

The Many Faces of Formative Assessment

Judith C. Stull
La Salle University

Susan Jansen Varnum, Joseph Ducette,
John Schiller, and Matthew Bernacki
Temple University

In this research paper we consider formative assessment (FA) and discuss ways in which it has been implemented in four different university courses. We illustrate the different aspects of FA by deconstructing it and then demonstrating effectiveness in improving both teaching and student achievement. It appears that specifically “what is done” was less important since there were positive achievement gains in each study. While positive gains were realized with use of technology, gains were also realized with implementation of nontechnology dependent techniques. Further, gains were independent of class size or subject matter.

The issues of assessment and accountability have gone beyond the classroom and entered the political arena. With this development they have become less nuanced as broad generalizations and policies are sought. What sometimes gets lost in many of these discussions is the fact that the educational sector is incredibly varied by grade, by subject, and by instructional format. Yet, at every level of instruction within this sector, the focus continues to be on improving instructor practices and raising student achievement. In this research paper we are going to consider an aspect of assessment that has been garnering increasing interest, specifically formative assessment, and consider different ways in which it has been implemented. All of the studies are in higher education and were subjected to statistical analyses. The goal is to illustrate the different types of formative assessment by deconstructing the concept and then demonstrating effectiveness in improving both teaching and student achievement.

Assessment

Assessments should define in measurable terms what instructors should teach and students should learn. Thus assessment, whatever form it takes, defines the playing field of academic interaction where the processes of teaching and of learning should be mutually reinforcing. However, in an era where accountability has become a driving force, certainly in the K-12 educational reform movement, the definition of how and what an instructor should teach and how and what a student should learn is becoming significantly narrower.

As usually understood, assessment is used by most instructors to determine what learning has occurred, and it serves as the basis for the assignment of grades. Such assessment is *summative* as it is the end point of the teaching-learning sequence. Assessment is *formative* when the evidence is used as an on-going process within the class to adapt the teaching to meet student needs as well as providing feedback to the students (Black &

Wiliam, 1998). Specifically, according to Heritage, Kim, and Vendlinski (2007), formative assessment is a systematic process to continuously gather evidence about learning. The data are used to identify a student’s current level of learning and adapt lessons to help the student reach the desired learning goal. In formative assessment, students become active participants with their instructors, sharing learning goals and understanding how their learning is progressing, what steps they need to take and how to take them. However, it is very difficult for instructors not to focus on summative assessment measures since the prevailing pressures for improved learning drive them inevitably in this direction. Some have indicated that the time has come when formative assessment, occurring within the learning process, needs greater prominence (Black & Wiliam 1998; Layng, Strikeleather, & Twyman 2004). In reality, both formative and summative assessment need to be incorporated into a total learning process.

Formative assessment informs both instructors and their students as to the degree to which the students have mastered the material. Feedback to the students serves two functions: to identify problem areas and to provide reinforcement of successful learning and achievement. Feedback to the instructor serves to identify the degree to which instruction was successful and to identify needed changes in instruction. It can be used to distinguish between individual and group problems that can then be used to suggest solutions: revision of instruction, specific group work, or individual remediation. The model, as shown in Figure 1, is a dynamic one recurring throughout the course. It is composed of the following stages.

1. The instructor constructs a lesson module and related assessments based on the perception of the students’ readiness and prior knowledge (Stage 1).
2. The instructor presents the lesson module (Stage 2)

3. The instructor administers an assessment (Stage 3).
4. The instructor considers assessment results. The student considers the assessment results (Stage 4).
5. Dialogue between the instructor and the student begins (Stage 5). Depending on dialogue with the instructor, the student adjusts learning style or proceeds with current style.
6. Depending on the dialogue, the instructor adjusts teaching or proceeds to the next learning module (Stage 6).

Although not stated, this model underpins much of the research that has been conducted thus far and makes explicit the connections between the role of the instructor and the role of the student. For the instructor, formative assessment generally implies frequent assessments that vary by: a) how formal the assessment is (exam, quiz, or class discussion), b) its length, c) depth of knowledge expected, and d) format, altered instruction based on assessments, instruction on the interpretation and use of the assessment results, and perhaps altered classroom interaction to increase student learning and engagement. For the student formative assessment means considering adjustments in studying and perhaps in classroom behavior in light of assessments (see Figure 1).

Wiliam and Black (2003) argue that formative assessment is the only way for which a strong *prima facie* case can be made for improving learning. While students across the achievement spectrum should benefit from the incorporation of formative assessment techniques, it has been argued that the effects should be more notable for the lower achieving students, and research has supported this position (Athanases & Achenstein 2003). Possible gains for higher achieving students could be limited by the fact that they most probably have already incorporated many of the student related formative assessment practices.

Wiliam and Black noted that they were able “to identify 20 studies that showed that innovations which included strengthening the practice of formative assessment produce significant and often substantial learning gains” (2003, p. 41). However, the research base on formative assessment and the efforts to demonstrate its effectiveness in improving teaching and learning have focused very heavily on K-12 classrooms and the professional development of in-service instructors, have generally focused on the role of the student and the student reactions, and have been based on small samples (Boston 2002; Ruston 2005; Taras 2002; Brookhart, Moss, & Long 2010). Aspects of FA that been researched have focused on students at all grade levels from early childhood (MacDonald, 2007)

to university students (Costa, Mullan, Kothe, & Butow 2010; Carrillo-de-la-Pena et al. 2009). Furtak, et al. (2008) present an impressive model of FA, yet it is focused entirely on the student.

The use of personal and online computer based feedback and student self-regulation systems has been researched (Pachler, Daly, Moore, & Mellar 2010; Wang 2006; Heinze & Heinze 2009; Ibabe & Jauregizar 2010; Miller 2009) with varying degrees of success found. Chen-Ming, and Ming-Chen (2009) present a very sophisticated and complex on-line system with embedded data mining, but it is only for student use. Other researchers have focused on the way use of formative assessment affected student behaviors irrespective of the delivery system used. While Carrillo-de-la-Pena et al. (2009) argued that there is a dearth of empirical studies of FA’s impact on achievement, they did find a positive effect on student achievement in their research. Lipnevich and Smith (2009) found that while feedback to students had a positive effect on learning, it did not matter whether the feedback was computer generated or from the instructor. Chin and Teou (2009) found use of concept cartoons effective with middle school aged students. Furtak and Ruiz-Primo (2008) found FA could be effectively used to improve students’ writing and discussion skills. Marcotte and Hintze (2009) found that a use of self-regulated learning environment had a moderate effect on students.

On the instructor side, research has been done on the ways in which FA has affected teaching. Shavelson et al. (2008) discussed the role of instructors in the development of materials that would then be provided to students for their self monitoring. Puddy et al. (2008) showed the way in which continuous monitoring and adjusting positively affected participants in a mental health program. Frey and Fisher (2009) document how teachers in one school collaborated over a four year period to embed formative assessment techniques in the curriculum, resulting in significant achievement gains. However, Luttenegger (2009) found that instructors were not skilled in implementing FA, and Heritage, Kim, Vendlinski, and Herman (2009) provided empirical evidence that instructors were better at drawing reasonable inferences about student levels of understanding from assessment information than they were at deciding the next instructional steps.

Further, mentors have been found effective in helping in-service and pre-service instructors implement formative assessment practices during their practicum courses (Ash & Levitt 2003; Athanases & Achenstein 2003). Ruiz-Primo and Furtak (2007) broadened the discussion of assessment to informal interactions, although more attention has been paid to formal, planned assessment contexts.

Formative assessment techniques are increasingly being conducted online (Gipps 2005). The online environment presents opportunities for formative assessment to be conducted more efficiently by decreasing student feedback time (Beatty et al. 2008) and facilitating peer-feedback and collaboration. It has been shown to positively affect achievement (Cassady, Budenz-Anders, Pavlechko, & Mock 2001; Chung, Shel, & Kaiser 2006; Henly 2003; Peat & Franklin 2002; Smith 2007; Wang, Wang, Wang, & Huang 2006), attitudes, and student/instructor interaction (Chung, Shel & Kaiser 2006). Tierney and Charland (2007) also identify a strengthening of student voices as critical to improving formative assessment. Online tools provide increased opportunity for students to initiate formative assessment by allowing them to interact with instructors virtually (Nichol & Macfarlane-Dick 2006).

Although student use of online formative assessment tools is limited, virtual office visits and chat room attendance have been positively related to increases in student achievement (Lavooy & Newlin 2008). It follows, then, that students who initiate formative assessment processes in addition to completing those created by instructors in their coursework may further increase the knowledge gained during a course. To date, empirical research has yet to determine whether student initiated formative assessment has a different effect on summative learning outcomes than teacher initiated formative assessment activities.

Another FA technique that is generating interest is the use of clickers. Mayer et al. (2009) situate the use of clickers in a theoretical context involving deep or generative learning. Specifically, they indicate that clickers facilitate students' use of self-questioning and foster what they term the "self-explanation effect." They hold that research on the self-explanation effect has shown that students perform better on a final test when they are encouraged to explain aloud to themselves as they read a textbook rather than simply read the text without self-explanation. While this statement refers to reading a textbook, the same logic has been applied to the type of behavior required in a clicker-augmented lecture. On the other hand, Hatch, Jensen, and Moore (2005) believe that the effectiveness of clickers resides in the fact that they require the students to pay attention to what is happening in class. As proof of their belief they report that the students who seem to most benefit from clickers are those who have mild to moderate degrees of attention deficits.

In sum, not enough attention has been paid to the fact that formative assessment can be operationalized in different ways. To advance the discussion, we are going to consider four different types of formative assessment, all at the university level. The courses

involved varied from chemistry to mathematics to physics to an educational assessment course with enrollments ranging from 19 to over 250 students. Taken together, these studies demonstrate the applicability of formative assessment to all or part of a university course.

Study 1: Formative Assessment Can Be Effective in the Large Lecture Setting

General Chemistry is the first course in chemistry and is a requirement for most science and health profession majors at a large urban university. The enrollments are large and the courses are composed of large lecture sections, smaller discussion sections, and laboratory sections. Exams tend to require factual recall and problem solving. It is a difficult course for many as it is their first exposure to what is expected of science classes at the university level. As a result it is also a course that traditionally has a high withdrawal rate and a high failure rate. The focus of the study was to capture the effect on student achievement of the incorporation of formative assessment techniques.

In the fall of 2005, a study was conducted to determine if formative assessment techniques could be successfully incorporated into this large lecture university science course. Two lecture sections were taught by the same instructor under two different conditions. Each of the sections had over 200 students enrolled. The students were not assigned to the sections, and demographic characteristics and mean ability levels as measured by the students' entrance SAT mathematics and verbal test scores were similar. Both sections were morning classes which met for three 50-minute periods a week. Pre/post achievement and attitudinal data were collected at the beginning and at the end of the semester in each section. Additionally course evaluation data were collected as well.

While the content and exams of the two sections remained the same, the sections did differ in the way they were taught. One section was traditionally taught (the non FA section) and the other had elements of formative assessment techniques embedded in it (the FA section). The students in the formative assessment courses were given weekly, small, content-based quizzes. The quizzes were graded, and any problem areas identified were discussed in the class day immediately following the "quiz day." Appropriate instructional modifications were made. However, once the quizzes had been discussed in the "formative assessment" course, all of the quizzes and answers were made available to all of the students in both sections on the course related web pages.

Statistical analyses of the data indicated that students in the FA section experienced a greater gain in achievement than did those in the non-FA section as

demonstrated in a regression analysis where the student's post-test score was the dependent variable and included the following independent variables: student's mathematics and verbal SAT scores as controls for prior knowledge, the number of hours the student reported doing homework and attending class, the student's age, a dummy for the student's desire for getting good grades, the student's score on the 7 point "like's science" scale, and a dummy for whether the student was in the FA section or not (see Table 1 for complete results). A regression captures the size of the effect and enables controlling for possible factors that could affect the results, in this case ability and attitude toward science. In this analysis, all other things equal, being in the section in which FA techniques were used added 51.2 points to a student's total point count for the semester. This gain represented a more than 6.4% increase as a function of the way the course was taught. In addition, graded on the same scale, 50% of the students in the FA section received an A or a B. In the non FA section, 39% received an A or a B. Also, 6% more students failed the non FA section. Further, FA section students gave higher rating to the course than non-FA-section students.

Gibbs and Lucas (1996) held that instructional methods need to vary by class format. This study supports that position in that it demonstrates the fact that formative assessment techniques can be successfully incorporated into the large lecture format with positive results.

Table 1
Study 1: Regression Analysis

	Regression Coefficient	Significance
In FA section	51.245	.017
Mathematics SAT	.238	.094
Verbal SAT	-.329	.019
Number hours/week spent on homework	-2.425	.026
Age	-9.411	.079
Want good grades dummy	72.714	.009
Like science scale (0-7)	10.834	.250
Constant	711.200	.000

Study 2: Formative Assessment in Differential Equations Courses

Sadler (1998), in an article about formative assessment, argued that grades may be counter productive to formative assessment in that they are focused on what has been accomplished and not what needs to be done. Taras (2002) argues that grades often have the unfortunate effect of distracting students from

what they should be focused on and that is learning. Further, according to Taras, "I reiterate that marks have a place even in formative assessment, but not in isolation and not before feedback and judgements have been interiorized" (p. 507). This study focuses on whether the students taking the quizzes also assume some control over their own learning, which will be measured by their performance on regular tests and the final exam.

In this study increased feedback to students was tried under different conditions in four sections of a differential equations course during two semesters at an urban university, two in the Spring 2007 and two in the Fall 2007 semesters. The university where the research was done is very large, thus reducing within semester and between semester contamination threats. The sections were generally of the same size (N=30 students) and did not differ in gender and race/ethnicity distributions, nor in their ability as measured by their entrance SAT scores.

The same materials and the same number of tests (4) were administered in each class. What differed was the weight of the quizzes. The course instructor, an experienced mathematician, opted to implement a number of quizzes in each course, but put only grades on some and detailed analyses on others, a strategy that had been found effective with younger students. There were three formative assessment sections, and one control section. Lastly, in addition to content-based pre- and post-tests, pre- and post-survey attitudinal and behavior data were collected as well. The number of students was reduced from 117 to 79 because of the need to have data from all of the different sources (pre-test results, post-test results, pre-survey results, and post-survey results). The students for whom data were complete were not different from those for whom the data were incomplete.

In these analyses, the dependent variable is the post-test content score. To control for confounding factors such as variability in the initial knowledge base, a regression analysis was performed (see Table 2 for details). The R Square indicated that 24 percent of the variation in the dependent variable was explained by the independent variables taken together. The F statistic was significant at the .000 level. The variable most strongly related to the dependent variable was the student's pre-test content score as evidenced by the Beta value of .41. However, controlling for differences in ability, being in one of the formative assessment sections added 10.30 points to the final score as shown by the regression coefficient, which is equivalent to a whole grade difference, that is, a "B," instead of a "C".

To assess how implementation of FA affected the students, an analysis of residuals was conducted. Here, the actual test score was subtracted from the predicted test score. A negative result meant that the student

Table 2
Study 2: Regression Analysis

	Regression Coefficients	Significance
Student in FA section dummy	10.30	0.028
Number of hours spent per week on going to class and doing homework	0.30	0.022
End of course academic self confidence dummy	5.09	0.082
Course pre-test results	0.17	0.002
Constant	40.55	

Table 3
Study 2: Number of Hours Per Week Devoted to Various Activities

	At the beginning of the semester				At the end of the semester			
	Overall mean	Students performing well above expectations	Students performing as expected	Students performing well below expectations	Overall mean	Students performing well above expectations	Students performing as expected	Students performing well below expectations
Watch TV	6.49	5.53	4.96	4.59	6.26	8.43	4.94	8.56
Play computer games	3.11	2.89	2.58	3.68	2.93	2.22	2.27	5.66
Talk to friends	12.26	13.11	11.44	10.55	9.88	11.09	9.83	8.50
Do household chores	5.88	4.11	7.13	7.15	6.53	4.72	6.88	7.60
Play sports	4.00	4.86	3.72	2.27	4.06	4.47	4.62	2.22
Work at a paying job	14.28	10.29	14.39	19.86	11.54	9.66	11.82	12.00
Go to class	18.32	18.43	18.00	15.63	17.67	17.19	17.99	17.34
Do homework	14.33	12.71	15.52	12.00	17.28	14.56	18.23	17.44

performed higher than expected, and a positive result meant that the student performed lower than predicted. In all, 58.2 percent of the students performed higher than expected. The difference score ranged from a student whose predicted score was 51.62 points higher than the actual score earned to a student whose predicted score was 31.81 points lower than what was actually earned. The first student performed below expectations, while the latter student performed above expectations.

Next the data were divided into three groups: those who achieved well above what was expected (80th percentile and above), those who achieved well below what was expected (20th percentile and below), and those in the middle percentiles. A student classified in the 80th percentile or higher on this difference score need not have achieved at the highest level, but certainly did achieve significantly higher than predicted. Also, it is possible for a student to have achieved a good grade, yet be in the 20th percentile or lower on the difference score. What would be true of such a student is that s/he achieved significantly less than predicted. The difference score is a value added model designed to capture the effects of what happened in the classes. While not statistically significant, a greater percentage of those achieving well above expectations were in the formative assessment sections

than was the case for those students in the control or non-formative assessment section.

At the beginning of the semester the students were asked to estimate the number of hours per week that they spent in eight areas. These items were included in the post survey as well. At the beginning of the semester, on average, 30.65 hours were spent per week on academic activities, going to class (18.32), doing homework (14.33), and talking to friends (12.26). At the end of the semester, the number of hours talking to friends declined and the number of hours doing homework increased. When the students are divided according to whether they performed well above what was predicted (80th percentile and higher), as expected (21st to 79th percentiles), or well below what was predicted (20th percentile and lower), interesting patterns emerge. It is apparent that those students who performed below prediction had time allocation problems from the start as they spent almost twice as much time at a paying job as those students who performed well above what was predicted (see Table 3 for details). This was at the expense of going to class, doing homework, and doing household chores. At the end of the semester, these students had reduced the number of hours working and increased the hours doing homework. It is apparent that getting a good start is crucial.

Integrating formative assessment techniques – in this case quizzes – into a university course during the semester did have a significant affect on student performance. In this case, students allocated their time differently. Those students who performed significantly above expectations devoted time early in the semester to their course work, while those who performed significantly below expectations did not. At the end of the semester, students in this latter group re-allocated their time and more than likely were playing “catch up.”

Study 3: Class Context: Assessing the Effects of Interactions

The use of formative assessment can be very time consuming both for the students and the faculty. In this study, the goal was to measure the effect size of participating in a class that was structured to facilitate the interaction component of formative assessment. The research question for this study was: To what extent is student achievement a function of differences in instructor/student interactions. The goal was to go beyond determining if there were a relationship between achievement and aspects of formative assessment and to quantify the difference if statistical significance were attained.

Among the formative assessment vehicles included in these analyses are online quizzes, instructor office visits, and email conversations with the instructor. Additional data that were collected include pre/post attitudinal and behavior surveys, pre/post subject knowledge tests, quiz taking history, and email and office logs of university students taking tests and measurement courses over two semesters.

Data were collected from upper division students enrolled in two sections of an Educational Assessment – Tests and Measurement course taught by the same instructor. The database is composed of student demographic and achievement items (gender, race/ethnicity, GPAs, SAT scores, course grades by components (tests, online quizzes, pre/post-test results, etc.), and attitudes and behaviors (electronic contacts and office contacts decomposed into FA-related and non-FA-related) and pre/post survey responses.

The students were told that they were participating in an NSF funded study and generally what its focus was, but specifics of the project were not discussed. The students were not paid stipends, but randomly selected students were given gift certificates to the university bookstore for their participation. The sections were generally of the same size (N=30 students) and did not differ in gender and race/ethnicity distributions, nor on their ability as measured by their entrance SAT scores. The same materials and the same number of tests were administered in each class. What

differed was the fact that online quizzes were incorporated into two of the sections and not in the other two.

Students in two sections completed an online quiz for each of 19 chapters prior to the scheduled session covering that chapter. Quizzes were available through a companion website (Luftig 2009) to the course text (Miller, Linn & Gronlund 2009) and were composed of 20-41 objective items per chapter. Students completed quizzes on their own and emailed the results to the instructor. Results sent to the instructor included percent correct and a log of answers to each item. To measure forms of student initiated formative assessment, the instructor kept a log of all student emails and office visits. Student- initiated contacts were coded as administrative or content-oriented. Contacts about schedule, syllabus, attendance, and grades were considered administrative issues. Requests for clarification on a procedure or concept and requests for assistance with assignments are two examples of content-oriented, student-initiated contacts.

In an analysis of the two sections for which online quizzes were available, the quiz average was not related to knowledge gain, but the number of quizzes taken was related. Additionally, in a regression analysis of those 46 students enrolled in the FA section, the percentage of contacts that were formative assessment was negatively related, and the percentage of electronic contacts was positively related (see Table 4 for further results). Thus, the findings indicate that complex relationships exist and that the attitudinal items need to be incorporated into the model being estimated.

Table 4
Study 3: Regression Analysis

	B	p
Pre-test (out of 100)	0.333	0.061
Number of quizzes completed out of 19	1.265	0.015
Total number of office visits	-1.221	0.026
Total number of non FA online contacts	1.514	0.031
Total number of online FA contacts	-4.947	0.046
Constant	26.621	0.028

It is apparent that integrating formative assessment techniques, in this case online quizzes, during the semester into a university course did have a significant effect on student performance. A number of issues still need to be addressed. Is this the only effect that the integration of formative assessment can have? Are some students affected more than others? Do some students need to be affected more than others? In a related study Stull, Schiller, Jansen Varnum, and Ducette (2008) showed that embedding in class formative assessment opportunities in mathematics courses prompted some students to study earlier in the

semester than others. Will this be the same with online formative assessment opportunities?

Study 4: The Use of Clickers in an Introductory Physics Class – Fostering Student Interaction as a Method of Formative Assessment

Use of personal response systems (primarily known as “clickers”) has become a widely recognized means of increasing student interaction in large lecture classes. These clickers allow the students to respond to various forms of instructor provided questions, usually in a multiple-choice format, and provide instantaneous feedback to the students and the instructor concerning the extent to which the students in the class have mastered the material. In effect, the use of clickers is a means by which instructors and students in large lecture classes can obtain the same kind of interaction that is available in small classes where instructor/student interaction is more feasible. As Duncan (2005) says, “... students press a button on a hand-held remote control device corresponding to their answer to a multiple choice question that is being projected on a screen, see the correct answer along with the class distribution of answers, and hear a description of the thinking that leads to the correct answer” (3).

Most of the research on clickers has focused on the perceptions of how useful and enjoyable students found these devices (Draper & Brown 2004; Duncan, 2005; Latessa & Mouw 2005; Campbell, Knight, & Zhang 2009). In general, this research has reported that students find clickers helpful in their attainment of course content. As some writers have commented, however, there is a clear possibility that the effectiveness of clickers may be due to some extent to the Hawthorne effect. Outside of these student opinion studies, however, there has been very little research investigating whether clickers have an impact on student achievement and attitudes. In addition, there has been no direct investigation of whether clickers are more or less effective for specific subgroups of students. The present study will fill some of these gaps in the literature by providing data from an introductory physics class in which clickers were used as one form of formative assessment.

This study was conducted at a large, public, urban university in the northeastern section of the country. The class in question was introductory physics, a course that meets the university’s requirement for a core science class as part of the general education requirements. The course is offered in both the fall and spring semesters, with approximately 150 students in each section. As part of a National Science Foundation Grant, the instructor agreed to offer the fall section of the course using the typical course format (large lecture with minimal class participation) and the spring

semester using clickers. In both classes the course content was as identical as possible under normal classroom conditions. Specifically, the same textbook was used in both classes, the course outline was identical, and all quizzes and assignments were identical. The only difference between the two classes was that the instructor used clickers as an integral part of his presentation in the spring course. For the most part, this involved the students responding to questions, usually in multiple-choice format, that covered the material already presented in the class. These data were then fed back to the instructor and to the students. If 75 percent or more of the class missed an item, the material was either immediately reviewed, or was taught in another format in a later lecture. The questions used for the clicker data were not presented again in any of the quizzes or the final.

In both classes a pre-survey was given to capture student’s attitudes toward, and previous experience with, science. The survey contained two types of questions: those focusing on content (e.g., “I’m sure I can understand the most difficult material presented in science class” and those that would be considered more “constructivist” in nature (e.g., “There is only one correct way to solve science problems” and “Learning science is mostly memorizing facts”). The same questionnaire was administered at the end of the course. In addition, both classes were given an identical set of course examinations consisting of three quizzes and a final exam. The classes were essentially the same in terms of their demographic profiles while their achievement results were not, as shown in Table 5 where the mean performance of the two classes of students, expressed as percentage correct for the three quizzes and the final are presented.

Table 5
Quiz and Final Exam Performance

	Quiz 1	Quiz 2	Quiz	Final
Clicker	48%	64%	76%	85%
Non	52%	51%	64%	77%

Since use of clickers is becoming more common, it is important that the impact of these devices be systematically studied. The results from this study offer support for clickers, but also indicate some areas of concern. It is significant that students in the clicker class obtained higher scores on the quizzes and final exam as compared to students in the non-clicker class. In an analysis of the attitudinal surveys, it appears that students in the clicker class seemed more confident in their ability to solve difficult problems. To some extent, however, these benefits may have been obtained at the cost of an over-emphasis on discrete and clearly demarcated outcomes. That is, the students in the

clicker class seemed intent on providing answers to the questions asked, and seemed less open to exploring and investigating physics. This was supported by the finding that students in the clicker class are more concerned with obtaining a good grade in the class. It is also interesting that clicker use seems most pronounced for students who have a moderate initial level of interest in the course. It is possible that students who had a high level of initial interest found that the clickers did not increase their understanding of the material, and they ultimately stopped paying as close attention as they should have. This is evidenced by the fact that the most pronounced difference in quiz performance occurred later in the course.

The data from this study suggest that the use of clickers can facilitate performance in physics, at least to the extent that this is measured by performance on objective quizzes and exams. It is also encouraging that the clickers seemed to enhance the students' sense of competency and mastery in dealing with the content. It is discouraging, however, that this enhancement seemed to be achieved by making the students over-emphasize concrete knowledge.

Conclusion

A number of points can be made about the use of formative assessment techniques. First, formative assessment clearly has a role to play in improving teaching and learning at the university level. While all of the instructors who participated in these studies were very skeptical of these formative assessment techniques at the outset, each has continued their use beyond the time frame of his or her study. Secondly, it appears that specifically "what is done" is less important since there were positive achievement gains in each study. Third, while some of the formative assessment techniques imply considerable instructor commitment (Study 1 and Study 2), positive student achievement effects were realized with lower levels of commitment (Study 3). Fourth, while positive gains were realized with use of technology, gains were also realized with implementation of nontechnology dependent techniques. Lastly, gains were independent of class size or subject matter.

References

- Ash, D., & Levitt, K. (2003). Working with the zone of proximal development: Formative assessment as professional development. *Journal of Science Instructor Education, 14*, 23-48.
- Athanases, S., & Achinstein, B. (2003). Focusing new instructors on individual and low performing students: The centrality of formative assessment in the mentor's repertoire of practice, *Instructors College Record, 105*, 1486-1520.
- Beatty, I. D., Feldman, A., Leonard, W. J., Gerace, W. J., St. Cyr, K., & Lee, H. (2008). Teacher learning of technology-enhanced formative assessment. Paper presented at the 2008 Annual International Conference of the US National Association for Research in Science Teaching (NARST), Baltimore, MD: April 2001.
- Black, P., & Wiliam, D. (1998). Inside the black box. *Phi Delta Kappan*. Retrieved from <http://www.pdkintl.org/kappan/kbla9810.htm>.
- Black, P., & Wiliam, D. (2003). In praise of educational research: Formative assessment, *British Educational Journal, 28*, 623-637.
- Boston, C. (2002). The concept of formative, assessment, practical assessment, research, and evaluation., Retrieved from <http://PAREonline.net/getvn.asp?v=8&n=9>
- Brookhart, S., Moss, C., & Long, B. (2010). Teacher inquiry into formative assessment practices in remedial reading classrooms. *Assessment in Education: Principles, Policy & Practice, 17*(1), 41-58.
- Campbell, J., Knight, A., & Zhang, H. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology, 34*, 51-57.
- Carrillo-de-la-Pena, M., Bailles, E., Caseras, X., Martinez, A., Ortet, G., & Perez, J. (2009). Formative assessment and academic achievement in pre-graduate students of health sciences. *Advances in Health Sciences Education, 14*, 61-67.
- Cassady, J. C., Budenz-Anders, J., Pavlechko, G., & Mock, W. (2001). *The effects of internet-based formative and summative assessment on test anxiety, perceptions of threat, and achievement* Paper presented at the Annual Meeting of the American Educational Research Association in Seattle, WA.
- Chen, C. M., & Chen, M. C. (2009). Mobile formative assessment tool based on data mining techniques for supporting web-based learning. *Computers & Education, 52*, 256-273.
- Chin, C., & Teou, L. (2009). Using concept cartoons in formative assessment: Scaffolding student's argumentation. *International Journal of Science Education, 10*(1), 1307-1332.
- Chung, G. K., Shel, T., & Kaiser, W. J. (2006). An exploratory study of a novel online formative assessment and instructional tool to promote students' circuit problem solving. *Journal of Technology, Learning, and Assessment, 5*. Retrieved from <http://www.jtla.org>.
- Costa, D., Mullan, B., Kothe, E., & Butow, P. (2010). A web-based formative assessment tool for Masters

- students: A pilot study. *Computers and Education*, 54, 1248-1253.
- Draper, S. W., & Brown, M. I. (2004). Increasing interactivity in lectures using an electronic voting system. *Journal of Computer Assisted Learning*, 20, 81-94.
- Duncan, D. (2005). *Clickers in the classroom: How to enhance science teaching using classroom response systems*. San Francisco, CA: Pearson/Addison-Wesley.
- Frey, N., & Fisher, D. (2009). Using common formative assessments as a source of professional development in an urban American elementary school. *Teaching and Teacher Education*, 25, 674-680.
- Furtak, E., & Ruiz-Primo, M. (2008). Making students' thinking explicit in writing and discussion: An analysis of formative assessment prompts. *Science Education*, 799-824.
- Furtak, E., Ruiz-Primo, M., Shemwell, J., Ayala, C., Brandon, P., Shavelson, R., & Yin, Y. (2008). On the fidelity of implementing embedded formative assessments and its relation to student learning. *Applied Measurement in Education*, 21, 360-389.
- Gibbs, G., & Lucas, L. (1996). Using research to improve student learning in large classes. In G. Gibbs (Ed.), *Improving student learning: Using research to improve student learning* (pp. 14-27). Oxford, UK: Oxford Center for Staff Development.
- Gipps, C. V. (2005). What is the role for ICT-based assessment in universities? *Studies in Higher Education*, 30, 171-180.
- Hatch, J., Jensen, M., & Moore, R. (2005). Manna from heaven or "clickers" from hell. *Journal of College Science Teaching*, 34, 36-39.
- Heinze, A., & Heinze, B. (2009). Blended e-learning skelton of conversation: Improving formative assessment in undergraduate dissertation supervision. *British Journal of Educational Technology*, 40(2), 294-305.
- Henly, D. C. (2003). Use of web-based formative assessment to support student learning. *European Journal of Dental Education*, 7, 116-122.
- Heritage, M., Kim, J., Vendlinski, & Herman, J. (2009). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practices*, 28(3), 24-31.
- Ibabe, D., & Jauregizar, J. (2010). Online self-assessment with feedback and metacognitive knowledge. *Higher Education*, 59, 243-258.
- Latessa, R., & Mouw, D. (2005). Use of an audience response system to augment interactive learning. *Family Medicine*, 37, 12-14.
- Lavooy, M., & Newlin, M. (2008). Online chats and cyber-office hours: Everything but the office. *International Journal on E-Learning*, 7, 107-116.
- Layng, T., Strikeleather, J., & Twyman, J. (2004). *Scientific formative evaluation: The role of individual learners in generating and predicting successful educational outcomes*. Paper presented at the National Invitational Conference on The Scientific Basis of Educational Productivity, sponsored by the American Psychological Association and the Laboratory for Student Success, May 13-14, Arlington, VA.
- Lipnevich, A., & Smith, J. (2009). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology*, 15(4), 319-333.
- Luftig, R. (2009). *Measurement and assessment in teaching*. Retrieved from http://wps.prenhall.com/chet_miller_measurement_10.
- Luttenegger, K. (2009). Formative assessment practices in reading instruction in pre-service teachers' elementary school classrooms. *Journal of Education for Teaching*, 3, 299-301.
- MacDonald, M. (2007). Toward formative assessment: The use of pedagogical documentation in early elementary classrooms. *Early Childhood Research Quarterly*, 22, 232-242.
- Marcotte, A., & Hintze, J. (2009). Incremental and predictive utility of formative assessment methods of reading comprehension. *Journal of School Psychology*, 47, 315-335.
- Mayer, R. E., Stull, A., DeLeeuw, K., Ameroth, K., Bimber, B., Chun, D., Zhang, H. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Education Psychology*, 3, 51-57.
- McFadden, A. C., Marsh, G. E., & Price, B. J. (2002). Computer testing in education: Emerging trends. *Computers in the Schools*, 18, 43-60.
- Miller, M., Linn, R., & Gronlund, N. (2009) *Measurement and assessment in teaching* (10th ed.). Upper Saddle River, NJ: Pearson.
- Miller, T. (2009). Formative computer-based assessment in higher education: The effectiveness of feedback in supporting student learning. *Assessment & Education in Higher Education*, 34(2), 181-192.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31, 199-218.
- Pachler, N., Daly, C., More, Y., & Mellar, H. (2010). Formative e-assessment: Practitioner cases. *Computers & Education*, 54, 715-721.
- Peat, M., & Franklin, S. (2002). Supporting student learning: The use of computer-based formative assessment modules. *British Journal of Educational Technology*, 33, 515-523.
- Puddy, R., Boles, R., Dreyer, M., Mailkranz, J., Roberts, M., & Vernberg, E. (2008).

- Demonstrating support for the formative and summative assessment paradigm in a school-based intensive mental health program. *Journal of Child Family Studies*, 17, 253-263.
- Ruiz-Primo, M., & Furtak, E. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching*, 44(1), 57-84.
- Ruston, A. (2005). Formative assessment: A key to deep learning? *Medical Teacher*, 27, 509-513.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(1), 145-165.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, (181), 145-165.
- Shavelson, R., Young, D., Ayala, C., Brandon, P., Furtak, E., & Ruiz-Primo, M. (2008). On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers. *Applied Measurement in Education*, 21, 295-314.
- Smith, G. (2007). How does student performance on formative assessments relate to learning assessed by exams? *Journal of College Science Teaching*, 36, 28-34.
- Stull, J., Schiller, J., Jansen Varnum, S., & Ducette, J. (2008). The use of formative assessment in university level mathematics courses. *Journal of Research in Education*, 18, 58-67.
- Taras, M. (2002). Using assessment for learning and learning for assessment. *Assessment and Evaluation in Higher Education*, 27, 501-510.
- Tierney, R. D., & Charland, J. (2007). Stocks and prospects: Research on formative assessment in secondary classrooms. Paper presented at the Annual Meeting of the American Educational Research Association, April, in Chicago, IL.
- Wang, K. H., Wang, T. H., Wang, W. L., & Huang, S. C. (2006). Learning styles and formative assessment strategy: Enhancing student achievement in web-based learning. *Journal of Computer Assisted Learning*, 22, 207-217.
- Wang, T. (2006). What strategies are effective for formative assessment in an e-learning environment? *Journal of Computer Assisted Learning*, 23, 171-186.
- William, D., & Black, P. (2003). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment. *British Educational Research Journal*, 22, 537-48.
- concentrated on statistics and research methodology. For the past 15 years she has been involved in research on improving teaching and learning. She has been a PI and co-PI on federal and state funded education grants. She holds a Ph.D. from Boston College.
- SUSAN JANSEN VARNUM is a Professor of Chemistry at Temple University. She has published widely in scientific journals and has produced 13 Ph.D. students. An analytical chemist, Dr. Varnum has directed numerous federally and locally funded projects designed to improve the educational opportunities of diverse student populations. She holds a Ph.D. from the University of Missouri.
- JOSEPH DUCETTE is a Professor of Educational Psychology at Temple University. Author of numerous articles in educational psychology, his research has focused primarily on diversity, urban education, and early intervention. He holds a Ph.D. from Cornell University.
- JOHN SCHILLER is an associate professor of mathematics at Temple University. He specializes in issues related to differential equations and has been involved in projects designed to improve the educational opportunities of minority students. He holds a Ph.D. from the University of Pennsylvania.
- MATHEW BERNACKI is completing his graduate studies in educational psychology at Temple University. His dissertation focuses on the use of educational technology in improving student learning.

JUDITH C. STULL is an Associate Professor of Sociology at LaSalle University. Her teaching has been