2009

# Peer assessment of oral presentations using clickers: the student experience

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

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

## Publication Details

Barwell, G and Walker, R, 'Peer assessment of oral presentations using clickers: the student experience', in Wozniak, H & Bartoluzzi, S (eds), 32nd HERDSA Annual Conference: the student experience, Higher Education Research and Development Society of Australasia, Milperra, Australia, 2009, p 23-32.

# Peer assessment of oral presentations using clickers: the student experience

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

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

This paper reports student reactions to the use of a personal response system (clickers) to provide peer assessment. Trials were conducted in three upper level seminar classes in two different subjects in an Arts Faculty, where students were required to give individual in-class presentations as part of their assessable work. Class members assessed the presenters using criteria based on those used by the tutor, but modified to make them appropriate for student use. At the end of the session some students in the trials discussed their experiences in focus groups. The comments of those focus group participants are analysed to reveal the key issues for the students. Their experience was generally positive, and the comments indicate that using clickers for the purpose of peer assessment can make a significant contribution to student engagement.

**Keywords**: personal response systems (PRS), oral presentations, peer assessment

## Introduction

Oral presentations had been assessed by one of the authors (Barwell) in his upper level undergraduate classes in Media and Cultural Studies and in English Literature over a number of years, but he wanted to find an efficient way of providing presenters with feedback from their audience in addition to that from their tutor. Models for this kind of response had been in the public eye for some years in popular television shows and televised political debates, so these modes were familiar. After some investigation of available technologies and the ways they had been used in educational settings, it was decided to trial a device, commonly called a clicker, which forms part of what is variously called a personal response system (PRS), audience, student, or classroom response system, or an electronic voting system.

With support from an Educational Strategies Development Fund grant from our home institution, we conducted trials of this technology over two sessions, gathering feedback from the student participants by means of focus groups conducted at the end of each session. This paper examines that feedback from the participants on their experiences as presenters and audience members.

Clickers have been part of the technologies available for teachers for some time, but have been used primarily in large classes, usually lectures, to enable the lecturer to assess how well students have understood particular points, to determine their thinking on specific issues, to conduct quizzes or even to track attendance. The discipline areas most commonly associated

with the use of this technology are in the sciences, engineering, medicine and commerce. A body of research has developed around this pattern of usage with several extended literature surveys (Simpson & Oliver, 2002; Judson & Sawada, 2002; Simpson & Oliver, 2007), two books devoted to the technology (Duncan, 2005; Banks, 2006a), and numerous journal articles, conference papers and online resources including bibliographies (e.g., Bruff, 2008). Relatively little of this research relates to humanities or social science teaching, though there are a few exceptions (e.g., Jenkins, 2007, for English literature; Stuart, Brown & Draper, 2004, for philosophical logic). While clickers have been used in teaching smaller groups (e.g., Durbin & Durbin, 2006), we have found only one scholar who has experimented with them for a purpose similar to ours, to provide student presenters with feedback (Banks, 2003; Banks, 2006b). While our results share something in common with those reported by Banks, we used a different approach in a more extensive trial and consequently can greatly expand upon his contribution to thinking about the use of this technology for the purpose of peer assessment.

## Methodology

Three third-year seminar groups of 15 to 20 students per group took part in the trial: two in an Electronic Culture subject in one session and one in a Early Modern English Drama subject in the following session. All classes had the same tutor (Barwell). Each student was required to make a presentation lasting approximately 15 minutes to the class, on a topic selected from those provided in the subject outline, which was distributed in the first week. The presentations began several weeks after classes first met in order to give students a chance to get to know each other and to prepare their presentations. The presentations then continued until the end of the session. Class discussion would follow each presentation, then the peer assessment process using the clickers. This peer evaluation took about 5 minutes.

Presentations were assessed by the tutor on criteria given in the subject outline with portions of the mark assigned to content, delivery, timing and the use of visual aids or handouts. A detailed report sheet was given to each presenter in the week following the presentation. Only the mark assigned by the tutor was counted towards the final mark in the subject, so the peer assessment mark was solely for feedback.

The criteria used for the peer assessment had been developed with a class the previous year and were made available to the trial groups via the subject's eLearning website as well as being displayed in class each week during the peer assessment process. The criteria were modelled on those used by the tutor but adjusted to be more suitable to the expertise of the student audience. For example, audiences were asked to assess the content in terms of whether the presentation offered a clear viewpoint, an adequate depth of response and level of research, provided suitable support for its claims, and was appropriately pitched and well organised. Delivery was assessed by whether the speaker appeared interested, spoke at an appropriate speed, addressed the whole class, and spoke clearly and audibly. The question on timing was omitted, since this was more easily measured by the tutor.

After class discussion early each session, it was decided that students would respond to the questions with a number ranging from 1 (very negative) to 10 (very positive), though the response to a question on the effectiveness of visual aids was changed to a simple yes/no answer during the third class, after students reported difficulties in assigning an evaluative number. In class, the tutor could see on the laptop screen whether all clickers had responded before moving to the next question. Clickers were distributed randomly, so all responses

remained wholly anonymous and students became familiar with their use in the week before presentations began.

The clickers resembled a television remote control and offered students the choice of ten buttons to push for their response. They could change their response up to three times. Their responses were sent wirelessly to the tutor's laptop via a receiver and software supplied with the clickers.

The software automatically produced a tabular list of responses from each clicker to each question at the end of the peer assessment. This list was transferred into an Excel spreadsheet which calculated both an average peer response to each question and a total weighted mark. The spreadsheet was required because the clickers were being used in a way which the designers of the supplied software did not seem to anticipate. After discussion with each class, it was decided that the numerical total would be added next to the tutor's total on the feedback sheet, with students given the option of getting on request a more detailed breakdown of their peer assessment via the spreadsheet.

| clicker id | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 | Q11/12 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (ignore) | | | | | | | | | | | | |
| 402618 | 9 | 8 | 8 | 7 | 9 | 7 | 10 | 8 | 8 | 8 | 2 | |
| 404098 | 8 | 8 | 8 | 8 | 8 | 7 | 9 | 9 | 9 | 8 | 2 | |
| 404155 | 8 | 7 | 7 | 7 | 8 | 7 | 7 | 9 | 6 | 9 | 1 | |
| 404221 | 7 | 8 | 8 | 6 | 8 | 7 | 8 | 9 | 8 | 9 | 1 | |
| 404724 | 9 | 8 | 8 | 8 | 8 | 7 | 10 | 10 | 6 | 10 | 2 | |
| 404756 | 7 | 9 | 9 | 8 | 7 | 8 | 7 | 8 | 7 | 6 | 1 | |
| 404784 | 8 | 10 | 9 | 10 | 10 | 8 | 10 | 10 | 10 | 10 | 2 | |
| 404815 | 7 | 7 | 6 | 8 | 9 | 7 | 9 | 8 | 10 | 9 | 1 | |
| 405147 | 5 | 5 | 5 | 4 | 5 | 5 | 8 | 7 | 8 | 8 | 1 | |
| 409137 | 8 | 7 | 8 | 8 | 8 | 7 | 9 | 9 | 10 | 9 | 2 | |
| 409140 | 7 | 7 | 7 | 7 | 7 | 7 | 6 | 7 | 7 | 7 | 1 | |
| 409624 | 9 | 8 | 9 | 9 | 9 | 7 | 10 | 9 | 9 | 9 | 2 | |
| 409695 | 10 | 8 | 9 | 8 | 9 | 7 | 8 | 9 | 8 | 9 | 1 | |
| | | | | | | | | | | | | |
| Average | 7.85 | 7.69 | 7.77 | 7.54 | 8.08 | 7.00 | 8.54 | 8.62 | 8.15 | 8.54 | | |
| | | | | | | | | | | | | |
| Questions 1-10 are of equal value; Q11/12 reported separately | | | | | | | | | Total respondents = | | | 13 |
| Q1-6 are on content, 7-10 on delivery | | | | | | | | | | | | |
| Sum of averages Qs 1-6; Qs 7-10 | | | | | | 45.92 | | | | 33.85 | | |
| % of 60 for content; of 40 for delivery | | | | | | 76.54 | | | | 84.62 | | |
| | | | | | | | | | | | | |
| Final total for peer assessment (= sum of averages for each question on content and delivery) | | | | | | | | | | | | 79.77 |

Figure 1: Spreadsheet showing peer responses for one presentation.

At the end of each session students in the trial classes were invited to take part in voluntary and anonymous focus group discussions led by the researcher who had not been their tutor (Walker), and about one quarter of the students took up the invitation. These discussions covered the students' reactions to the technology, its use for the provision of peer assessment of presentations and the place of peer assessment in final marks for a subject. The students' comments (in italics) are considered in the next section of this paper.

## Student comments

Students found the clickers easy to use, commenting that they "fit in your hand nicely", were "fun" and "something different". Those who had experienced other methods of getting peer assessment were generally positive about the advantages of this method.

> I think it easier for the peers, because all we have to do is press a button instead of like writing a sentence or a few words . . . [With] a mark on a scale, it's easier to just go 'yeah I think that this is just an eight'.

It was recognised that pressing a button meant that "you can't be too critical with the technology", but the advantages of anonymity were significant.

> In one of our other subjects [the tutor] asked what we liked or didn't like about our own presentations and then the class also responds. So it is more like a general discussion about what happened, which is kind of good. Except you have to stand up there and think of things to say, and you don't want to be all 'yeah, I was great'.

Anonymity was advantageous compared to the open discussion of group presentations where the feedback is not directed towards one individual.

> I think doing it in this way . . . is more honest. Because when you've got a group of six people who have just put in 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?

The literature on personal response systems frequently notes the advantages of anonymity in producing more honest feedback, often in student groups where loss of face is a significant issue (e.g., Zhao, 1998). While loss of face is not the determining factor here, the anonymity and confidentiality of the responses mean that students are much more candid. We suggest that this is why our methodology in this trial did not produce the results Banks (2006a) observed, where his students were required to openly discuss an individual's results as the feedback on the presentation was made available to the whole class. Banks noted that in these circumstances the students tended to "talk about the data, rather than the individual that the data related to" (p. 375), even though the feedback under discussion had been provided anonymously.

The methods we adopted in our use of the technology produced some interesting and incisive comments from students, not all positive. The software could deal with responses to only one question at a time, so, in order to have the same number of respondents to each question and thus maintain a consistent process, it was necessary to wait until every audience member had responded before moving to the next question. This caused some adverse comment, with the system of coloured lights on each clicker indicating whether the response had been received not being as effective as students would have liked. Students who were used to texting on mobile phones found the process irritatingly slow. In one class the audience questions were read out each week to enable a sight-impaired student to take part, and this was recognised in the focus group discussion as a factor in the speed of the assessment process.

Consistency of process with set criteria was critical to ensure fairness in the trial and most students recognised it was important that there was some correlation between the questions answered by the tutor and the student audience, even if this produced a certain rigidity. As one student said:

> I don't think you could do it with less questions. I'd add more, but I don't think there is anyway of getting around taking up time in class.

It was also useful for the students to see the questions and criteria each week since this aided their learning.

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

The approach used meant that some of the behaviours reported in the literature were also reported by the students. One student confessed to a form of "button fatigue" (Banks, 2006b. p. 379), but rather than the requirement to input too much data, cited laziness as the cause and wondered if:

> everyone is like me and kind of picking a number at random, pressing anything on the thing because they get lazy.

Another revealed the pattern described as "response set" (Banks, 2006b. p. 384), as a result of finding the task of assessment too difficult.

> I think in general it got, like, it was just too much in the end. Like, this was probably really bad…we had to mark out of 10 but I just picked one overall and said it for each question.

But most of the participants in the focus groups reported taking great care over the assessment of their peers.

> I mean, I did try to be quite conscientious about how I graded people, because this aspect interested me because I am a great big nerd.

Not unexpectedly, given that a numerical answer was required for each question, some students found it difficult to decide how to allocate their marks. Although they had been consulted on the nature of the questions before the presentations began, it nevertheless proved a challenge when they came to make their assessments each week. If they made this feeling known during the session it was possible to make adjustments to the questions, as happened with one question in the third class. As a result, the students in this class felt more ownership of the assessment process.

> Yeah, [the question] was worded really weirdly, so we changed it to a yes/no. It was like rate the effectiveness of the visual aid from 1-10, and we were like, is it effective or not, so we changed it.

The methodology we followed did cause some students concerns over possible manipulation of the assessment by the class.

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

There was no evidence that this happened, possibly because the peer assessment did not count towards the final mark. The disadvantage of being the first presenter was also a concern, as shown in this statement by one student (A) and retort from another (B).

> (A) It would be a bit harder if you're the first person presenting because you haven't really been able to see the criterion that you are being marked for. But later on you get used to it, so you say, right, these are the things.
> (B) But all of those questions, they are in the subject outline.

The decision to deliver peer feedback to individual students in the form of an overall numerical score with the option of the spreadsheet on request was arrived at in discussions with each class at the start of session. This had the advantage of not burdening students with detail, but some presenters felt reluctant to request the peer evaluation spreadsheet, though the tutor was happy to provide it.

> It seems to have been a hassle to ask. I would have been interested but I didn't want to seem too keen.

Others wanted it only if the peer assessment was to contribute to their final mark. Whatever reservations they had about the methods, the focus groups were all agreed on the value of the peer assessment.

> 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.

The process was seen as a valuable counter to any bias.

> The other thing is .... you can get a good tutor or you can get a nasty tutor. Whereas when you have a [peer assessment component], because there are more people, it is more fair.

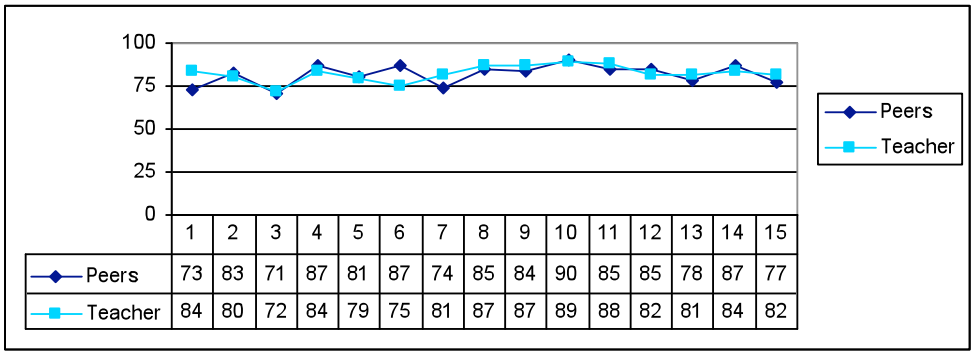
The students anticipated that the peer results could be fairer than a tutor's, and would therefore be more useful in encouraging improvement.

> So it is good to get hear back 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.

Students were particularly pleased when they found the peer assessment total was similar to the mark assigned by their tutor.

> I thought it was good to have the feedback and 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.

As it happened, we found over the three classes that the variation between the tutor's mark and the peer assessment mark was very low. In two classes the average peer marks were slightly lower (by 0.7% & 0.6%); in the third they were slightly higher (1%). The typical nature of the difference is shown in Figure 2 below, where the average peer mark was 81.7%, the average tutor mark was 82.3%, and the greatest variation between the tutor and the peers in the marks for the 15 presenters was a difference of 12 marks (presenter 6 misunderstood the question but still gave a lively presentation).




| | 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 |

Figure 2: Comparison of peer and teacher results for a sample class.

It was clear that the peer assessment process had a powerful effect on most students. It made them think more about their own practice as presenters.

> But, like, from the one side of being a presenter, it pushes you to find out how can I engage people more, how can I get them listening.

It had a similar effect to their sense of themselves as assessors.

> You have to actually think about the presentations a bit more because you know you are going to have to answer questions about it so obviously, you are . . . thinking of a few things, and after a while you remember what you are supposed to be marking them on, so you are thinking, alright, are they speaking clearly and things like that.

Another student agreed, pointing out that they felt guilty if they did not participate fully:

> It encourages you to actually pay attention. At other times you just can't be bothered. You are at uni for . . . 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.

Having an attentive audience was beneficial to the whole process.

> And I think that watching everyone look really engaged was good . . . In other classes they just sit there and stare at the roof.

And as one student observed, it was valuable feedback in their development of a communication skill they would need outside university.

> Like, in terms of the working environment - if you have to address a whole group of people you want to know that you are reaching the whole group rather than just one person.

Despite the recognition of the value of the exercise, there was a fairly constant anxiety expressed in all three groups about the capacity of the audience to assess their peers. On the one hand students felt comfortable about a general response:

> I don't think that there was anyone in our class who gave a presentation that was not good…. I mean even if I hadn't have been doing the clicking thing, I would have said, yes, that was a really good presentation. I don't think anyone did something that meant that we all just sat there and said, well, that was crazy.

But concerns over assessing specific aspects of the content of presentations were common. One question asked about the level of research displayed.

> Well, first of all, we don't know how much research they did and they just throw up an internet image, or someone might think that if there's heaps of information on the PowerPoint presentation, they might think that's research….

Some students were able to articulate clearly how they would judge this.

> Yeah, so I'd give people a high mark for their research if they have acknowledged 'this is the idea I've had and this is where I've got that from'. Whereas for people who just talked constantly without saying where they got this piece of information from…well, you assume that in their head they've referenced where they've got it from, and in the piece of paper that they are reading from it is separated. But when they are speaking it is not so apparent.

Even if they regularly made judgements about a fellow student's presentation without necessarily basing this on explicit criteria, such as referring in passing to a presentation being well-researched, their lack of confidence in their assessment was noticeable.

> We're not professionals, we are just students.

What they were happy to judge was what they referred to as the entertainment value of the presentation, and this is a feature of their approach to peer assessment that we will address in a follow-up paper.

Their reluctance to have faith in the combined judgement of the class, together with the fear of lazy or biased marking by their peers, probably made the group discussions unanimous in their rejection of the peer assessments forming part of the final mark. This was true even when the similarity between the peers' and tutor's marks was pointed out by the facilitator, as evident in this exchange.

> (A) But I wouldn't want the marks to count.
> (B) You still have a lot of people who would just give any mark.
> (Facilitator) But you started out by saying the marks all averaged out between tutor and peers?

(C) But we were a pretty good group, so we could trust each other. In another class . . . maybe you wouldn't like a girl, or a boy wouldn't like you ….

This exchange, while disappointing, did not necessarily undermine the trial. It did, however, underline the ongoing concern students felt at the prospect of peer assessment, despite their earlier positive comments about the value of this activity.

## Conclusions

Although we have not covered the issues this trial raised for curriculum and subject design, for staff and for the institution, the comments in the focus group discussions show that clickers can be used very effectively in small classes for the provision of peer assessment. They can clearly contribute to greater engagement of students in the process of giving presentations and improving their skills as communicators and listeners. They are easy to use and relatively unobtrusive. Students responded very positively to their use, finding them novel and different, though this could change if they were adopted widely in an institution and used in every class.

As Banks (2003) found, close involvement by students in the development of their questions helps establish ownership of the process. In future, we will spend more time workshopping the questions at the start of the session, in order to improve students' confidence in their ability to assess their peers fairly. Nevertheless, the close correlation between the tutor's and the peer marks was significant. It may not be possible to change students' reluctance to include peer assessment in their final mark until they become more familiar with clicker use, or their anxieties about peer assessment are alleviated. There is a danger here that widespread use in a degree program, while making the technology familiar, might also give rise to more extensive examples of some of the lazy patterns of clicker use we found. Teachers would need then to take steps to monitor and address such behaviour.

We found that the anonymity and confidentiality of the process was beneficial in producing candid feedback from the student audience. The close correspondence of the peer and tutor marks in the trial shows that the clickers were successful in producing valid peer assessment. While the way we used the clickers to produce this feedback is not well established in the literature on clicker use and produced some reservations in the students, we believe that using the technology for this purpose has great potential to increase student engagement in an important aspect of their learning in small classes.

## Acknowledgements

We wish to thank the students for their participation in this project and the anonymous referees for their comments on a draft of this paper.

## Reference

Banks, D.A. (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* (pp. 37-46). Adelaide: ASCILITE.

Banks, D.A. (Ed.) (2006a). *Audience response systems in higher education: Applications and cases*. Hershey, PA: Information Science Publishing.

Banks, D.A. (2006b). Reflections on the use of ARS with small groups. In D.A. Banks (Ed.), *Audience response systems in higher education: Applications and cases* (pp. 373-86). Hershey, PA: Information Science Publishing.
Bruff, D. (2008). *Classroom response system ("Clickers") Bibliography.* (Retrieved 05/11/2008). http://www.vanderbilt.edu/cft/resources/teaching_resources/technology/crs_biblio.htm.
Duncan, D. (2005). *Clickers in the classroom: How to enhance science teaching using classroom response systems.* San Francisco: Pearson Education.
Durbin, S.M. & Durbin, K.A. (2006). Anonymous polling in an engineering tutorial environment: A case study. In D.A. Banks (Ed.), *Audience response systems in higher education: Applications and cases* (pp. 116-26). Hershey, PA: Information Science Publishing.
Jenkins, A. (2007). Technique and technology: Electronic voting systems in an English literature lecture. *Pedagogy,* 7(3), 526-533.
Judson, E. & Sawada, D (2002). Learning from past and present: Electronic response systems in college lecture halls. *Journal of Computers in Mathematics and Science Teaching,* 21(2), 167-181.
Simpson, V. & Oliver, M. (2002). *Using electronic voting systems in lectures.* UCL Internal report. (Retrieved 05/11/2008) http://www.ucl.ac.uk/learningtechnology/assessment/ElectronicVotingSystems.pdf.
Simpson, V. & Oliver, M. (2007). Electronic voting systems for lectures then and now: A comparison of research and practice. *Australasian Journal of Educational Technology,* 23(2), 187-208. Also online. (Retrieved 05/11/2008) http://www.ascilite.org.au/ajet/ajet23/simpson.html.
Stuart, S.A.J., Brown, M.I., & Draper, S.W. (2004). Using an electronic voting system in logic lectures: One practitioner's application. *Journal of Computer Assisted Learning,* 20(2), 95-102.
Zhao, Y. (1998). The effects of anonymity on computer-mediated peer review. *International Journal of Educational Telecommunication*, 4 (4), 311-345.