CORE
Metadata, citation and similar papers at core.ac.uk
Provided by University of Hertfordshire Research Archive

# Marking complex assignments using peer assessment with an electronic voting system and an automated feedback tool

**Trevor Barker and Steve Bennett, University of Hertfordshire**

## Abstract
*The work described in this paper relates to the development and use of a range of initiatives in order to mark complex masters' level assignments related to the development of computer web applications. In the past such assignments have proven difficult to mark since they assess a range of skills including programming, human computer interaction and design. Based on the experience of several years marking such assignments, the module delivery team decided to adopt an approach whereby the students marked each other's practical work using an electronic voting system (EVS). The results of this are presented in the paper along with statistical comparison with the tutors' marking, providing evidence for the efficacy of the approach. The second part of the assignment related to theory and documentation. This was marked by the tutors using an automated feedback tool. It was found that the time to mark the work 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. Feedback was delivered to all within three weeks of the test submission date.*

## Keywords
Electronic voting system (EVS); peer assessment; automated feedback.

## Introduction
It has proven extremely difficult in the past to assess assignments on the Multimedia Specification Design and Production masters' level module. Assignments on this module assess a range of skills including computer programming, human computer interaction and visual screen design. The module has four summative assignments. The first is an individual online multiple-choice test covering the principles of multimedia design. The work leading up to the first assignment is intended to prepare learners for the latter three assignments. The 55 student on the module then split into 22 groups of between two and four persons. The second group assignment relates to the development of a prototype Flash website in which students, in groups, produce a minimal content software prototype: essentially the basic structure of the website together with an animation to promote it. This is submitted along with documentation relating to the website's information architecture, the goals and mission of the site, grouping and labelling of content, tree structure diagrams and design ideas for the visual appearance of the site. The third assignment relates to students' evaluations of other groups' websites and finally, in the fourth assignment, the groups reform and redevelop their website based on the feedback obtained both from the tutors and from the other students on the module. The feedback they receive from their work on the second assignment (the minimal content website) is therefore extremely important as this is used to guide the development of the full content website for assignment four.

In order to provide a flexible approach to marking and feedback, it was decided to use a combination of peer assessment using an electronic voting system (EVS) in class to mark the practical component of assignment two. The theoretical component and documentation would be marked using an automated feedback system developed at the University of Hertfordshire, based on previous research (Barker 2009; Lilley et al. 2005). The learning outcomes for assignments on computer science modules such as this one are complex, as the assignments assess many diverse programming, computational, design and other skills. For such assignments it is important to develop flexible approaches to marking and the provision of feedback. This is a challenge not only for the markers, but also for students who must understand and make use of the feedback provided. It was hoped that the approach adopted here would go some way to achieving this aim. In the following section we present a brief overview of approaches to the delivery feedback to learners. We then go on, in later sections of this paper, to describe our use of peer assessment and how we used an EVS in order to attempt to improve the quality of engagement and feedback to learners. We also describe briefly the development and use of an automated feedback system which was intended to improve the speed, efficiency and effectiveness of feedback to learners.

## Automated feedback

Providing feedback that is detailed, useful and timely is often difficult in Higher Education (HE) today (Barker 2009). The results of the recent national students survey (NSS) revealed that a major problem in HE today is that of assessment and student feedback. For several years we have been involved in research to develop methods of providing fast and effective feedback related to objective testing (Barker 2009; Lilley et al. 2005). The results of this work have been encouraging. Evaluation of the software with more than 500 students and 100 university staff has shown that the systems we have developed are highly valued by learners and are seen by staff as an important addition to the methods available for providing feedback. Indeed it was seen by many to be better than the traditional methods used in the university, which are often too slow and too general to be of use to learners. Despite the reported benefits of the computer-aided assessment approach, high staff/student ratios mean that tutors are often unable to provide learners with feedback on assessment performance that is timely and meaningful. Freeman and Lewis (1998) amongst others have reported on the importance of feedback as a motivator for student learning.

Sadler (2010) stresses the importance of effectiveness and efficiency of feedback. According to Sadler, the type of feedback provided within available resources ’especially time’ is also a concern. Feedback should also be individual to be effective.

Given Sadler’s concerns regarding resources, our approach was to look towards automating the feedback process. Thus, we decided to develop a software application that would enable the provision of timely, individual and meaningful feedback to those learners who are assessed via computer-aided assessment applications. Objective testing is important in HE today and providing good feedback for such tests is essential. Objective tests lend themselves to automated approaches to providing feedback (Lord 1980; Pritchett 1999; Wainer 2000). It is also important to provide feedback for other forms of assessment. Our work on adaptive and automated approaches to feedback provision for objective tests has been in use for

several years at our university (Lilley and Barker 2002; Lilley and Barker 2003; Lilley and Barker 2004; Lilley et al. 2004). Recently we have extended this work to include automated approaches to the provision of feedback for general assignments.

Peer assessment

It has been claimed that peer assessment is an important method to engage and motivate students. Van Berg (2006) suggests that peer assessment not only engages learners but also leads to improvements in written work and interaction. Other advantages have been described by Zariski (1996) and Race (1998), including ownership of assessment, taking responsibility for learning and reflective skills useful in lifelong learning and depth of engagement. Rust (2001) suggests that peer assessment is valuable for students who are able to develop judgment skills as well as saving time for tutors. Li (2001) suggests that peer assessment is an effective way of grading individual contributions to group working. Li also suggested that it was possible that unfair marking would be likely to skew the grades awarded by peers. It would therefore be important to make sure that if this approach were to be used in a summative assignment, that peer assessment was a fair and valid method in the context of this module.

Electronic voting systems (EVS)

Electronic voting systems (EVS) have the potential to enhance learning and motivation, as well as providing variety and engagement within lectures according to Simpson and Oliver (2006). Often they are used in multiple-choice tests or interactive quizzes as a means of enhancing learning and teaching (Russell 2008) and as a means of enhancing the opportunity for deeper learning (Draper 2009). In this study it was decided to use a combination of EVS and peer assessment to mark the practical element of assignment two, the minimum content website application.

**Practical assessment with EVS**

The practical part of assignment two, the students' content free website was marked in class by tutors and students using an EVS. A component of the final (summative) mark was awarded by peer assessment. Students were required to evaluate the quality of the website generally on a scale of 1 to 5 and to rate the quality of the animations within the web pages on a scale of 1 to 5. In order to achieve this, each group was required to present their work to the class to be marked by their colleagues and by the module delivery team. The score they obtained from that would be the result of 40% each from the two tutors, and 20% as an average from the student cohort themselves. The award of 20% of the mark by peer assessment was considered to be a reasonable measure, yet still permitted learners to have a significant contribution to the grade achieved on the assignment. The arbitrary figure of 20% for the student contribution to the mark was arrived at after discussion between the tutors, based on their experience of assessing this module over several years.

There were some concerns expressed between the tutors that there might be a possibility of tactical marking for friends etc. Therefore, in order to promote 'honest' marking, we also offered twenty bonuses of 5% for their assignment score as a whole, for those 20 students whose marking pattern was nearest to the average marks of the tutors.

Ten days before the marking event, students were given a rehearsal session where a selection of work submitted in previous instances of the course were marked and

commented upon. In this session the websites were displayed on a screen and discussed, marked, and then discussed, explaining the marking criteria and exactly why the marks had been awarded in the way they were. This helped us rehearse using the EVS software, it helped the students more particularly understand how they would be marked, but equally, it also helped the students to have a greater understanding of the marking criteria that would be applied to their website. For the marking session itself the logistics themselves were the most complicated thing to resolve. The course has approximately 57 students formed into 22 groups. Allowing for approximately ten minutes per group presentation meant that the whole marking experience would take approximately four hours. It would have been impractical to ask the students to sit through every presentation so we created two sessions and stated that students had to mark the work of at least ten other groups in order to be eligible for the 5% bonus. This meant that students definitely had to attend the session where their group was presenting but not the whole four hours. The room itself had two projector screens which was absolutely essential. On the first screen students demonstrated their work and on the second the peer-assessed scores for the students' work was presented. After each presentation, lasting approximately five minutes the students had to give grades based on the following questions shown in table 1 below.

*Table 1: Marking criteria for the minimum content website*

| **Category** | **Percentage** | **Criteria** |
|---|---|---|
| The site generally | 25 | Level to which the navigation system is implemented and works. Clarity of signposting: what is implemented and what isn't |
| Think about the (a) the idea behind the site (b) how well the prototype is implemented (c) the expected content:<br>5) Great idea behind the site, everything works and looks of professional standard, the sorts and amount of content that it will contain is just right.<br>4) Good idea behind the site, everything looks good, the kind of content is right.<br>3) Not bad, shows promise. Mostly good, might lack something either in the idea behind the site (perhaps too general, or too obvious), the implementation (some buttons juddering etc), or lack of content (not enough screens).<br>2) Some bugs or very ugly pages, juddering, maybe the idea behind the site is not good enough, might have too little content.<br>1) Really poor, very many bugs, loads of things not working. | | |
| Animation | 25 | Level to which the structure as envisaged in the design documents is implemented. |
| Think about the animation particularly (a) how good it looks and (b) how much effort or skill was involved in making it:<br>5) Animation is really beautiful, and obviously uses some really clever techniques.<br>4) Animation is very good.<br>3) Animation is average – either not enough content, or not nice enough content.<br>2) Animation is poor – only very simple Flash used.<br>1) Animation is extremely poor – looks like a complete Flash beginner did this. | | |

The session was conducted as follows. Each group would present their work on the screen and talk about it. Tutors would ask questions. The first criterion (the website generally) would be marked, then after 30 seconds the result shown on the marking screen, then the second criterion (animation) would be marked, and then the results for that criterion displayed on screen. Figures 1 and 2 below provide an example of the screens shown to the students after each presentation, with the marking criteria. The bar charts shown appear after the marks have been awarded by the students.

The tutors' marks do not appear on the screen. The marking criteria are clearly presented to the students prior to marks being awarded.

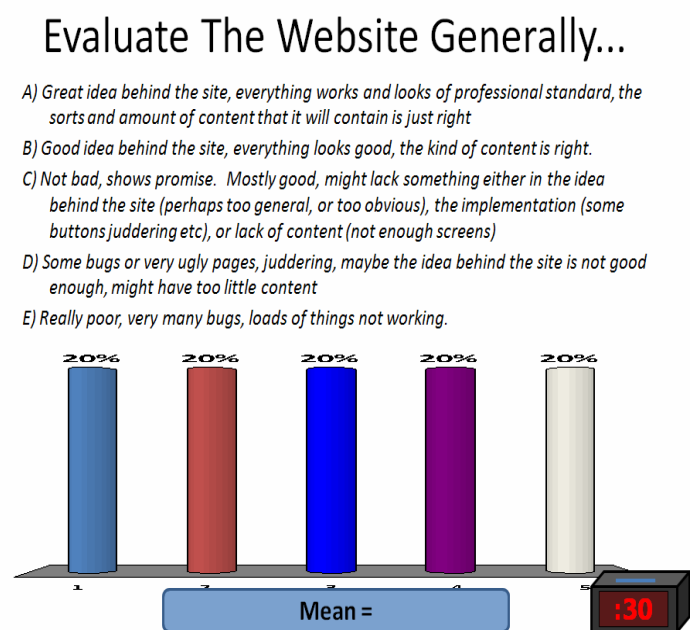


*Figure 1: Marking criteria displayed for assessing the website generally*

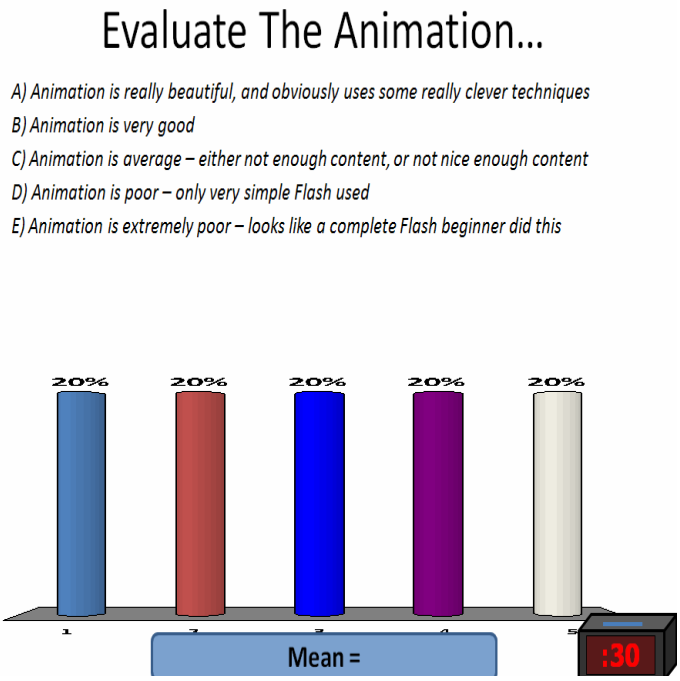


*Figure 2: Marking criteria displayed for assessing the animation*

During the session, tutors made notes and asked questions. The use of questions was important not only to clarify points related to the practical work of a group, but also as a way of providing additional and specific feedback related to aspects of a group's presentation. It was important that the tutors' questions were asked in a non-judgemental way in order not to bias the students' marking. It should be noted that the actual mark awarded to the students' artefacts in this assignment has in the past

been entirely performed by the module delivery team with the students not present to explain their work. It has also proven difficult in the past to deliver feedback on the written component of the assignment sufficiently quickly in order to inform the development of the later stages of learners' project work. In order to improve this situation it was decided to use an automated approach to improve the speed and efficiency of feedback to learners. It was hoped also that the quality of feedback delivered by the systems would be high.

**The automated feedback tool**

The automated feedback tool 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 from the tutor and deliver feedback. The system was developed using a prototyping system development method. Details of the development of the system are described by Barker (2010). The output from the system was a database 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.

A few days after the EVS session, one of the tutors downloaded the documentation related to the practical work. A detailed marking scheme was prepared containing criteria for awarding marks and also feedback comments related to the marks awarded in each section:

- goal and mission of site; audience research and scenarios;
- suggested grouping and labelling of content;
- tree structure diagram with explanations, navigation principles;
- design ideas for the visual appearance of the site.

An example of the automated feedback comments for 'goal and mission of site; audience research and scenarios' is shown in table 2 below.

*Table 2: Example of feedback comments related to marks awarded for one section of the documentation*

| ***Mark awarded*** | ***Feedback comment*** |
|---|---|
| 5 | The goal and mission is superbly clear, unique and specific. Scenarios describe in depth credible potential users with credible search strategies. Analysis of competitor sites is deep. This will be of industrial standard. |
| 4 | Very good goal and mission. Good scenarios and good analyses of competitor sites. This will be high standard. |
| 3 | Unimaginative but credible goal and mission. Scenarios a bit generic but with some local colour. Analysis of competitor sites not great but generally correct. |
| 2 | Uninteresting goal and mission. No detail in the scenarios – or ones taken from a very post facto point of view. Analysis of competitor sites done only cursorily. |
| 1 | Uninteresting goal and mission. Scenario not credible or not there. Analysis of other sites purely descriptive without analysis |

The marking scheme was rapidly converted by the tutor into a format that could be read by the feedback tool and used for marking. During the marking sessions, the tutor was able to add and modify comments related to the completeness of the submission and also make general feedback comments to add to or modify the automated feedback provided by the system. After marking, feedback reports were prepared by the feedback tool for distribution after checking to ensure there were no errors. Feedback documents were then distributed to students automatically either via electronic mail using a template specifically designed for the purpose or handed out in class.

**Results of peer marking**

Immediately after the end of the marking session for the practical work, marks were collected and collated. A comparison was made between the average marks awarded by the tutors and the average marks awarded by the students for each of the group submissions for the general website quality and for the animation. A Pearson's correlation was performed on the data summarized in table 3 using the SPSS software package to test the significance of any relationship between the marks awarded by the students and the tutors for each group. It was interesting to note that there was a significant relationship between the students' and tutors' marks in all but three of the group presentations. It was noted anecdotally that the tutors were more likely to give higher marks and also occasionally lower marks than the students. It is suggested that the tutors were more likely to award higher and lower marks as they were more confident in using the marking criteria than were students.

*Table 3: Marks awarded by groups and tutors in the EVS peer assessment session*

| Group * | Mark % mark achieved for website | Mark % mark achieved for animation | Average % mark overall | Average mark awarded by students for website | Tutor mark awarded by tutor for website | Average mark awarded by students for animation | Tutor mark awarded by tutor for animation | Pearson's correlation between students' and tutors' marks for each group (df=43) |
|---|---|---|---|---|---|---|---|---|
| 1 | 61 | 44 | 52.5 | 3.26 | 3.00 | 2.91 | 2.00 | 0.51 * |
| 2 | 71 | 71 | 71.0 | 3.69 | 3.50 | 3.69 | 3.50 | 0.57 * |
| 3 | 88 | 79 | 83.5 | 4.00 | 4.50 | 3.96 | 4.00 | 0.63 * |
| 4 | 79 | 61 | 70.0 | 3.63 | 4.00 | 3.26 | 3.00 | 0.57 * |
| 5 | 62 | 88 | 75.0 | 3.56 | 3.00 | 4.08 | 4.00 | 0.39 * |
| 6 | 71 | 53 | 62.0 | 3.79 | 3.50 | 3.35 | 2.50 | 0.01 |
| 7 | 55 | 61 | 58.0 | 3.64 | 2.50 | 3.25 | 3.00 | 0.28 |
| 8 | 98 | 98 | 98.0 | 4.45 | 5.00 | 4.91 | 5.00 | 0.74 * |
| 9 | 52 | 43 | 47.5 | 2.95 | 2.50 | 2.79 | 2.00 | 0.54 * |
| 10 | 42 | 30 | 36.0 | 2.43 | 2.00 | 1.45 | 1.50 | 0.32 ** |
| 11 | 69 | 60 | 64.5 | 3.24 | 3.50 | 2.88 | 3.00 | 0.51 * |
| 12 | 70 | 70 | 70.0 | 3.41 | 3.50 | 3.39 | 3.50 | 0.69 * |
| 13 | 88 | 78 | 83.0 | 3.88 | 4.50 | 3.38 | 4.00 | 0.30 ** |
| 14 | 62 | 70 | 66.0 | 3.44 | 3.00 | 3.59 | 3.50 | 0.15 |
| 15 | 53 | 51 | 52.0 | 3.31 | 2.50 | 2.82 | 2.50 | 0.85 * |
| 16 | 96 | 88 | 92.0 | 3.91 | 5.00 | 3.91 | 4.50 | 0.02 |
| 17 | 61 | 43 | 52.0 | 3.32 | 3.00 | 2.73 | 2.00 | 0.46 * |
| 18 | 79 | 71 | 75.0 | 3.74 | 4.00 | 3.65 | 3.50 | 0.31 ** |
| 20 | 52 | 41 | 46.5 | 2.93 | 2.50 | 2.37 | 2.00 | 0.65 * |
| 22 | 60 | 80 | 70.0 | 3.12 | 3.00 | 3.88 | 4.00 | 0.67 * |
| 23 | 70 | 80 | 75.0 | 3.48 | 3.50 | 4.00 | 4.00 | 0.56 * |
| 24 | 63 | 71 | 67.0 | 3.86 | 3.00 | 3.79 | 3.50 | 0.47 * |

*There was no group 19 or 21. (Significance * = $p<0.01$; ** = $p<0.05$)

The Pearson's correlations between students' and tutors' marks for each group shown in table 3 above were interesting and important. This result suggested that for the most part learners were using similar criteria to teachers when assessing the work of their peers and they were agreeing with the tutors' marks. There were only four exceptions to this rule, most of the relationships being significant at the ($p<0.05$) level or better. In addition to the individual group correlations shown in table 3, Pearson's correlations were performed on the average marks awarded by all the student groups and the tutors' marks both for the general quality of the website and the animation shown in table 3. Figure 3 shows the relationship between the tutors' marks and the students' marks for the quality of the website.

A Pearson's correlation was performed to test the significance of any relationship between the data summarized in figure 3. A significant relationship was identified ($R=0.78$; df 21; $p<0.01$). A paired T test was also performed to test the significance of any difference between the mean marks awarded by the students and the tutors. There was no significant difference found between the means ($p=0.19$).

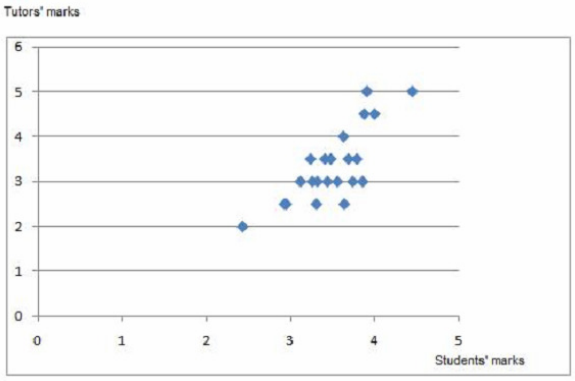


*Figure 3: Scattergram showing the relationship between the student group marks and tutors' marks awarded to each of the groups' presentations for the website qualit*

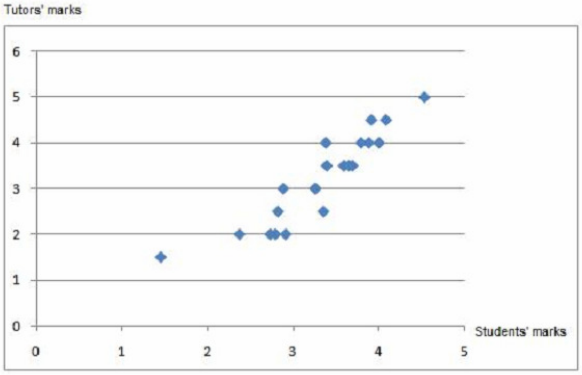


*Figure 4: Scattergram showing the relationship between the student group marks and tutors' marks awarded to each of the groups' presentations for the quality of the animation*

Figure 4 above shows the relationship between the tutors' marks and the students' marks for the quality of the animation.

A Pearson's correlation was performed to test the significance of any relationship between the data summarized in figure 4 ($R=0.91$; df 21; $p<0.01$). A paired T test was also performed to test the significance of any difference between the mean

marks awarded by the students and the tutors. There was no significant difference found between the means of the marks awarded by tutors and students (p=0.25). These two results show that there was a significant relationship between the marks awarded by tutors for both the quality of the website and the quality of the animation as shown by the correlation results. The T test showed that there was no significant difference between the mean marks awarded by tutors and learners for both the quality of the website and the quality of the animations.

It was also decided to investigate the significance of any relationship between the scores awarded by *individual* students in the practical assessment when they assessed their peers and the closeness of their mark to the tutors' marks. The closeness of their mark to the tutors' marks was measured by a Pearson's correlation of the students' *individual* marks awarded and the tutors' marks. It was hypothesized that high scoring students would have a clearer idea of the marking criteria than those performing less well and would be more likely to mark closely to the tutors' mark. Figure 5 below shows the relationship between the marks awarded by tutors and the closeness of an individual student's mark to the tutors' mark.

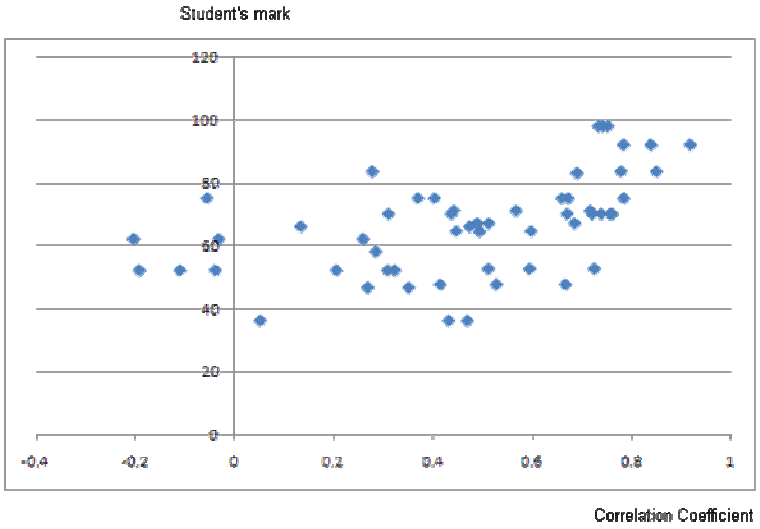


*Figure 5: The relationship between the marks a student obtained and the closeness of their marks to the tutors' marks*

The results of the correlation showed that there was indeed a significant relationship between score obtained and the tutors' marks awarded to the groups ($r = 0.52$; df. 54; $p < 0.01$), suggesting that those performing well were more likely to mark closely to the tutors than were those performing less well.

**Discussion**

The tutors involved in marking the assignment were convinced that the peer marking exercise was beneficial both to the learners and to the tutors themselves. It was decided to use the method to mark assignment four, the full content website, as it had been such a success. Reflection on the process by tutors brought up one or two interesting points. Tutors were aware that students were attempting to match their mark closely to the tutors' own marks. This had an effect on the tutors' marking it was felt. It was felt by both markers that tutors were less likely to give very high or very low marks as compared to previous years. Both expressed this quite independently. It was also interesting to note that on no occasion was there a

difference of more than one point between the marks awarded by the two tutors to the groups. A Pearson's correlation showed a very strong significant correlation between the marks awarded by each of the two tutors for the presentations ($r=0.97$; df. 43; $p<0.001$). The scattergrams displayed in figures 3 and 4 show very clearly the close relationship between the average marks awarded by the tutors and the students. This is an interesting result and suggests that on average peer assessment is a fair and valid means of involving learners in assessing their work.

A slight concern however is the significant correlation between marks awarded by individual learners and the tutors' marking. Those performing less well have less of an idea of appropriate marking criteria. Indeed some students exhibited slight negative correlations (not significant), suggesting that they valued precisely those aspects of the website applications that the tutors did not. This was despite the dedication of a trial marking session to the process and also a significant amount of theory time discussing the assignment and the marking criteria. The marking scheme was available to learners prior to the assessment and was discussed fully with them. It was also refreshing and no small relief for the two tutors to note the closeness of their marking to each other's. One important issue for the tutors was the granularity of the criteria presented in table 1 above. The criteria had multiple components which rather complicated the marking for students to some extent. It would have been possible to make the criteria clearer for learners, more discriminatory and therefore more detailed by increasing the number of categories. After some considerable discussion between tutors it was decided that this was likely to be time consuming and self-defeating and that a better alternative was to reduce the number of categories by combining similar criteria. It was hoped that the question and discussion sessions would clarify any uncertainties resulting from this decision. The results of the study suggest that this was a reasonable assumption on our part.

Our concerns initially were that learners would be inclined to mark tactically in some way, either awarding more marks for friends or less marks for enemies. This was not upheld by the results. The main problem was that some learners did not completely share the same marking criteria as the tutors. However, the exercise in itself provided a fantastic opportunity for learners to see the work of others, to comment on it and to hear the comments and questions of tutors. This was a very positive outcome of the exercise. We believe that the peer assessment was a good exercise and was rewarding for all involved, tutors as well as learners. Not least, it is important to note that feedback on practical work was fast and direct. In previous years this had not been the case.

The use of peer assessment coupled with the directness of EVS was seen as an important step in helping students to improve their performance. Sadler (1989) suggests that the capacity to monitor the quality of a learner's own work during actual production is important in skills development. In this way, students are able to appreciate the quality of their work in comparison to that of others. The directness of our feedback approach with peer evaluation and EVS we suggest goes some way to addressing Sadler's recommendations.

The automated feedback tool was also important and useful in this exercise. The system was shown to be fast and efficient and useful to both students and staff using the system. Marking of the documentation was completed more than 30% faster than in previous years. Feedback was delivered to learners within a few days

of submission. The evaluation of the system by the module delivery team and the external moderator suggested that the feedback quality was good and was delivered quickly and effectively. This was an improvement when compared to the manual methods used previously which were often slow. In the context of this assignment, fast, good quality feedback is essential as it is used by students when completing part three and especially part four of the assignment. This was achieved. It was not possible to measure formally student attitude to the peer assessment process or to the feedback provided on their theory work, however the tutors were given a strong feeling from informal discussions with those taking part that the experience was by and large an enjoyable one and that learners felt it was useful to them. Marking complex assignments such as the one described here is a challenge to tutors. Students need timely and good feedback and the skills to understand it. In this project this was largely achieved in an interesting and rewarding exercise. We will certainly extend this work in the future and concentrate firstly in measuring learners' attitudes to the peer assessment session and also their attitude to the automated feedback provided by the tool.

The use of three approaches in this project, peer assessment, EVS and automated feedback, was important in that it provided fast, efficient and effective feedback to learners in such a way as to increase engagement and motivation. It was necessary to use a triad of approaches due to the complexity of the projects being evaluated and the need for fast feedback in order for it to be effective. The use of peer assessment was important in engaging learners. They were able to see their own work in the context of that of others. Not only is this important in learning, it is an important professional skill that learners need to acquire. The requirement to present their work in front of colleagues and tutors added additional pressure for them to engage. It was more than just handing in a piece of coursework. The EVS approach was also important as feedback was fast and direct. We suggest that the approach was beneficial and useful to learners. Not only did they see the work of others, but they also had to respond directly to questions from others on their own work, and to engage with others, commenting on their work. Previously this had not been the case. For these reasons we see the provision of a wider experience as an important outcome of this work. The use of the automated feedback tool was perhaps less direct, but by no means less important. The speed of feedback delivered by the system meant that learners were able to respond in a timely manner to the initial feedback on their written documentation and improve their final project work as well as their final documentation. In the past it had proven difficult and sometimes impossible to get feedback to learners in time to inform the next stage of the assignment.

The approach was so successful that we have decided, in addition to its use on the MSc programme, to use it with larger cohorts in computer science modules on the BSc programme at levels 1 and 2. Recently we used a modified version of the EVS tool and automated feedback in the assessment of approximately 300 BSc first year learners undertaking practical assessments. Results from this have been extremely encouraging.

## References

Barker, T. 2010, Developing and evaluating an automated marking and feedback system. To appear in *Proceedings of London International Conference on Education (LICE-2010),* London.

Barker, T. 2009. An automated feedback system based on adaptive testing: extending the model. In *Proceedings of the Interactive Computer Aided Learning Conference, ICL2009*, 23 - 25 September 2009. Villach, Austria.
Draper, S.W. 2009. Catalytic assessment: understanding how MCQs and EVS can foster deep learning. *British Journal of Educational Technology* 40, 2: 285-293.
Freeman, R. and R. Lewis. 1998. *Planning and implementing assessment*. London: Kogan Page.
Li, Lawrence K. Y. 2001. Some refinements on peer assessment of group projects. *Assessment & Evaluation in Higher Education* 26,1: 5-18.
Lilley, M. and T. Barker. 2002. The development and evaluation of a computer-adaptive testing application for English language. In *Proceedings of the 6th CAA Conference*, 169-184. Loughborough: Loughborough University.
Lilley, M. and T. Barker. 2003. Comparison between computer-adaptive testing and other assessment methods: An empirical study. In *Proceedings of the 10th International Conference of the Association for Learning Technology (ALT-C)*, Sheffield: University of Sheffield.
Lilley, M. and T. Barker. 2004. A computer-adaptive test that facilitates the modification of previously entered responses: An empirical study. In 'Proceedings of the 2004 Intelligent Tutoring Systems Conference', *Lecture Notes in Computer Science* 3220: 22-33.
Lilley, M. T. Barker and C. Britton. 2004. The development and evaluation of a software prototype for computer adaptive testing. *Computers & Education Journal* 43, 1-2:109-123.
Lilley, M., T. Barker and C. Britton 2005. The generation of automated learner feedback based on individual proficiency levels. In 'Proceedings of the 18th International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems', *Lecture Notes in Artificial Intelligence* 3533.
Lord, F. M. 1980. *Applications of item response theory to practical testing problems*. New Jersey, Lawrence Erlbaum Associates.
Pritchett, N. 1999. Effective question design. In *Computer-assisted assessment in Higher Education* eds. S. Brown, P. Race and J. Bull. London: Kogan Page.
Race, P. 1998. Practical pointers in peer assessment. In *Peer assessment in practice* (SEDA paper 102), ed. S. Brown, 113-122. Birmingham: SEDA publications.
Russell, M.B. 2008. Using an electronic voting system to enhance learning and teaching. *Engineering Education* 3, 2: 58-65.
Rust, C. 2001. *A briefing on the assessment of large groups.* York: LTSN Generic Centre.
Sadler, D. R. 1989. Formative assessment and the design of instructional systems. *Instructional Science* 18: 119-144.
Sadler, D. Royce. 2010. Beyond feedback: developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35, 5: 535-550.
Simpson, V. and M. Oliver. 2006. *Using electronic voting systems in lectures*, http://ucl.ac.uk/learningtechnology/assessment/ElectronicVotingSystems.pdf.
Van Den Berg, I. 2006. Peer assessment in university teaching: Evaluating seven course design. *Assessment and Evaluation in Higher Education* 31, 1: 19.
Wainer, H. 2000. *Computerized adaptive testing (A Primer.* New Jersey: Lawrence Erlbaum Associates.
Zariski, A. 1996. Student peeraAssessment in tertiary education: Promise,

perils and practice. In *Teaching and learning within and across disciplines' : Proceedings of the 5th Annual Teaching Learning Forum*, eds. J. Abbott and L. Willcoxson. Perth: Murdoch University.