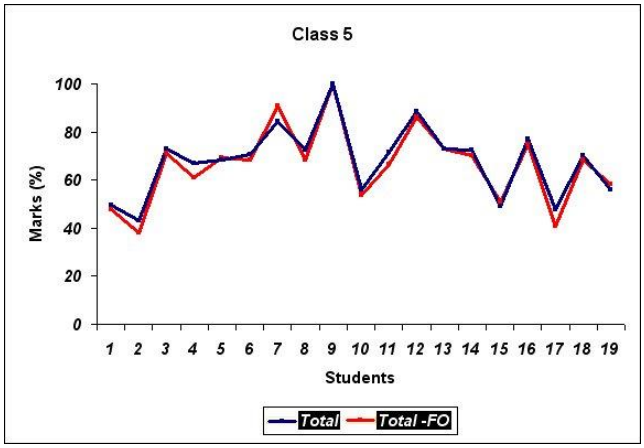

Figure 5 – a comparison of marks for full blended assessment (blue line) with marks excluding follow-on questions (red line).

In this assessment, 14 out of 19 students had their marks increased by between 0.35% and 6.78% when marks for the follow-on questions are included. Four students show a decrease ranging from -0.6% to -6.3. One student's mark remained the same.

Figure 6 shows the results where overall class marks were excellent for a class of 15 students:

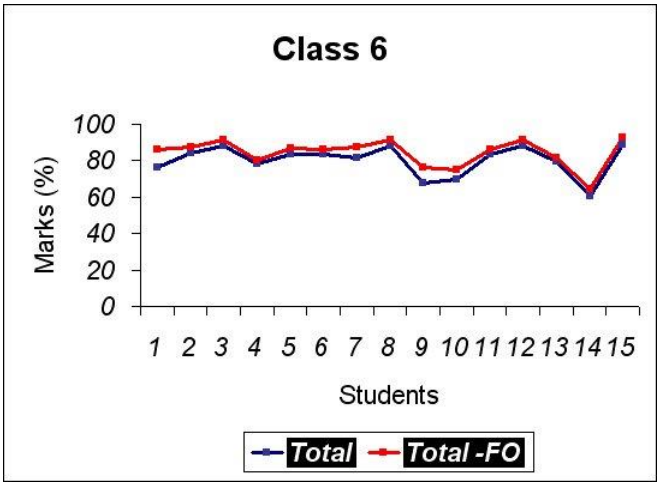

Figure 6 – a comparison of marks for full blended assessment (blue line) with marks excluding follow-on questions (red line).

In this assessment, all 15 students had their marks decreased by between 1.9% and 9.9% when marks for the follow-on questions are included.

While in all experiments the results were closely matched, they do show that in five out of six tests, the majority of students gained extra marks in the follow-on questions and benefited overall from this method of assessment. This is demonstrated in the above tests for a range of overall performance.

The fact that in all six experiments above, the percentage results including or excluding the follow-on questions, were close together in value could also be used to validate and MCQ test. Had results varied significantly, the examiner may have to review the MCQ part of the assessment if students are scoring high in the MCQ component, but poorly in the follow-on component. Conversely, if the scores are high in the follow-on, and low in the MCQ – the MCQ component may need reviewing. Further research with larger samples and more variation in the numbers of questions asked needs to be done to evaluate this possibility.

Future of Blended Assessment

The drawback in using the above type of blended assessment is that there is a still a written component to be manually marked. There is a lack of tools that could be used in the on-line environment to aid this extra marking. In order to succeed in on-line environment, question authoring tools would have to be created to allow for a number of follow-on questions to be inserted at the author's choosing. This mixture of on-line and hand written responses can be confusing to students, so instructions must be very clear - in patricular the amount of hand written material to be submitted. Our practice is to hand out prepared sheets with fixed-sized boxes for student response. Practice tests may be necessary where students are not familiar with this type of assessment.

As blended learning increases its hold, educators will need to continue to look at different ways of assessing learning. Blended and balanced assessment is a necessity in this new landscape. Dochy (2005) believes that students might find "new modes of assessment motivating and interesting not certainly because of the characteristics of such new modes, but perhaps because they severely dislike and are de-motivated by traditional testing methods." Many courses today feature e-Learning, classroom based learning, blogs, wikis, podcasts, discussion groups, and many other technologies. Richardson (2006) states, "The coming years will be marked by a flood of new innovation and ideas in teaching." Educators need to plan for changes in the way we assess students in this changing blended learning landscape. New blended assessment techniques, tools, and methodologies will be necessary to respond to this challenge.

References

Bersin, J. (2003*) What Works in Blended Learning.* Learning Circuits – ASTD. Available at: www.learningcircuits.org/2003/jul2003/bersin.htm. (Accessed 15th August, 2007)

Bonk, C. J. (2004). *The Perfect E-Storm emerging technology, enormous student demand, enhanced pedagogy, and erased budgets*. http://www.publicationshare.com/part1.pdf (Accessed November 14, 2006)

Dochy, F. 2005. *'Learning lasting for life' and 'assessment': How far did we progress?* Presidential address, EARLI 2005 at the 20th anniversary of the European Association for Research on Learning and Instruction.

Driscoll, M. (2002) *Blended Learning: Let's get beyond the hype*. LTiNewsline. Available at: http://elearningmag.com/ltimagazine/article/articleDetail.jsp?id=11755. (Accessed 15th August, 2007)

McCabe, E. M. (2006) *B-Assessment for B-learning - Blended, Bundled or Bungled?* First Annual Blended Learning Conference "Blended Learning - Promoting Dialogue in Innovation and Practice" University of Hertfordshire. Available at: http://perseus.herts.ac.uk/uhinfo/library/l80356_3.pdf (Accessed 15th August, 2007)

Richardson, W. (2006). *Blogs, Wikis, Podcasts, and Other Powerful Web Tools for Classrooms*. Corwin Press. Thousand Oaks.

Rosenberg, M. J. (2006). *Beyond e-Learning: Approaches and Technologies to Enhance Organizational Knowledge, Learning and Performance*. San Francisco, CA: Pfeiffer.