Collaborative Assessment Using Clickers

By Maha Bali [1], Heather Keaney[2]
[1]Center for Learning and Teaching, American University in Cairo bali@aucegypt.edu
[2]Department of History, American University in Cairo hkeaney@aucegypt.edu

OVERVIEW

The American University in Cairo (AUC) started piloting the use of clickers (also known as personal response systems) in fall 2006. Most instructors piloting this new technology were from economics, science or engineering disciplines, as it was found easier to create multiple-choice questions for their classes. In spring 2007, however, clickers were piloted in a history class. We present a unique case of a one-off use of *technology in assessment* - a graded activity in which students worked in teams to answer questions using clickers. Prior to this activity, instructors at AUC had succeeded when following "best practice" of using clickers for non-graded formative assessment only, however, this graded activity showed improvement in student learning, and 95% of them said they would like to use clickers in a similar activity again.

Keywords
Technology-supported assessment, Personal response systems (PRS), clickers, history, collaborative assessment

INFORMATION ABOUT THE CLASS

AUC is an American liberal arts university and is the most prestigious higher education institution in Egypt. Because AUC is a liberal arts university, all undergraduate students are required to take classes from various disciplines. One of these requirements is that all students are required to take a minimum number of courses relating to Arab World Studies (this includes a choice between various history and literature courses), and such a class will have students of various ages and disciplines.

The innovation occurred in one such history class entitled "The Making of the Modern Arab World", which covers selected issues in nineteenth and twentieth century Arab history. Students therefore come from various disciplines, as this is part of their core curriculum.

The class runs for a full semester (about 15 weeks) where students meet on campus twice a week for an 80-minute lecture. This semester, the class was taught by an American instructor, and had twenty-five[1] students enrolled (a normal range for AUC), seventeen of whom were Arab (mostly Egyptian) and eight of whom were non-Arab (mostly American study-abroad students).

DESCRIPTION OF THE CASE

Pre-Activity Planning

The instructor wanted to provide the students with an engaging activity, different from the usual essay exams, and so came to the Center for Learning and Teaching (CLT) to explore the possibility of using clickers for the first time. After a thorough discussion, it was decided that clickers would help achieve the instructor's objective at this point in the

[1] On the day of activity, one student was ill, so only 24 participated

semester of providing an activity that enabled students to work together in teams to prepare and conduct the activity, but at the same time enabled the instructor to give students individual scores and avoid free-riders benefiting from the brighter students in the team. Students were told that part of their grade would be individual and part of it would be team-based, but the exact breakdown was not decided in advance.

The instructor created a set of thirty questions, of varying degrees of length and difficulty, some of which asked straightforward factual questions but the majority of which were conceptual or required that students had done deep reading.

The instructor divided students into teams of 5, and gave each team a "color". Students were asked to prepare for a graded activity by studying together material that had been covered so far in the semester. Students knew that during the activity, they would answer each question twice using clickers: once individually and a second time after a discussion with their teammates; students were given a time limit to answer each question. The design of the activity encouraged students to work together for two reasons: first, their individual grade would be affected by the team grade, and second, studying together would enable them to reach consensus in their team answers faster if they had studied and prepared together (e.g. one student could tell the others "don't you remember? It was in chapter three of the book"). Because the instructor did not want to encourage free-riding, she also assured students that part of their grade would be based on their initial individual responses before the team answer.

A week before the actual activity occurred, members of CLT explained to the students how the technology worked and allowed them to test it to familiarize themselves with the clickers and assure them they would have support during the activity. Students were shown how the teams were recorded on the software, how to use the clickers, how to ensure their responses had been recorded, how to change their answers, and were also shown how the software created charts showing the frequency distribution of each answer. They were told that during the graded activity they would not be shown the charts.

The instructor also emphasized that any technological glitches would be dealt with, and any technical issues outside the students' hands would not affect them negatively.

Implementation

The activity ran as follows: students sat in their teams, and a multiple-choice question appeared on a PowerPoint presentation on a large screen at the front of the classroom. First, students had 60 seconds to respond individually (using the clickers) with the choice they thought best answered the question. This answer was recorded on the software, then they were given another 60 seconds to discuss the question with their teammates and answer it one more time, again each using the clickers. Each individual in the team could choose to give an answer different from the rest of the team members if s/he was not convinced by what the others had said in the discussion. A team score appeared after each question had been answered twice, and students would know which team was leading. The team score is automatically calculated by the software that comes with the clickers: for each question, the average number of correct answers per team is given as a percentage. Two scores are calculated by the software: a team score (which can be displayed at any time) and an individual score (which can be reported upon in detail after the activity has ended). Both of these scores accumulate each student's individual and team answers such that each time a student answers a question it counts towards his/her own score and the team score.

Initially, the activity had been planned such that students would run through the questions knowing their team scores but not knowing each question's specific answer frequency distribution (i.e. how many students chose each of the answer choices), nor the correct answer per question. However, as the activity unfolded, the students started calling out and asking to know the correct answer, so the instructor frequently decided to go back and show the students the answer distributions for each question's individual and team answers

(this functionality existed in the software), then give the correct answer, commenting on whether the team discussion helped improve the rate of correct answers.

Post-Activity

After the activity was over, a member of CLT analyzed the results and presented them to the instructor to decide on the final grading structure. It was decided to assign 70% of each student's grade on the initial individual response, and 30% of their grade on the second team response (note that individuals' second team response might not have been the same as their entire team's response, but it was the case in the majority of questions).

RATIONALE IN TERMS OF EDUCATION IDEAS

Clicker-use lends itself to active learning strategies, as they allow all students in the class to be actively engaged and to receive immediate feedback if needed. The software that comes with clickers also allows for detailed analysis and reporting during and after an activity, allowing instructors to analyze and grade the outcome of the in-class activity.

The instructor was aiming to promote cooperative learning and student engagement with this activity, thus following Chickering and Gamson's (1987) good practice principles of active and cooperative learning. Since students were asked to prepare together, the entire team had a part to play in each individual's score. Moreover, well prepared teams would be more likely to do well the second time the question was asked, because the experience of having studied together would make the discussion more focused. It was hoped that preparing together in advance would be an opportunity for "peer instruction" - students to learn from each other, and to improve their own understanding by having to explain to their peers. The opportunity to answer each question twice allowed students to self-correct after consulting with their peers, although the time period for this allowed little reflection. The instructor decided to use teams and a competition to further motivate students.

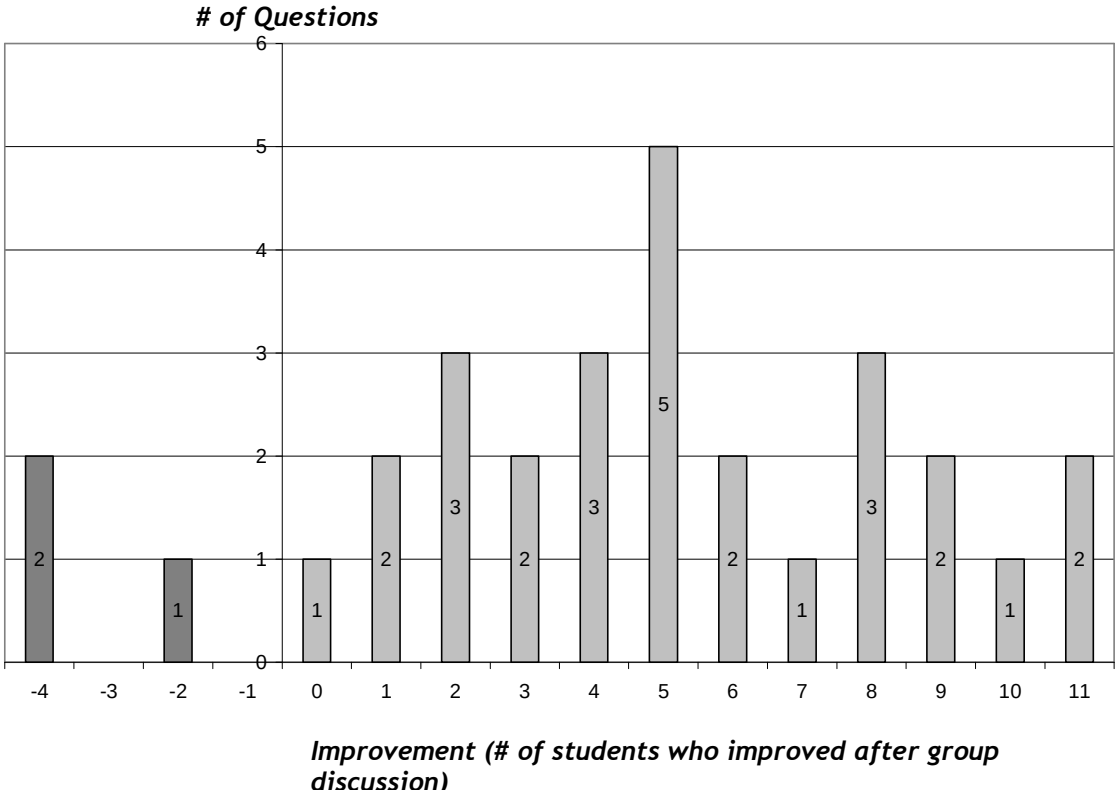
Individual and group accountability was ensured by having both an individual and a team score. This ensured that hard-working students' grades would not be unnecessarily pulled down by free-riders, and that free-riders' grades would not be pulled up too much by their team members. It also ensured that the team would work together because all would be affected by each other's grades.

Although not initially planned as such, the activity also allowed for immediate feedback, since the instructor was able to go over the answer distribution for each question and quickly discuss the correct answer if most students had answered incorrectly. This is something that is almost impossible in a written exam or assignment, because of the time required for grading. This aspect of the activity therefore implemented Chickering & Gamson's (1987) good practice principles of encouraging student-faculty contact and giving prompt feedback.

EVALUATION

The success of this case rests on two main aspects: student engagement and student learning.

Figure 1: Improvements in Team Responses versus Individual Responses:



Student engagement was clear: when asked, 95% of the students wanted to conduct a similar activity again. During the activity, students were animated, actively involved in discussing the answers with their team when it was time to do so, cheering when their team score went up, and interacting with the instructor to understand what the correct answer was and why.

In terms of student learning, the majority of questions (26/30) showed more correct answers in the team response versus the individual response, one question showed no improvement, and only 3/30 showed negative improvement (i.e. more incorrect responses in team response, denoted in darker grey in Figure 1), showing that the team discussions helped students arrive at the correct answers.

The total score of team answers was greater than or equal to the total score of individual answers. Figure 2 shows the percentage of correct answers by team (this is the sum of correct answers of team members in individual and group responses). It is clear that all the teams fared similarly except for the blue team. Figure 2 also shows the average percentage improvement per team, between initial individual response and team response. It is clear that all teams improved on average, although some individuals in certain teams improved very little (0-3.3% is 0-1 questions). These individuals were the ones with the highest scores in their teams, and either had most of the answers correct to start with, or became convinced to change their answer to the correct one after team discussion. There was only one of these people in each team. The rest of the team members showed bigger improvement.

Figure 2: Team Score and Average, Highest and Lowest Percentage Improvement by Team

Team	Team Score (%)	Average Improvement (%)	Highest Improvement (%)	Lowest Improvement (%)
Pink	66.0	16.0	30.0	0.0
Yellow	66.7	22.7	30.0	3.3
Blue	56.7	11.7	26.7	0.0
Green	63.0	19.3	40.0	3.3
Orange	67.7	22.0	46.7	3.3

One could look at these numbers and consider that the existence of a hard-working or brighter student in the team may have resulted in the other team members simply following his/her answer after the discussion. However, a closer look at one of the teams (the one that fared best) shows otherwise. For example, the "orange team", had an average improvement of 22% on the team score vs. the individual score, with individual improvements ranging from 3.3% (the student who was doing well from the start) to 46.7% (for the student who was doing the worst from the start). All of the team members had a similar team score, and closer inspections shows that the majority of the team members did give the same team answer the second time each question was asked, except for one question. However, one could also speculate that the team always followed "brighter" student's answer (let's call her student A), thus always doing better the second time based on only one student's effort. Closer inspection shows that this is not the case. As seen in Figure 3 below, student A benefited from the rest of the group in 4/30 questions, and all or most of the group were initially correct in 11/30 questions. In fact, student A confused the group into choosing an incorrect answer in 5/30 questions. Even when student A helped the team with the correct answer (9/30 questions) she was often not the only one with that particular answer initially.

Figure 3: Changes in Team Answers Compared to Student A's Initial Response (special case of the Orange Team)

Team answer changed to	Correct by following student A	Incorrect by following student A	Correct because majority was initially correct	All team already initially correct	Student A correct, benefiting from team
# of questions	9	5	7	4	4

This shows that all students benefited, albeit to different extents from the team work and team discussions in this particular group. Although students often followed student A, it was not an overwhelming majority of the questions.

LIMITATIONS AND CONCLUSION

Obviously, one limitation of this kind of activity is the necessity of coming up with multiple-choice questions that can test students beyond simple factual recall. The questions also need to be controversial or indirect enough to be "discussable" so that the team discussion would actually be fruitful. Design of such questions required a lot of thought and preparation from the instructor, and according to this particular instructor, would not have been possible later on in the semester when assessment would have required deeper analytical involvement of students.

Although results showed that overall, students' responses improved after the team discussion, the instructor still harbored concerns that the "brighter" students would have felt "used" until she asked them in person. They said they enjoyed teaching their

colleagues, and felt that their own knowledge was reinforced when they had to explain their answer. One way to ensure such attitudes is to explain to students, in advance, that this would be one of the positive outcomes of the teamwork.

In conclusion, it seems that overall, student engagement and learning benefited from this activity. Further analysis of student learning as well as an opinion survey could give more insight into benefits of the activity, and how well the groups worked together for preparation as well as during the activity.
The instructor implemented this activity in the fourth week of the semester, where creating multiple choice questions on historical topics was possible; it may be less applicable for other courses, more advanced courses, or even later on in the same semester.

REFERENCES

Chickering, A.W. and Gamson, Z.F. (1987). Seven Principles of Good Practice in Undergraduate Education. *The Wingspread Journal*, June 1987, special insert. Racine, W.I.: Johnson Foundation. Accessed April 25, 2007 http://www.johnsonfdn.org/Publications/ConferenceReports/SevenPrinciples/SevenPrinciples_pdf.pdf