THE PROVISION OF AUTOMATED FEEDBACK FOR PRACTICAL AND ESSAY TESTS

TREVOR BARKER, University of Hertfordshire

Abstract The recent National Students Survey showed that feedback to students was an ongoing problem in Higher Education. This paper reports on the extension of our research with objective testing, into the provision of automated feedback for practical and essay questions. Recent research showed that learners and tutors valued our automated feedback approach based on objective tests and Computer Adaptive Testing. The automated feedback system has been extended to include practical testing and essay type questions. This system, which can be used within any subject area, is based on a simple marking scheme created by the subject tutor. As marks are awarded for each question by the teacher an individual feedback file is created automatically for each learner. Teachers may also add and modify comments to each learner and save additional feedback to the database for later use. Each individual feedback file was emailed automatically to learners. The development of the system is explained in the paper and testing and evaluation with 350 first year 120 second year and 100 final year undergraduate Computer Science students is reported. It was found that the time to mark practical and essay type tests was reduced by more than 30% in all cases compared to previous years. More importantly it was possible to provide good quality individual feedback to learners rapidly. In end of module tests it was very beneficial indeed as it had proven difficult to provide feedback in the past after modules have ended. Examples of the feedback provided are presented in the paper. Staff teaching on these modules and a sample of students took part in an evaluation of the system. The results of this evaluation were very positive and are reported in the paper. The provision of fast effective feedback is vital and this system was found to be an important addition to the tools available.

Introduction

Despite the reported benefits of the computer-aided assessment approach, high staff/student ratios often mean that tutors are often unable to provide learners with feedback on assessment performance that is timely and meaningful. Freeman & Lewis (1998) amongst others have reported on the importance of feedback as a motivator for student learning. Thus, there is an increasing demand for the development of software applications that would enable the provision of timely, individual and meaningful feedback to those learners. In our previous work we have reported on the use of automated feedback systems related to Computer Adaptive Testing (CAT) (Barker 2009 & Lilley et al, 2004). We were able to show in this research that the systems we developed had several benefits. It was tested and evaluated by staff and students and shown to be effective and valued by all (Barker, 2009). For feedback to be effective, it is argued, it should be individual for each learner and timely. The use of our feedback system based on objective computer-based adaptive testing was shown to be effective, but limited as it could only be applied to CATs. The CAT applications that were used in our feedback test systems have been reported by Lilley and colleagues Lilley & Barker, 2002, 2003, 2004 & Lilley et al, 2004, 2005). The application comprised a graphical user interface, an adaptive algorithm based on the Three-Parameter Logistic Model from Item Response Theory (Hambleton 1991,

Lord 1980 & Wainer, 2000) and a database of questions. This contained information on each question, such as stem, options, key answer and IRT parameters. Subject experts were employed for question calibration to ensure that the item database was valid. The subject experts used Bloom's taxonomy of cognitive skills (Anderson & Krathwohl, 2001 & Pritchett, 1999) in order to perform the calibration. Questions were first classified according to the cognitive skill being assessed. After this initial classification, questions were then ranked according to difficulty within each cognitive level. Objective adaptive testing and individual feedback based on them are important tools available to teachers. In our system students' abilities in the subject domain and in the cognitive domain according to Bloom's taxonomy were assessed. However it is important that fast and effective feedback be provided for a wider range of tests. The work described in this paper relates to the development, testing and use of an automated marking and feedback system for essay and practical assessment.

Requirements of the system

It was decided to use an iterative prototyping system in order to develop the automated marking and feedback system. The reason for this choice related to the necessary speed of development and also because a user-centred design and evaluation of the system was thought necessary. In complex domains such as teaching and learning, the evaluation of implemented systems by stakeholders at all stage of the development process is absolutely vital as explained by Barker & Barker (2002). Teachers and learners as well as other stake-holders would be expected to have a significant input into the direction the prototype would take. The initial requirements of the proposed system therefore, were arrived at as follows. Based on the evaluation of previous automated feedback system (Barker, 2009) by both staff and students a list of desirable functional requirements for the proposed system was produced. This list was then modified after discussion with colleagues to produce a ***basic*** set of functions for the design of a first stage prototype. The list of ten requirements was selected and is presented in table one below.

Table 1: list of first set of functional requirements

Function	
1	The system should be a computer-based marking system.
2	Simple to install and useful for a range of assessments and assessment types.
3	It should be able to mark both practical and essay type questions.
4	It should provide fast feedback.
5	The list of students and email addresses to be read in from university admin system in order to minimize work for the teacher.
6	Teachers would be able to enter five levels of feedback for each question.
7	General feedback would be allocated for each question based on the mark awarded.
8	The system would collate marks and produce feedback records for

	teachers.
9	Individual feedback and marks for each learner to be saved to a database file.
10	Feedback and marks to be distributed via electronic mail after checking.

The above ten functions shown in table one, were considered to be the *minimum* set necessary in order to develop the first stage prototype. This prototype would be developed and the implementation tested and evaluated in a real context. It was intended then that the results of this evaluation would enable the prod ction of an improved set of requirements based on this experience.

Development of the first prototype

The first prototype was developed using a standard Microsoft event driven programming language. This was decided upon mostly for speed and for ease of installation and testing. The system consisted of three main parts. A feedback file that contained the general feedback for each question, a student file that contained the list of students and their details, provided by the university admin system and a graphical user interface that read in the feedback and student files in order to allocate marks and feedback. The output from the system was a file which contained marks and feedback suitable for distribution via electronic mail. This was achieved by using a simple mail merge application within a Microsoft word processing application that read the file and applied it to a mail merge template developed for this purpose.

Figure one shows the first prototype developed in this study.

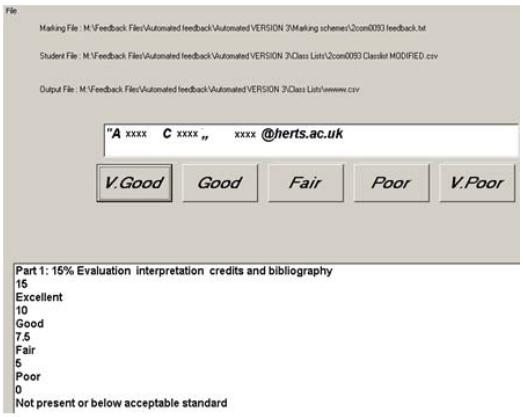

Figure 1: first version of the automated feedback and marking system

The prototype shown in figure one above was used to mark a summative practical test for a set of approximately 350 first year computer science students. This test was taken under supervised conditions in a computer laboratory. Practical work was uploaded to the University's managed learning environment (MLE) for marking. The test consisted of 16 questions. The text box below the buttons presents performance indicators for the mark. The buttons are used to allocate feedback for each question and the actual mark awarded for each section was entered

by the marker after the appropriate button had been selected. An example of the feedback provided for each mark range is shown in table 2 below.

Format of the feedback

The feedback file created for use in the prototype was developed based upon a marking scheme for the assignment. For each question in the assignment five general feedback statements were written for excellent, good, fair, poor and absent or below acceptable standards. After discussion between markers these were manually written into file which suitable for reading by the prototype. Table two below shows an example of the feedback provided for one question

Table 2: example of feedback range provided for one question

Part 4: 10% Sequence of still photographs
10: Excellent sequence of stills with excellent subject matter and high quality images
7.5: Good sequence of stills with good subject matter and good quality images
5: Fair sequence of stills with some issues with either subject matter or the quality of images
2.5: Poor sequence of stills with considerable issues with either subject matter or the quality of images
0: Sequence of stills was not present or below acceptable standard

The number in each section is a guide to the marks relating to each of the feedback comments.

Use of the prototype

The first stage prototype shown in figure one above was used to mark approximately 350 practical assignments over a one week period. After the marking was completed marks were transferred to a spreadsheet in order to check that no errors had been made in marking, markers were consistent and that the mean and other statistical measures for the test were similar to other tests on the module. A sample of marked work and feedback was then passed to an external marker to be moderated and his comments were received for later analysis.

Once the course team was satisfied that the test had been marked fairly and accurately, marks and feedback were released via electronic mail to individual learners. In previous years it had proven difficult to achieve this timescale with smaller groups of approximately 250 students within a six week period. On this occasion we were able to release the marks three weeks after the end of the assignment for a group of 350 learners. Markers reported that the marking itself was faster and more efficient, taking approximately 30% less time to mark the work than previously. The greatest saving on time was related to writing and distributing feedback. On some occasions in the past it had not been possible to deliver feedback until after the end on the course itself and on one occasion feedback was not delivered at all since students were on their summer vacation by the time feedback was ready for distribution.

Evaluation of the first prototype

Approximately one week after the marks and feedback had been distributed, markers met to discuss and reflect upon the exercise. Comment from the external marker were also distributed and considered. Fifteen students were selected quasi-randomly to answer a short questionnaire. Selection was based on their scores obtained in the test. It was important that students with a range of scores in the test had an opportunity to comment on the feedback provided by the system, so five students were selected in each on the performance ranges, under 50, between 50 and 75 and above 75 marks. Table three below presents a summary of their responses to the questionnaire.

Table 3: learners' responses to questionnaire and score achieved in the test.

Likert Scale (1 to 5) 1=disagree, 3=neither agree or disagree, 5 = agree	Student responses (n=15)		
Statement	score > 75	score 50-75	score < 50
Feedback was useful to me	4.0	3.8	3.2
Feedback was fast	4.8	4.6	4.6
Feedback was delivered conveniently	4.8	4.2	4.4
Feedback was fair	4.4	3.8	3.8
The amount of feedback was good	4.2	4.0	4.2

Table three suggests that students felt that the feedback was in general timely, fair and useful and delivered in an acceptable manner. There did appear to be a correlation between attitude to feedback and score achieved on the test, though in general attitude was good for all groups. Markers considered that the marking and feedback system was a good idea and a valuable tool to help in the rapid delivery of marks and feedback. All agreed that it had operated flawlessly and that marking time had been reduced considerably. In all only 9 of the 350 students reported any problems with the marks awarded to them, none of which related to the performance of the prototype itself.

Several issues however were raised by markers and the external moderator relating to the feedback and functions available in first stage prototype. Perhaps the most important related to the flexibility of the feedback provided. Although markers considered it extremely useful it was considered to be very inflexible. Feedback comments of the type shown in table two above were considered to be too general and inflexible by the markers. For example in table two the feedback statement *"Excellent sequence of stills with excellent subject matter and high quality images*" relates to image quality and subject matter. It would be an improvement to separate this into two sections and provide a mark and appropriate feedback for each

Markers also wanted to add their own feedback comments on the assignment related to aspects of the performance overall. It was also considered useful to include some feedback related to

the completeness of the work handed in. These two features were added to the list of functional requirements presented in table one to be used in the development of the next stage prototype. In order to make the feedback more relevant, marking schemes were re-written in such a way that feedback could be related more specifically to performance. This was achieved by breaking each of the questions into smaller parts and writing the marking scheme and feedback comments to reflect this more closely. In this way a larger amount of more directed feedback could be written relating to each section of a question.

Another suggested improvement was the replacement of the rather inefficient way in which marks and feedback were allocated using the buttons shown in figure one. It was suggested that marks be awarded each section of a question and that feedback would be presented based on the mark awarded. In the previous system feedback was awarded by selecting the appropriate button and the mark entered later. This modification would do away with the buttons altogether and make the use of the system more efficient. It was also decided to produce additional introduction and summary screens to show in the first place the submission requirements for each assignment and also a final screen summarizing the marks and feedback for each learner.

Development of the second prototype

The modifications were made to the system as outlined in section 4 above and a second stage prototype of the system was developed and tested prior to use with students on an assignment. The modified version of the system allowed markers to comment on the completeness of the hand-in for the assignment as shown in figure two. In this version, the hand-in information is presented to the marker who may then make additional comments on the completeness or nature of the hand-in.

Figure 2: modified version of the system, showing hand-in information and tutor-entered feedback.

Each of the questions and question parts were presented by the system and the marker was simply required to enter a mark. Appropriate automated feedback was determined by the system based on the mark awarded in each section of a question, reading it from the database file for the assignment. After all the question sections had been marked, the system presented a final summary screen so that the marker could check that the marks had been awarded accurately. If

this were not the case, the marker could cancel the entry and mark the student again. An example of the summary screen is shown in figure three below.

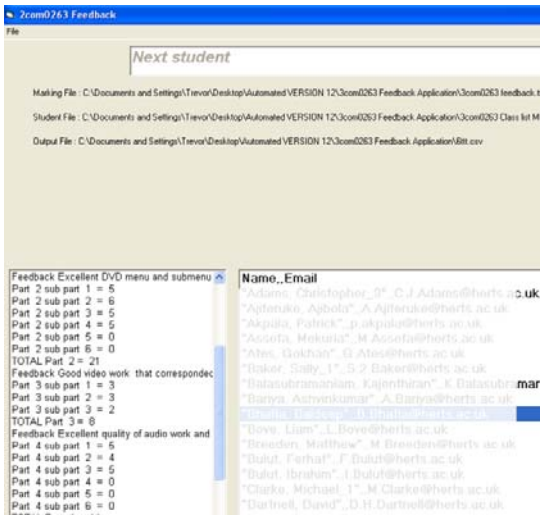

Figure 3: Modified version of the prototype system, showing final summary screen

Evaluation of the second prototype

The second prototype was used to mark three assignments, a second year practical summative assessment for 125 students and two final year summative assessments for approximately 100 students. These assignments were complex in that they contained both practical and theoretical elements and were completed over a six week period. In the past it had been extremely difficult to achieve consistent marking in this type of assignment. They were slow to mark and feedback delivered in terms of quality and quantity was fairly inconsistent between markers.
After the assignments had been marked, moderated and subject to quality measures, marks and feedback were released as before. In this case, a sample of 15 students only from the second year module was selected using the same sampling method as before in order to test student attitude to the feedback using the same questionnaire as before. The results of this are shown in table four below.

Table 4: second year learners' responses to questionnaire and scores achieved in the test.

(Likert Scale 1 to 5) 1=disagree, 3=neither agree or disagree, 5 = agree	Student responses (n=15)		
Statement	score > 75	score 50-75	score < 50
Feedback was useful to me	4.8	4.2	4.0
Feedback was fast	4.8	5	4.6
Feedback was delivered conveniently	5	4.8	4.8

Feedback was fair	4.6	4.4	4.4
The amount of feedback was good	4.8	4.6	4.4

As before, the attitude of learners to the feedback was positive. It was not possible to perform a statistical analysis on these results due to the sample size. However cautious comparison of the results suggests that learners were slightly more satisfied with modified prototype than with the earlier version.

Tutors were also invited to discuss their experiences of using the second prototype system and these were taken along with comments from the external marker. It was generally agreed that the system was improved by the modifications in terms of efficiency of marking and the quality of feedback presented to learners. As before the system performed faultlessly and there were no problems with installation or the automated distribution of feedback via electronic mail. One issue that surface is that although learners were in general satisfied with the feedback awarded, there was a greater number of students prepared to challenge their mark. It is suggested that this was due to the more detailed and specific marking scheme provided along with more detailed feedback relating specifically to each question part. Errors in marking made by markers were more readily identified by students and these were naturally more likely to be questioned. This issue was considered to be a positive feature by markers of the system, leading to greater accuracy and fairness of marking. The tutor-entered feedback on the hand-in and the general comments were considered to be a good feature of the modified version. The external marker considered the system to be extremely useful and made several highly supportive comments related to the system. The quality and quantity of feedback provided was considered to be much batter in the second prototype than in the first.

Suggested improvements

Based on the findings from the evaluation of the second stage prototype it was decided to produce a third version with new modifications suggested by the markers and external moderator. It was suggested that tutors should be able to add to or modify the automated feedback at each stage of marking as well as adding general comments at the end. It was further suggested that these additional comments be saved to a database file so that they might be used again later. In addition an option to show an image of the student was suggested, although this suggestion did not receive universal support from all present as some favoured more anonymous marking.

Development of the third prototype

In order to allow for the comments from markers and the external moderator a third stage prototype is currently being developed. In this version several improvements are being made.
A more robust and secure database system is being employed which will not only store the automated feedback for each question section but will also store additional feedback comments that are added to each question to supplement the automated feedback. A list of these is then available when the question is again marked for another learner and can be added to the feedback with a single click.

Developing a high quality feedback database is an important part of the system and at present this is a time consuming and demanding task. In order to simplify this task, the system is being modified in order to make it simpler for the tutor to enter their marking scheme and feedback statements and create a feedback file that can be used by the system. At present the teacher

must create a text file containing a great deal of meta-information relating to the format of the questions along with the feedback for each option. In the new system the tutor need only enter the marking scheme and the feedback. The feedback database file is then formatted by the system.

An option to display an image of each student has also been added, although this may be switched on or off. It was also suggested that it be possible to modify marks without needing to re-enter the complete set of marks for a student and this feature was added. Finally the feedback distribution has been improved so that it is no longer necessary to use a word processor to distribute the feedback via electronic mail merge. This function is now achieved from within the application directly. Figure four shows an example screen of the latest prototype. In this version the user interface has been improved. Teachers are able to set up assessments and feedback much more easily.

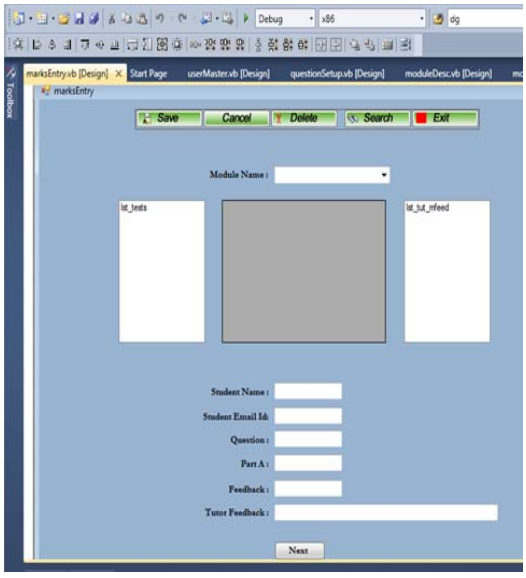

Figure 4: The latest version of the prototype showing the modified user interface.

Discussion

The research described here has presented the development, testing and evaluation of an automated feedback and marking system. The system was shown to be efficient and useful to both students and staff using the system. The evaluation of the system by staff and students suggests that the feedback quality was good and was delivered quickly and effectively using electronic mail. This was a vast improvement when compared to the manual methods used previously. The quality of feedback is vital and this was improved through modifications to the system, using the best features of automation and also by allowing tutors to add and save their own comments for later use. The use of a user-centred iterative prototyping approach involving staff and students was vital to the development of the system. It was important that the system developer was able to understand the detailed requirements and functions of the system based upon the thoughts and opinions of a range of users. In this way the system was more likely to be accepted by colleagues and external examiners and more likely to be beneficial to all. Currently the third stage prototype is being tested prior to use with learners.

It is hoped that the next version will, in addition to practical and mixed practical/theory examinations, be used in a pure essay type theory test, where it is expected to be especially beneficial. It is hoped that in later stage prototypes feedback comments will be inserted directly into the documents being assessed at an appropriate place in the text. This idea is currently under consideration for prototype version four.

References

Anderson, L.W. & Krathwohl, D.R. (2001) *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Longman, New York.

Barker, T. (2009) An automated feedback system based on adaptive testing: extending the model, Proceedings of the Interactive Computer Aided Learning Conference, ICL2009, 23 - 25 September 2009, Villach/Austria

Barker, T. & Barker, J. (2002) "The evaluation of complex, intelligent, interactive, individualised human-computer interfaces: What do we mean by reliability and validity?", *Proceedings of the European Learning Styles Information Network Conference*, University of Ghent, June 2002.

Freeman, R. & Lewis, R. (1998) *Planning and Implementing Assessment*, Kogan Page, London.

Hambleton, R.K. (1991) *Fundamentals of Item Response Theory*, Sage Publications Inc, California.

Lilley, M. & Barker, T. (2002) "The Development and Evaluation of a Computer-Adaptive Testing Application for English Language", *Proceedings of the 6th Computer-Assisted Assessment Conference*, Loughborough University, United Kingdom, pp. 169-184.

Lilley, M. & Barker, T. (2003) "Comparison between Computer-Adaptive Testing and other assessment methods: An empirical study", *Proceedings of the 10th International Conference of the Association for Learning Technology (ALT-C)*, University of Sheffield, United Kingdom.

Lilley, M. & Barker, T. (2004). "A Computer-Adaptive Test that facilitates the modification of previously entered responses: An empirical study", Proceedings of the 2004 Intelligent Tutoring Systems Conference, *Lecture Notes in Computer Science 3220*, pp. 22-33.

Lilley, M., Barker, T. & Britton, C. (2004) "The development and evaluation of a software prototype for computer adaptive testing", *Computers & Education Journal 43(1-2)*, pp. 109-123.

Lilley, M., Barker, T. & Britton, C. (2005) "The generation of automated learner feedback based on individual proficiency levels", Proceedings of the 18th International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems, *Lecture Notes in Artificial Intelligence 3533*, pp. 842-844.

Lord, F.M. (1980) *Applications of Item Response Theory to practical testing problems*. Lawrence Erlbaum Associates, New Jersey.

Pritchett, N. (1999) "Effective question design" In S. Brown, P. Race & J. Bull (Eds.), *Computer-Assisted Assessment in Higher Education*, Kogan Page, London.

Wainer, H. (2000) *Computerized Adaptive Testing (A Primer)*, Lawrence Erlbaum Associates, New Jersey.

Biography

Trevor Barker is a principal lecturer and researcher in Computer Science at the University of Hertfordshire. He obtained his PhD for research into personalized adaptive multimedia systems. He is a National Teaching Fellow and Head of Systems and Software Research Group at the University. His research interests include adaptive educational software development, student modeling and three dimensional virtual reality system development and evaluation.