

# Promoting Student Ownership of Learning Through High-Impact Formative Assessment Practices

Susan M. Brookhart

Connie M. Moss

Beverly A. Long

*Duquesne University*

**ABSTRACT:** The most important instructional decisions, those with the greatest influence on student success, are made by learners themselves (Stiggins, 2008). Formative assessment, done well, contributes to student ownership of learning more than any other classroom-based instructional or assessment practice (Bloom, 1984). It is an economical, highly effective, and uniquely flexible method that can improve learning (Leahy, Lyon, Thompson, & Wiliam, 2005). Simply put, the teacher's purpose in formative assessment is to give students the means, motive, and opportunity to take control of their own learning. And, through their involvement in formative assessment, students develop self-efficacy for specific learning and, more generally, they develop skills that contribute to increased self-regulation and self-assessment of learning. In order for students to be meaningfully involved in formative assessment, they must be guided by teachers who hold the beliefs, knowledge and skills that engender active student engagement in the learning process. This paper highlights interim findings from a five-year professional development initiative involving the Armstrong School District, a large, rural school district in Western Pennsylvania, and the Center for Advancing the Study of Teaching and Learning at the Duquesne University School of Education. The initiative rests on the fusion of formative assessment, teacher-student communication, and student ownership of learning. The professional development program employs online modules, peer study groups, classroom walkthroughs, and teacher inquiry into their classroom practices and the beliefs that drive them. The program explores seven formative assessment components: 1) Identifying and Clearly Communicating Learning Targets, 2) Feedback that Feeds Forward, 3) Student Goal-Setting, 4) Student Self-Assessment, 5) Strategic Questioning, and 6) Formative Discourse. All components are linked to specific aspects of student motivation: intrinsic motivation, self-efficacy, self-regulation, goal-setting, and student attributions. The paper describes the impacts of formative assessment on student ownership of learning, student achievement, motivation, and active engagement as well as provides insights into teachers' experiences with student involvement. Findings show that not only have the teachers come to value and promote student ownership of learning using high impact formative assessment strategies, but that their efforts have resulted in high student engagement in learning and increased student achievement.

**KEYWORDS:** *formative assessment; teacher professional development; motivation; feedback; student self-assessment*

A strong research base supports the efficacy of formative assessment practices. Assessment practices in general

have long been known to have profound effects on student motivation, learning, and educational decisions (Crooks, 1988;

Natriello, 1987). Research suggests that, in particular, achievement increases when students experience formative assessment (Black & William, 1998; Meisels, Atkins-Burnett, Xue, & Bickel, 2003; Newman, Bryk, & Nagaoka, 2001; Rodriguez, 2004). Frequent, informational feedback is only one of the important components of effective formative assessment (Butler & Winne, 1995; Hattie & Timperely, 2007; Kluger & DeNisi, 1996). In the literature currently, “formative assessment” effect sizes are reported in the .40 to .70 range (based on several sections of the Black & William’s, 1998, review).

## Student Ownership of Learning

Formative assessment, as the term is used here, traces its roots back to instructional theory laid out by Sadler (1983, 1989). Scriven (1967) coined the term “formative evaluation,” meaning program evaluation carried out during program development, while there was the opportunity to use the evaluation results to improve the program or innovation. Bloom and his colleagues (1971) quickly saw the application of the formative concept to the classroom, concentrating on teachers’ formative use of evaluation information in selecting instructional methods and materials. It was Sadler (1983, 1989) who laid out the concept of formative assessment in the sense it is used today. Historically, the switch of terms from “formative evaluation” to “formative assessment” signaled the change in emphasis from a teacher-directed activity to a shared student-and-teacher activity (Brookhart, 2007). The formative assessment practices that have been established as so effective for student achievement rely not only on the teacher understanding where a student stands in his or her trajectory towards important learning goals, but also on the teacher utilizing methods that allow the student to clearly understand the goal, his or her current place in relation to it, and actions

that he or she could take to move closer to the learning goal. In fact, Bloom (1984) discovered this as he and his students engaged in “the search for methods of group instruction as effective as one-to-one tutoring” (p. 4). Bloom and his students found that Bloom’s mastery learning, with its teacher-directed use of formative evaluation, alone was not enough to approach the learning gains of tutoring, but when information available to the student was also enhanced, classroom learning gains did approach desired levels. Today we would call this the addition of formative assessment practices.

Perhaps the most influential research report to date on formative assessment has been the review by Black and William (1998). Their review was in several sections, because as they pointed out, studies of many different specific things—curriculum-based assessment, feedback, formative evaluation, and others—had proved relevant, and it was not possible to amalgamate them into one discussion. Nevertheless, the study was based on 250 publications, is widely cited, and provides broad support for a set of practices, following the definition by Sadler (1989), that collectively have come to be known as “formative assessment:

- clearly communicating the learning goal to the student
- providing information, via both teacher feedback and student self-assessment, about the location of current student work in relation to the goal
- providing strategies for moving closer to the goal and supporting the student in actually taking those steps.

## The Professional Development Agenda and Design

A rural school district in Pennsylvania has been partnering with the Center for Advancing the Study of Teaching and Learning (CASTL) at Duquesne University's School of Education for ten years for professional development, some of which is accomplished in an online environment. CASTL employs a process of systematic and intentional inquiry (*Teaching as Intentional Learning* [TIL], Moss, 2001, 2002), allowing teachers to examine their classroom practices and the beliefs and assumptions that drive those practices. The online learning environment provides convenience and flexibility for the school district, whose buildings are widely separated over its 436-square-mile area. In addition, the online environment organizes an abundance of resources that are available to educators twenty-four hours a day, seven days a week.

The CASTL online learning environment includes a library of selected and vetted documents and links, organized both by topic and by key principles of learning, as well as

tools for synchronous and asynchronous communication. These resources are available to participants in any one of several professional development programs or focused topic studies known as a CASTL Path of Inquiry. All areas of CASTL's online learning environment are organized to advance professional inquiry into a specific topic, issue, or concept and in the beliefs and assumptions teachers hold regarding them. For a more thorough treatment of the assumptions, theoretical framework, and design principles behind the professional development, see Moss (2002) and Moss and Goldbach (2000) and the CASTL website: [www.castl.duq.edu/Castl\\_pgrms\\_TIL\\_Ovrw.htm](http://www.castl.duq.edu/Castl_pgrms_TIL_Ovrw.htm).

Table 1 summarizes the development of the professional development program in formative assessment, begun in 2005, from a small seminar to a structured set of online modules, with assignments and individual feedback. These changes represent refinements based on evaluations of successive years and design changes to accommodate the increase in number of participants.

Table 1  
Summary of Professional Development Program

Year	Participants	Program Design
Year One (2005-06)	6—selected early literacy teachers	Intensive seminar Action research via the TIL process Online resources and postings
Year Two (2006-07)	18—early literacy teachers	Content delivered via face to face workshops Classroom tryouts and reflection supported by TIL process Online resources and postings
Year Three (2007-08)	68—early literacy & induction year teachers	Content delivered via seven online modules Classroom tryouts and reflection reported by assignments that included student work Individual feedback delivered online Online resources and postings

*Note:* For a description of the results for Years One and Two, see the respective technical reports (Brookhart, Moss, & Long, 2007, 2008) at [www.castl.duq.edu/Castl\\_TechReports.htm](http://www.castl.duq.edu/Castl_TechReports.htm).

In Year Three (2007-08), the district scaled up the project to include all early literacy teachers funded by Title I monies, as well as Induction Year (first-year) teachers, taking the number of participants to sixty-eight at the beginning of the year. The teachers engaged in seven online modules, each of which required them to review some material, engage in specific formative assessment strategies in their classroom, and file a report that included answering some questions and supplying samples of student work and instructional decisions. These assignments promoted systematic and intentional inquiry into what was happening in their classrooms, deepened their understanding of the content of each module, and became the core of the professional development. A few face-to-face visits with teachers were made for purposes of discussion and to facilitate participation in the modules, but the visits were not the mode of delivery of content. Each of the seven module reports, from each participant, received individual feedback. And one of the authors made eight site visits.

The schedule and topics of the online modules were as follows:

1. October, 2007: Overview of formative assessment and intrinsic motivation
2. November, 2007: Communicating learning targets and productive self-views
3. January, 2008: Feedback I: Feedback that feeds forward and self-efficacy
4. February, 2008: Feedback II: Feedback strategies and self-regulation
5. March, 2008: Student goal setting and self-assessment and self-attributions
6. April, 2008: Strategic questions and rich discourse and the ARCS model
7. May, 2008: Wrap-up

Each module presents content about formative assessment, structured with online

design principles that maximize visual impact by using appropriate images, everyday language, and relevant element interaction to reduce cognitive load (e.g., Paas, Renkl, & Sweller, 2003). A “misconception alert” feature is included in each module, as are guided practice experiences and self-check questions with built-in feedback.

Other aspects of the modules represented refinements in and extensions of the professional development based on findings about teacher knowledge and practice from Years One and Two. The focus remained on students. To provide material for the teachers as well as the researchers to analyze, each module except the first asks teachers to complete a “connection to practice case.” These assignments assist teachers in moving beyond comprehension of the surface features of formative assessment as a whole and each module in particular to a deeper understanding of how each topic impacts student achievement and motivation to learn. The cases presented teachers with authentic tasks that enabled them to demonstrate their competence in a tangible way. Each case includes answering some questions about their practices in regard to the topic of the module (communicating clear learning targets, providing high quality feedback, and so on) and including two student work samples as illustrations. One work sample is from a student who was successful on the learning target for the assignment, and one is from an unsuccessful student. Teachers describe these as part of their work. This serves two purposes. It supports the teachers’ inquiry into their professional knowledge, decisions of practice, beliefs and assumptions about both, and inquiry into specific aspects of students’ work. In addition, it allows the research team to evaluate the level of teacher knowledge and skill exemplified. The final feature of the formative assessment modules is the provision of individual feedback on a

“connection to practice case” each teacher submits with each module.

### *The Logic Model*

A logic model (see Figure 1) posited that teachers’ participation in professional development in formative assessment would lead to changes in their knowledge and practice, which would in turn eventually lead to changes in student achievement.

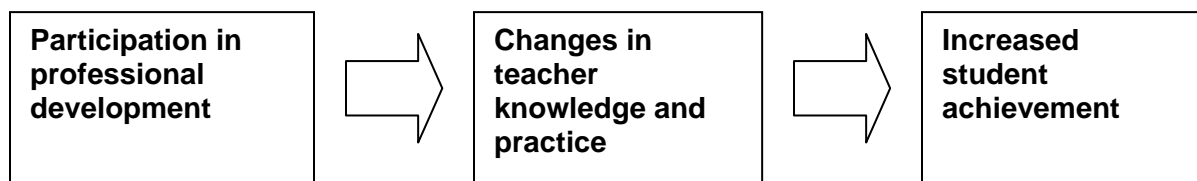


Figure 1. Basic Logic Model

Of course this logic model is an oversimplification that ignores building and classroom contexts, teacher and student backgrounds, and a host of other factors. Nevertheless, it functioned well as a guide for professional development design (the program aimed to help teachers change knowledge and practice in such a way as to impact student achievement) and data collection (indicators of changes in teacher knowledge and practice and indicators of student achievement were required).

## Method and Results

The purpose of this paper is to present results for Year Three. The methodology reflects district realities and not optimal research-design choices. Following the logic model, the research questions were as follows:

1. What changed in teachers’ knowledge and practice of formative assessment?
2. What effects on student achievement could be documented?

### *Teacher Knowledge and Practice*

*Survey of teacher confidence.* Before entering the first module, and as part of the final module, teachers complete a Formative Assessment Questionnaire (FAQ). The FAQ was patterned after the Classroom Assessment Confidence Questionnaire (Stiggins, Arter, J. Chappuis, & S. Chappuis, 2004), using a four-point scale—I’m uncertain about my confidence, I am not very confident, I’m somewhat confident, I’m very confident—scored for this analysis as 0 to 3.

Five scales were written to match the content of the formative assessment modules. For example, five items were written to measure confidence in “knowing my assessment purpose,” the topic of the first module. To illustrate, Item 1 read, “I am confident that I understand the various users of assessment information, including students, and can accommodate their need for assessment OF and FOR learning.” Scale scores were calculated as the mean of each of the items on each topic.

Table 2  
Formative Assessment Questionnaire Reliability ( $\alpha$ ) of Scales

Scale	<i>n</i> items	Pre <i>n</i> = 68	Post <i>n</i> = 42
Knowing My Assessment Purpose	5	.81	.58
Sharing Learning Targets & Success Criteria	7	.74	.74
Providing High Quality Feedback that Feeds Forward	10	.90	.78
Using Assessment to Actively Engage Students	8	.87	.77
Fostering Rich Dialogue and Using Strategic Questions	8	.89	.81

Cronbach's alpha was calculated to estimate internal consistency reliability. Scale information is presented in Table 2. Reliability was acceptable; lower reliability for the post-administration reflects decreased variability in posttest scores compared with pretest scores.

Teachers' confidence in their abilities to use these formative assessment practices

increased approximately one standard deviation for four scales (Knowing Assessment Purpose, Sharing Learning Targets, Providing Feedback, and Using Assessment to Actively Engage Students), and over half a standard deviation for Fostering Rich Dialogue and Using Strategic Questions.

Table 3  
Formative Assessment Questionnaire Pre-Post Analysis

Scale ( <i>n</i> = 42)	Pretest Mean (SD)	Posttest Mean (SD)	Change <i>df</i> = 41, <i>p</i> = .00
Knowing My Assessment Purpose	2.12 (.51)	2.53 (.31)	+ .41 <i>t</i> = 5.68
Sharing Learning Targets & Success Criteria	1.71 (.53)	2.32 (.43)	+ .60 <i>t</i> = 7.18
Providing High Quality Feedback that Feeds Forward	2.06 (.51)	2.49 (.32)	+ .44 <i>t</i> = 6.01
Using Assessment to Actively Engage Students	1.92 (.57)	2.36 (.35)	+ .45 <i>t</i> = 6.11
Fostering Rich Dialogue and Using Strategic Questions	2.15 (.55)	2.46 (.39)	+ .32 <i>t</i> = 4.03

*Teachers' reflections on their practice.* Teachers' final reflections were structured with a set of questions that matched the modules' content. Thirty-two teachers gave permission for their teacher reflections to be used in research; these teachers completed the entire set of Modules, assignments, and reflections.

Reflections were analyzed by summarizing each teacher's responses in a table, with a column for each question and a row for each teacher. Reading down columns provided a sense of the set of responses for each content area. Table 4 summarizes the main themes for each question.

Table 4  
Teacher Reflections on Main Themes

Question (abbr.)	Main Theme(s)	Quote Illustrating Main Theme
Describe your current understanding of classroom formative assessment.	Importance of student involvement; Importance of descriptive feedback	<ul style="list-style-type: none"> <li>▪ I see my students as more capable for assessing their own learning needs. Initially I felt it was my responsibility to set the learning targets. As the school year progressed, I released some of this power to the students, encouraging them to self-evaluate and determine what, how, and why they needed to learn. <i>Grades 4-6 Title I Reading</i></li> <li>▪ I became more descriptive with my feedback when I am doing an oral assessment with a student. Instead of just letting them know what is right or wrong. <i>Grade K</i></li> </ul>
What have you learned about motivation?	Feedback and knowing what is expected caused (formerly external) motivation to be internal to the student, and increased motivation.	<ul style="list-style-type: none"> <li>▪ I've seen students exhibit observable motivation at the promise of receiving a sticker or piece of candy upon the completion of some task or requested behavior. There always seems to be an eagerness on their part to comply in whatever is asked in order to receive that extrinsic reward. However, the quality of work is usually less ... typically the student rushes through the process, does the minimum necessary, and then wants to know when they will receive their prize. In contrast, when a student is shown his/her specific weak areas in reading and receives an explanation that their difficulty with that question is why the group will be working on a specific type of question or reading strategy that day, there is a more authentic and deeper motivation that is observable. <i>Grade 5 Title I Reading</i></li> </ul>
What have you learned about communicating learning targets to students?	Students need and want to know the learning targets; They become more motivated and also achieve more.	<ul style="list-style-type: none"> <li>▪ At first I thought this was an obvious point, but as I began thinking about my own teaching, I began to notice that although I stated the target, some of the students didn't internalize it to understand how that target pertained to them as a learner. Now I make it a point to state the target and discuss the importance. I also make sure to reiterate the target when speaking one-on-one to a student. <i>Grade 5 Mathematics</i></li> </ul>
What have you learned about giving students effective feedback?	Students want feedback and, for the most part, use it; Descriptive feedback works best.	<ul style="list-style-type: none"> <li>▪ Students can learn from both their successes and failures if the feedback is effective and timely. They remain positive because they learn what will help them achieve a goal and what sometimes might stand in their way to achieving that goal and what they can do about it. My feedback is only worthwhile if it is descriptive and timely. Students need to know specifics so they can achieve targets and gain confidence. – <i>Grade 1</i></li> </ul>
What have you learned about student self-assessment and goal-setting?	Self-assessment must be taught, but students can learn it (major theme); Giving over control of learning was an issue for four people.	<ul style="list-style-type: none"> <li>▪ ...the process of setting a goal seems almost like making a promise to yourself. <i>Grade 5 Title I Reading</i></li> <li>▪ I learned that most of the time I had to have students write their goals down or give them some time to think about their goals in class or else they would not set goals. <i>Grade 9 Mathematics</i></li> <li>▪ It is important for the students to set their own goals for learning. Sometimes, I need to step back and allow learning to happen by the students and not for the students. <i>Grade 11 Social Studies</i></li> </ul>

What have you learned about raising the quality of questioning and sharing ownership of questioning with students?

Higher level questions and discussion come when students are encouraged to question (surprising some); Students liked asking questions; Students had to learn how to ask higher-order questions; Wait time is important for teacher questions.

- Students need to be able to practice creating good questioning. By asking more probing questions, a teacher gets to see a clearer picture of what the child truly knows and understands. *Grade 3*
- My students surprised me. They were able to participate in questioning on a higher level than I anticipated. – *Grade1*
- I learned that my students have the ability to generate good thought-provoking questions if given the chance. I learned that it is very important to model good questioning strategies before expecting my kids to just be able to generate higher level questions. I also learned that wait time is important in generating and answering questions. *Grade 6 Title I Reading*

## *Student Achievement*

*Anecdotal evidence and claims of cause.* Teachers reported in their reflections that some students' achievement was a result of (caused by) changes in their formative assessment practices. Here are two examples.

### Grades 4-6 Title I reading:

An example of increased motivation through the use of formative assessment occurred this year after I showed each student a bar graph of the types of questions they were consistently missing on two 4Sight Benchmark tests from September to December. I asked students for their observations as they looked over their graphs, and I immediately witnessed their desire to improve in certain areas. For example, one student remarked that she really needed practice with inference questions because she noticed on her graph that she had missed several of these types of questions on the two tests. She next commented that she just doesn't "get" those types of questions. I saw this as a prime opportunity to point out to her that many students seemed to be having difficulty with that type of question, and we would be learning a new strategy called, "It Says, I Say, and So" that can really help with answering inferences. Conversations like this have occurred across the three grade levels (Grades 4-6) in which I teach.

### Grade 2 Title I reading:

My children are asked to learn 300 sight words during the year. In past years, I would have a treat or allow a trip to the prize box whenever 100 of the words were learned. I would assess the words a few different times during the year, allowing most

of the learning to be done through the regular classroom or incidentally in my classroom. This year, we worked on learning the sight words almost every day in my classroom. The 100 word lists were broken into groups of 30, 30, and 40. We practiced on word cards and lists, highlighting words green if we knew them, yellow if we weren't sure, and red if we were stopped by the word. We took those lists home and practiced; we practiced with a peer at least two times a week. The students tracked their progress by moving a magnet with their name under the different numbered lists. When the students knew they were ready to read to me, they moved their magnet to the card that said, Ready to Check. The students were anxious to see their name moving from under the card indicating that they were working on words 1-30 to under the card for 31-60 and so on. They would come into class and tell me, "I'm ready to read to you." I would hear them say to one another, "which list are you working on; I can't get the word 'were'; I'm using a sentence to help me remember."

The children motivated themselves to continue learning the 300 words and only two of the 28 students that I have serviced all year did not complete the task. The children were satisfied taking a new list home to practice and moving a magnet with their name to show their progress. Success bred success while we practiced and learned words. By having a small attainable goal the children were motivated to continue working so that they reached the end goal.

### *Achievement test evidence: Measures and analyses.*

Identifying appropriate achievement measures was difficult for this project. In fact, while the teachers feel strongly that many incidents like the two above occurred and that students did benefit from their teachers' increased knowledge and use of student-centered

formative assessment strategies, we have not been able to document that using standardized achievement evidence. Of the 62 teachers who completed at least the majority of the modules, 38 taught reading or English language arts, including Title I-funded reading teachers, elementary classroom teachers who did most of their formative assessment professional development in reading or writing, and a middle school learning-support English teacher. Sixteen taught mathematics or science, including elementary classroom teachers who did most of their formative assessment professional development in math or science and middle and high school mathematics and science teachers. Eight taught a variety of other subjects and grade levels, including social studies (3), emotional support, health and physical education, technical education, consumer science, and business and communications.

We decided to focus on elementary reading measures for two reasons. First, there were enough teachers and students at this level for reasonable sample sizes in elementary reading. Second, a standardized measure of classroom-based achievement was available (DIBELS, see below) at the appropriate “grain size.” The standardized measure available for other grade levels was the Pennsylvania System of School Assessment (PSSA). The PSSA measures achievement at the level of state standards, while the formative assessment practices in this professional development were exercises at the level of daily classroom learning targets.

Identifying appropriate comparison groups was also difficult for this project. In elementary reading, all the Title I-funded reading teachers were participants and had been working on formative assessment practices. Only some classroom teachers participated, and some of those did so because it was required as part of their induction year activities. Any benefit from formative assessment might have been offset

by the first-year teachers’ adjustment to their new careers.

Given the problems with identifying measures and identifying appropriate analysis methods, we did the best we could. Despite the qualitative evidence that learning occurred, however, the standardized achievement results are not definitive. There was some good news in Kindergarten, especially, but no causal attribution is possible for that, and in first and second grade, overall students of participants in the program performed about as well as other students in 2007-08.

*DIBELS assessments.* Dynamic Indicators of Basic Early Literacy Skills (DIBELS) measures are designed for benchmark testing in early literacy. These measures do not encompass all reading achievement constructs of interest. However, they were the indicators available from the district and used in district decision making.

The primary focus of instruction in Kindergarten was the learning of the letter names. First grade concentrated on blending sounds to make words (simple decoding) and oral reading fluency. Second grade concentrated on oral reading fluency.

- All students in Kindergarten ( $n = 382$ ) were tested in September, January, and May using the DIBELS Letter Naming Fluency (LNF) measure.
- First-grade students ( $n = 419$ ) were tested at the same intervals using the DIBELS Phoneme Segmentation Fluency (PSF) measure. First-grade students were also tested in January and May using the DIBELS Oral Reading Fluency (ORF) measure.
- Second-grade students ( $n = 444$ ) were tested in September, January and May using the DIBELS Oral Reading Fluency (ORF) measure.

Repeated measures analysis of variance with three (beginning, middle, and end-of-year) or two (middle and end-of-year) levels were used to analyze student progress in reading in each grade. Between-subjects factors compared groups (see grade level sections for a description of groups).

*Kindergarten.* In Kindergarten, Title I identified students (whose teachers were program participants) were compared with non-Title I students (whose teachers were not program

participants), overall and in an analysis using just those students whom the DIBELS identified as at-risk. The first analysis described the context of the whole class but confounded program participation with student level; the second analysis more appropriately compared students who were at-risk in reading. “At-risk” was defined as students having a beginning year LNF score less than 8 (University of Oregon, 2008).

Table 5  
Kindergarten DIBELS LNF Results

	Beginning Mean (s.d.)	Mid-year Mean (s.d.)	End-year Mean (s.d.)	Sample size
Title I / FA group	11.60 (12.37)	36.70 (17.04)	51.89 (17.39)	149
Non-Title I, Not in FA	24.13 (13.65)	43.97 (14.39)	55.39 (15.49)	233
Total	19.25 (14.50)	41.13 (15.86)	54.02 (16.33)	382

In Kindergarten, all students improved during the year (main effect of time,  $F(2,760) = 1353.96, p = .00, \text{partial } \eta^2 = .78$ ). Difference in improvement was steeper for students of Title I, formative assessment group teachers (interaction effect of time X group, ( $F(2,760) = 21.33, p = .00, \text{partial } \eta^2 = .05$ )). While on average all students learned their letters by the end of the year, students of

Title I, formative assessment group teachers improved more to get there.

In this analysis of Kindergarten at-risk students, all students improved during the year (main effect of time,  $F(2,208) = 361.40, p = .00, \text{partial } \eta^2 = .78$ ). There was no main effect by group, and no interaction.

Table 6  
Kindergarten DIBELS LNF Results – Students at-risk (Beginning LNF < 8)

	Beginning Mean (s.d.)	Mid-year Mean (s.d.)	End-year Mean (s.d.)	Sample size
Title I / FA group	2.49 (2.26)	27.96 (15.51)	46.57 (16.89)	77
Non-Title I, Not in FA	4.10 (2.34)	30.21 (13.91)	46.10 (13.72)	29
Total	2.93 (2.38)	28.58 (15.06)	46.44 (16.02)	106

*Grade 1.* In Grade 1, the comparison group was not so clean. All of the Title I identified reading students had at least one teacher (their Title I teacher) in the program. For some students, their regular classroom teacher was also in the program and working on formative assessment; however, some of these were

experienced regular classroom teachers who participated voluntarily and some were inductees (first year teachers) whose participation was required. Some of the non-Title I students had a classroom teacher in the program, and some did not. We constructed four groups. For this analysis, we coded

students as “exposed” to formative assessment in the regular classroom if their teacher participated voluntarily, excluding the first-year teachers whose participation was required, reasoning that their inexperience would confound any program effects.

- Group 1 = Exposure to two program teachers (Title I students whose regular classroom teacher was also in the program voluntarily)
- Group 2 = Exposure to one program teacher (Title I students whose regular classroom teacher was not in the program)
- Group 3 = Exposure to one program teacher (non-Title I students whose regular classroom teacher was in the program voluntarily)
- Group 4 = Exposure to zero program teachers (non-Title I students whose regular classroom teacher was not in the program)

In this section, we report only the results for the analyses of students at-risk for poor language and reading outcomes, as measured by PSF and ORF, respectively. Doing this four-group comparison and using all students results in an unbalanced design (an overwhelming majority of the students are in group 4), and confounds reading ability with group membership. Using only at-risk students, it seems reasonable to say that all students in this analysis were low readers, even those who were not receiving Title I reading services.

Table 7  
Grade 1 DIBELS PSF Results: Students at-risk (Beginning PSF < 35)

	Beginning Mean (s.d.)	Mid-year Mean (s.d.)	End-year Mean (s.d.)	Sample size
Group 1 – Title I & CT	24.25 (12.84)	43.88 (13.94)	51.38 (12.55)	8
Group 2 – Title I only	16.83 (10.69)	46.21 (14.16)	47.79 (12.33)	24
Group 3 – CT only	23.50 (10.81)	44.00 (10.76)	51.63 (8.33)	8
Group 4	21.06 (11.06)	43.12 (21.94)	43.41 (20.07)	17
Total	20.07 (11.23)	44.65 (16.14)	47.53 (14.68)	57

Title = Title I reading student, whose Title I teacher participated in the formative assessment professional development program.  
CT = Classroom Teacher participated voluntarily in the formative assessment professional development program.

In this analysis of Grade 1 PSF for at-risk students, all students improved during the year (main effect of time,  $F(2,106) = 98.90, p$

$= .00$ , partial  $\eta^2 = .65$ ). There was no main effect by group, and no interaction.

Table 8  
Grade 1 DIBELS ORF Results – Students at-risk (Mid-year ORF < 20)

	Beginning Mean (s.d.)	Mid-year Mean (s.d.)	End-year Mean (s.d.)	Sample size
Group 1 – Title I & CT	na	11.50 (4.84)	29.45 (14.86)	20
Group 2 – Title I only	na	12.99 (4.39)	35.67 (13.35)	69
Group 3 – CT only	na	14.77 (3.35)	48.23 (17.38)	13
Group 4	na	14.71 (3.75)	43.73 (16.85)	66
Total	na	13.63 (4.25)	39.07 (16.15)	168

Title = Title I reading student, whose Title I teacher participated in the formative assessment professional development program.

CT = Classroom Teacher participated voluntarily in the formative assessment professional development program.

In this analysis of Grade 1 ORF for at-risk students, all students improved during the year (main effect of time,  $F(1,164) = 391.56, p = .00$ , partial  $\eta^2 = .71$ ). The groups also differed in overall performance (main effect of group,  $F(3,164) = 7.54, p = .00$ , partial  $\eta^2 = .12$ ), and there was an interaction ( $F(3,164) = 6.40, p = .00$ , partial  $\eta^2 = .11$ ).

Graphing these means makes these effects easier to see. The two Title I groups, as might be expected, increased less than the other two groups, but otherwise the order of groups' improvement is not easily interpretable.

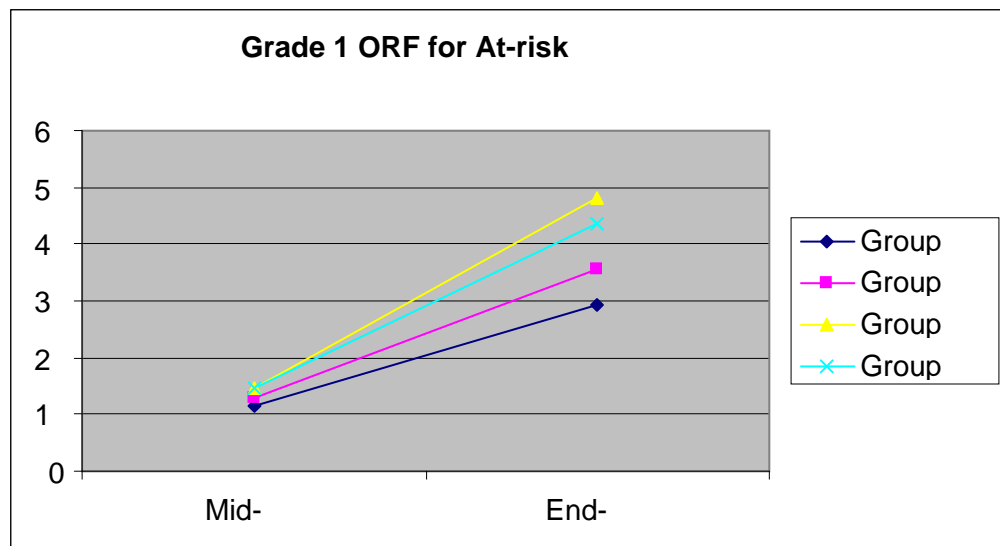


Figure 2. Grade 1 ORF for At-Risk Students

Another point to note is that for first-grade students in this district, the group sample sizes of students identified as “at risk” by the DIBELS PSF measures were much smaller than the group sample sizes of

students identified as “at risk” by the DIBELS ORF.

Grade 2. In Grade 2, for the same reasons as for Grade 1, we constructed four groups:

Group 1 = Exposure to two program teachers (Title I students whose regular classroom teacher was also in the program voluntarily)  
 Group 2 = Exposure to one program teacher (Title I students whose regular classroom teacher was not in the program)  
 Group 3 = Exposure to one program teacher (non-Title I students whose regular classroom teacher was in the program voluntarily)  
 Group 4 = Exposure to zero program teachers (non-Title I students whose regular classroom teacher was not in the program)

As for Grade 1, we report only the results for the analyses of students at-risk for poor language and reading outcomes, as measured by ORF.

Table 9  
 Grade 2 DIBELS ORF Results: Students at-risk (Beginning ORF < 44)

	Beginning Mean (s.d.)	Mid-year Mean (s.d.)	End-year Mean (s.d.)	Sample size
Group 1 – Title I & CT	27.50 (7.72)	53.30 (16.50)	70.80 (11.84)	10
Group 2 – Title I only	30.06 (6.02)	52.16 (12.74)	73.06 (17.59)	80
Group 3 – CT only	28.33 (13.22)	49.50 (22.01)	59.50 (24.83)	6
Group 4	33.63 (8.51)	57.88 (19.156)	77.02 (26.16)	52
Total	31.07 (7.63)	54.14 (15.99)	73.75 (21.12)	148

Title = Title I reading student, whose Title I teacher participated in the formative assessment professional development program.

CT = Classroom Teacher participated voluntarily in the formative assessment professional development program.

In this analysis of Grade 2 ORF for at-risk students, all students improved during the year (main effect of time,  $F(2,288) = 230.97$ ,  $p = .00$ , partial  $\eta^2 = .62$ ). There was no main

effect by group, and no interaction. The graph below shows that all groups had approximately the same learning curve except for Group 3, which only had six students in it.

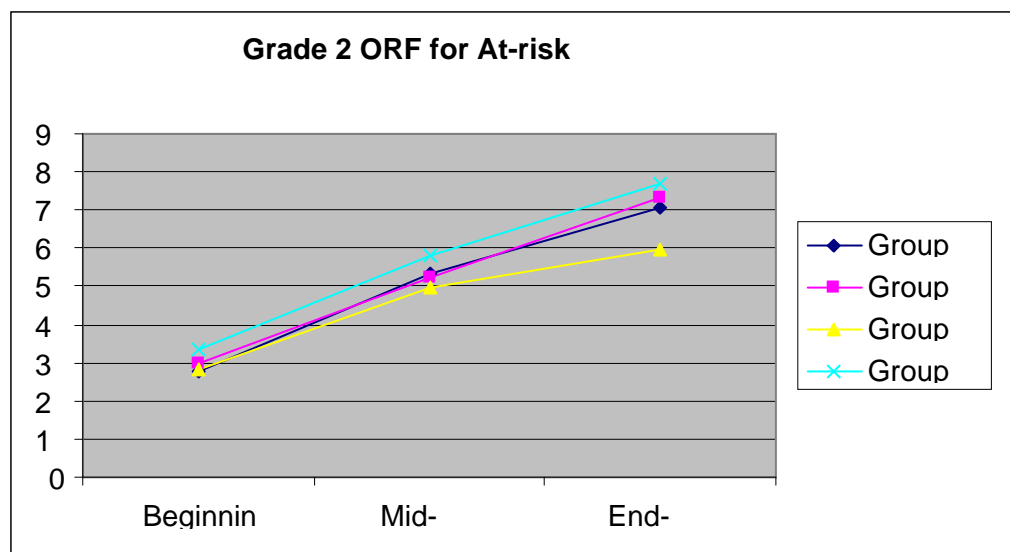


Figure 3. Grade 2 ORF for At-Risk Students

## Conclusion

The Assessment FOR Learning professional learning agenda carried out between CASTL at Duquesne University School of Education and the school district has had mostly positive outcomes, interpretable according to the logic model. There was evidence of teacher learning. Teachers' involvement in professional learning experiences in formative assessment showed them the importance of specificity in feedback, of systematically comparing student work with learning targets, and of student involvement in their own goal-setting and self-assessment (Ames & Archer, 1988; Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Black & Wiliam, 1998; Butler & Winne, 1995; Kluger & DeNisi, 1996; Meece & Miller, 1999; Meisels, Atkins-Burnett, Xue, & Bickel, 2003; Newman, Bryk, & Nagaoka, 2001; Rodriguez, 2004). Major changes in teacher perspectives included observation of deeper motivation as students realized they had information to control their own improvement and surprise at the capability even of very young students to do self-assessment and to ask higher-order questions—when these things were taught. Further, these outcomes were accomplished in the context of (a) a rural district with no unusual advantages, (b) where buildings are widely spread, which makes collegial communication difficult, (c) in contexts that might be expected to constrain (reading teachers with highly scripted curriculum and novice teachers, respectively).

Evidence that these changes in teacher knowledge and practice led to student achievement was limited to teacher observations of their own students. Changes did not show through in standardized achievement measures. However, the appropriateness of both available measures and available comparison groups is questionable. In particular, formative assessment is hypothesized to work by supporting student self-regulation and learning autonomy, and the DIBELS measures are based on a more incremental, discrete building-block

view of reading. The reported changes in teacher knowledge and practice, coupled with a significant step toward closing the achievement gap between Title I and non-Title I students, has led the district in this study to expand this process to all teachers, regardless of grade level or subject area. Believing that student achievement gains for Title I students were more dramatic than what was demonstrated through the DIBELS testing, the federal programs coordinator has selected additional standardized measures of reading comprehension and phonics skills that can be used to track student progress.

An additional factor that might contribute to the limited evidence of student achievement could be the schedule of the professional development itself. Since we worked with a large group of first-year hires, we decided it best to allow these new teachers a few months (August through September) to become acclimated to their classrooms and the school district. Therefore, the monthly schedule of modules did not begin until October of 2007 and there was no module in December. Each module lasted one month to provide teachers ample time to explore new content, connect new learning to learning from previous modules, implement new learning in their classrooms, and submit materials for each Connection to Practice Case. In effect, teachers were introduced to formative assessment in October and November, began to explore formative feedback in January and February, worked on student goal-setting and self-assessment in March, and moved into using strategic teacher questioning and encouraging student questioning in April. And while our study looks for the impact of the professional development on student achievement at the end of the 2007-08 school year, it may have been premature. Implementing an element or two of the formative assessment process—like communicating learning targets and the criteria for success or using more descriptive feedback—may not be enough to impact student

achievement. In fact, we would argue that all of the elements in our model must work together with consistency and high degrees of fidelity in order for student achievement to be effected.

Finally, our study does not consider the inconsistencies present in the quality and uniformity of teacher engagement (differences among teachers who implemented high quality formative assessment practices and those who did not, or teachers who consistently incorporated formative assessment into their daily practice and those who were unable to successfully integrate the process).

What we did find, however, is that when teachers are given the chance to explore formative assessment in their classrooms over time, teachers are able to use the formative assessment process to focus their self-improvement efforts in intentional and belief altering ways.

## References

- Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology, 80*, 260-267.
- Bailey, A. L., & Drummond, K. V. (2006). Who is at risk and why? Teachers' reasons for concern and their understanding and assessment of early literacy. *Educational Assessment, 11*, 149-178.
- Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*, 213-238.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*, 7-74.
- Bloom, B. (1984). The search for methods of group instruction as effective as on-to-one tutoring. *Educational Leadership, 41*(8), 4-17.
- Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill.
- Brookhart, S. M. (2007). Expanding views about formative classroom assessment: A review of the literature. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp.43-62). New York: Teachers College Press.
- Brookhart, S. M., Moss, C. M., & Long, B. A. (2007). A cross-case analysis of teacher inquiry into formative assessment practices in six Title I reading classrooms. *CASTL Technical Report Series No. 1-07*. Retrieved June 7, 2009, from [www.castl.duq.edu/Castl\\_TechReports.htm](http://www.castl.duq.edu/Castl_TechReports.htm)
- Brookhart, S. M., Moss, C. M., & Long, B. A. (2008). Teacher inquiry into formative assessment practices among Title I reading teachers and literacy coaches. *CASTL Technical Report Series No. 2-08*. Retrieved June 7, 2009, from [www.castl.duq.edu/Castl\\_TechReports.htm](http://www.castl.duq.edu/Castl_TechReports.htm)
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*, 245-281.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*, 438-481.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81-112.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254-284.
- Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership, 63*(3), 18-24.
- Meece, J. L., & Miller, S. D. (1999). Changes in elementary school children's achievement goals for reading and writing: Results of a longitudinal and an intervention study. *Scientific Studies of Reading, 3*, 207-229.
- Meisels, S., Atkins-Burnett, S., Xue, Y., & Bickel, D. D. (2003). Creating a system of

- accountability: The impact of instructional assessment on elementary children's achievement scores. *Educational Policy Analysis Archives*, 11(9). Retrieved June 7, 2009, from epaa.asu.edu/epaa/v11n9/
- Moss, C. M. (2001). *A resource guide for the teacher scholar*. Pittsburgh, PA: School of Education, Duquesne University. Retrieved Jun 7, 2009, from www.castl.duq.edu/res\_guide/
- Moss, C. M. (2002). Professional learning on the cyber sea: What is the point of contact? In R. all (Ed.), Special topic issue: World Wide Web and education, *Journal of CyberPsychology and Behavior*, 1(3), 41-50.
- Moss, C. M., & Goldbach, D. (2000). Designing a learning environment that learns: Design decisions for teaching as intentional learning. *CASTL Technical Report Series No. 1-00*. Retrieved June 7, 2009, from www.castl.duq.edu/Castl\_TechReports.htm
- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, 22, 155-175.
- Newmann, F. M., Bryk, A. S., & Nagaoka, J. K. (2001, January). *Authentic intellectual work and standardized tests: Conflict or coexistence?* Chicago: Consortium on Chicago School Research.
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, 38(1), 1-4.
- Rodriguez, M. C. (2004). The role of classroom assessment in student performance on TIMMS. *Applied Measurement in Education*, 17, 1-24.
- Sadler, D. R. (1983). Evaluation and the improvement of academic learning. *Journal of Higher Education*, 54, 60-79.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39-83). Chicago: Rand McNally.
- Stiggins, R. J. (2008). *Student-involved assessment FOR learning*. Upper Saddle River, NJ: Pearson Merrill Prentice Hall.
- Stiggins, R. J., Arter, J. A., Chappuis, J., & Chappuis, S. (2004). *Classroom assessment for student learning: Doing it right—using it well*. Portland, OR: Assessment Training Institute.
- University of Oregon Center on Teaching and Learning. (2008y). *DIBELS benchmark goals: Three assessment periods per year*. Retrieved June 7, 2009, from dibels.uoregon.edu/docs/benchmarkgoals.pdf