RESEARCH ONLINE

University of Wollongong
Research Online

Deputy Vice-Chancellor (Education) - Papers

Deputy Vice-Chancellor (Education)

2009

Click or clique? Using educational technology to address students' anxieties about peer evaluation

Ruth Walker
University of Wollongong, rwalker@uow.edu.au

Graham C. Barwell
University of Wollongong, gbarwell@uow.edu.au

Publication Details

Walker, R and Barwell, G, Click or clique? Using educational technology to address students' anxieties about peer evaluation, International Journal for the Scholarship of Teaching and Learning, 3(1), 2009, p 1-20.

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

RESEARCH ONLINE

Click or clique? Using educational technology to address students' anxieties about peer evaluation

Abstract
Peer bias is recognised as a primary factor in negative student perceptions of peer assessment strategies. This study trialled the use of classroom response systems, widely known as clickers, in small seminar classes in order to actively engage students in their subject's assessment process while providing the anonymity that would lessen the impact of peer pressure. Focus group reflection on the students' impressions of the peer evaluation process, the use of clickers, and their anxieties about potential peer bias were analysed in the light of the results of teacher and class evaluations of each individual student presentation. The findings revealed that students recognised the value of peer assessment in promoting class engagement and active learning, despite their ongoing resistance to the practice of peer review. An unexpected finding suggested that the clickers, selected as an educational technology for their appeal and ease of use by the 'digital native' student already familiar with a variety of mobile communication and gaming devices, reinforced student perception that the peer review process was akin to a popularity contest.

Keywords
evaluation, anxieties, address, educational, clique, click, peer, about, students, technology

Disciplines
Arts and Humanities | Social and Behavioral Sciences

Publication Details
Walker, R and Barwell, G, Click or clique? Using educational technology to address students' anxieties about peer evaluation, International Journal for the Scholarship of Teaching and Learning, 3(1), 2009, p 1-20.

This journal article is available at Research Online: http://ro.uow.edu.au/asdpapers/229

International Journal for the Scholarship of Teaching and Learning
http://www.georgiasouthern.edu/ijsotl
Vol. 3, No. 1 (January 2009)
ISSN 1931-4744 @ Georgia Southern University

Click or clique? Using Educational Technology to Address Students' Anxieties About Peer Evaluation

Ruth Walker
University of Wollongong
Wollongong, New South Wales, Australia
rwalker@uow.edu.au

Graham Barwell
University of Wollongong
Wollongong, New South Wales, Australia
gbarwell@uow.edu.au

Abstract

Peer bias is recognised as a primary factor in negative student perceptions of peer assessment strategies. This study trialled the use of classroom response systems, widely known as clickers, in small seminar classes in order to actively engage students in their subject's assessment process while providing the anonymity that would lessen the impact of peer pressure. Focus group reflection on the students' impressions of the peer evaluation process, the use of clickers, and their anxieties about potential peer bias were analysed in the light of the results of teacher and class evaluations of each individual student presentation. The findings revealed that students recognised the value of peer assessment in promoting class engagement and active learning, despite their ongoing resistance to the practice of peer review. An unexpected finding suggested that the clickers, selected as an educational technology for their appeal and ease of use by the 'digital native' student already familiar with a variety of mobile communication and gaming devices, reinforced student perception that the peer review process was akin to a popularity contest.

Keywords: Clickers, classroom response systems, peer evaluation, small classes, higher education, peer pressure, presentations

Introduction

While peer evaluation strategies in higher education classes have been widely recognised as a positive strategy to promote active learning, concern about the potential for bias based on personality is at the heart of student resistance. This study was conducted in small classes in a Humanities program at the University of Wollongong, Australia, in 2008 and investigated the potential use of classroom response systems, or 'clickers', to engage students in the evaluation of class presentations, while alleviating their anxieties about peer pressure and personality bias. The original organising questions for this research study were whether clickers could be effective in engaging students in the evaluation of their peers' presentations in small classes, and whether a comparison of the class and the teacher results might reveal disparities that would prevent the future use of peer evaluation as a formal assessment method. This paper will focus on the feedback from students who had participated in the clicker trial, which drew attention to their ongoing concern about possible personality bias from their classmates, their unease at their own role in evaluating their peers, and their impressions of the use of an educational technology as a learning tool. The students' familiarity with mobile communication technologies like mobile telephones and gaming devices meant that they quickly adapted to the more limited level of interactivity offered by the clickers, but also raised interesting questions about whether the clickers themselves served to alleviate

their anxiety about peer evaluation, or reinforced their impression of the process as a popularity/talent contest.

Three seminar classes, with an average enrolment of 17 students, were selected for our trial use of clickers as part of the students' individual class presentations. For this trial, the peer results were not factored into the formal assessment, but students were given marking criteria (developed in collaboration with a previous class) that mirrored that of their teacher's, and at the end of each individual presentation, were guided through the criteria and asked to submit their responses anonymously via handheld clickers. Each student's response was collated by a central computer, so that the teacher could later compare the overall class mark with their own assessment of the individual student presentation. The week following the presentation, each individual student was given the breakdown of the teacher's marks for each component of the marking criteria for class presentations, as well as the results from the peer evaluation from the class using the clickers. This paper will discuss the relationships between the peer and teacher results, and will then analyse the student feedback about this trial of peer assessment of presentations using clickers gathered from the focus groups. This focus group discussion offers a qualitative, user-focused perspective on the selected educational technology as well as of the peer assessment process, which will illuminate the more quantitative aspect of the data analysis generated by the summative result comparison of teacher and peer evaluation of each individual presentation.

Literature Review

The evolution of mobile learning devices has lagged behind students' everyday experience with new media technologies. However, the use of clickers has had some notable success in the higher education context, particularly in large classes or lectures, where they have been celebrated as an educational tool that appeals to the 'digital native' generation of student (Prensky 2000 cited in Brill and Park, 2008 p.71). Clickers, which are wireless hand-held devices, encourage student-teacher and potentially student-student interaction during class time, and require very little technology readiness on the part of the student (Yourstone et al 2008), as the 'typical' student these days will already own at least one portable communication device, such as a mobile phone, laptop or handheld gaming device.

Clickers are considered appealing to the media literate generation who 'are both familiar with technology' and 'reluctant to suffer impassive learning silently' (Murphy, 2006, p.1). Most of the literature on clickers investigates teacher/student perceptions, with the focus tending to target the 'affective benefits' of clickers, which include greater student engagement, increased student interest, and heightened discussion and interactivity (Stowell and Nelson, 2007). Typically, educational researchers who use clickers in their classes focus their reports on the functionality of the technology in encouraging student participation and active learning in class time (Nelson and Hauck, 2008). Guthrie and Carlin (2004) point out that modern students are primarily active learners, and that lecture-based courses may be increasingly out of touch with how students engage their world. Clickers, it is generally argued, move beyond the teacher-focused lecture model to allow engagement in large classes through interactivity from students to teachers, and potentially between students. This potential for increased student engagement, however, depends on the development of creative pedagogical activities within the class and with the technology, as it has been pointed out that clickers may simply 'feign interactivity' (Socol, 2008). At heart, they are necessarily single-directional communication systems; the teacher is always in control of the central database, and although students give feedback, it is generally only in response to pre-scripted questions and answers.

Nevertheless, clickers are increasingly in demand, particularly for large classes (Guess, 2008). But while the popularity of clickers rest on their appeal to the media literate student, many scholars are dismissive of this popular cultural factor, even as they comment on the positive aspects of the clickers' 'game approach' that engaged students more than class discussion (Martyn, 2008 p.71) or their similarity to the 'electronic voting systems' of television programs like *Who Wants to Be A Millionaire* (Jenkins 2007, p.528). Indeed, Banks recounts disparaging reactions by other scholars to his trials of clicker technology as 'infotainment' or 'edutainment' (2006, p.383). Although there are of course dangers associated in introducing popular cultural media or approaches into the classroom, they can be educationally rigorous if they are introduced with pedagogical principles in mind, since student enthusiasm 'may be predicated on novelty, fun, richer learning environments, or a clearer sense of being a member of a learning community' (Campbell, 2007, p.385). The 'average' student would be familiar with the process of grading presentations using clickers, as the technology is similar to the procedure on mainstream television shows like *American Idol*, where audience members vote to support their favourite contestants, or with the real-time tracking of focus group responses to televised political debates. It was perhaps not surprising, given this mainstream use of audience response strategies in television, that students equated peer review of presentations in the classroom with popularity contests and were anxious about possible peer bias, or the formation of 'cliques' based on personality, impacting on the validity of the results.

The focus of this study was on incorporating the use of clickers into seminars or tutorials, with the intent to more involve students in class presentations and peer evaluation strategies which would then become a more active part of their own learning and assessment experience. Previously, very little attention has been paid to the impact of clickers in small classes, where students are not simply checked for their attention or knowledge-acquisition, and where the technology is used to aid critical evaluation. A few articles report some experimentation with the parameters of the use of clickers in a variety of forms of active learning, such as case studies, group work, and demonstrations. Ribbens (2007), for instance, moved from using clickers to deliver pop quizzes at the end of his lectures, to interspersing quizzes alongside his lecture material to check on students' ongoing understanding, and then to re-formatting the quizzes to allow for basic problem solving and critical analysis in his small biology classes.

Only one other study was found in a review of the literature around clickers in higher educational contexts, that puts 'control' in the hands of the students, as with our trial of peer evaluation of presentations. This other study reported on a trial of a version of clickers called a 'keypad-based group process support system' in the UK with Information Systems students in 1995 (Banks, 2003; Banks, 2006). Banks had set up this use of clicker technologies in response to his students' reported unease at having to make short presentations as part of their assignment work: 'they had carried out similar presentations in other subjects but felt that a simple mark for the previous presentations had not allowed them to learn from the process and consequently improve their presentation skills' (2003, p.39). Interestingly, although Banks commented that 'despite their obvious uses in aiding pacing and content management in face-to-face lecture environments, their main benefits may lie in smaller group work' (p.45), very few other scholars have taken up his challenge to explore the use of clickers small classes.

The key differences between the case study reported by Banks and our research project was that his solicited formative public peer review, rather than summative private, or anonymous, peer review. Otherwise, our conclusions are similar: that clickers 'offer an opportunity to build on modern educational paradigms which emphasise collaborative learning rather than the knowledge transfer from teacher to learner' (Banks, 2003,

p.45). The literature on the use of clickers discusses a number of advantages to their use, including active learning, providing feedback, increasing attention span and motivation (Nelson and Hauck, 2008). These are all aspects that are equally important for large classes (lectures) and small classes (seminars or tutorials), and are particularly highlighted when combined with peer assessment strategies. Peer assessment can give learners opportunities to dynamically practice applying criteria, giving and receiving feedback, comparing their work with others, while also providing a framework for clearer goals and expectations. The more 'active learning' aspects of peer assessment are understood to develop students autonomy and critical abilities (Falchikov and Goldfinch, 2000). Gauthier (2004) points out that learning situations in the classroom are often treated as if the only interactions happening are the ones between the teacher and student: 'the student-student relationships are often ignored for many of our learning objectives' (p.7). Even though a student in a classroom situation is most often assessed on their individual performances by their teachers, they can also benefit and be challenged by their peers. Gauthier therefore argues that 'involving students in the assessment process can promote self regulated learning and collaboration without jeopardising existing assessment structures in the system' (p.12).

Campbell et al (2001), in their study of undergraduate student presentations in a business faculty at an American university, found that 'the use of peer evaluations certainly encourages more active participation in the learning process' (p38). They went on to look at the correspondence between instructor and peer evaluation of presentations, questioning whether self or peer evaluation is a valid representation of performance when the instructor is used as the benchmark. There was a high rate of agreement with instructor ratings when students acted as 'peer raters' rather than 'self raters' (p.37). Campbell et al concluded that self evaluation was a relatively poor substitute for instructor feedback, but that peer evaluations 'offered considerable promise' (p.37), assuming sufficient training or marking criteria explanation was given. They argued that peer feedback would aid in bringing overall presentation quality (both holistic and analytic) up to the level expected by the audience, and would have a strong impact on future presentations. The outcome of the study (that peer evaluation was more valuable than self evaluation) augments the outcomes of MacAlpine's earlier study of peer evaluation of presentations in an engineering class, where he encouraged students 'to see peer assessment as a learning experience in exercising judgement and assessment, and a first step in developing their ability to assess their own work and their own strengths and weaknesses in a realistic manner' (MacAlpine, 1999, p.17). MacAlpine concluded that the enthusiasm and greater confidence in marking generated by peer feedback activities 'should increase the motivation, confidence and ability to be a lifelong learner'(p.24) in students.

There does seems to be a general understanding in the literature that students appreciate the need for peer assessment as an opportunity to develop a transferable personal skill that will be of use in the workplace upon graduation, even though they may have difficulties in implementing it in practice during their studies. Boud (1990) argued that both self and peer assessment is 'fundamental to all aspects of learning; as it encourages the development of the reflective student' (cited in Greenan et al 1997, p71). The overall benefits of peer assessment are understood to include the promotion of higher order thinking and cooperative learning. Topping's (1998) review of peer assessment strategies highlighted the positive effects of peer assessment in higher education, as well as its generally positive impact on students' subjective perceptions. But while Topping reported evidence that peer feedback is usually valid and reliable, most of the literature on peer assessment maintains that there are issues that can seriously affect the perceived reliability and validity of peer ratings, which are based on the social relationships within a given group.

There is a general expectation in the literature that students will 'over-mark', that is, give higher marks than their teachers. This has been identified as the most important issue that will affect the validity of the peer assessment process (Wadhwa, 2003, p1). It is generally understood that perceived or anticipated inconsistencies in peer evaluations – whether they actually occur or not - can affect the learning benefits expected from the peer assessment process. Reasons for over or inconsistent marking include: 'friendship marking' where peer assessors over-mark due to social pressure (Topping et al, 2000; Lu and Bol, 2003); 'rater's style' where misunderstandings about the numerical rate or grade lead to differences in results (Stefani, 1994); 'marking criteria' where different peer assessors will understand the marking criteria differently (Falchikov and Goldfinch, 2000); the ability of the peer assessor, where competence will affect the peer marking (Stefani, 1994); implicit bias according to factors like race or gender (Langan et al, 2005; Falchikov and Goldfinch, 2000); and student recalcitrance or laziness in giving feedback to their peers (Nilson, 2003). It has also been pointed out that students find it extremely difficult to give negative feedback to classmates because they worry about hurting others' feelings or of damaging personal relationships (Schaffer, 1996; Topping, 1998). To alleviate these concerns, researchers advocate using electronic communication technologies that allow for anonymous peer reviewing, which will avoid the possible embarrassment students may experience in face-to-face interaction (Zhao, 1998). This was the methodological approach adopted in our study, as outlined below.

Methodology

Participants and Setting

The research team included a learning advisor, who conducted the focus groups of the trial use of the clickers for the peer evaluation of student presentations, and the teacher who launched the use of clickers in his small classes, for which he also was the subject coordinator. Three of his classes were selected to trial the peer evaluation of individual student presentations using clickers. All of these classes were delivered in the Faculty of Arts at the University of Wollongong: the first two were in an Electronic Cultures subject, and the third was a Shakespeare subject. Both of these subjects were 3rd level, which meant that the majority of students were in their final year of their undergraduate studies, as they were enrolled in three year undergraduate bachelors programs in either Communication and Media Studies or in Arts. The two subjects concerned were electives in the majors the students were undertaking, and these majors were among a range, broad in the case of the Arts degree, which students could choose from in making up their degree programs.

All three of the classes were small seminars or tutorials, rather than large lectures, the more traditional venue for using clickers in an educational setting. In the case of the Electronic Cultures subject, the students taking part in the clickers trial comprised about half of the cohort in the subject in the autumn session, the other half having their seminars run by casual tutors. All the students in the Shakespeare subject took part in the trial. In all classes in the trial, there were many more female than male students (ratio of about 5:1 or greater), most students were in their 20s, and international students comprised around 5% of the class total. In one class there was a sight-impaired student.

The class presentations that were peer evaluated in this trial were worth 20% of the overall assessment. Because this was a trial of both the educational technology and of peer evaluation of the presentations, the resultant peer marks were not formally recorded as part of the assessable results, and the tutor took care to record his final mark before reviewing the responses from the peers, in order to ensure that he was

not influenced by the class response. Students were given their individual presentation marking sheet from their teacher in the week following their presentation, as well as the results from their peers. The clickers were used over an eight week period, about a month after the start of the lectures and tutorials, when students started their individual presentation tasks.

Marking criteria

There were some minor differences between the peer and the teacher marking criteria for the class presentations. The peer marking criteria for the presentation focused primarily on components for content, delivery and use of audio-visual aids. Similar but slightly more detailed components, which required specialist knowledge, made up the teacher's marking criteria, with the addition of a question on timing, as this last technical detail was felt to be the responsibility of the teacher to record. Each student was given a time limit of 15 minutes for their individual presentations.

Both sets of marking criteria deliberately had very similar wording. Overall, there were 12 criteria for the students to respond to with the clickers in class following the presentations. The topic areas for these broad marking criteria listed in the table above were broken down into sub-questions. For instance, for the criterion 'content', there were a number of sub-questions including: focus (whether the presentation addressed the selected topic directly); quality of content (the depth of research, use of quotes); relationship of content to audience (did the presentation encourage discussion); structure (organisation of material; did the presenter make it clear where the presentation was going). The broad criterion 'delivery' included questions about whether the speaker appeared interested in their topic, whether they spoke clearly, at an appropriate speed, and whether they addressed the whole class and not just the tutor.

Data collection

After each 15 minute student presentation, and class discussion about the topic, students were guided through a list of questions that made up the marking criteria for these presentations. As each student responded to questions by clicking on a scale of 1-10 on their clickers, their individual results were sent to a central database: in this case, the teacher's laptop. Although it would be possible to determine individual responses by matching the students to their individual clicker codes, for this trial the clickers were distributed randomly and students were reassured that their responses were anonymous. The teacher could keep track of whether all the students were participating by noting the total number of individual responses per question, easy enough to do in small seminars. The teacher first recorded his mark for each individual student presentation, then used the software supplied with the clickers to generate a list of responses to each question. This list was then copied into a spreadsheet to tabulate and calculate a numerical value for the peer response.

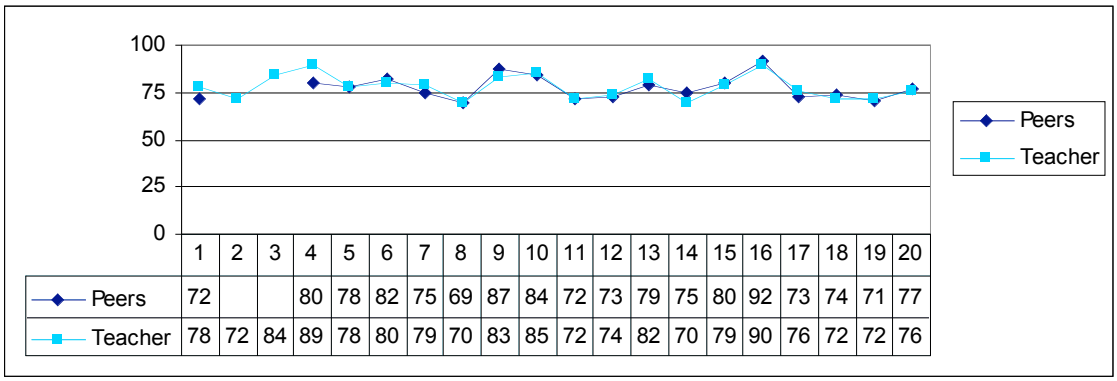
Data results

There were 20 students in the first class trial, 15 in the second, and 16 in the third. The marks for each class have been tabled chronologically in the following figures, from the earliest to the latest weeks of presentations, in order to identify any possible trends or patterns in the comparison between the peer and the teacher results.

In the first class trial, although students successfully used the clickers to evaluate the presentations in the first week, 2 of their results were not properly recorded, as the teacher was still getting familiar with the software. Even with these 'missing' peer results, it is clear from the comparison of results in the Figure 1 below that there is a tight correlation between the peer and the teacher results for each student presentation.

Vol. 3, No. 1 (January 2009)
ISSN 1931-4744 @ Georgia Southern University

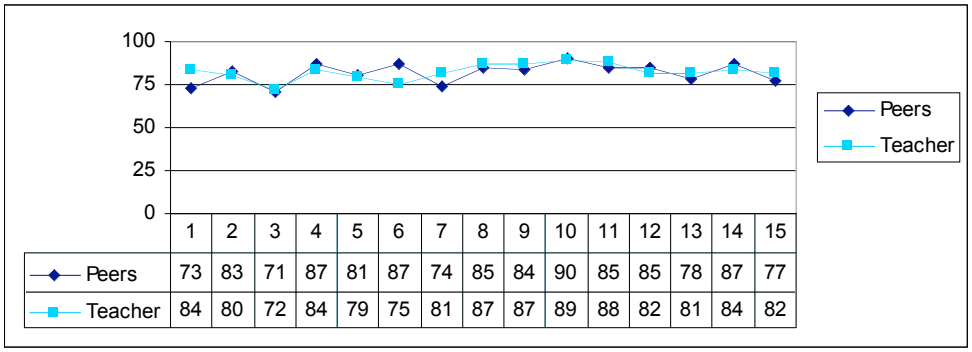
Figure 1: comparison of results for the first class trial



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Peers	72			80	78	82	75	69	87	84	72	73	79	75	80	92	73	74	71	77
Teacher	78	72	84	89	78	80	79	70	83	85	72	74	82	70	79	90	76	72	72	76

The table above shows that in this first class of 20 students, the students returned a mark higher than the teacher for 7 presentations, lower than the teacher for 9, the same as the teacher for 2, and responses for 2 were lost. The greatest variation between the two for an individual presentation was nine marks (with the teacher's mark higher than the students') for presentation 4. The average student mark was 77.3% and the average teacher mark was 78%.

Figure 2: comparison of results for the second class trial

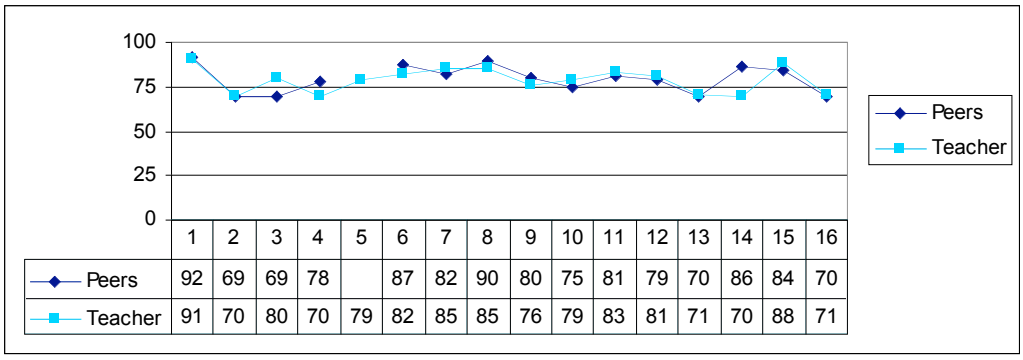


	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Peers	73	83	71	87	81	87	74	85	84	90	85	85	78	87	77
Teacher	84	80	72	84	79	75	81	87	87	89	88	82	81	84	82

In the second class of 15 students, the students returned a mark higher than the teacher for 7 presentations, and lower for 8. There were two really noticeable discrepancies between marks: when the students returned a much lower mark of 73 to the teacher's 84 in the first week of the trial, then five weeks later the students returned 87 to the teacher's 75. It seems apparent that there is no notable pattern of lower or higher marks. Instead, the variation between student and teacher results seems to be consistently inconsistent, even though the marks generally correlate. Overall, the average peer mark was 81.7% and the average teacher mark was 82.3%: as in the first trial, the average peer mark was slightly lower than the teacher's.

Vol. 3, No. 1 (January 2009)
ISSN 1931-4744 @ Georgia Southern University

Figure 3: comparison of results for the third class trial



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Peers	92	69	69	78		87	82	90	80	75	81	79	70	86	84	70
Teacher	91	70	80	70	79	82	85	85	76	79	83	81	71	70	88	71

The third class showed a very similar pattern of tight correlation between peer and teacher results, particularly in the early weeks. The inconsistent pattern of higher and lower marks continued for this class too, with 9 peer returns lower than the teacher's, 6 higher and one not counted (unfortunately, one student's peer mark was not successfully recorded at week 5).

However, the Figure 3 above shows the greatest variation across the trial in all of the three classes, for the 14th presentation, with a peer return of 86 to the teacher's mark of 70. This turned out to be a classic case of peer over-marking, as the teacher reported that the student in question had misinterpreted the task: rather than following the instructions to select a single topic to present on, the student presented on all three, and did not manage to keep to time. The class marked the presentation highly, despite the basic mistakes that meant a lower teacher's mark. The average teacher mark of 78.4 for this class was slightly lower than the student mark at 79.4, reversing the trend of the first two classes for a lower average peer result.

Data analysis

There is no obvious week-by-week pattern to the variations in peer and teacher results in the 3 groups: one presentation the peer result might be higher than the teacher's, but the next week the teacher's result would be higher. Overall the students gave very similar results to the teacher's across the session. This bunching of marks was no doubt due to the careful allocation of a proportion of marks to different components in the marking criteria, which meant that students were unlikely to deviate too greatly from the teacher's mark. It is worth pointing out that all three of these classes were of elective subjects involving 3rd year students, so the presentations by these students were already likely to be of a high standard, as they had had time to develop their presentation skills and knowledge of the discipline.

While there was an observable trend for slightly lower results for the peer evaluation by students in the first two classes, the close correlation of marks across all three classes reinforces the impression of the validity of peer evaluation. This is interesting in light of the comments volunteered by students in the focus groups (which will be analysed in greater detail in the following section) that they feared that their peers would over- or under-mark and therefore invalidate the peer evaluation process.

Student Focus Groups: Feedback and Analysis

From the beginning of the session, students were informed that the class presentations and use of clickers were the subject of a research project. They were invited to consider participating in a focus group at the end of the session to reflect on their experience, and

reminded that this would not form part of their class participation mark, or was in any way an assessable component for their subject. The focus groups were organised in the last few weeks of the teaching session, for one hour immediately after their class, and were entirely voluntary: there was a relatively small turn out for the focus groups of 3, 5 and 9 participants for the three classes. The focus group discussion was facilitated by the non-teaching member of the research team who recorded the student comments for later transcription, to ensure that the student reflections remained anonymous and reassure them of the lack of impact on their class participation or presentation grades.

The focus groups raised some interesting points regarding their anxieties about peer assessment, which ranged from concern about their own capacity to evaluate their peers to their fears about their presentation skills being judged harshly. The majority were in favour of the peer evaluation of presentations in particular, and appreciative of the use of clickers in this process. However, it was interesting that the discussion circled around recognition of the value of peer assessment for their learning experience, and outright dismissal of the proposal that this be used for a formal assessment task in the future.

Parity of Results

The above breakdown of the overall class marks was not given to students before the focus groups, so they did not have evidence of the parity of the results. A few students did comment that they had noted, and been surprised by, the similarity in their results:

> *My mark was actually pretty similar to the teacher's so it was nice to know that the class is having the same opinions as the teacher.*
>
> *A lot of people had really similar marks – the class's was the same as the tutor's. It was interesting that it really did seem to average out.*

A common concern in the literature with peer assigned marks is that peer assessors have a tendency to over-mark, to assign higher marks relative to the assessor (Boud and Holmes, 1995; Wadhwa, 2003). It has been argued that this inconsistency in peer marking may affect the validity of the peer assessment process. However, this study has revealed a tight correlation between the peer and the teacher mark – if anything, and bucking the trend of previous studies, the peers have returned more lower marks than the teacher.

Student Anticipation of Peer Bias

The majority of students in the focus groups reported fearing that there would be a larger discrepancy between the peer and the teacher results. This was attributed to a number of factors, one of which was student solidarity:

> *You want to be kind to your peers so you want them to do well…I tended to give out higher marks because I didn't want to wreck their marks.*

Another, less generous factor, was the potential misuse of the peer evaluation process. As one student confessed:

> *You could manipulate the whole thing. A clique could just get together and just stick together and all give each other very high marks and raise the marks. That's the first thing that came to my mind when I heard about it.*

Other students were more concerned with student laziness – 'you have a lot of people who would just give any mark' – and anxieties about peer friendships skewing the results:

> *Everyone has their own, you know, friendships, how you felt about an individual. There was a lot of pressure. It was like, if they don't like me, they're just going to give me a bad mark.*

There was a strong sense that peer evaluation would be impacted by the relative popularity of the participants, although nobody could point to an instance when that actually happened, or where they themselves felt biased against or had marked a friend higher. One student qualified the point:

> *But we were a pretty good group, so we could trust each other. In another class, or a larger group, I don't know, maybe you wouldn't like a girl, or a boy wouldn't like you, so…*

This speculation about potential bias according to popularity didn't seem to have affected any of the participating students in the focus groups. Indeed, one student who reported receiving a lower mark from her peers immediately rationalised this result as a positive function of the peer evaluation process:

> *I was quite surprised that my peer assessment was lower then the mark [the teacher] gave me…And I thought, well, people were being honest there. Because I am not someone who is confident in doing presentations and I wouldn't mark myself very highly and so I thought, oh well, then that's good that people are being honest.*

The intriguing shift to note here is that while there may be concern that other people will get over-marked because they are liked, if an individual is under-marked, they internalised it as not a lack of peer support, but because of the poor quality of their work.

Perceived Value of Peer Evaluation

One student volunteered early in their focus group discussion that, while they enjoyed the process of the peer evaluation using clickers, the results themselves didn't have an impact:

> *I didn't take the peer assessment mark into account very much. It wasn't something that bothered me or really interested me.*

This reported lack of interest could have been because the students were told at the beginning of the session that the peer evaluation results would not be part of the formal assessment, and students were repeatedly reminded that the use of clickers for peer evaluation was a trial-run. It is also possible that the student could also have been dissembling lack of interest to their peers in the focus groups. However, the majority of students participating in these focus groups were very interested in the peer results, for a variety of reasons. As one student said, 'the more feedback the better', and another pointed out that the peer results worked to reinforce the teacher's mark – the marks 'were really close, so I was like, that was bad' (the implication being that she might not have as readily accepted the low result as indicative of a poor quality presentation simply from the teacher). Other students were more interested in the results from their peers as they commented on their presentation skills, rather than for the more scholarly benchmarks the teacher might be using:

> *When you get a class point of view, I feel like you value that more, because they are interested in more than what the tutor thinks, and I am just worried about how I'm coming across to who I am talking to.*

Another student in a different focus group made a related point, that the peer marks served to reinforce the teacher feedback, which might otherwise be disregarded:

> *It is good to get feedback from your peers because you know that your tutor is not being ridiculously unfair by saying, like, you didn't speak to everyone. Because if everyone is saying it, then maybe that is something you can work on.*

In fact, a number of students thought that peer evaluation marks could ward against any potential unfair bias on the part of a teacher. As two different male students in different focus groups pointed out:

> *If you had a tutor that you thought was really biased against you, or something like that, or didn't get along with you, if you had the peer assessment that could kind of help you.*

> *When you are assessed by a tutor, you can change from subject to subject. And you can get a good tutor or you can get a nasty tutor. Whereas when you have a class, because there are more people, it is more fair…it averages out.*

The fact that two male students made this unsolicited point was interesting. While there was a much higher percentage of female to male students - both in the classes and in the focus groups - there did seem a perceivable trend for the female students to be less concerned about possible manipulation or misuse of the peer review process, and more concerned with supplying supportive comments or fear that other students might not like them.

It has previously been noted that presentations are particularly useful for studying the influence of presenter qualities such as gender or race (Falchikov and Magin, 1997; Langan et al, 2005). Admittedly, this was a very small sample group, and more research about the particular higher education context of this trial in a Humanities program in Australia, as well as more directed questions to the focus groups, would be needed to come to a strong conclusion about the impact of gender. However, the overall impression from the three focus groups is that the female students were convinced that the popularity factor of other classmates would result in their higher peer evaluation, while the male students invoked the concept of student solidarity to account for the anticipated higher peer results. In this study, both male and female students talked about not wanting to fail their peers, although the male students set this up as a preventative measure, as they 'hoped they wouldn't fail you', while the female students wanted 'everyone to do well'. It is also interesting that the girl who earlier had expressed disinterest in the trial (as one of the first presenters, she had received poor results from her peers, and reported just clicking the same button throughout the rest of the trial) also said that she clicked a high score of 9 (from a range of 1-10) each time. Although disengaged, she did not consider taking revenge on her classmates, who had given her low peer scores.

Another female focus group participant worried about the cumulative effect of negative comments from the entire class:

> *For someone with low self-esteem that would be devastating, to hear all the same things about whatever issue it is that you are struggling with.*

However, another student pointed out that the clicker technology used in this trial of peer assessment, with its limited range of numerical responses to a series of questions that made up the marking criteria, protected against overtly critical peer feedback:

Vol. 3, No. 1 (January 2009)
ISSN 1931-4744 @ Georgia Southern University

> *On the other hand, you can't be too critical with the technology. It doesn't allow you that.*

In this instance, the student was commenting both on the limited range of responses afforded by the clickers (only a numerical response and not a qualitative comment), and also on the way that this protected class-mates from too overtly negative comments, that could potentially be hurtful.

Clickers as a tool for peer assessment

The students saw the clickers as a useful tool for peer assessment. Most reported enjoying using them, and made a number of comments about their novelty and ease of use:

> *I like the clickers, they were fun.*
>
> *It went pretty quick – click click click.*

The students found that the clickers were a good medium for peer assessment compared to other peer reviewing techniques, positively identifying factors such as anonymity and functionality:

> *It was just easy for the peers and easier for the presenter.*
>
> *It was the most functional way to get feedback from the class.*
>
> *I thought it was a bit confronting at first, but it was a better and more efficient way of doing it. It was a lot better than writing surveys.*

Although many of these third year students had experienced peer assessment before, either of group projects or presentations, they had used paper-based reports or surveys to give their feedback, which they did not like (although arguably a more effective way of soliciting qualitative feedback in the form of critical commentary). A couple of students had experienced peer feedback in the form of on-the-spot commentary and discussion immediately following a student's presentation, which they found very confronting. While any peer evaluation after a presentation was considered 'nerve-wracking', the anonymity afforded by the clickers was particularly attractive:

> *I think doing it this way is more honest. Because when you've got a group of six people who have just put a whole heap of work to do something and you are going to have to say ' it wasn't very good'…you don't really say that do you? Because they are standing there and you know you're saying something that is really…but when you are pressing a number… I can say that I didn't really like it but they are not going to know.*

An earlier study of peer evaluation techniques in small classes reported publically displaying the peer results after each presentation, but identified a problem as 'it was observed that students tended to talk about the data, rather than the individual that the data related to' (Banks, 2006, p.375). However, the student quoted in the comment above found the anonymity encouraged more 'honest' and subjective responses, a finding supported by Wadhwa's (2003) conclusions that anonymity in a peer review situation helps relieve social pressure and inhibition, and is expected to encourage accurate, honest and critical response. On the other hand, other studies have shown anonymity to be a 'double edged sword in collaborative learning' (Zhao, 1998, p.311), pointing out that while anonymity might reduce social pressure, it also reduces

responsibility, resulting in careless responses, or 'ruthless' feedback (Bostock, 2000). The majority of students in our study, however, were convinced that the anonymity afforded by the clickers, as well as the detailed marking criteria that organised their individual evaluation of the presentations, guided them towards a more objective response:

> *It moves away from the more personal response. It is probably a lot easer because it is more removed. You are not trying to think of things to say like 'conscientious effort'. You are asked a specific question and how they did it.*

This point about the clickers being easier was reiterated across the focus groups. While this was equated with the speed of the process, the relative ease of use was also an aspect of the limited functionality of the clickers, which only allowed for a numerical response from 1-10, rather than a sentence or comment:

> *You don't have to write comments about factors, you're not straining your brain to think of things to say.*

With this type of comment, students moved away from a discussion about how they felt at the receiving end of the peer evaluation, to a consideration of their role as evaluators:

> *I think it is easer for the peers, because all we have to do is press a button. Instead of like writing a sentence or a few words – this is going to be really bad – but you have to really think of something to say. Whereas a mark on a scale, it's easier to just go, 'yeah, I think that is just an eight'. It's easier, and the quick pace of doing it is easier than having to write a paragraph.*

While the majority of students appreciated the marking criteria framework for the peer evaluation, some students found the numerical responses to be too limiting, and suggested that they would have welcomed the opportunity to add more commentary, either in an anonymous survey or in class presentation of the discussions.

> *This was pretty much pre-made questions and I thought that my thoughts on the presenter's style didn't always fit into this. So, if I wanted to say, you know, the presentation was actually laid out nice, but the information was bad, there wouldn't be an actual vehicle where I could really voice my opinion on that.*

> *It was quite rigid, the technology.*

Clickers and the 'Button Etiquette'

The feedback from the focus groups for our trial identified relief at the possibilities of being 'honest' where previously, with face-to-face feedback, they would be reluctant to be critical. However, some students reported that they were taken aback at the culture of silence that developed around the clickers and the peer evaluation, or what was dubbed by one as 'button etiquette', where everyone was really 'hush hush' about their clicker responses, and some students even made efforts to hide the clickers under the table so their peers could not track how they were responding. This raises some interesting questions about anonymity and risk taking, and follows through with what has been called the 'politeness effect' of using interactive media in the classroom (Wang et al 2008), which argues that better learning outcomes are yielded when subjects receive indirect feedback that is non-threatening and promotes learner 'face'. Generally this indirect method of participation offered by the clickers was considered by the focus group be a positive factor:

> *You know how there are a lot of people who don't talk in class, don't feel able to contribute? This is one technique where students have some sort of power.*

It was considered a good way of engaging the majority of students in the presentations and the process of peer evaluation:

> *Everyone is hesitant about asking questions but this was a way of just getting into it.*

However, the anonymity afforded by the clickers came at the expense of more open or direct discussion in the class about the presenter's performance and was considered a lost opportunity for more qualitative commentary:

> *How it went was: presentation, discussion [of the topic], peer assessment. There was no discussion of the presentation itself, there was no talking about the assessment questions or marks.*

The Peer Evaluation Process as a Learning Experience
The students reported that the peer evaluation process encouraged ongoing and active engagement with the class presentations:

> *Certainly in this subject I paid more attention to the presentations than in any other subjects.*

> *It encourages you to actually pay attention. At other times you just can't be bothered, you are at uni for only just one day and you go in and you think, I know that this person put a lot of effort in but I just can't pay attention today. But with this you have to mark them, so you feel too guilty not to pay attention.*

It was considered a positive for the presenters, too, who had a class that was actively listening to them while they gave their presentations:

> *I think that watching everyone look really engaged was good – even though they have to in this class it was good. In other classes they just sit there and stare at the roof.*

Being reminded of the marking criteria each week also became an advantage for individuals who began to plan for their own presentations:

> *I liked the way you could see the [marking criteria] each week and be reminded about what you were going to be marked on. That was really good.*

Some students also saw that the feedback from their peers about their presentation skills would be useful for their future working careers:

> *Like, in terms of the working environment, if you have to address a whole group of people you want to know if you are reaching the whole group rather than just one person. And if you know that the tutor is the only person who cares in giving the marks, then you might just address them rather than everyone else.*

One student reflected on the learning experience afforded by the peer review process,

as alternately a presenter and an assessor:

> *I did try to be quite conscientious about how I graded people, because this is an aspect that interests me because I am a great big nerd. But, like, from the one side of being a presenter, it pushes you to find out how I can engage people more, how can I get them listening. And from an audience point of view, thinking that I actually have to be paying attention to this person, not just reading the paper, so that I can be doing my peers justice.*

This understanding of the value of developing presentation skills follows through on the general recognition of the importance of oral presentation skills to university graduates, and the potential value of getting habituated to peer review techniques (Campbell et al 2001). It indicates that students had an underlying awareness of the value of this task as formative activity, that enhanced the process of learning. However, although the clickers afforded the opportunity for anonymous and critical feedback, none of the students reported giving below 50%. One student justified this by pointing out that the overall quality of the presentations was excellent. However, as a whole the focus groups reported a reluctance to give poor marks:

> *When working with numbers you see a 5 as 50% and you're not going to fail a classmate.*

The clickers themselves reinforced the division between pass and fail marks, as the numbers on the face of the device were organised in columns from 1-5 and 6-10 and divided by a line. This influenced students to concentrate their marks within the higher range:

> *You just wouldn't go there. The lowest I'd go is 5. It's the yellow line – you don't go below the line!*

While the physical layout of technology influenced student responses into a higher range, some students queried the value of the resulting collated percentage marks in response to the marking criteria, instead of the qualitative comments afforded by written responses:

> *Because 'wasn't clearly organised = 70%' doesn't tell you what you did wrong, or how to get there. Just if the group liked it or not.*

Students' Concern About Their Role as Evaluators

Many students were concerned about their capacity to evaluate, even though as third year students, they had a lot of experience in giving and hearing presentations and a familiarity with the kind of marking criteria they were expected to meet. This lack of confidence was evident in a number of comments across all three of the focus groups:

> *My thought on that would be that the class isn't expert on that, the teacher is, so it would be like I think that their opinion should be weighted less than the teacher's would be, because that's their job and we are just here for a tutorial class.*

> *I don't think that we are the best people to ask about the other stuff about content because … if someone hadn't done the reading, the class wasn't going to notice it.*

> *I think that we were students and had no teaching or marking knowledge.*

> *Really we don't understand all the quantitative method behind marking so it was difficult, so a lot of us just vent on pure feeling.*
>
> *Yeah, which ones were more enjoyable.*
>
> *How can we tell if they had really read the article, or just reading it out now? Either way, we can just say it was interesting.*

They suggested that instead of responding to the content questions that made up half of their marking criteria, the students should just mark on 'entertainment' or overall 'enjoyability' of the presentations:

> *Just use our marks on the entertainment factor*

This was intriguing in the light of earlier reported concerns about potential peer bias when focusing on whether 'the group liked it or not'. The students had already voiced their anxieties about personality being a factor in higher marks for some of their classmates, and their fear of losing out in a popularity contest – an aspect that would surely be reinforced by focusing on the entertainment value of individual presentations. Some students did recognise the paradox inherent in their desire for simpler marking on the level of personal engagement, and the dangers of bias affecting the peer results:

> *You could listen, you could like the topic and you could like the person, and your mark might not reflect the work that went into it or the research. Like, a really well-researched paper might be a bit dull and the person didn't engage that well, you would tend to mark lower.*

Peer Evaluation by Clickers as Formal Assessment

At the end of the focus group discussion, students were asked if they would recommend using the peer assessment process using clickers as part of the formal assessment practice in future iterations of their subjects. After such a thoughtful discussion of the value of the peer assessment process, it was somewhat surprising that the students' immediate response was in the negative:

> *I think that in the same way – that is, not linked to your marks – I wouldn't mind. But if it was, I would really hate it.*
>
> *I know that I would not like to have liked to have it contribute to my overall mark*

When reminded of their earlier positive feedback of the clickers, of how the peer evaluation improved their engagement with the class presentations, and of the reported parity of between student and teacher results, students persisted in their reluctance to formally assess their peers:

> *I think it would be good to have the responsibility, but I wouldn't want to have the marks count, for some reason.*

This reluctance could be simply attributed to students' 'resistance to change' (Nelson and Hauck, 2008 p.60), either of educational technologies or assessment regimes, but it is clear that students' anxieties about peer evaluation are difficult to overcome.

Conclusion

Overall, the focus group discussion revealed that students found the peer evaluation of presentations a rewarding activity, as it encouraged class engagement, critical evaluation, student-to-student interaction, and the improvement of individual's presentation skills. Students appreciated the use of the clickers in the context of this peer assessment activity for their functionality, novelty and, most particularly, anonymity. Given all these positive outcomes, the students' resistance to making peer assessment a formal component of their grades was surprising, although it mirrored reports in the literature of widely held concerns that peer marks are not accurate enough to be used as formal grading procedures, despite studies showing that concerns of over-marking are groundless (Stefani, 1994; Falcovich and Goldfinch, 2000; Campbell et al, 2001). The parity between the peer and teacher results in this trial, acknowledged at the start of each of the focus groups, did not overcome students' ongoing concerns about peer bias, or about their own lack of expertise as evaluators. This anxiety stemmed from the core belief that students are tempted to mark on what they found entertaining, rather than the 'real' value of the presentation. Recognition of this factor did not prevent students in the focus groups from recommending that the marking criteria be adapted to focus less on content and more on 'enjoyment', which risked circling back to the solicitation of responses based on the implicit popularity of individuals in the group.

For the teacher the trial was a valuable exercise. The presentations were pedagogically effective when combined with peer assessment and the focus groups confirmed this. While some minor adjustments can be made to the process, peer evaluations have an important role in an upper level class, even if only to provide detailed feedback to presenters. These are good reasons to continue the practice and recommend it to colleagues, though there are some caveats to bear in mind, particularly if the use of clickers was to be widely incorporated in a major or degree in order to develop students' skills as presenters, listeners, and evaluators. The students' caution in recommending the use of clickers as a part of the summative assessment process suggests that its effectiveness rests on its future use, as part of the process of learning, as a formative learning activity.

This trial also highlighted the challenges with assuming the neutrality of educational technologies. The clickers were originally chosen for their ability to engage students in peer evaluation in an anonymous and user-friendly fashion, with the side benefits of allowing for an easy collation, analysis and distribution of the peer responses to the marking criteria for each individual presentation by the teacher. Their appeal, as an educational version of more familiar mobile communication devices, was anticipated to contribute to students' engagement with the process. However, the neutrality of the clickers in this peer evaluation context was undermined by their very familiarity, as students had been habituated to the process of voting using communication technologies in mainstream television popularity, talent, or game shows. This popular culture contextualisation potentially reinforced student anxieties about personality bias in the peer evaluation process. One proposed solution to overcoming these concerns could be to give students more opportunities to use clickers in peer evaluation activities so that they are familiarised with both the process and its outcomes, which include a significant parity between teacher and class results, as well as an improved and risk-free engagement by students in the class presentations. This, however, threatens to diminish the novelty factor which had a strongly positive student response. Another solution to alleviating student anxieties about peer evaluation would be to show the results of research studies like this, where the trial use of an educational technology to evaluate

class presentations revealed a strong correspondence between class and teacher results. Discussing the outcomes of such studies will not only strengthen student awareness of the nexus between teaching, learning and research, but will lead to more confidence in the positive learning outcomes for activities involving peer evaluation.

References

Banks, D (2006). Reflections on the use of ARS with small groups, *Audience Response Systems in Higher Education: Applications and Cases*, Information Science Publishing, South Australia, Pennsylvania & London, 373-386.

Banks, D (2003). Using keypad-based group process support systems to facilitate student reflection. In G. Crisp, D. Thiele, I. Scholten, S. Barker and J. Baron (eds) *Interact, Integrate, Impact: Proceedings of the 20th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education*, Adelaide, 7-10 December.

Bostock, S (2000). Student peer assessment, The Higher Education Academy PDF, Retrieved Nov 8th 2008 from: http://www.palatine.ac.uk/files/994.pdf

Boud, D (1990). Assessment and the promotion of academic values, *Studies in Higher Education,* 17 (2), 101-11.

Boud, D and Holmes H (1995). Self and peer marking in a large technical subject. In *Enhancing Learning through Self Assessment*, edited by D. Boud. London: Kogan Page.

Brill JM and Park Y (2008). Facilitating engaged learning in the interaction age: taking a pedagogically-disciplined approach to innovation with emergent technologies, *International Journal of Teaching and Learning in Higher Education,* 20 (1), 70-78.

Campbell JE (2007). *Increased Learning In College Lecture Classes: Is It Just A Click Away?* PhD thesis, Department of Psychology, University of California.

Campbell KS, Mothersbaugh D, Brammer C and Taylor T (2001). Peer versus self assessment of oral business presentation performance, *Business Communication Quarterly*, 64 (3), 23-42.

Falcovich N and Goldfinch J (2000). Student peer assessment in Higher Education: a meta-analysis comparing peer and teacher marks, *Review of Educational Research*, 70, 287-322.

Falcovich N and Magin (1997). Detecting gender bias in peer marking of students' group process work, *Assessment and Evaluation in Higher Education*, v22 n4: 385-396.

Greenan K, Humphreys P and McIlveen H (1997). Developing transferable personal skills: part of the graduate toolkit, *Education and Training*, 39 (2/3), 71-78.

Gauthier, G (2004). *Running Head: Web based Self and Peer Assessment*, PhD thesis, Department of Educational and Counselling Psychology, McGill University.

Guthrie RW and Calin A (2004). Waking the dead: using interactive technology to engage passive listeners in the classroom *Proceedings of the Tenth Americas Conference on Information Systems,* New York.

Jacobs, RM, Briggs DH and Whitney DR (1975). Continuous-progress education: student self evaluation and peer evaluation, *Journal of Dental Education* v39 n8: 535-541.

Jaschik, S (2008). Clicker U, *Inside Higher Education*, Retrieved Nov 3rd 2008 from: http://www.insidehighered.com/news/2008/11/03/clicker

Jenkins, A (2007). Technique and Technology: Electronic voting systems in an English Literature Lecture, *Pedagogy*, 7, 526-533.

Langan M, Wheater CP, Shaw EM, Haines BJ, Cullen WR, Boyle JC, Penney D, Oldekop JA, Ashcroft C, Lockey L and Preziosi RF (2005). Peer assessment of oral presentations: effects of student gender, university affiliations and participation in the development of assessment criteria, *Assessment and Evaluation in Higher Education* 30 (1), 21-34.

Lu and Bol (2003). A comparison of anonymous versus identifiable e-peer review on college student writing performance and the extent of critical feedback, *Journal of Interactive Online Learning*, 6 (2), 100-115.

MacAlpine, JMK (1999). Improving and encouraging peer assessment of student presentations, *Assessment and Evaluation in Higher Education*, v24 n1: 15-25.

Martyn M (2007). Clickers in the classroom: an active learning approach, *Educause Quarterly,* 30 (2), 71-74.

Murphy, B (2006). Convergence of learning experiences for first year tertiary commerce students: are personal response systems the meeting point?, *The Journal of American Academy of Business*, 10 (1), 186-191.

Nelson ML and Hauck RV (2008). Clicking to learn: a case study of embedding radio frequency based clickers in an introductory management information systems course, *Journal of Information Systems Education*, 19 (1), 55-64.

Nilson LB (2003). Improving student peer feedback, *College Teaching*, v51 n1: 34-38.

Ribbens E (2007). Why I like clicker personal response systems, *Journal of College Science Teaching*, 37(2), 60-62.

Socol, ID (2008). The instant anachronism, Retrieved on 14th June 2008 from: 2008speedchange.blogspot.com/2008_02_01_archive.html

Stefani, LAJ (1994). Peer, self and tutor assessment: relative reliabilities, *Studies in Higher Education,* 19, 69-75.

Stowell JR and Nelson JM (2007). Benefits of electronic audience response systems on student participation, learning and emotion, *Teaching of Psychology*, 34 (4), 253-67.
Topping KJ (1998) Peer assessment between students in colleges and universities, *Review of Educational Research,* 68 (3), 249-276.

Topping, KJ, Smith EF, Swanson I, and Elliot A (2000). Formative peer assessment of academic writing between postgraduate students, *Assessment and Evaluation in Higher Education,* 25 (2), 149-169.

Wadhwa, G (2003). *Effects of Anonymity and Peer Accountability During Peer Assessment in a Graduate Web-Based Education Research Methods Course,* Masters thesis, Faculty of Education, Memorial University of Newfoundland.

Wang N, Johnson WL, Mayer RE, Rizzo P, Shaw E, and Collins H (2008). The politeness effect: pedagogical agents and learning outcomes, *International Journal of Human Computer Studies*, 66, 98-112.

Yourstone SA, Kraye HS and Albaum G (2008). Classroom questioning with immediate electronic response: do clickers improve learning? *Decision Sciences Journal of Innovative Education*, 6(1), 75-88.

Zhao, Y. (1998). The effects of anonymity on computer-mediated peer review. *International Journal of Educational Telecommunication*, 4(4), 311-345.