Observations on Peer Evaluation using Clickers

Kari-Jouko Räihä, Saila Ovaska
Unit for Computer-Human Interaction (TAUCHI)
Department of Computer Sciences
FIN-33014 University of Tampere, Finland
{kari-jouko.raiha, saila.ovaska}@cs.uta.fi

David Ferro
Weber State University
Ogden, Utah 84408
USA
dferro@weber.edu

ABSTRACT
Clickers and paper forms were used in a senior level HCI class for peer evaluation of student project presentations. Scores given by teachers and students were compared and subjective opinions of students analyzed. Students gave higher scores than teachers and scarcely used the higher and lower ends of the scale. The medium used for evaluation did not have a significant effect on the scores. Students preferred the clickers to paper because of their novelty, possibility of immediate feedback, and easier preservation of privacy. Concerns were raised over lack of individual feedback in the clicker device and limited added value over traditional pen and paper.

Categories and Subject Descriptors
K.3.1 **[Computer Uses in Education]**: Collaborative learning, Computer-managed instruction; H.5.2 **[User Interfaces]**: Input devices and strategies; H.5.3 **[Group and Organization Interfaces]**: Synchronous interaction

General Terms
Measurement, Performance, Design, Experimentation, Human Factors

Keywords
Clickers, peer evaluation, interaction, privacy.

1. INTRODUCTION
While recent advances in educational technology, for instance web-based instruction, have shown potential in engaging students in active learning outside the classroom, lectures are still seen as the prevalent method of teaching. However, lectures focus more on content transmission to students than scaffolding their active learning [10].

Clickers (also called audience response systems (ARS), personal response systems (PRS), and student response systems (SRS)) are one approach to employing active learning in the classroom [6, 8].

Studies on clicker systems seem to converge in their findings: the introduction of clicker technology in the classroom both improves learning and makes the students more active in participating in class, thus giving the teacher additional feedback on their progress. However, reviewers of the field studies also agree that the studies have not been systematic in their evaluation, and thus it is not certain whether the positive outcomes are due to the adoption of clicker technology, or some other improvement in the lecture settings and teaching. [3]

Based on our literature review of the use of interactive technology in computer science instruction, there are examples of programming classes adopting clicker technology. For instance, Java classes in a university [7], and among high school students [11] have used the technology for eliciting the correct answers from students on various concepts.

The use of the technology in Human-Computer Interaction (HCI) instruction may require a different approach. While the content of HCI lectures varies from concepts and theories to critique on existing designs and small design exercises, often there is no single correct answer like there would be in an algorithms class. Thus, applying clickers in HCI lectures seems to call for other meaningful questions that can be asked as part of the lecture. It was surprising to find so few directly relevant papers on clicker technology in the field of HCI.

We used clickers in a senior level HCI class for evaluating project presentations. Students provided peer evaluations and three teachers did the same evaluation for comparison. A control evaluation had students use pen and paper instead of clickers. We were interested in seeing whether and how the scores would differ, whether the medium would have an effect on the scores, and how the student felt about the peer evaluation, e.g. concerning privacy.

We present related work on clickers in Section 2 and our course and data collection in Section 3. The results of the analysis are given in Section 4 and discussed in Section 5. We conclude by summarizing the findings.

2. RELATED WORK
The technologies that can be adopted in the classroom vary considerably. Of those technologies, clickers are some of the most limited – they only allow sending numeric feedback to the questions posed by the teacher. A clicker commonly denotes a small wireless device with a numeric keypad. Clicker-type questions for collecting student feedback have been implemented with more advanced technology as well: for instance, laptops. Pargas and Shah [9] have implemented classroom "clicker" questions as single-choice, multiple-

option radio button controls in a web application that is used by the students in the classroom.

However, other kind of student feedback is often longed for in lectures, and some researchers have especially targeted that need with systems far more powerful than clickers, so that even those students who do not want to interrupt the flow of the lecture can ask questions. For instance, in the Classroom Feedback System [1] students who have problems understanding the slide content can write questions and other feedback anonymously on the slide being discussed or the previous one. The annotation does not show up in the public presentation but only on the lecturer's own display. Since the slide forms a persistent and shared frame of reference to both the lecturer and the students, the question will be shown in the right context and it can be very effective in guiding the lecturing [1].

In their recent review of the research literature in the field of classroom response systems, Fies and Marshall [6] point out that the studies generally emphasize individual feedback instead of small group use of response technology. Another finding of the review was that while the studies often entail comparisons between traditional practice and the new response systems, they do not consider issues like anonymity of responses that have become possible with the help of the individual handsets.

Several commercially available clicker systems have been implemented (see [2] for a review of six of the most popular models). The detail and presentation format of the results and their summary capabilities vary across systems [2]. However, the functionality offered in them is often similar [10]. The student selects one of the options to answer the question posed, and the results can be displayed to the students at once if the teacher so desires.

Clickers are used widely in all kinds of classes, from groups of 15 students to large audiences with more than 200 participants [3]. Lots of practical advice on how to use them has been accumulated in various field studies [2, 5], even on the level of what kind of questions are effective as clicker questions [3].

Peer assessment has shown high correlations in the resulting scores given by the peers and the teachers, thus indicating validity of peer assessment [11]. However, in most of the peer assessment studies reviewed by Tseng and Tsai [11] the assessment was conducted with pen and paper, not electronically. Furthermore, the study by Tseng and Tsai focuses on an on-line programming course where peer feedback was collected and given in several phases and the students worked in different roles in the phases, and their focus is on the changes in the student projects that are done based on the outcomes of the peer feedback.

Peer assessment can be taken outside the lecture hall so that the students have more time to work on their feedback. If written assessment is allowed, the students can give corrective feedback to point out some errors, reinforcing feedback to praise the solutions' good parts, suggestive feedback to point out possible improvement, and didactic feedback which relies on the theories discussed in class but not directly relating them to the problem areas needing improvement. [11]

One of the goals of collecting peer assessments is to reduce the teacher's work load. When giving and receiving peer assessment, the students learn about the others' work, and become more skillful in their own projects. [11]

In one algorithms and data structures course, peer assessment is collected, for instance, with small in-class exercises. All students need to submit a solution and the solutions are then peer reviewed with several clicker questions: are they correct or not, are they understandable or not, and should they be discussed in class. Since the teacher sees this feedback, the remaining class segment can be directed to topics that are of interest – e.g., those solutions that receive the most varied responses from the students. In addition to peer assessment and reviews, similar clicker questions can be posed based on the algorithms presented by the lecturer. For instance, the students can try to convince their neighbors about the correctness of their own answer in peer discussions, or the teacher can take up some of the solutions to a larger group discussion. Such an approach to teaching the content gives the teacher relevant feedback of how well the students are learning that can be taken into account before proceeding to the next topics. Furthermore, the students find the clickers useful in their own learning. [9]

3. DATA COLLECTION

3.1 Course Description

Our experiment was carried out in connection to a course on New Interaction Techniques. It is a senior level course in HCI at the Department of Computer Sciences in the University of Tampere, Finland. The goal of the course is to give the students an overview of some active research themes in HCI. Additional objectives are that they should improve their skills in critical reading of research papers and in putting new research results into context.

The course covers four main themes: (1) gaze-based interaction, (2) ubiquitous computing, (3) tangible interfaces, and (4) interface design for small and large displays. The first theme is heavily based on research done at our department, the third theme is mainly based on published studies done elsewhere, and the second and fourth theme are a mixture of both.

The course is given yearly, but the teaching methods vary somewhat from year to year. It is being taught by one person (the first author) over seven weeks (the second of the four teaching periods of the academic year). In 2007, the coursework consisted of four elements: (1) four introductory 2-hour lectures, each giving an introduction to one of the main themes of the course; (2) student projects, where students worked in groups of one to three persons to produce a Wikipedia-style web page on a specific topic; (3) 15-minute presentations in class by the student groups on the topic of their project; and (4) an exam based on required readings consisting of six original articles in the four themes of the course. The introductory lectures were not required but were, nevertheless, well attended. In grading, the project counted for 50%, the presentation in class for 20%, and the exam for 30% of the final grade.

The course is part of a Master's Program in Interactive Technology offered in English. Consequently, the participants came from eight different countries. Out of 36 participants, 23 were Finnish and 13 came from abroad.

For more information on the course, see http://www.cs.uta.fi/nit.

3.2 Project Work

The students were offered 24 topics for projects to choose from, six in each of the four themes of the course. In addition, they had the option to propose their own topic, but only one student made use of that possibility. 16 projects were carried out. The topics varied from cutting-edge themes with just a handful of material to draw from (e.g., "Environmental control by eye tracking") to more survey-type topics (e.g., "Multi-touch interfaces").

The instructions for the web pages produced in the projects asked the students to produce material that would be suitable for self study of their topic. They were asked to base their work on recent material available on the web and in the literature. They were asked to include links, references, pictures, and examples in their web pages.

The process of choosing a topic and working on the project was supervised through discussion forums in Moodle. Each group had its own forum (visible to all course participants). They were required to post an initial message describing the main sources they were going to cover in their project, and also an outline for their presentation in class. The teacher had to acknowledge that the group was on the right track before they could proceed to work on the topic in detail.

Discussion in Moodle was encouraged, but was not active beyond the required initial screening step. In only one case the group had misunderstood the topic and had to radically revise its plans; in two other cases the teacher provided pointers to material that would be more up-to-date or relevant than the group itself had found. Altogether, there were 145 postings in the project-specific discussion areas. 119 were from the students and 26 from the teacher. Most postings were of administrative nature, concerned with formation of groups and making changes to the presentation schedule, in addition to the messages related to the checkpoint for proceeding.

This lack of discussion and lack of seeking advice from the teacher (which was offered) was not unique to this course. Other teachers teaching courses for the same body of students have had the same experience: the students have preferred to work on their own and submit a finished project for evaluation, instead of exposing preliminary versions to the teacher.

3.3 Evaluation of Project Presentations

Since the goal of the course is to provide a window to cutting-edge research, the content has to be updated yearly. The course has been given twice before without the project work element. It was observed that while the students appreciated the fact that the teacher had made available the latest information in structured form, they many times had good suggestions themselves on recent material that would have been useful for the themes covered. Moreover, it is certainly not a modern style of teaching to have the teacher pour the information into the students' heads – making them find and learn the material themselves is more beneficial for their education. The challenge in the projects that were introduced for this purpose is to make the information gathered available to all students, so that they learn from the work of all groups, not just their own.

The project presentations in class were introduced to serve this goal. They were to last for 15 minutes, including a few minutes for questions. The students were encouraged to make the presentations interesting by including videos and illustrations in their presentations.

The projects were presented in four meetings, and students were allowed to skip at most one of the four. At the end of each class, each presentation was evaluated by the teacher and by every student. The last two classes were attended by two additional teachers (the last two authors) who also graded the presentations independently. Of the three teachers, two were Finnish and one was American. Our evaluation is based on the last two classes where all three teachers were present.

Each presentation was graded on two aspects: content and presentation quality (called simply just quality in the sequel). For content, evaluators were asked to base the evaluation on the following.

- Was it informative – did you learn anything?
- Was it on the topic that the title promises?
- Did there seem to be a reasonable effort in finding material for the presentation?
- Was it timely, not outdated sources?

The quality, in turn, was judged by the following criteria:

- Was it understandable?
- Was it interesting?
- Was the time used well (not exceeded and balanced well between the various parts)?

Each evaluator (students and teachers) gave each presentation (except their own) two scores, one for content and one for quality, between 1 and 5. They were asked to use the following scale:

1. Fair, but could easily have been improved
2. Reasonable, some small problems
3. OK, on the expected level
4. Good, some nice points or features
5. Excellent, great effort

Students were told that their fellow students would not see the scores, but that the teacher would know who gave which score to each presentation. This was done to encourage the students to take the scoring seriously, and to encourage them in being honest in their ratings and using real judgments. Students were told that the final score in grading the course would be based on the scores given by the teacher, and not on, for example, averages of scores given by students.

3.4 Technology

In the first three classes for project presentations, students gave their scores at the end of the class using clickers produced by Turning Technologies (see Figure 1).

The clickers work by sending an RF signal, picked up by the dongle inserted into the USB port of the voting supervisor. In class, after a test phase where the students were asked to press any button on their clicker to verify that it worked, they were

shown two voting slides for each presentation, one for the content and another for the quality.

When voting started, votes were collected until the vote count had stopped to grow. Students could abstain from voting if they wished, and this was not controlled. In practice, almost all students seemed to vote. Unfortunately, a couple of malfunctions prevent an exact analysis of vote counts vs. students attending the class.

Figure 1. TurningPoint clicker by Turning Technologies (with dongle for the server in the background).

By default, after the voting is closed, TurningPoint shows the distribution of votes on the screen (by popping up a bar graph on the slide that displayed the voting options; see Figure 2). This screen was fast forwarded by the teacher so that students could not see the distribution of votes. This was done so that project groups would not suffer or benefit from the fact that the students could have seen the vote distributions for previous groups.

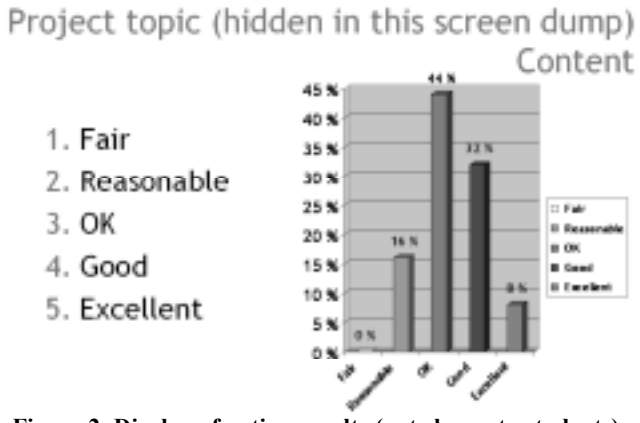

Figure 2. Display of voting results (not shown to students).

At the end of the class, each group that had presented its project was given the chance to see the distribution of the votes given to it (but not to the other groups). Some, but not all, groups made use of this possibility.

After the class, each student returned the clicker. A sheet they filled out showed the id of the unit that they were using.

The fourth class with project presentations was carried out without the use of clickers. Instead, students gave the scores using a scoring sheet that was distributed to them in the beginning of the class and collected at the end of the class.

3.5 Subjective Opinions

To scaffold student opinions on using clickers in class, the final exam had a bonus question worth 4 bonus points (the three elements described in Section 3.1 – project, presentation, and exam – were worth 100 points). The students were asked for extensive feedback on the experiment with the following questions:

a) Both clickers and a paper form were used for evaluating the student presentations. Which one did you prefer? Why?

b) Did you like the way the clickers were used in this class? Why or why not?

c) Were you interested in seeing the voting results for your own group? Why or why not?

d) Would you have liked to see the voting results for all groups?

e) Were there any privacy concerns for either the clickers or the paper form? Did you try to hide your voting behavior from those sitting next to you? If so, how did you do it? If it was not a concern, explain why you were not worried about it.

f) How do you feel about the idea of having students give scores to the presentations of their fellow students (independently of the medium used for scoring)?

g) Compare use of the clickers to other means of class interactivity – specifically online discussions and oral in-class discussions. Which do you prefer for what purposes?

h) Would you like to have clickers used in a normal lecture situation to increase interactivity? What (if any) would be interesting and useful ways to use them? If you don't like them, explain why.

21 students (13 Finns, 8 international students) answered the bonus question and shared their opinions. Only one student did not use this opportunity. The rest of the students in the course will take the make-up exam in the spring.

4. RESULTS

4.1 Evaluations

The average scores given to the 10 evaluated presentations studied in the analysis are shown in Figure 3. Presentations evaluated using clickers are denoted by C1–C5, and presentations evaluated on paper are denoted by P1–P5.

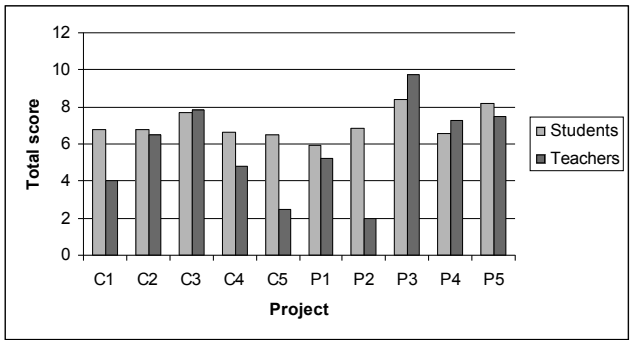

Figure 3. Average scores given to each presentation by students and teachers.

As can be seen in Figure 3, P3 was the most highly ranked presentation by students and teachers, but after that the rank orders differ. However, the student scores are so close to each other, with averages ranging from 5.95 (for P1) to 8.38 (for P3) that they do not really bring out the differences between the presentations.

Overall, the average of all scores given was 3.6 by students and 2.8 by the teachers. Figure 4 shows in more detail the distribution of scores. The figure shows the distributions separately for Finnish students, international students, and the teachers.

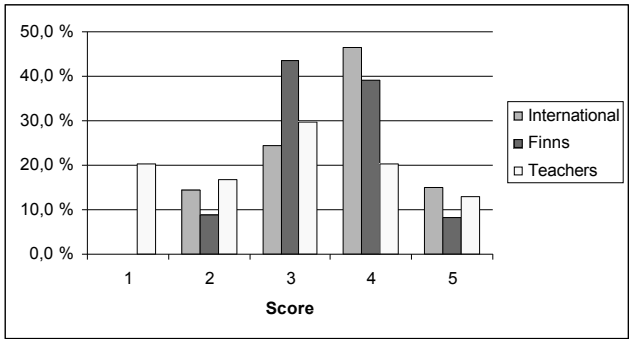

Figure 4. Distribution of scores by all evaluators.

For teachers, the distribution of scores for content and quality were fairly similar (averages of 3.00 and 2.85, respectively). The same held true for students, the difference being that the averages were higher (averages of 3.74 vs. 3.55). The differences in scores for the two criteria were not significant: in general, a presentation that was ranked high or low on content was ranked similarly also on quality of presentation by teachers and students alike. The average difference between the two scores was .56 both in the scores given by students and in the scores given by teachers.

Figure 5 shows the average scores given using clickers and paper evaluation (teachers used the paper form in all cases).

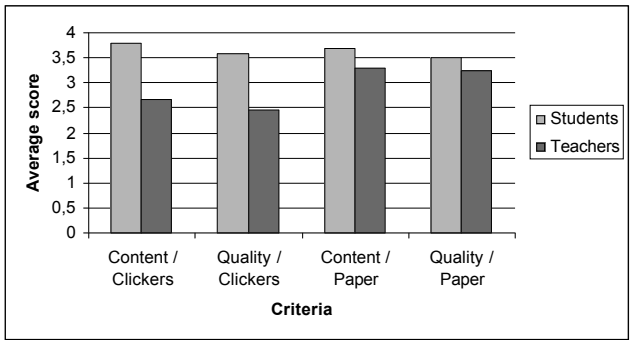

Figure 5. Scores for content and quality given using clickers and paper evaluations.

4.2 Subjective Opinions

The question of which evaluation technique – clickers or paper form – was preferred ended in a tie: both got 10 votes, with one undecided. Among those who preferred clickers, a few people mentioned that they needed to take notes anyway and it was redundant to use the clicker: "The clicker became another object that I had to interact with". A few mentioned that the clickers did not work as well as they could have or, editorializing, that they could have been used in a better way in the class – but no detailed suggestions were given. Without prompting, a few mentioned that the novelty probably helped them pay more attention to the presentations. Several students appreciated the possibility of getting feedback immediately after the class, but equally many complained about the lack of feedback on their own voting.

When asked for opinions on how the clickers were used in class, some considered the simple numerical scoring too restrictive; they would have preferred more flexibility – but did not indicate how that could have been accomplished. Another suggestion from several students was to use a wider scale, e.g. 0 to 9 instead of 1 to 5. Almost half of the students commented that they would have liked to vote immediately after the presentation.

Students unanimously commented that they were interested in seeing the scores for their own presentation although not everyone made use of the possibility (usually because they had to rush to the next class). Fourteen students (a majority) would also have liked to see the scores for the other groups, four were neutral, and two were fairly strongly against it. A reason for not seeing the scores cited by a couple of students was that seeing low scores for yourself would give rise to a revenge mentality and feared that students would be tempted to give low scores to others as well. Some students did not like the increased competitiveness that revealing the scores would have resulted in, and one expressed fair amount of empathy: anticipating feeling remorseful if a fellow student got really low scores.

More than half of the students replied that they tried to maintain their privacy while voting – not necessarily because of being afraid to show the score they gave, but more as a matter of principle. It was generally felt that it was easier to maintain privacy using the clickers than with a paper form. However, one student commented on the difficulty of seeing your own vote if the clicker was hidden under the table to maintain privacy.

5. DISCUSSION

5.1 Evaluations by Students and Teachers

The difference in the scales used by the students and the teachers is striking, and in clear contrast with the results of Tseng and Tsai [11]. In particular, none of the students gave the score of 1 to any of the presentations. As seen in Figure 4, the distribution of scores by the teachers used the full scale, whereas scores given by students peaked at 3 and 4. This was also anticipated by the students themselves, who felt that friendships and fear of revenge by other students might drive their scores up. That is likely why, in response to question (f), most students preferred the scores given by the teacher over the evaluations of their peers.

Figure 3 shows that the students did not rate any presentation really low. The lowest total score (for content and quality) given for any presentation by the students was 4, which happened only 7 times in the total of 178 evaluations. In contrast, in 3 out of the 10 presentations the average score by the teachers was 4 or below.

The most striking difference in scoring is for presentation P2, which got an average score of 2 from the teachers and 6.86 from the students. Only one student rated that presentation with a total of 4 (2+2), but that same student gave a 5 (2+3) to presentation P3, which was highly rated by everybody else. In general, just by looking at the individual scores it is not possible to find a pattern that would help in understanding the differences in student and teacher scoring.

To shed more light on the differences, it would have been helpful to use a more fine-grained evaluation: instead of just content and quality, various attributes could have been evaluated. This, however, would have slowed the procedure down considerably. Using a more fine-grained scale, as suggested by the students, could also have helped in making the scores by students a little more wide spread.

Although using the peer evaluations may not reveal the differences between student projects reliably for course grading, it has other benefits. Several students commented that having to give a score to each presentation made them attend to the presentations more alertly, and it also made them reflect on what makes a presentation good, helping them develop their own performance for the future. On the other hand, this was partly attributed to the novelty of the clickers, implying that the effect might wear off with time and no longer serve that purpose.

Figure 4 shows an interesting difference between evaluations given by Finnish and international students: the median for Finns is 3, and for non-Finns it is 4. Finns have a reputation of being more reserved than international students, and this shows up very clearly in oral discussions in class. We hypothesized that using the clickers would enable them to express their opinions more freely. In addition, the international students have a wider distribution of scores – more frequently using the score of 2 and 5 – although not nearly as distributed as the Teachers. The observed difference in the distributions motivates further study of this issue. However, in the subjective comments there was no nationality-related difference in preference for clickers or paper.

5.2 Using Clickers for the Evaluation

5.2.1 When to Vote?

Did the medium used for scoring affect the scores given? Figure 5 indicates that this could be the case. The teachers used paper in all cases and scored the presentations in the first group significantly lower (with an average of 2.6) than the presentations in the second group (average 3.3). However, in the student scores there is no similar difference (average 3.6 in both cases). Thus, if we consider the evaluations by teachers as objective measures of quality, it seems that the clickers made the students give relatively higher scores.

However, the medium used for giving the scores is not the only factor that was different in the two groups. When clickers were used for scoring with the first group, all scores were given at the end of class, whereas the paper scores could be jotted down immediately after each presentation. Making notes during the presentations in the first group was encouraged, but most students seemed not to make use of this possibility. Consequently, they admitted in their subjective comments that by the time of voting they had somewhat forgotten the differences between the presentations, and this tended to drive the scores together, closer to the middle scores.

The decision to do voting by clickers at the end of the class was based on the desire to give the students some basis for their evaluation, so that they would not shoot the scores through the roof (or bottom) with the first presentation, and making it possible to better compare the presentations. In retrospect, this choice was counter-productive; if clickers are used for evaluation, it is better to use them without delay. Similar observations were made in the study by Denning at al. [4].

5.2.2 Clicker Technology

Several comments were made on the perceived unreliability of the technology. The slide that was shown during voting shows the total number of votes cast. The indicator jumped quickly in the beginning, but the last votes drifted in slowly, even though all students presumably voted promptly. It took the server some time to collect all the signals, and in all cases some were missing. It was normal that there were fewer votes than the number of people in the class since students were not supposed to vote on their own presentation (a fact that was verified by the teacher after the class). However, the apparent unpredictability of the software in counting the votes may have created doubts in the voters' minds.

Another fact contributing to voter doubt was that there was not much feedback on the clicker handset itself. When the student clicked on a key, a green led indicator lit up to show that a key had been clicked – but it did not indicate which key. Moreover, not all students noted the indicator light. The students could also click again if they were uncertain of whether they clicked correctly, since only the last click counted. Nevertheless, there was still no indication of which score the software had registered for the voter. Providing more feedback on the clicker device itself would increase trust in the system.

Despite the reservations, clickers were, as noted earlier, overwhelmingly favored over paper for those concerned with overcoming boredom. For instance, one student mentioned that "they weren't monotonous and boring as the paper evaluation method was."

5.2.3 Privacy

The clickers were overwhelmingly thought to be better at creating privacy for voting – even by one student who did not hide her voting behavior. The urge to hide was impressed upon one international student who initially did not attempt to hide his voting but then observed others hiding. This shows how dynamic the notion of the need for privacy was in the class. Comments made by the students included: "Feels a little odd to vote on people when sitting right next to them," "If I gave a bad score to a team that deserved it, they might revenge it on my group if they noticed that," and "I hid my paper form a little, not to hide how I voted, but in order not to affect the voting of others."

Keeping the scores private was not specified as necessary, but many felt it should be anyway. There was no significant difference in the comments by Finnish and international students.

5.2.4 Increasing Interactivity

Many felt that the clickers could be used to add to class interactivity and to keep people involved, but only within limited settings such as the one used in this class – voting on presentations. A few noted that they could be the catalyst for increased class oral discussion.

One interesting result of our open-ended question (h) was that almost 1/3 of the students realized that clickers could be used to provide feedback on the way the course was conducted and thus steer the course in real-time. Somewhat optimistically, one student noted: "If the lecture was getting boring or if the student thought the material could be presented in a better way then the lecturer could change his enthusiasm or personal motivation for the class." Others noted that PowerPoint presentations could be slowed down or speeded up given, respectively, more or less class understanding of the material. One noted that a professor might take more notice of questions in the class if the questions flashed on their screen.

5.2.5 Comparing Other Techniques for Interaction

Many students felt that the clickers were good for voting but only one thought they would work for more free-ranging discussions – and then only as a catalyst. Online and face-to-face discussions were judged better for free-ranging discussions although the students often had opinions on the effectiveness of those techniques as well – noting that face-to-face doesn't work as well for "introvert personalities" and that online should have a limited number of participants for some kind of conversational cohesion.

Only one individual thought "moderated online discussions are the best" because "you have more time to tell your opinion, no one can interrupt you." A small number of students (both international and Finnish) volunteered that online discussions are difficult ("unless forced") because they "become a task instead of something I am willing to talk about." And two students stated that unless forced, "only a few people participate." Alternatively, one indicated that online discussions are not natural especially if they are forced. Some also noted that outside of the classroom they are preoccupied with other topics and are less interested in interacting online – even if they like the subject matter.

Interestingly, some students made little differentiation between online and face-to-face discussions regarding shyness and being forced to participation. However, differentiating between clickers and online/face-to-face seemed not to pose the same problem for any of the students. One noted that it was a flawed comparison since they are technologies whose "purposes are clearly divergent." Other students could not imagine using the clickers for discussion at all. Nevertheless, many students noted the benefit of the clickers as "immediate feedback" and at least half were excited about their use whatever their purpose in the classroom.

5.2.6 Peer Evaluation

Despite instructions to the class participants that their peer evaluations would not have any influence over their grade, some students still expressed that concern. Almost a third felt that students were "unqualified" to grade and that they could not trust the results if "people are being nice to their pals." A number of students also mentioned that they couldn't trust how seriously their fellow students were taking the responsibility for evaluation.

Despite these reservations and qualifications, most of the students liked peer evaluations. They noted "evaluating skills need practice," "if it is a good presentation you can learn more," "it is a good way to improve attention in class," and that it was good practice for presenting: "it gives the feeling that you are really presenting for the people in the class and not only for the teacher."

6. CONCLUSIONS

Clickers are not a new technology for increasing classroom interaction, but there has been slow adoption for major use. The technology has remained relatively expensive. The value-add of using clickers needs to be clear for teachers and departments to make the investment.

We carried out a small scale study of using clickers for peer evaluation. Our observations on the evaluation are in line with previous results on peer evaluation. However, past work has used different media; ours is among the first studies on the effect of clickers on evaluation behavior. Our results show that the timing of when to use clickers needs to be planned carefully. If the voting cannot take place immediately, the scores given by students tend to converge too much. In our experiment, students did not trust the peer evaluations to be sufficiently objective to be used in the actual grading of the projects. However, they felt that having to evaluate the student presentations made them more attentive and helped them reflect on their own work.

As expected, the study confirmed that privacy was valued by the students. Clickers were felt to preserve privacy better than the paper forms. Surprisingly many students, however, mentioned avoiding revenge mentality among fellow students as the main reason; otherwise they would have been happy to share their thoughts on the other presentations in public. Most students would have liked to see the summaries of voting at the end of the class, but some mentioned again the retaliation factor, and some also felt that it would unduly increase the competition in class.

In conclusion, clickers have advantages and potential. The setup in this class was very restricted. Other uses suggested by students included rating interface designs, using voting to illustrate the controversial nature of the topic under discussion, making choices on what to discuss next, asking for further explanation on a particular (numbered) slide – and simply waking people up.

We realize that our questions and conclusions might benefit from a larger sample size. We are currently engaged in examining more classes and further refining our understanding of the variables we have set forth and uncovered, such as cultural differences and controlling for the 'revenge' mentality in peer evaluation. In addition, some of the suggestions made by the students – for example, rating interface designs – would be particularly interesting to examine in an HCI class.

7. ACKNOWLEDGMENTS

We are grateful to all participants on the course for participating in the experiment and especially for all the verbal comments given.

8. REFERENCES

[1] Anderson, R., VanDeGrift, T., Wolfman, S. A., Yasuhara, K., and Anderson, R. 2003. Interaction patterns with a classroom feedback system: making time for feedback. In *Extended Abstracts on Human Factors in Computing Systems* (*CHI '03*). ACM, New York, NY, 880–881. DOI= http://doi.acm.org/10.1145/765891.766047

[2] Barber, M., and Njus, D. 2007. Clicker evolution: Seeking intelligent design. *CBE-Life Sciences Education*, 6(1), 1–8. http://www.lifescied.org/cgi/reprint/6/1/1.pdf

[3] Caldwell, J. E. 2007. Clickers in the large classroom: Current research and best-practice tips. *CBE-Life Sciences Education*, 6(1), 9–20. http://www.lifescied.org/cgi/reprint/6/1/9.pdf

[4] Denning, T., Kelly, M., Lindquist, D., Malani, R., Griswold, W. G., and Simon, B. 2007. Lightweight preliminary peer review: does in-class peer review make sense? In *Proceedings of the 38th SIGCSE Technical Symposium on Computer Science Education* (*SIGCSE '07*). ACM, New York, NY, 266–270. DOI= http://doi.acm.org/10.1145/1227310.1227406

[5] Draper, S. W., Cargill, J., and Cutts, Q. 2002. Electronically enhanced classroom interaction. *Australasian Journal of Educational Technology*, 18(1), 13–23. http://www.ascilite.org.au/ajet/ajet18/draper.html

[6] Fies, C. and Marshall, J. 2006. Classroom response systems: A review of the literature. *Journal of Science Education and Technology*, 15(1), 101–109.

[7] Lopez-Herrejon, R. E. and Schulman, M. 2004. Using interactive technology in a short java course: an experience report. In *Proceedings of the 9th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education* (*ITiCSE '04*). ACM, New York, NY, 203–207. DOI= http://doi.acm.org/10.1145/1007996.1008051

[8] Martyn, M. 2007. Clickers in the classroom: An active learning approach. *EDUCAUSE Quarterly*, 30(2), 71–74. http://connect.educause.edu/Library/EDUCAUSE+Quarterly/ClickersintheClassroomAnA/40032

[9] Pargas, R. P. and Shah, D. M. 2006. Things are clicking in computer science courses. In *Proceedings of the 37th SIGCSE Technical Symposium on Computer Science Education* (*SIGCSE '06*). ACM, New York, NY, 474–478. DOI= http://doi.acm.org/10.1145/1121341.1121489

[10] Simpson, V. and 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. http://www.ascilite.org.au/ajet/ajet23/simpson.html

[11] Tseng, S.-C. and Tsai, C.-C. 2007. On-line peer assessment and the role of the peer feedback: A study of high school computer course. *Computers & Education*, 49(4), 1161–1174. http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=EJ773943