# An articulated approach to the development and evaluation of automated feedback for online MCQ quizzes in Human Biology

**Jan Meyer[1], Sue Fyfe[2], Georgina Fyfe[2], Mel Ziman[3], Kayty Plastow[1], Kathy Sanders[1]** and **Julie Hill[1]**, [1]School of Anatomy and Human Biology, The University of Western Australia, Australia; [2]Curtin University of Technology, Australia; [3]Edith Cowan University, Australia
kjplastow@anhb.uwa.edu.au ksanders@anhb.uwa.edu.au jhill@anhb.uwa.edu.au
jmeyer@anhb.uwa.edu.au m.ziman@ecu.edu.au
S.Fyfe@curtin.edu.au G.M.Fyfe@curtin.edu.au

Keywords: feedback; evaluation; expectation; confidence; paid employment

***Abstract**: This paper describes an articulated programme of development and evaluation of automatically-presented explanatory feedback comments for online, enriched-multiple choice style quizzes in Human Biology for first year university courses. The degree of articulation of the separate components of the programme arose almost unintentionally from the inclusion of common sets of demographic questions in several of the components of the work, and from continuity of logon identities, but proved to be a powerful means of reaching an understanding of the dynamics of student engagement with the online learning process and of the effectiveness of the product we were testing. In particular, links were established between expectations of academic performance and the amount of paid employment in which students were engaged, and between expected and achieved levels of performance. Students who expected lower levels of performance at the outset were also less convinced of the potential of feedback to help them with their studies. Analysis of the patterns of use of the online test revealed a serious disadvantage to working students of current accessibility to online summative assessments, and that the standard duration of the summative tests was approximately three times the preferred online work span of the younger students. 'Dose' and 'decay'-graded selective improvements in end of semester assessments in the topics covered by the feedback comments could be demonstrated.*

## Introduction

This paper describes the process of development and evaluation of a set of explanatory feedback comments to be integrated into enriched multiple choice question (MCQ) style online quizzes in Human Biology already employed in first year courses at three Western Australian Universities. While the MCQ question banks had grown independently over the years with contributions from many course co-ordinators at each institution, feedback comments for all question banks were 'retrofitted' over two semesters by an eight person Carrick Project Grant team, in consultation with the current course co-ordinators. The annual cohort of students enrolled in the three units offering the quizzes is in excess of 2000. At least 1100 of the 2006 cohort contributed to this study. While the testbank of MCQ questions with feedback was developed primarily for use in Human Biology courses, many sections are equally applicable to introductory courses in cell biology, preclinical medicine and general biology. Many thousands of items with attached feedback comments are now contained in the testbank. Since it was clear from the outset that the construction of this resource was going to be very labour intensive, it was felt to be important that the type of feedback developed be as effective as possible.

The first stage of the programme of development was, therefore, devoted to determination of the nature and circumstances of the intended users and of their perception of, attitudes towards and conscious expectations of feedback (Figure 1). The next involved the determination of the difficulties and desired outcomes to be addressed by the feedback. A set of principles governing the construction of the feedback comments was developed in the light of the information gathered in these two stages, and advice from the literature.

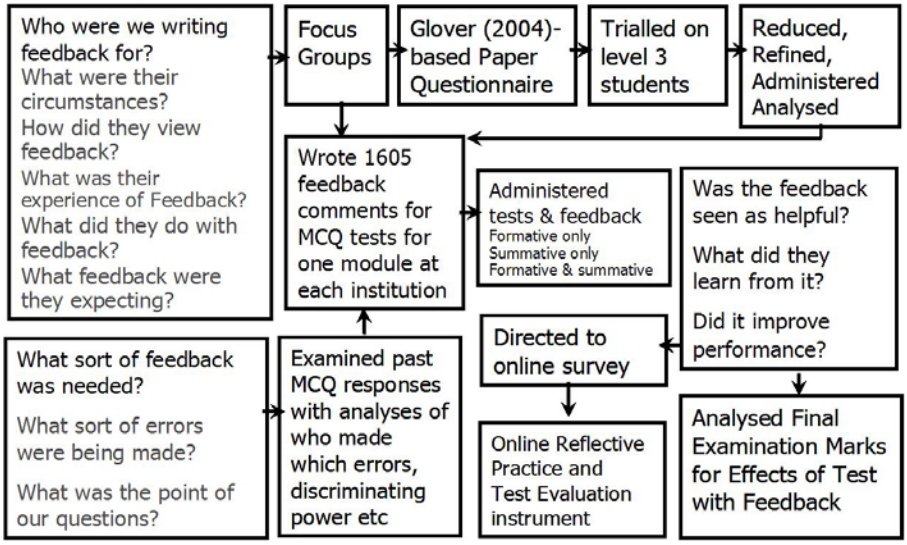


**Figure 1**. The relationship between components of the programme of development and testing of feedback comments for enriched multiple-choice style quizzes in Human Biology

## Methods

Investigation of the needs and circumstances of the students took the form of a survey questionnaire. The range of issues to be addressed in the questionnaire was assayed in a series of 15 focus-group sessions spread across the three institutions involved in the study. Since the issues raised in these sessions converged upon those addressed in the survey instrument developed by Glover et al. (2004) in their study of attitudes to feedback at Sheffield- Hallam and the Open University, this instrument was used as the basis for this phase of our study. A trial of the survey on a group of 30 third year Human Biology students led to a reduction in the number of questions used to approximately one third, and a focus almost entirely restricted to experience with and attitudes towards online feedback. The usual demographic issues were canvassed as part of this survey, and additional questions included concerning the number of hours in paid work per week, mode of enrolment (full-time, part-time or external), expected mark, language spoken at home and declared disability status. A paper-based version of the survey, which took approximately 15 minutes to complete, was presented to first year students in the third week of semester and 1099 responses obtained.

The issues to be addressed by the feedback comments were determined by analysis of student responses to several years' worth of end-of-year multiple choice examinations. Attention was paid to the areas in which errors occurred most frequently, to the overall grades of students making errors in each question, to the form of the question (identification of correct or incorrect elements) and to the context of statements in which errors were made (whether associated with other statements from the same general area, or closely related distracters). Each question was also classified according to its intention, or the outcome it addressed. Not only was a disturbing proportion of statements found to address not much more than terminology, but the error-rate on such statements was relatively high. Since this did seem to indicate a real area of difficulty, all but the nonsensical terminology-based statements were left in the testbanks, though fewer were selected for inclusion in the sets of questions to be drawn upon for each test, at the discretion of the current course co-ordinators. Particular care was taken with the writing of feedback comments for these statements to give reasons for the need to know those particular terms. Typographical and logical errors were corrected as they were encountered and near-duplicate and irredeemably poor questions eliminated, with the permission of the course co-ordinators.

The set of principles governing the construction of feedback comments developed in the light of the information gained from these sources was to:

- limit comments to 1-2 sentences;
- address the 'framework' rather than the specific – the 'why', not the 'what';
- explain why correct responses are correct, as well as why incorrect are incorrect;
- use plain language, especially in relation to explaining terminology;
- include the meaning of key terms in parentheses as encountered;
- use the personal voice for correct responses, the impersonal for pointing out error (the student is right, the statement is wrong);
- affirm correct responses;
- re-iterate in the affirmation the need to choose an incorrect statement ('You are right – that statement was untrue');
- remind of the type of choice required in a general comment, and that where identification of a correct response is required, that most of the statements the student has just read are incorrect; and
- give mental images to remember by, but avoid meaningless mnemonics.

The quizzes were presented to the students in different modes at each of the three institutions involved in the study, according to their usual practice before the addition of the feedback comments. Multiple choice questions, whether online or on paper formed a minor part of course assessment at each institution, the rest consisting of a variable mix of essays, practical examinations, portfolios and in-class quizzes. At one institution the online quiz was freely available 24 hours a day for one month as a formative learning exercise, though questions from the same test bank formed 17% of the marks for the course. At the second it was presented as a 30 item summative test available for 40 minutes under strictly secure conditions in a central computing facility open from 9am to 5pm. This test was one of four for the semester, each worth 6% of the final course mark. At the third the quiz with feedback comments was freely available for one week as a formative exercise, then 30 items from the same test bank were presented again in moderately secure conditions without feedback as one of a series of 6 summative tests worth 2% each of the course mark. Questions from the same test bank appeared paper-based final exam, contributing a further 3% to the overall course mark. Eight hundred and fifteen students at least logged on to the quizzes.

The effectiveness of inclusion of automatically-presented feedback comments for online multiple-choice style quizzes was examined in an online survey of student opinion of its value. The same survey also asked students to reflect upon whether their test score was as they had anticipated, whether it accurately reflected their understanding, where they might improve in terms of content, test and study techniques, and what strategies for improvement they might employ before the next test. Most of the demographic questions from the first paper-based questionnaire were also included in the online survey, which was completed by 315 students.

The advantage gained by exposure to the online quizzes with feedback was assessed by comparing the proportion of available marks in sections of the end of semester examination covered in the quizzes with the proportion of available marks gained in other parts of the examination for students who had or had not accessed the quizzes.

## Results

The most notable feature of the first year cohort participating in this study was the proportion engaged in paid employment. Ninety seven percent of students in paid work were enrolled on a full-time basis. The lowest rate of engagement of students in paid employment at the three institutions was 65%, the highest 75%, at the institution which offered the test only as a formative learning exercise. One quarter of all students at this institution spent more than 20 hours a week in paid employment, largely in paramedical areas. At the two institutions where the test was freely available approximately 40% of students logged on to use it outside of the hours of 9 am to 5 pm.

That the mismatch between preferred and imposed times of access had an impact on student performance was indicated by the behaviour of a group of externally-enrolled, but locally resident students of the institution offering only the summative task. This pattern of enrolment is most commonly encountered amongst students in full-time work, but is also employed by women with heavy family commitments. Despite obtaining above average grades on other aspects of the course, not one of these students presented themselves for testing, and thereby forfeited almost one quarter of their course marks.

The marks that students expected to obtain at the outset of their course fell as the hours per week in paid employment rose (Figure 2a). Amongst the students whose only exposure to the test was as a mid-semester summative assessment those in paid work obtained lower test scores, whether these met, were higher or lower than they had expected (Figure 2b). Despite a lack of difference in the proportions of students feeling their scores fell below their expectations, significantly more students in paid work felt that their results had not accurately reflected their understanding of the subject.

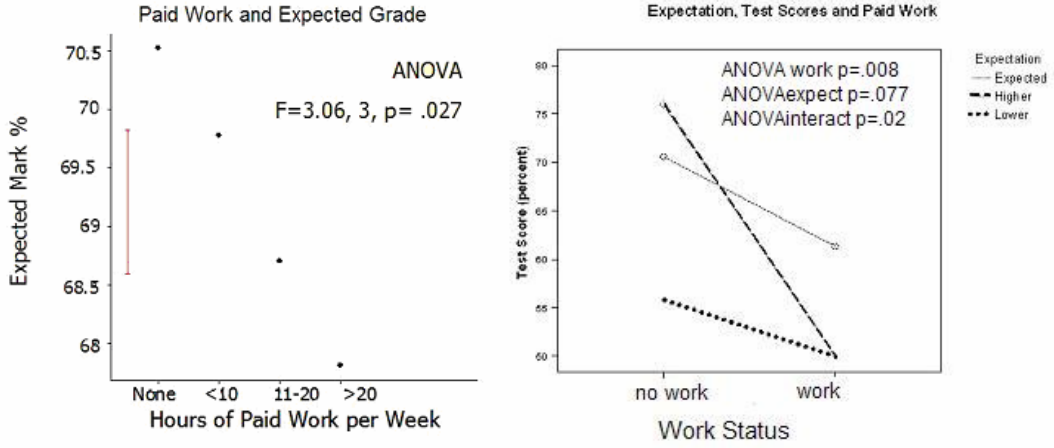


**Figure 2a.** (left) The relationship of marks expected at the outset of the semester and hours spent per week in paid employment (bar = standard error of difference of means)
**Figure 2b**. (right) Mid-semester test scores in relation to the meeting of expectations and involvement in paid work

Overall, students who expected to get no more than a fail or a simple pass grade expected to receive less feedback, claimed to have had less experience of online feedback and were more likely to be dismissive of the prospect of online feedback assisting their learning than students expecting higher grades. They were more likely to make free comments about the negative impact of feedback on self-esteem, such as:

*'…because it does not help the person and can sometimes make the person feel down on themselves'*
*'…it hurts me on the inside'*
*'…comments that make me feel stupid…… makes me don't wanna learn'*

Older and more experienced students placed a higher value on feedback.

Examination of the pattern of submission of questions in the 30-question formative task at the institution which also presented them in a summative test revealed that only 39% of those who logged on to the formative task completed all 30 questions in one session. Another 35% completed at least one third of the questions without finishing the full test, while 13% remained logged-on for days, eventually completing the test in a series of short bursts. In the evaluation survey attached to the test, many complained of the tiring nature of the task. The fall-off rates in valid responses across a paper-based survey of attitudes to feedback conducted in the first weeks of semester confirmed the

impression that, especially for younger (16-18 year old) students, meaningful engagement with test-structured tasks lasted no more than 10 minutes. This is one third of the intended length of our current summative online tests.

The commonest area of difficulty in Human Biology tests identified by students at all three institutions was terminology (10.5% of all students). Free comments also indicated that considerable difficulty was also being experienced with non-technical language.

*'We were told that the information we needed to know was in the lec notes, yet the way the questions were worded did not reflect this…'*
*'…hard to find the correct statement among all the jargon…'*
*'…it used words i was unfamilar with and it confussed me and i lost cncentration all together…*
*'…just that some multiple choice answers were similar, it kinda confused me. therefore, i took a guess of the two…'*

The student response to the feedback comments was overwhelmingly positive – more than 95% agreed that it helped their understanding of both their correct and incorrect responses, and 65% that the comments helped them understand questions other than those with which they first appeared. Comments included:

*' they are a brilliant idea for the people who want to learn and understand their problem areas. it is a useful aspect of understanding the unit and the feedback for the correct answers make you feel good about yourself too'*
*'My approach to multiple choice tests is usually to first try it, then learn the answers and try again. I like it when there is extra information with the answers because then rather than just memorising the correct answer (whilst sometimes still not understanding much about it) I actually learn the concept and better understand it. '*
*'Found them helpful for the questions I got wrong as I could find out right there and then how I went wrong. Before, I had to try to remember what the question was to then be able to look it up in the book. Please keep them for next year.'*

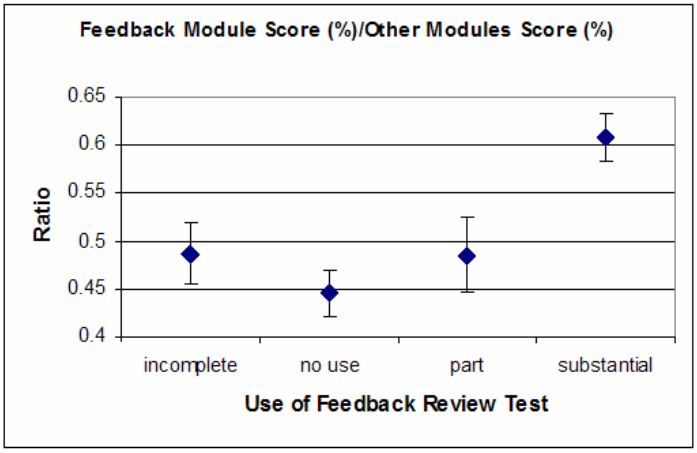


**Figure 3**. End of course performance on feedback-associated topics compared with other topics amongst students with varying degrees of engagement with the online quiz as a formative learning exercise at the university presenting it both formatively and summatively. ANOVA of feedback topic score with nonfeedback topic scores as covariate $F=8.31$, 3&1 df, $p<.001$.

The areas of the course addressed by the quizzes with feedback were those which students had previously experienced most difficulty. While scores for these areas remained relatively low, those students who had fully engaged with the feedback-enriched quizzes gained a clear advantage in the

areas covered by the quizzes in separate end of course paper-based multiple choice examinations. Their marks in these areas were higher than those expected from their marks in other areas and the general relationship of marks in the feedback-related area to those in other parts of the course. The greater their use of the online tests with feedback, the greater was their advantage (Figure 3).

## Summary and conclusions

This study was able to trace the impact of demographic variables from the beginning of semester right through to the end of semester examinations. This was achieved in part by planning, through the inclusion of common demographic items in all surveys evaluating the addition of explanatory feedback comments to online multiple-choice style quizzes and partly by chance, when consistent logon identities allowed the tracing, though not identification, of individuals through different online activities. The insights gained provided guidance for modification of current practice. The necessity to find more flexible ways of presenting summative online assessments to a student body increasingly committed to higher and higher levels of paid work outside university became apparent, as did the need to present shorter tasks to students just beginning the transition from secondary school to university.

The use of consistent, coded identifiers for each student at all stages of this work would have enabled the teasing out of the impact of levels of paid work and of confidence on final performance, and of the extent to which use of the formative tests with feedback may have compensated for their effects. They would, for example, have allowed for follow-up investigation of the reasons why 'metropolitan external' students failed to attend online test sessions, and the linking of initial attitudes towards feedback to use of the online feedback-enriched tests and examination performance. It is our intention to pursue this option.

### Reference

Glover, C. (2004) *Report of research carried out at Sheffield Hallam University for the formative assessment in science teaching project (FAST) for the period 2002-2003.* Sheffield: Hallam University.