Developing meaningful feedback to large class sizes via short answer assessment

William Gibson[1], Jon Yorke[2] and Heath Wilkinson[3]
[1]*School of Physiotherapy, Curtin University;* [2]*Office of Assessment, Teaching and Learning, Curtin University;* [3]*Business Systems (CITS), Curtin University*

Student learning is known to be enhanced by the provision of formative assessment opportunities. The term 'formative' necessitates an element of 'feedback' provision to the student. A substantial body of research indicates for optimal effect this feedback has to be timely, individualised and provided in such a manner that near-future learning addresses the feedback enabling the student to make a change in their understanding and their performance levels. Many courses have moved to computer based automatically marked assessments which in large part comprise multiple choice questions (MCQs). However computer automated marking of MCQ assessment has a number of inherent difficulties which may diminish the 'return' in terms of student learning. 'Short answer' (SA) formative assessments have a number of advantages over MCQs however they take longer to mark and provide feedback needed. The 'Short Answer Feedback System' (SAFS) addresses this problem by replicating human marking/feedback provision across a database of student responses. This paper compares SAFS to manual SA marking, exploring whether SAFS may be realistically used to provide rapid and 'just in time' feedback in support of learning activities within the semester.

Keywords: feedback, formative, assessment, automated, SAFS

Introduction

Increasingly, assessment is recognised as playing a sophisticated and integral role in student learning and achievement. Formative assessments (with associated appropriate feedback) are used to facilitate student learning while traditionally summative assessments have been employed to rank, certify or measure student learning or achievement (Gibbs & Simpson, 2005; Knight, 2002; Nicol & Owen, 2009). The importance of *feedback* associated with formative assessment cannot be overstated. Associations between feedback and formative assessment with improved grades has previously been outlined (Cook, 2001; Smith, 2007) while Boud & Associates (2010) stress this importance of feedback in relation to assessment by including it within their "Seven propositions of assessment reform in higher education" (Boud, 2010).

Feedback provided as part of formative assessment may be thought of as an opportunity for learners to assess their current performance against desired personal (intrinsic) or externally required standards (extrinsic).

> "Few physical, intellectual or social skills can be acquired satisfactorily simply through being told about them. Most require practice in a supportive environment which incorporates feedback loops" (Sadler, 1989).

Formative assessment empowers learning by providing these feedback loops to the learner.

This enables them to "close the gap" (Sadler, 1989) between their current and their intrinsically or extrinsically motivated *desired* performance levels (Nicol & Macfarlane-Dick, 2006; Sadler, 1989). Clearly then, *feedback* is of paramount importance in a students journey through their learning.

However, feedback as an entity has to be further examined. If feedback associated with formative assessments is to be provided, then it should be provided optimally as it has been demonstrated that students do not always use feedback or feel that feedback is helpful to their learning (MacLellan, 2001). Sadler (1989) makes the point that for feedback to be of value in learning it must be utilised to influence future learning. Data from a recent study reinforces this, with feedback that cannot be incorporated into students learning or *changing* learning reported as "useless" (Poulos & Mahony, 2008). This implies that optimal feedback provided to students following formative assessment should be provided in such a fashion that it is pertinent to the individual and can be 'scaffolded' into the ensuing weeks learning activities. The ability to do this requires feedback to be delivered in a rapid, timely manner.

Timeliness of feedback is also recognised as being crucial in the effectiveness of feedback (Gibbs & Simpson, 2005; Poulos & Mahony, 2008) to engender or allow changes to a students learning. This leads directly to a conflict within the higher education system as staff time and resources are under pressure from increasing student cohort sizes (Nicol & Owen, 2009), meaning the time required to mark and provide meaningful, individualised, timely feedback may simply not be available to academic staff.

Perhaps in response to the inherent tension between increasing class sizes and the need for timely meaningful feedback to formative assessment activities, there has been a move toward automated computer based formative (and summative) assessments. There are a number of assessment formats employed although it is recognised that often multiple choice question (MCQ) assessments can end up being employed due to their ease of deployment and marking (Harris et al., 2007). MCQ formative assessments do have a number of drawbacks to consider. If feedback on these assessments is a grade or mark of correct scores only then students have little to employ in order to effect the change in learning mentioned previously. In addition automated MCQ formative assessments leave little scope for allowing 'partial credit' at the level of the individual question (Sim, Holifield, & Brown, 2004), whilst simultaneously posing the risk of students passing by simple chance due to probability functions being derived from the number of questions and number of possible answers (Brown, 2001). Interestingly, it has been reported that students perceive MCQ assessments as being more about knowledge and recall and tend to employ a more 'surface' learning based strategy when using them (Scouller, 1998). Lastly, a significant level of "hazardous ignorance" has been reported with MCQ assessment (Dory, Degryse, Roex, & Vanpee, 2010). In this study the authors suggest "hazardous ignorance" as being the situation in which a student selects a wrong answer but is very confident they are right (Dory et al., 2010). Computer based MCQ formative assessment leave little room to explore this aspect of incorrect response with MCQ testing, meaning students may persist with erroneous concepts without getting the feedback necessary to effect a change in their learning. Clearly then, simple automated marking of MCQs is not optimal for recurrent formative assessment and feedback provision.

Short answer (SA) questions "require the respondents to construct a response in natural language and to do so without the benefit of any prompts in the question" (Jordan & Mitchell, 2009). This is an important point of difference between SA and MCQs where the correct answer is always on display, potentially testing nothing more than answer recognition. SA

questions not only remove the possibility of results due to 'chance' but can also be used to assess higher orders of cognitive function (Wood, Jones, Sargeant, & Reed, 2006). Additionally, clearly erroneous thinking leading to answers that are wrong will also be attended to via feedback provided. The drawback with SA questions in formative assessment is the time required to mark and provide feedback. With the move to increasing class sizes, manual marking and feedback provision becomes highly time-consuming such that this type of question in formative assessment becomes impractical.

With this in mind, 'proof of concept' of an automated "Short Answer Feedback System" (SAFS) designed to provide rapid, individualised feedback to SA type questions has recently been presented (Yorke, Gibson, & Wilkinson, 2010) . In brief, the system replicates tutor marking decisions and feedback provision across the database of answers supplied by the students. The students complete the assessment via a custom-built web-site and their responses are stored on a database. The system collates the answers to the SA questions, grouping matching answers supplied by students in terms of descending frequency. The tutor marks the response, providing feedback as appropriate (whether the response is correct or not) and the computer then applies this same marking/feedback response to every matching response. This process is continued across all provided responses to that question. The tutor then moves on to the next question. Thus, the system in essence automates and replicates *human* marking decisions. When all questions and responses are marked, the system will (in time) automatically retrieve the marking/feedback responses for each individual student and re-assemble this into a complete set of apparently individual responses. The advantages of this system lie in the fact that it has been created to speed up the process of marking and feedback. This allows for a more comprehensive feedback process with more time spent providing feedback to wrong answers as the correct answers are very quickly dispatched. Essentially it allows feedback to the student cohort *and* the tutor with respect to areas of weakness which may require further attention. The speed of delivery of marking/feedback should allow these areas of concern to be 'scaffolded' into the following weeks learning activities, thereby explicitly allowing students to *change* their learning in response to the feedback they received from their formative SA assessment.

The current paper describes the use of this system to carry out a full-scale trial of marking/feedback of a formative SA assessment via the SAFS system when compared to manual marking/feedback provision time. This will identify whether the SAFS system can function as a viable option to provide formative assessment within semester to undergraduate students. Pilot testing of the initial SAFS system (Yorke et al., 2010) had resulted in learning with respect to improving question design such that larger numbers of respondents could now be realistically invited to take part. The project was undertaken with a volunteer group of first year anatomy students as first year students are a likely future target cohort for delivery of enhanced feedback (Nicol & Owen, 2009) and due to the fact anatomy is highly amenable to questioning in a SA format.

Methods

Fifty eight first year anatomy students responded to a SA formative assessment test placed on the appropriate learning system website. Students were aware this was a formative test and they would receive feedback on their submitted work. The project was undertaken in two phases separated by 5 months. The assessment was delivered via a "Word" document and students responded by emailing their completed assessments to the investigator.

Phase 1.
One investigator manually marked and provided feedback to all students SA formative assessments. Feedback was provided as an overall mark and as individual comments to each response whether right or wrong. The time taken to do this was carefully noted for each student and a tally was kept to allow determination of overall time spent on the manual marking phase of the project.

Phase 2.
Five months later, student responses were entered into the SAFS database and the same marker then undertook marking/feedback provision of the same set of responses via SAFS. An extended time period was allowed between marking in an attempt to reduce 'carry-over' of from marking in the initial phase (although it is recognised that some 'learning' effect in marking cannot be entirely ruled out). The marking and feedback provision was done in the same manner with an overall grade and individual comments to each response being generated. The feedback provided (again whether right or wrong) was written at the time of marking and was unique to this marking occasion. A post marking comparative review indicated that the complexity and detail contained in feedback responses were broadly similar across both phases of the project. Again, the time taken to complete the entire Phase 2 exercise was carefully recorded.

Results

Manual marking/feedback of the fifty eight assessments required 459 minutes. The second phase of the project completed five months later via SAFS took 92 minutes to provide the same level of marking/feedback response. Additionally, the tutor was able to pick out common mistakes as the marking engine presented the answers provided in terms of frequency of response. As such, an overview of the entire assessment was very easily ascertained.

A number of questions did generate multiple 'unique responses' which initially appears to undermine the concept. However in these cases most unique responses resulted from slight variations on the same theme eg "lateral rotation of the hip" vs "external rotation of the hip". Ultimately the answer is the same and so the same feedback can be replicated across these responses. The very fact of having answers automatically collated and ranked in descending frequency order was very helpful in grading and providing feedback quickly.

The above results demonstrate SAFS is a viable option for marking/feedback of formative assessments within semester. This may prove to be an additional valuable 'tool' within the many assessment types currently employed. The previously mentioned positive benefits of formative assessment in the learning process may be realised for larger class- sizes via SAFS without having to resort to MCQs with their multiple attendant difficulties. The fact that marking/feedback provision to a relatively large sample size was completed in 92 mins is very encouraging when considering translation into higher numbers. The relatively large sample size did not translate into a myriad of answers (potentially negating the effect of the system) to each question. As expected during question design, students responded in a number of predictable ways to any given question. Thus it is anticipated that doubling or even tripling the current sample size would not result in increased marking time. The figure of 92 minutes for marking an assessment such as this would seem entirely consistent with marking within semester and as such implies SA formative assessments marked via SAFS could play a realistic and integral role in student learning as the feedback may be scaffolded very quickly

into the following weeks learning activities.

Further testing across differing sample sizes and study areas is now warranted for SAFS. The system will continue to be improved; entry and collation of student responses will be automated in full while auto -emailing of results to students will be developed. Reduction of the hidden admin costs of e-assessment, combined with the marking/feedback decision replication afforded by SAFS means formative SA assessment may be developed in a sustainable manner. Future studies employing this system will also include student perceptions elicited via interview and focus groups.

Acknowledgements

Many thanks are extended to Ms Ieva Bergman for her aid in facilitating use and deployment of SAFS in this project.

References

Boud, D. A. & Associates (2010). Assessment 2020: seven propositions for assessment reform in higher education. *Sydney: Australian Learning and Teaching Council*. Available from: *http://www.iml.uts.edu.au/assessment-futures/Assessment-2020_propositions_final.pdf* (Accessed 12/11/10)

Brown, R. W. (2001). Multi-choice versus descriptive examination. *31st Annual Frontiers in Education Conference Impact on Engineering and Science Education Conference Proceedings (Cat No 01CH37193) FIE-01*.

Cook, A. (2001). Assessing the use of flexible assessment. *Assessment & Evaluation in Higher Education, 26*(6), 539-549.

Dory, V., Degryse, J., Roex, A., & Vanpee, D. (2010). Usable knowledge, hazardous ignorance - Beyond the percentage correct score. *Medical teacher, 32*(5), 375.

Gibbs, G., & Simpson, C. (2005). Conditions under which assessment supports students' learning. *Learning and teaching in higher education,* 1(1), 3-31.

Harris, K.-L., Krasue, K., Gleeson, D., Peat, M., Taylor, C., & Garnett, R. (2007). Enhancing Assessment in the Biological Sciences: Ideas and resources for university educators. Available from: *http://www.bioassess.edu.au/key-issues/engaging-large-classess-through-assessment* (Accessed 12/11/10)

Jordan, S., & Mitchell, T. (2009). e-Assessment for learning? The potential of short-answer free-text questions with tailored feedback. *British journal of educational technology, 40*(2), 371-385

Knight, P. T. (2002). The Achilles' Heel of Quality: the assessment of student learning. *Quality in Higher education, 8*(1), 107-115.

MacLellan, E. (2001). Assessment for learning: the differing perceptions of tutors and students. *Assessment & Evaluation in Higher Education, 26*(4), 307-318.

Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in higher education, 31*(2), 199-218.

Nicol, D., & Owen, M. C. (2009). Quality Enhancement Themes The First Year Experience. Transforming assessment and feedback: enhancing integration and empowerment in the first year.

Poulos, A., & Mahony, M. J. (2008). Effectiveness of feedback: the students' perspective. *Assessment and Evaluation in Higher Education, 33*(2), 143-154.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science, 18*(2), 119-144.

Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education, 35*(4), 453-472.

Sim, G., Holifield, P., & Brown, M. (2004). Implementation of computer assisted assessment: lessons from the literature. *ALT-J, 12*(3), 215-229.

Smith, G. (2007). How Does Student Performance on Formative Assessments Relate to Learning Assessed by Exams? *Journal of College Science Teaching, 36*(7), 28-34.

Wood, M. M. G., Jones, C., Sargeant, J., & Reed, P. (2006). Light-weight clustering techniques for short text answers in HCC CAA *Proceedings of the 10th International Conference on Computer Aided Assessment,* Loughborough, UK.

Yorke, J., Gibson, W., & Wilkinson, H. (2010). Towards sustainable marking practices and improved quality of feedback in short answer assessments. *ATN Assessment Conference*, University of Technology, Sydney, Australia.