

Assessing the Impact of a Tablet-PC-based Classroom Interaction System

Kimberle Koile *MIT Center for Educational Computing Initiatives*
kkoile@mit.edu

David Singer *MIT Department of Brain and Cognitive Sciences*
singerd@mit.edu

1. ABSTRACT

This paper describes a strictly controlled study of the impact that Classroom Learning Partner, a Tablet-PC-based classroom interaction system, had on student performance. The goal of the study was to test the hypothesis that the use of such a system improves student learning, especially among the poorest performing students. This paper describes our validation of that hypothesis, and the controls, performance metric, and assessment methodology that we developed in the course of our study.

2. PROBLEM STATEMENT AND CONTEXT

Tablet PCs hold great promise for improving student interaction and learning. Indeed, much recent work has focused on development and deployment of Tablet-PC-based systems for use in classrooms, e.g., [1, 2]. With the increased interest in Tablet PCs, comes an increased need to determine, as with any other kind of technology, whether the promised benefits can be realized. For the last couple of years, we have been investigating the benefits of a Tablet-PC-based classroom interaction system [4, 5]. Our hypothesis has been that such a system will improve student learning, especially among students who might otherwise be left behind. Our goal has been to test this hypothesis by rigorously assessing student learning in controlled studies involving deployment of the system. We report here the findings from our most recent and most valid study to date.

3. SOLUTION EMPLOYED

We conducted a study in MIT's introductory computer science course in Spring 2007. The course had an enrollment of 236; students taking the course met in 50 minute classes of various sizes several times a week—lectures for all students twice a week, recitations for 15 to 25 students twice a week, and tutorials for five to seven students twice a week. Students were randomly assigned to recitations and tutorials. Our study was conducted in recitation classes taught by the first author. Surveys, classroom observations, interviews, and data analysis were carried out by the second author. The technology used in the study consisted of a network of Tablet PCs running a software system called Classroom Learning Partner (CLP), developed by the first author's research group [5]. CLP is built on top of Classroom Presenter [1], and like Classroom Presenter, it allows students to wirelessly and anonymously submit digital ink answers to in-class exercises. An instructor chooses student submissions—both correct and incorrect—to be used as the basis for class discussion. CLP extends Classroom Presenter in several ways: It adds instructor presentation authoring tools, a login procedure, a central repository for storing student submissions, an instructor interface that groups student submissions by exercise number, and components for interpreting and aggregating student

submissions. (The interpretation and aggregation components were not relevant to the study reported here.)

3.1 Controls

The study was run with one control class and one experimental class. Students in the experimental class used Tablet PCs running CLP; students in the control class used a blackboard and/or paper handouts. The study employed the following strict controls.

(1) Teaching style: We controlled for teaching style by having the same instructor (the first author) teach both the control and experimental classes. In addition, the instructor was independently observed by the second author, to insure that the same teaching style was employed in both classes. Controlling for teaching style is critical, as even an instructor's affect can influence student performance, e.g., [3].

The instructor began each class with a review of material, lecturing and writing on a blackboard, Tablet PC, or referring to class handouts. The instructor spent the majority of class time (between 75% and 90%) engaged in high levels of teacher-student interaction: Students asked and answered oral questions, worked written problems individually or in small groups, participated in class discussions of problem-solving approaches and solutions, and worked at their own pace on extra problems when they wanted to. As a result, the students spent most of class time in two ways: *processing information* by solving problems and answering questions, and getting *immediate feedback* on responses to problems and questions. The wireless network of Tablet PCs running CLP greatly facilitated both processing, by letting students easily handwrite answers, wirelessly and anonymously submitting them to the instructor; and feedback, by allowing an instructor to choose submissions for public display and class discussion, often "inking" directly on the submissions. In the control class, the students spent the same amount of time processing information and getting feedback, but at the loss of anonymity and/or discussion of incorrect answers (since students were reluctant to share incorrect answers).

(2) Class material and exams: The students in both the experimental and control classes received the same information and problems. In addition, all 236 students attended the same lectures, took the same exams, and completed the same problem sets and project assignments.

(3) Time of day: The control and experimental classes met at approximately the same time of day. The control class met at 11am; the experimental class at 12pm. In this way, we expected to mitigate the problem of students not attending early morning or late afternoon classes.

(4) Assignment to class: This study took place in two of ten recitation classes to which students were randomly assigned for the introductory computer science course. With random assignment we did not bias our sample by asking for volunteers, who may have had a predilection for using Tablet PCs. Students were given the opportunity to switch recitations if they did not want to participate in the study. None switched.

(5) Student characteristics: We only included students who were taking the class for a grade; no listeners or pass/fail students were included, since such students may not have been as motivated as for-grade students. We also only included freshmen and sophomores. No graduate students or upperclassmen were included because we felt that they might have had better study habits or might have taken other courses that would have benefited them in the current course. Finally, we determined that the students in the control and experimental classes were comparable academically: We delayed deployment of the technology until close to the time of the first

exam; we then compared the two classes' exam scores and found the difference in scores to not be statistically significant.

(6) Attendance: Recitation attendance was not mandatory, so we did not include in the study all students assigned to the experimental and control classes. We only included students in the experimental class who attended at least 67% of the classes during which the technology was used. This standard was set in order to insure that the students had enough exposure to and involvement with the technology, while at the same time insuring that the number of students involved would be sufficient for analysis. The control class attendance was based upon the same corresponding days on which the technology was used; i.e., to be included in the study, students in the control class must have attended 67% of the classes whose topic had been taught using technology in the experimental class.

(7) Technology: The Tablet PCs were associated with the class, rather than having students borrow them for the term. In this way, we were able to keep the machines identical.

3.2 Data

We collected the following four types of data.

(1) Amount of technology use: We recorded the number of minutes that technology was used in the experimental class. Inherent in our hypothesis of improved learning is the idea that the amount of time spent learning a task is correlated with the amount learned [3].

(2) Performance metric: We used final exam score as the performance metric, since it was a direct summative evaluation of the material taught in the course and was a direct measure of the problem-solving pedagogy that was used in the recitation classes. This performance metric is to be distinguished from other performance metrics in the course, such as problem sets or projects, both of which the students did outside of class.

(3) Interaction metric: In the experimental class, we used the number of answers wirelessly submitted by each student for each problem as a quantitative measure of interaction in the class. We compared the number of answers expected with the number actually submitted and computed an average daily submission fraction for each student. Our goal was to see if this measure of interaction would correlate with performance scores. No such metric was easily computed in the control class, so our analysis was limited to the experimental class.

(4) Learning preferences and interests: Data on learning preferences, self-perceptions, and levels of interest and satisfaction was collected by evaluating questions asked of students in two surveys, one survey given at the beginning of the term, a second at the end. The evaluations involved T-Tests on questions posed using a seven point Likert Scale. We only considered scores from 1 to 2 (“disagree”) and 7 to 8 (“agree”). Although we reduced the number of cases, we insured that the learning preferences were more validly reflected in the statistical results. Multiple timed five-minute observation periods of students and short after-class interviews with students validated or clarified observed learning preferences and individual surveys.

4. EVALUATION

(1) Technology use: The technology was used in 16 of 23 classes by both design and circumstance. The technology was not used during the first four classes, in order to obtain baseline data on the performance of the students on the first exam. The technology also was not used during the instructor's absence (two classes) or when it was not working sufficiently well (one class). During each class that technology was used, we did not count the minutes used for

administrative procedures, such as login; or time spent fixing technology glitches, such as interference with wireless connectivity. At the end of the term, we tallied 672 minutes of technology use over the 16 days, which accounted for 84% of available class time.

(2) Attendance: As mentioned above, we only included in our study students who attended at least 67% of recitation classes. We started with 25 students in the experimental class, 24 in the control class. After disqualifying students who dropped the course (6), failed the attendance criterion (16), attended both classes (1), were upperclassmen (2), or outliers (1), we ended up with an N of 13 in the experimental class and an N of 10 in the control class.

(3) Performance metric: When comparing final exam scores for the experimental and control classes, we saw highly statistically significant differences in the scores ($p > .001$). The mean score for the entire class (N=236) was 77.8 (a C), with a standard deviation of 7.0; for the experimental class the mean was 81.5 (a B); for the control class the mean was 73.0 (a C). We looked at this performance data in several different ways.

The higher performance of the students in the experimental class was evident when we looked at the *performance distribution*. (See Figure 1.) Six of eight final exam scores (75%) within one standard deviation above the entire class mean were in the experimental class, while all four scores (100%) that were between one and three standard deviations above the mean were in the control class. In contrast, the distribution of scores below the class mean largely reflected final exam scores of students in the control class. Specifically, four of six final exam scores (66.7%) within one standard deviation below the mean were from the control class, and four of five scores (80%) between one and two standard deviations below the mean were scores from the control class.

When we *ranked* final exam scores for both classes, we saw that eight of ten scores in the control class were below the lowest two scores in the experimental class. (See Figure 2.)

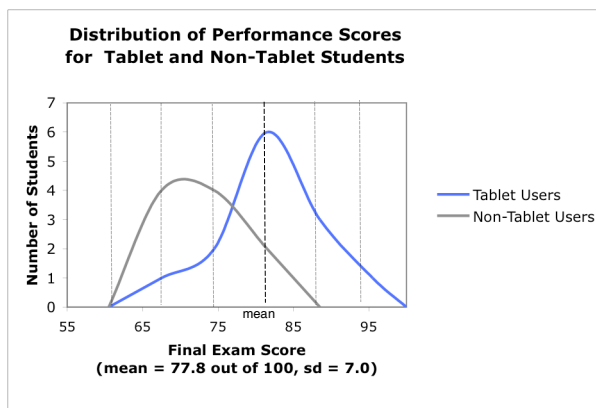


Figure 1. Distribution of final exam scores

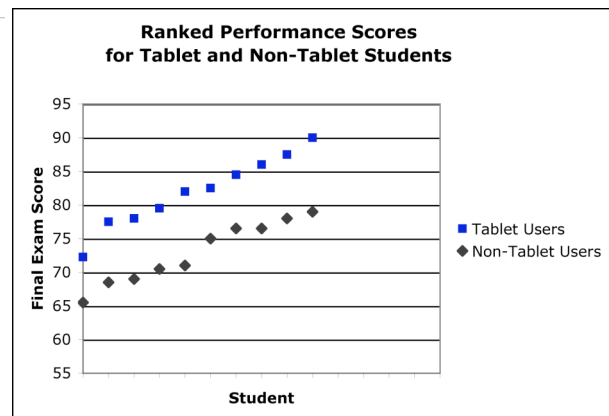


Figure 2. Rank ordering of final exam scores¹

Finally, we looked at *percentiles* to get a sense for how the students in the experimental and control classes performed with respect to the entire class (N=236). We found that more students in the experimental class than expected were in the top 30% of the class (five instead of four). More strikingly, we found far fewer students than expected in the bottom 30% of the class (zero instead of four). In contrast, we found fewer students in the control class than expected in the top 30% (zero instead of three), and one more than expected in the bottom 30% (five instead of

¹ To simplify the graph, we averaged nearby scores to reduce experimental N to match control N of 10.

four). This result, in combination with the overall significant difference in performance between the experimental and control classes, suggests that the greatest influence of the Tablet PC and CLP (and Classroom Presenter, by extension) may be in improving performance among students who might otherwise perform poorly.

(5) Interaction metric: When we counted students' submitted answers in the experimental class, we found no significant correlation between a daily average fraction of submissions and final exam score. When we looked at the top 25% and the bottom 25% of final exam scores, however, we did find a very significant difference in the interaction metric ($p > .008$): The mean final exam score for top 25% (3 students) was 89.2, bottom 25% (3 students) was 74.0; the daily average fraction of expected submissions for top 25% was 0.78, bottom 25% was 0.56. The N is small, but the results are nonetheless statistically significant. This result suggests that the opportunity to work problems, submit answers, and receive immediate feedback can be a factor for high performing students.

(6) Learning preferences and interests: Preliminary noteworthy findings about the relationship between learning preferences and performance support the idea that matching technology with learning preferences benefits students. Students who prefer to work problems in class, for example, or who benefit from being asked questions, did significantly better on the final exam when they had the use of a Tablet PC and CLP (82.5 vs. 73.5, $p > .015$; 85.0 vs. 73.1, $p > .013$, respectively). We also found that students who benefit from classes that review written material performed better with a Tablet PC and CLP (84.0 vs. 72.5, $p > .008$), as do students who do not prefer to work by themselves (85.0 vs. 64.7, $p > .013$). Preliminary findings regarding interests and performance are noteworthy: We found no correlation between performance on the final exam and a student's interest, enjoyment, or perception of difficulty of the subject matter; or their desire to try out a Tablet PC.

5. CURRENT AND FUTURE WORK

This work makes two important contributions: a sound assessment methodology and validation of learning gains among students using our Tablet-PC-based classroom interaction system, especially low performing students. The instructor's teaching style matched the technology well in that it emphasized student problem-solving and immediate feedback. This past term, we deployed the technology in MIT's introductory chemistry course for several weeks and currently are analyzing the results. In future research efforts, we intend to investigate the use of such technology in other undergraduate subjects and in K-12 classrooms.

6. REFERENCES

- [1] Anderson, R., et. al. Experiences with a Tablet PC Based Lecture Presentation System in Computer Science Courses, in *Proceedings of SIGCSE 2004*.
- [2] Berque, D., Bonebright, T., and Whitesell, M. Using Pen-based Computers Across the Computer Science Curriculum, In *Proceedings of SIGCSE 2004*.
- [3] Bransford, J.D., Brown, A.L., and Cocking, R.R. Eds. *How People Learn: Brain, Mind, Experience, and School*. National Academy Press, Washington, D.C. 1999.
- [4] Koile, K. and Singer, D. Development of a Tablet-PC-based System to Increase Instructor-Student Classroom Interactions and Student Learning. In *Proceedings of WIPTE 2006*.
- [5] Koile K., et. al. Supporting Feedback and Assessment of Digital Ink Answers to In-Class Exercises. In *Proceedings of IAAI 2007*.