The use of electronic voting and peer assessment to encourage the development of higher order thinking skills in learners

Steve Bennett and Trevor Barker, University of Hertfordshire

Abstract
In the research reported here, electronic voting and peer assessment were used with 215 first year computer science and information technology undergraduates undertaking an electronic media design module. In previous years it proved difficult for learners following this module to develop good quality practical skills resulting in poor results in practical tests. In this study, the module delivery team used electronic voting and a form of peer assessment in order to motivate learners and to engage them more deeply in learning. It was hypothesised that this would improve higher order thinking skills and lead to improved performance in practical work. A significant improvement in performance of 6% was achieved (p<0.001) as compared to previous years. We were able to show that the most likely reason for this was the use of electronic voting systems and peer assessment. In the concluding section of the paper we discuss some issues related to this finding and our teaching approach.

Key words
Electronic voting systems; EVS; peer assessment; higher order thinking skills.

Introduction
E-assessment is currently of major concern to educational institutions in the UK and worldwide. There has been a great deal of interest in e-assessment in FE and HE for several years. A survey of e-assessment in the UK, providing an overview and vision of the future landscape is provided by Whitelock and Brasher (2006). This work, based on a survey of experts in the field, funding bodies and accrediting bodies identifies current e-assessment practice and provides an overview of key findings. In this paper we present our findings from a study into the use of electronic voting systems (EVS) and a form of peer assessment intended to improve the higher order thinking skills of learners. Higher order thinking skills are related to the ideas of Bloom (see Anderson and Krathwohl 2001), who suggested that some forms of learning require a greater depth of cognitive understanding than others. Bloom's taxonomy sets out a hierarchy of increasing cognitive skills, increasing depth of understanding and engagement with learning. It is our experience that skills described at higher levels of Bloom's taxonomy such as application, analysis, synthesis and evaluation are much more difficult to teach and assess than simple knowledge and understanding (Bloom 1956). Our hypothesis was that by using peer assessment and EVS we would engage learners sufficiently in order to develop and apply higher order skills. Our findings from previous research (Barker and Bennett 2010), suggested strongly that this approach would be likely to be effective.

Background
An interesting possibility for the module delivery team was the potential of the EVSs to not only provide feedback, but to open the possibility for use in a peer assessment process. Electronic voting systems 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 to provide the opportunity for deeper learning (Draper 2009). In this study it was decided to use a combination of EVS and a form of peer assessment to engage learners in order to improve their higher order thinking skills. This approach was intended to motivate learners to engage more deeply in their own learning, since it not only provides immediate feedback for learners, but is also a way of ensuring learner engagement in the process of assessment. Electronic voting systems have the potential to enhance learning and motivation, as well as providing variety and engagement within lectures (Simpson and Oliver 2006).

It has been claimed that peer assessment is an important method to engage and motivate students. Van Den Berg (2006) suggested that peer assessment not only was able to engage learners but also led 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) suggested that peer assessment was valuable for students, being effective in helping them in the development of judgment skills. It was also important in facilitating significant time saving for tutors. The work of Li (2001) has shown that peer assessment was an effective way of grading individual contributions to group working. Our own work (Barker and Bennett 2010) was able to show that the approach we adopted and used in a summative assignment was a fair and valid method in the context of this module.

The use of peer assessment and EVS in the development of higher order thinking skills was of major importance in our research. There has been a large amount of previous research on the value of using peer assessment in education. Van Den Berg (2006) suggests that peer assessment not only engages learners but also leads to improvements in written work and interaction. There have been two significant meta-analyses of this research done in the past (Topping 1998; Falchikov and Goldfinch 2000). More recently there has been a more systematic analysis of this area by Zundert and colleagues (2010) which incorporated peer assessment beyond the tertiary sector and also attempted to establish causal relationships between the interventions made and outcomes obtained in the study. The motivational benefits of peer assessment have been investigated. For example, Van Den Berg (2006) was able to show that peer assessment not only engaged learners but also led to improvements in study skills, in written work and interaction. Zariski (1996) and Race (1998) describe other benefits of peer assessment, including personal ownership of assessment, learners taking responsibility for their own 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 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.

The fairness of our approach was of major concern to us as this research was undertaken in a pragmatic summative context. There are usually two models

employed to assuage fears of unfair marking by inexpert markers: either the 'reduced percentage model', where the students' evaluations only contribute towards a small proportion of overall marks (Wen and Tsai 2006), or the 'tutor arbitration model' where marks are wholly awarded by peers, but with provision for appeals in case of perceived unfair marking (Davies 2000). In these studies a great diversity of subject disciplines is covered. Essay writing and academic English has proven to be a rich ground, subjects allied to medicine also are well represented in the literature, owing perhaps to the deep culture of peer assessment in that discipline. The types of artefacts for which peer assessment has been used include essays, lab reports and oral presentations.

One common point of comparison seen in the literature is the correlation between the marks awarded by tutors and those by students. The majority of results exhibit highly positive significant correlations with correlation coefficients greater than 0.80. Stefani (1994) using peer assessment to evaluate lab reports in biochemistry found correlation coefficients between peer assessment and the marks awarded by tutors to be 0.89. Hughes and Large (1993) using peer assessment to evaluate pharmacology students' communication and presentation skills found correlations of 0.83. Smith et al. (1990) using peer assessment to evaluate the debating skills of psychology students found correlations of 0.80. A less successful intervention is described by Bostock (2000). Here computer science students were required to evaluate draft multimedia artefacts created by peers. The correlations achieved in this case were only 0.45. In our own research we found that the correlation between tutor and learner was significantly related to the performance of individual learners. Those performing well had strong positive correlations and those performing less well had much lower correlations.

We therefore hypothesised that peer assessment and EVS would promote increased learner engagement with practical work and lead to an improvement in marks. Many researchers have reported increased learner engagement with courses as a consequence of participating in this peer assessment. Stefani found that 85% of her students believed the exercise helped them to learn more - notwithstanding the fact that 75% of the learners reported the assessment to be difficult. A number of studies have demonstrated initial scepticism by students before use of peer assessment but later exhibited subsequent acceptance and enjoyment of the practice (Smith et al. 2002; Venables and Sumit 2003; Wen and Tsai 2006). However, a number of studies, despite finding that students had a positive attitude to the process of peer assessing, found that many still remained very sceptical about the ability of their peers to assess correctly (Wen and Tsai 2006). This is an important issue as should learners be unhappy as to the validity of the assessment, this would likely counter the reported motivational benefits. We shall return to this point in the discussion section of the paper. In this study we used a combination of peer assessment and electronic voting systems (EVS) to engage learners. In the next section we summarise our previous research with EVS and peer assessment.

Previous research with EVS and peer assessment
In previous research by the authors reported at CAA in 2010 we found peer assessment with EVS to be useful, fair and beneficial to learners (Barker and Bennett 2010). The research reported in that paper related to a study with a cohort of MSc computer science students working in groups on complex multimedia project work. The students presented their work to the entire class and were required to

mark the presentations of their peers according to set criteria. As this was a master's course the assessment, as might be expected, had a rubric aimed very much at higher order evaluation of usability and information architecture. We found that the experience seemed to galvanise the students to much higher levels of effort and performance. The findings of this study suggested that this approach had enormous potential for extension to include other cohorts. For this reason we decided to use this approach in order to address problems with the delivery of a first year electronic media undergraduate module.

The module

One particular module had proven to be problematic in the past in that many students were not performing well on their practical work despite significant effort on the part of the tutors. The module is related to the development of a range of skills in the area of electronic media design. The aims of the module were to understand what motivates design decisions, to appreciate the importance of creating systems that are fit for their intended purpose and for learners to make straightforward design decisions of their own.

The learning outcomes (knowledge and understanding) were to:

• understand the reasoning behind some of the decisions that have been made in the design of existing techniques and technologies for storing, transmitting and processing information;
• understand the relationship between form, function, content and aesthetics, and their importance in the design of documents and the systems that manipulate them;
• understand some of the options that are available to those designing and implementing systems for the storage, transmission and presentation of information.

The skills to be developed included:

• being able to make straightforward design decisions, taking into account the relationship between form, function, content and aesthetics;
• making an informed choice between different means of representing, transmitting and processing information.

In order to achieve these objectives, learners were required to follow a programme of lectures and readings related to the design, implementation and evaluation of electronic media, supported by practical work related to designing a multimedia website. The learning outcomes were assessed by three theory 'in-class' tests and a practical assignment delivered under exam conditions.

In previous years, we found that performance on the final practical assignment, the design and implementation of a website, was less than we expected in terms of quality. Although we had devoted a great deal of effort into the delivery of the necessary skills and understanding, some learners were failing to take on board exactly what was required of them. This was despite the module delivery team exposing the marking criteria and guidelines in lectures prior to the test and devoting practical sessions explicitly to the delivery of the required skills. It was our hypothesis that the reason for these problems related to the inability of some learners to internalise the assessment criteria required for the creation of a well designed website. In short they were in possession of the necessary individual skills to perform

the task but were unable to put them together in order to produce a coherent website with suitable form, function, content and aesthetics as required in the learning outcomes. Based on our findings from the MSc course reported previously (Barker and Bennett 2010) we decided to use a combination of electronic voting system (EVS) and a modified form of peer assessment in order to approach this problem. It was not possible due to the size of the cohort to use direct peer assessment as we had done previously. However we argue that the *process* of peer assessment was the important aspect of our research rather than the actual mark awarded by peers. For this reason we employed a system whereby learners reviewed the work of the previous year's learners. This we suggest, had all the advantages of direct peer assessment, i.e., exposure to the marking criteria, direct feedback from tutors and peers and also increased engagement, yet the process was manageable for a large group of learners. The module had three summative assignments which are explained in the next section.

The assignments
Assignment 1 was a multiple-choice test intended to test the theoretical component of the module. Assignment 2 was an assessment of the practical work of the previous cohort. Marks were awarded based on the closeness of a learner's mark to that of the tutor's. This was intended to reinforce in learners the need to assimilate the tutors' criteria for the quality of the web sites. In order to do well, learners were required to understand details of the marking criteria, analyse, synthesise and apply the criteria in new contexts and to evaluate design decisions. These may be seen as corresponding to Bloom's levels 3, 4, 5 and 6 (Bloom 1957). Assignment 3 related to the design of a website according to a brief. Students were to prepare part of the website in advance and then complete the website later under examination conditions.

In this module learners are taught elements of screen design, usability and the evaluation of software prior to developing the website. Despite this, there are several common errors that continually crop up in their websites. These common errors seemed very much to relate to misunderstandings as to the necessary criteria for good website design, rather than a lack of basic technical skills. We hypothesised that many learners were not relating the theoretical work to the practical problems of website design. This may be a problem of meta-cognition, the inability of some learners to look upon problems holistically. It has long been our experience that in our subject area some learners have good technical ability yet are unable to synthesise complex artefacts despite them being well within the compass of their technical skills. For example, students may have the ability to create working buttons that link a database to a website. They may create a fully functioning database. However they are often unable to produce a usable and useful application that incorporates these basic skills.

Delivery of the module
In previous years, students attended twelve practical sessions where they developed the necessary skills to develop their website. Learners also attended six lectures on the basic theory of electronic media and multimedia production and six lectures on aspects of software usability, screen design and software evaluation. Each week learners were set compulsory set reading as well as some optional readings. In addition learners had one lecture supporting developing the website, including making explicit the marking scheme for the final practical test. This included the

presentation and discussion of examples by tutors. Learners were then given the brief for the final assignment, practical test 3, where they were required to develop a website based on the brief.

This year learners followed exactly the same practical and theory lectures and readings as in previous years. A major difference was the substitution of the lecture supporting developing the website with an EVS session evaluating last year's websites. Soon after the EVS session the students had a modified theory test instead of the previous year's coursework 2. In this test they were required to peer review a sample of last year's websites, using the tutors' marking scheme.

During the EVS session, students worked in small groups of about four. The tutors presented sample websites from the work of the previous cohort and displayed them on screens. Students marked the websites according to the criteria used on the marking scheme using the EVS. After each presentation, the tutors then discussed their marks for the presentation and compared them to the marks of the students. Much discussion occurred during the session and afterwards on the discussion forum of the Managed Learning Environment. The large lecture theatre contained more than 200 students and it was obvious to the tutors that the groups of learners were actively engaged in marking and discussion among themselves and with the tutors. It was apparent from the questions asked and the comments made that learners were actively applying the marking criteria not only to the work of their peers, but also to their own work. At times there were quite significant differences of opinion between the student cohorts and the tutors, which we considered to be a positive aspect of the initiative. The purpose of this session was not to get a totally accurate reading of the class's opinion, but rather to initiate a discussion of the assessment, and how marks were allocated. In order to achieve this we would demonstrate a particular website then ask the class to vote according to the criteria sheet given to mark the assessment, and then we would give our own opinion - and subsequently discuss, even argue about it. This forced students to clarify their idea of the meaning of the various criteria and compare it with our own.

Assessing the impact
In order to assess the impact of the intervention we made a comparison with learner performance in the previous year where no EVS was employed. In the previous year learners followed exactly the same theoretical and practical work. In place of the EVS session learners had a lecture where all the assessment criteria were explained with examples. The marking criteria were explained in exactly the same way as in the EVS session. The only difference was that EVS was used. It was decided to compare the performance of learners for coursework 1 and coursework 3 for the two cohorts. The comparison of coursework 1 (a theory multiple-choice test) was intended to act as a control. Any differences in performance for the two cohorts with regard to coursework 3 (the practical assessment) would then likely be related to the intervention with EVS). Table 1 below shows the performance by learners on coursework 1 and 2 for the two cohorts (2009 and 2010).

The results of an ANOVA performed on the data summarised in table 1 below show that the 2009 cohort performed significantly better than the 2010 cohort ($p<0.01$) on coursework 1, the control but that the 2010 cohort performed significantly better than the 2009 cohort on coursework 3 ($p<0.001$). The 6% difference in the mean test score we argue, suggests that the intervention with EVS and the peer review of previous coursework was effective in engaging learners at higher cognitive levels, as

Table 1: Performance by learners

Test	Mean% 2009 Cohort	Mean% 2010 Cohort	Probability
Coursework 1	59% (n=290) (SD 16.6)	56% (n=215) (SD 17.5)	p <0.01
Coursework 3	58% (n=277) (SD 18.4)	64% (n=218) (SD 14.3)	p <0.001

being tested by the assignment. The difference in performance (2009 cohort > 2010 cohort) on coursework 1 indicates that the improved performance on coursework 3 (cohort 2010 > cohort 2009) was not due to differences in the abilities of learners between the two cohorts. It is true to say that in studies with large numbers (n = 200+) such as in this study, even quite small differences in means may yield significant statistical differences, however, we are confident that our findings presented in table 1 represent a real improvement in the performance of learners. A difference of 6%, while not large in absolute terms represents an improvement by two grades, which is of great value to learners. It is also the highest mean score obtained for this assignment on this module for the six years during which it has been running. It is of course possible that the marking of coursework 3 was biased. To allow for this a random sample of the work of 20 learners was second marked by an independent assessor. The results of this moderation process suggested strongly that the marking was fair and accurate.

Discussion
Apart from the significantly improved marks the module delivery team was in agreement that the quality of the practical work this year has been very much improved over previous years. We are convinced this is because students had engaged deeply with the problem and internalised the relevant criteria at a deeper level, requiring the use of meta-cognitive and higher order thinking skills. The mark given by a tutor to a student in this type of assignment is to a greater or lesser degree a subjective judgement, rather than being either correct or incorrect. The ability to make such judgements is a skill that is difficult to teach yet it is important to develop in learners.

It was interesting to note that learner engagement went beyond the EVS session, and included the interaction of learners on the discussion forum of our MLE and also contact with lectures via email and face to face. Some learners were clearly sharing ideas about the peer review session and applying it to their own ideas regarding implementation of the criteria. This discussion was a vast improvement in quality and quantity from previous years and became quite heated at times. The exposure of our criteria in this way certainly encouraged engagement and also criticism and questioning of our marking. This was especially true after the marks for coursework 3 were released. The number of students questioning their marks increased from a very few in previous years to several tens in the current year. All learners were offered face to face meetings to discuss their grade with the tutor. At these meetings the criteria were discussed and the coursework reviewed. It was noticeable that it was mostly moderate to high performing learners that requested meetings. Learners
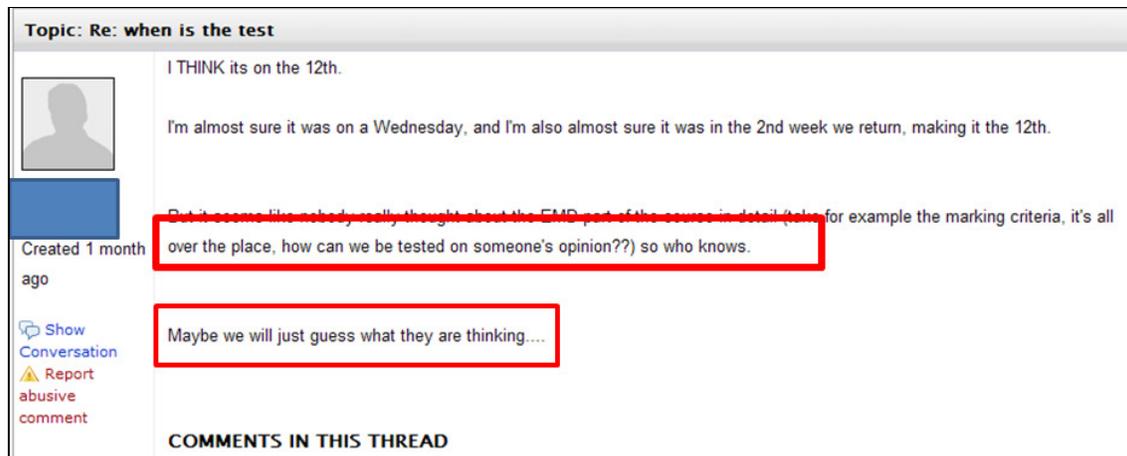
Figure 1: Discussion between two learners

were mostly satisfied with their grades after discussion. Figure 1 above shows an example of part of a discussion between two learners.

In this example, after discussing the date of coursework 3, they go on to discuss the criteria. In this case the learners are questioning the way in which the assignment was to be assessed.

> Learner 1: (… for example the marking criteria, it's all over the place, how can we be tested on someone's opinion??) so who knows.

> Learner 2: Maybe we will just guess what they are thinking.

From this short interchange it appears that the two learners are trying to second guess the tutor rather than engage with the criteria to solve their problem. Trying to guess what the tutor is thinking and being marked on someone's opinion are seen as being an unfair or insurmountable obstacle placed there to frustrate their efforts. Some learners like simple yes/no type solutions where marks are awarded based on simple criteria. This indicates a naïve understanding of the subject area. It is difficult for some learners to accept that there are many things that do indeed rely on opinion rather than purely simple criteria that can be ticked off. Even in programming and mathematics there are good and bad solutions to any problem. We try to emphasise to learners that it is not so much a case of guessing what the tutor is thinking, but rather adopting the same criteria as the tutor. This requires learning some of the higher order thinking skills that the tutors are using when they assess such work. The above exchange between learners suggests that those learners were thinking strategically rather than at a deep level. Although such exchanges were a little frustrating for tutors there was sufficient evidence that our intervention was successful, some learners were internalising our criteria and that this made it worth the additional effort. We are convinced that the process of using EVS and peer assessment was instrumental in achieving this.

The findings from this research were very encouraging. It will be important to find out if the encouraging improvement we achieved is maintained over a longer period of time. We shall therefore be repeating this work in the coming year. Other questions we have relate to the attitude of learners to the use of EVS and peer assessment in summative assessment. In order to find out more about this we intend to conduct a survey with the next cohort, using in the first place a questionnaire. A questionnaire

has been designed to this end. Based on the findings from this questionnaire we shall select a sample of high, medium and low achievers to interview in order to gain a richer understanding of learner attitude to the approach.

References

Anderson, L.W., and D.R. Krathwohl, eds. 2001. *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Barker, T., and S. Bennett. 2010. Marking complex assignments using peer assessment with an electronic voting system and an automated feedback tool. In *Proceedings of CAA 2010 Conference.* Southampton: Southampton University.

Bloom, B.S., ed. 1956. *Taxonomy of educational objectives, the classification of educational goals – Handbook I: Cognitive domain.* New York: McKay.

Bostock, S. 2000. *Student peer assessment,* http://www.keele.org.uk/docs/bostock_peer_assessment.htm (accessed May 24, 2011).

Davies, Phil. 2000. Computerized peer assessment. *Innovations in Education and Training International* 37, 4: 346-355.

Draper, S.W. 2009. Catalytic assessment: Understanding how MCQs and EVS can foster deep learning. *British Journal of Educational Technology* 40, 2: 285-293.

Falchikov, N., and J. Goldfinch. 2000. Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research* 70, 3: 287-323.

Hughes, I.E., and B.J. Large 1993. Staff and peer-group assessment of oral communication. *Studies in Higher Education* 18: 379-385. DOI:10.1080/03075079312331382281

Li, Lawrence K.Y. 2001. Some refinements on peer assessment of group projects. *Assessment & Evaluation in Higher Education* 26, 1: 5-18.

Race, P. 1998. Practical pointers in peer assessment. In *Peer Assessment in Practice: SEDA paper 10,* ed. S. Brown. Birmingham: SEDA Conference.

Russell, M.B. 2008. Using an electronic voting system to enhance learning and teaching, *Engineering Education* 3, 2.

Rust, C. 2001. *A briefing on the assessment of large groups: LTSN Generic Centre assessment series no.12*. York: LTSN Generic Centre.

Smith, H., A. Cooper and L. Lancaster. 2002. Improving the quality of undergraduate peer assessment: A case for student and staff development. *Innovations in Education and Teaching International* 39, 1: 71-81.

Stefani, L.A.J. 1994. Peer, self and tutor assessment: relative reliabilities. *Studies in Higher Education* 19, 1: 69-75.

Simpson, V., and M. Oliver. 2006. *Using electronic voting systems in lectures*, http://ucl.ac.uk/learningtechnology/assessment/ElectronicVotingSystems.pdf
Topping, K. 1998. Peer assessment between students in colleges and universities. *Review of Educational Research* 68, 3: 249-276.

Van Den Berg, I. 2006. Peer assessment. *University Teaching: Evaluating Seven Course Designs, Assessment and Evaluation in Higher Education* 31, 1: 19.

Venables, A., and R. Summit. 2003. Enhancing scientific essay writing using peer assessment. *Innovations in Education and Teaching International* 40, 3: 281-290.

Wen, Meichun and Chin-ChunTsai. 2006. University students' perceptions of and attitudes toward (online) peer assessment. *Higher Education* 51, 1: 27-44. DOI: 10.1007/s10734-004-6375-8. http://dx.doi.org/10.1007/s10734-004-6375-8.

Whitelock, D., and A. Brasher. 2006. Developing a roadmap for e-assessment: Which way now? In *Proceedings of 10th CAA Conference.* Loughborough: Loughborough University.

Zariski, A. 1996. Student peer assessment in tertiary education: Promise, perils and practice. In *Teaching and Learning Within and Across Disciplines: Proceedings of the 5th Annual Teaching Learning Forum,* ed. J. Abbott and L. Willcoxson. Perth: Murdoch University.

Zundert, M. van, J. Sluijsmans and D. van Merriënboer. 2010. Effective peer assessment processes: Research findings and future directions. *Learning and Instruction* 20, 4: 270-279.