The Effects of Professional Development in Formative Assessment on Mathematics Teaching Performance and Student Achievement

Maggie B. McGatha
Assistant Professor of Mathematics Education, University of Louisville

William S. Bush
Professor of Mathematics Education and Director of the Center for Research in Mathematics and Science Teacher Development, University of Louisville

Christopher Rakes
Doctoral Fellow in Mathematics Education, University of Louisville

ABSTRACT: This article describes a year-long professional development program on formative assessment for middle school mathematics teachers and its impact on teachers' use of formative assessment during instruction. The focus of the program was to support teachers in (a) embedding formative assessments into their instruction, (b) interpreting student responses and performance on formative assessments, (c) using effective questioning strategies and peer assessment, and (d) identifying student errors and misconceptions and implementing instructional strategies to address them. Teacher and student performance data were collected on participating teachers, control teachers, and their students during the year-long professional development. The professional development had an impact on teachers' cognitive level of questioning, use of peer assessment, and types of questioning strategies. Results of student performance data revealed differences in performance on some types of items among students of participating teachers and control teachers.

KEYWORDS: formative assessment, mathematics, professional development, student achievement, teacher attitudes, teacher efficacy

Formative assessment, as a viable and effective process to improve mathematics learning at all grade levels, has received increased attention from school administrators and teachers in recent years. Also called assessment *for* learning, it is a cyclic instructional process in which teachers continuously gather information about what students know and plan and implement instructional activities accordingly. The activities may be directed to individual students, groups of students, or whole classes. It is a way of thinking about and organizing the teaching-learning process that has proven to promote successful learning (Black and Wiliam, 1998; Guskey, 2003; Stiggins, 2005).

What is Formative Assessment?

Implementing formative assessment effectively requires teachers to reconceptualize their role as teachers, the roles of students, and their interactions with students (Black & Wiliam, 2005). Wiliam and Thompson (2008) discuss these changes through five key strategies for implementing formative assessment effectively:

1. Clarifying, sharing, and understanding goals for learning and criteria for success with learners.
2. Engineering effective classroom discussions, questions, activities, and tasks that elicit evidence of student learning.
3. Providing feedback that moves learning forward.
4. Activating students as owners of their own learning.
5. Activating students as learning resources for one another. (p. 64)

From their extensive work with classroom teachers, Black, Harrison, Lee, Marshall, and Wiliam (2004) found that it was more productive for teachers to implement formative assessments by focusing on one area of change at a time because "wholesale changes can be too risky and demanding" (p. 20). They also discovered that having teachers collaborate with other teachers who attempt similar assessment changes increases the likelihood of success. Other researchers have found that implementing formative assessment that results in improved student achievement is time consuming and requires sustained professional development and support (Black & Wiliam, 1998; Black, Harrison, Lee, Marshall, & Wiliam, 2004; Stiggins, 2004).

Research Support for Formative Assessment

Black and Wiliam (1998) conducted a meta-analysis of 250 research studies on the effects of formative assessment and found that, when implemented with fidelity, it "produced significant and often substantial learning gains" (p. 140). The studies also indicated that formative assessment was especially effective with low-achieving students and in reducing achievement gaps. In a more recent analysis of the research, Wiliam and Thompson (2008) found that formative assessment produced more significant increases in student achievement than reducing class size or increasing teachers' content knowledge. Finally, research by Wiliam, Lee, Harrison and Black (2004) showed that formative assessment increased student involvement in their own learning and teachers' professional satisfaction.

Research on Effective Professional Development Practices

Strategies that promote effective professional development in mathematics education have recently undergone important shifts. In the past, the predominant strategy for professional development was to use outside experts to increase teachers' knowledge of content, a particular teaching approach, or a specific program. Howe and Stubbs (1997) suggest that this passive approach to professional development is unlikely to affect individual teachers or schoolwide change. More recently, research has shown a more purposeful approach to professional development that includes job-embedded practices and ongoing opportunities for professional growth and systemic change is more effective in improving teaching (Guskey, 2003; Loucks-Horsley, Love, Stiles, Mundry, & Hewson, 2003; National Council of Teachers of Mathematics, 2000; National Staff Development Council, 2001; Sparks, 2004). Heck, Banilower, Weiss, & Rosenberg (2008) report that recent research reveals the following essential elements of effective professional development: (a) situating the activity in classroom practice, (b) actively involving teachers, (c) treating teachers like professionals, and (d) focusing on disciplinary content.

Formative Assessments in Middle Mathematics Professional Development

A team of faculty, staff, and former teachers in the Center for Research in Mathematics and Science Teacher Development at the University of Louisville designed and implemented a year-long, sixty-contact-hour professional development program to encourage and assist middle school mathematics teachers in integrating formative assessment strategies into their instructional practices. The program, Formative Assessments in Middle Mathematics (FAMM), used two interaction formats: (a) thirty contact hours in the summer (six hours each day for five days) at a central location and (b) monthly meetings (an additional thirty hours) throughout the school year. The monthly meetings included a combination of after-school large-group meetings (all teachers and leaders) and small-group regional meetings (five or six teachers from several schools and one leader). In designing the program, the leaders attempted to incorporate strategies that have proven effective through previous research on professional development. For example, teachers engaged in small-group activities and discussions, analyzed video and written case studies of mathematics teaching, analyzed written student work and videos of student interviews, completed self-assessments of their own formative assessment practices, and wrote professional growth plans to guide their work during the school year. During the five-day summer meetings, activities focused on

- clarifying and classifying student performance objectives and outcomes
- analyzing mathematics content embedded in standards and assessment items
- analyzing depth of knowledge (DOK) of standards and assessment items (DOK level 1—recall; DOK level 2—understanding skills/concepts; DOK level 3—strategic thinking) (Webb, 1997)
- writing; finding, and adapting clear and appropriate assessment items
- embedding formative assessments into instruction
- interpreting student responses and performance on formative assessments
- identifying student procedural and memory errors and misconceptions and implementing strategies to correct them
- identifying student difficulties in reasoning and problem solving and implementing strategies to correct them;
- discussing and analyzing teacher cases about classroom assessment
- discussing and analyzing questioning interaction patterns (Initiation-Response-Feedback (IRF)—teacher offers evaluative feedback to students' response; Funneling—guiding students to desired outcome based on teachers' thinking; Focusing— guiding students to desired outcome based on students' thinking) (Herbal-Eisenmann & Breyfogle, 2005)
- meeting in cohorts at designated times during sessions.

During the school year, activities focused on

- analyzing assessments that teachers used during mathematics lessons
- analyzing and discussing actual student work
- incorporating peer- and self-assessment strategies
- discussing challenges and problems faced by teachers in the classroom.

Each teacher received a set of readings and reference materials that included *Mathematics Assessment: A Practical Handbook for Grades 6-8* (Bush & Leinwand, 2000) from the National

Council of Teachers of Mathematics and *Elementary and Middle School Mathematics: Teaching Developmentally* (Van de Walle, 2007).

In designing the professional development experiences, the leaders decided at the outset to focus the mathematics content of activities primarily on rational numbers. This topic permeates the sixth-, seventh-, and eighth-grade mathematics curricula, so they believed that supporting teachers' knowledge and teaching expertise specifically with respect to rational number content would provide additional benefits to the teachers. Therefore, all demonstration lessons, samples of student work, session activities, and videos of teaching centered on rational numbers in some way.

Research Questions

This study sought to answer the following questions regarding the effects of the year-long professional development experience on teachers' implementation of formative assessment practices:

1. To what extent did the professional development experience have an impact on teachers' attitudes toward teaching mathematics and on teachers' self-efficacy?
2. To what extent did the professional development experience have an impact on teachers' knowledge of mathematics, particularly with respect to rational numbers?
3. To what extent did the professional development experience have an impact on teachers' implementation of formative assessment strategies in their classrooms?
4. To what extent did the professional development experience have an impact on their students' knowledge of rational numbers?
5. What strategies of formative assessment were most frequently implemented by teachers participating in the professional development experience?

Description of Study

This study was conducted in the context of a professional development experience conducted with twenty middle school mathematics teachers, seven of whom taught sixth-grade mathematics, six of whom taught seventh-grade mathematics, and seven of whom taught eighth-grade mathematics

Methodology

The unit of analysis in this study varied according to the research question. For the first three research questions, teachers were the unit of analysis. The twenty experimental teachers were from sixteen schools from twelve districts located in urban, rural, or suburban areas. Teachers in the control group were matched with experimental teachers by grade levels taught and types of classes taught. In most cases, control teachers were in the same school as experimental teachers; however, two schools with experimental teachers did not have teachers able to be matched by the aforementioned criteria, so four control teachers from different schools were matched to these four experimental teachers. The last two research questions addressed the amount of variation in student achievement explained by teacher assessment behaviors. For these questions, the mean achievement scores for students in control and experimental classes were compared using analysis of variance techniques.

Teacher Data

Three types of data were collected from the forty experimental and control teachers and their students. Data from teachers included: (a) pre- and posttest scores from an attitudes inventory and an efficacy inventory, (b) pre- and posttest scores from a teacher assessment on rational numbers, and (c) observation data from two lessons of each teacher.

Teacher attitudes about mathematics were measured using the Attitudes Toward Mathematics Inventory (ATMI). This measure has a reliability coefficient alpha of .97 and four subscales: self-confidence, value, enjoyment, and motivation (Tapia & Marsh, 2004). It has been shown to be valid for multiple populations including middle school teachers (Tapia & Marsh, 2002). Teachers' efficacy was assessed using the Teachers' Sense of Efficacy Scale (Tschannen-Moran & Woolfolk-Hoy, 2001). This measure has three subscales (efficacy in student engagement, efficacy in instructional strategies, and efficacy in classroom management) that together assess the extent to which teachers feel that they are effective motivators, instructors, and managers in the classroom. The construct validity of this measure has been confirmed with factor analysis, and the internal reliability is typically high (Tschannen-Moran & Woolfolk Hoy, 2001).

Teachers' knowledge of rational numbers was assessed through the Diagnostic Teacher Assessment of Mathematics and Science (DTAMS), a reliable and valid assessment developed by the University of Louisville Center for Research in Mathematics and Science Teacher Development (Saderholm, Ronau, Brown, & Collins, in press). This assessment measures the breadth and depth of middle school teachers' mathematics content and pedagogical content knowledge. The assessment is composed of twenty items—ten multiple-choice and ten open-response—and satisfies acceptable standards for validity, internal reliability, equivalency reliability, and interscorer reliability with regard to scoring open-response questions (see Bush, 2006, or Saderholm, Ronau, Brown, & Collins, in press, for specific validity and reliability data.)

Teacher instructional strategies were coded during fall and spring classroom observations by two retired teachers. Each observer visited half of the paired experimental teachers and control teachers. A classroom observation tool was developed that focused on how often teachers engaged in the following formative assessment practices:

- provides quiz or task(s) to assess student learning
- uses other strategies to assess learning
- gives student answers, solutions, or explanations
- offers students productive hints, cues, or prompts
- indicates how students can improve performance
- uses student self-assessment strategies
- uses peer-assessment strategies
- identifies errors in facts, skills, or procedures or misconceptions about concepts.

In addition, observers coded questioning interaction patterns (Initiation-Response-Feedback, Funneling, Focusing) (Herbal-Eisenmann & Breyfogle, 2005) and the depth of knowledge (DOK) of questions asked during lessons (1—recall; 2—understanding skills/concepts; 3-strategic thinking) (Webb, 1997).

Student Data

Pre- and post-data from students included scores on an assessment containing released fourth- and eighth-grade rational number items from the National Assessment of Educational Progress (NAEP). The assessment included items at all three DOK levels. DOK I questions are procedural or recall questions such as, "Ground beef costs $2.39 per pound. What is the cost of 0.78 pound of beef?" DOK II questions focus on understanding skills or concepts such as, "Explain why 4/5 is greater than 2/3." DOK III questions focus on strategic thinking or problem solving such as, "Find the diagonal measurement of a television screen given the width and length."

Findings

Data were analyzed in order to seek answers to the five research questions.

Effects on Teachers' Formative Assessment Practices

Multivariate analysis of variance (MANOVA) was used to compare the observed formative assessment practices among experimental and control teachers. As mentioned before, each pair of teachers was observed twice during the school year by the same observer (one of two). For analysis purposes, data from these two observations were averaged to produce a single score for each component of the observation form for each teacher.

Statistical assumptions. Three assumptions were examined before proceeding with the analysis. The assumption of independence of observations was met through two characteristics of the study—all teachers operated independently of other teachers in the study, and data from the two observations of each teacher were averaged to produce a single score for each criterion. Because many of the observations contained null values, Box's Test of Equality of Variance/Covariance Matrices was unable to be performed, so Levene's Test of Equal Variances revealed that four observation criteria did not meet the assumption: use of depth-of-knowledge level 3 questions, use of focusing questioning patterns, use of funneling questioning patterns, and use of peer assessment strategies. However, with equal sample sizes, the threat to validity was considered to be small. Finally the Shapiro-Wilk statistics of normality revealed that the data met the normality assumption necessary to conduct the MANOVA.

Differences in experimental and control teachers. The MANOVA revealed multivariate significance, $F(16, 23) = 2.432$, $p = 0.025$. The multivariate pairwise comparisons are shown in Table 1.

Table 1
Analysis of Variance of Teacher Assessment Practices

Group Differences on Assessment Criteria	F[a]	Effect Size[b]
Use of focusing questioning patterns	1.108	0.333
Gives student answers, solutions, or explanations	1.484	0.386
Identifies errors in facts, skills, or procedures	1.548	0.394
Identifies errors in thinking or reasoning	0.562	0.237
Identifies misconceptions about concepts	0.439	0.210
Indicates how student can improve performance	1.108	0.333
IRF	0.015	0.039
Offers student productive hints, cues, or prompts	1.970	0.444
Provides quiz or tasks to assess student learning	0.153	0.124
Use of DOK1 questions	7.525	0.869**
Use of DOK2 questions	6.223	0.790*
Use of DOK3 questions	3.416	0.585
Use of funneling questioning patterns	21.916	1.482***
Use of other strategies to assess learning	4.46	0.669*
Use of peer assessment strategies	4.733	0.689*
Uses student self-assessment strategies	0.091	0.096

[a] Degrees of Freedom for all criteria (1,38); [b] Computed effect sizes represent the standardized mean difference.
*$p < 0.05$
**$p < 0.01$
***$p < 0.001$

Experimental teachers demonstrated significant differences from control teachers in using five formative assessment strategies: DOK level 1 questions ($F = 7.525$, $p < .01$), DOK level 2 questions ($F = 6.223$, $p < .05$), funneling questioning patterns ($F = 21.916$, $p < .001$), other strategies to assess learning ($F = 4.46$, $p < .05$), and peer assessment strategies ($F = 4.73$, $p < .05$). The observed effect sizes ranged from moderate to large. The use of funneling questioning patterns resulted in the most significant differences.

Student Achievement

Using posttest scores as the dependent variable, analysis of covariance (ANCOVA) was used to compare students' growth of knowledge of rational numbers in the classes of experimental teachers with those in the classes of control teachers. A comparison of pretest and posttest scores on two different versions of the rational number assessments was used to determine growth in student achievement levels.

Statistical assumptions. All three assumptions for use of ANOVA were met. The assumption of independence of observations was met through secure testing procedures. The assumption of homogeneity of variance was met using Levene's Test of Equal Variances ($F(1, 653) = 1.202$, $p = 0.273$). The assumption of normal distribution of errors examined through the Shapiro-Wilk test, skewness coefficient, and normal probability plots was met. The assumption of linearity tested by examining the plot of the residuals versus predicted values was met. The homogeneity of regression assumption was met through analysis of scatterplots of both groups.

Results. The ANCOVA indicated that there were significant differences in the student achievement levels between the experimental and control groups, controlling for pre-existing differences as measured by pretest scores and grade level, $F(2, 648) = 24.681$, $p < 0.001$. The standardized mean difference effect size for this interaction effect was 0.39, $p < 0.05$. A closer examination of the group-by-grade interactions revealed that, although growth occurred in all the grades and groups, the experimental group had the greatest positive effect in the seventh-grade classes (Table 2). In fact, both sixth- and eighth-grade students with experimental group teachers were outperformed by students with control group teachers.

Table 2
Descriptive Statistics by Grade Level

	Teacher Sample Size (*N*)		Std. Mean Achievement Gain (SE)			
Grade	Control	Experimental	Control	Experimental	*df*	*t*
6	6	7	3.172 (0.560)	0.061 (0.512)	184	-5.75**
7	7	6	2.083 (0.422)	4.472 (0.377)	222	-5.97**
8	7	7	3.674 (0.318)	2.734 (0.353)	146	-2.80*

**$p < 0.01$
***$p < 0.001$

In order to determine why the seventh-grade experimental group students might have performed better than the sixth- and eighth-grade experimental students relative to their control counterparts, we re-examined the teacher observation data, this time looking for significant differences between the experimental teachers and control teachers for each specific grade. MANOVA results indicated no multivariate significance of the group by grade interactions, so individual variables were examined for univariate differences (Grade 6: $F(16, 21) = 0.797$, $p > 0.5$; Grade 7: $F(16, 21) = 1.044$, $p = 0.456$; Grade 8: $F(16, 21) = 1.089$, $p =$

0.421). We used ANOVA rather than t-tests to avoid possible inflation of Type I error. Only one observation criterion showed a significant difference between seventh-grade experimental teachers and seventh-grade control teachers—the use of DOK 3 (reasoning) questions. In seventh grade, experimental teachers asked significantly more DOK 3 level questions (Table 3). This difference in the number of DOK 3 questions asked corresponds to a moderately large effect size.

Table 3
Analysis of Variance of Teacher Observations by Grade Level and Experimental Group

Source	*F*	*p*	ES
Grade 7*Participant			
DOK III	4.371	0.044	0.68
Student Self Assessment	3.895	0.056	-
Funneling	3.496	0.070	-
Grade 6*Participant			
DOK III	1.362	0.251	-
Grade 8*Participant			
DOK III	0.394	0.534	-
Identify Errors in Skills/Procedures	4.172	0.048	0.66

Note: Computed effect size represents the standardized mean difference.

Eighth-grade experimental teachers identified errors in skills and procedures significantly more than eighth-grade control teachers, yielding a moderately large effect size. The sixth-grade experimental teachers, on the other hand, showed no significant differences in any of the observation criteria from the sixth-grade control teachers. While this result is only preliminary, the data suggest that the use of DOK level 3 questioning may have an important impact on student achievement.

Teacher Knowledge of Rational Numbers

We began examining teacher knowledge of rational numbers using a multivariate analysis of covariance (MANCOVA), however, the data indicated no multivariate significance, $F(4, 27) = 0.305$, $p > 0.5$, so each posttest outcome was examined using an ANCOVA model (Table 4). The five assumptions for ANCOVA were tested for these data and were met.

Table 4
Analysis of Covariance of Teacher Rational Number Content Knowledge

Group Differences on Assessment Criteria	F^a	*p*
DOK I	0.147	0.705
DOK II	0.014	0.906
DOK III	0.147	0.704
Pedagogical Content Knowledge	0.198	0.660

[a] Degrees of Freedom for all criteria = (1, 26)

The results indicated no significant difference in growth in rational number concepts through participation in the formative assessment professional development experience.

Teacher Attitudes

Teacher attitudes were assessed before and after the professional development experiences with the Attitudes Toward Mathematics Inventory (Tapia, 1996). As with teacher knowledge of rational numbers, we began our evaluation of the data using a MANCOVA, but the data indicated no multivariate significance, $F(4, 31) = 0.548$, $p > 0.5$. An ANCOVA was therefore used to assess change in attitudes between control and experimental teachers.

Assumptions. As with the previous data, all five assumptions of ANCOVA were met. However, examination of the normal probability plots revealed that all subscales followed a normal distribution, but that the statistical tests for the three non-normal subscales were thrown off by a single outlier in each case. Upon further examination, the particular teacher (from the control group) exhibited unusually low scores on each of the three subscales in question. However, the Cook's distance for these unusual observations did not exceed 1 for any of the subscales, so the threat of non-normality to the statistical validity was considered to be small.

Results. The ANCOVA of teacher attitudes revealed no significant differences in attitude change (Table 5). The change in teachers' sense of efficacy was close to significance ($p < 0.10$).

Table 5
Analysis of Covariance of Teacher Attitudes

Group Differences on Assessment Criteria	*F*	*p*
All points included[a]		
Efficacy	2.968	0.096
Enjoyment	0.443	0.511
Self confidence	0.963	0.334
Value	1.209	0.281
Outlier point excluded[b]		
Efficacy	0.481	0.494
Enjoyment	4.483*	0.044
Self confidence	3.455	0.074
Value	0.163	0.689

[a] $df = (1, 29)$
[b] $df = (1, 27)$
* $p < 0.05$

In order to understand how the one outlier point might have affected the results of the ANCOVA, the same analysis was run with the point removed. Without the outlier, the data showed significant differences in attitude change as a result of the workshop in enjoyment and near significance in self confidence levels ($p < 0.10$). Removing the single observation reduced the differences in efficacy, but the teacher was not an outlier on this scale.

Based on these results, the professional development experience had an impact on teacher attitudes, especially self-confidence and enjoyment. It did not appear to affect teachers' efficacy and value of the profession.

Limitations of the Study

The greatest limitation of the study was time. Although we were pleased with the few positive outcomes obtained, we realized from the outset that one year is insufficient time to expect significant changes in teaching practices and particularly student achievement. As we mentioned earlier, incorporating formative assessments effectively into instruction requires teachers to shift beliefs and philosophies about teaching. Furthermore, according to Black, Harrison, Lee, Marshall, & Wiliam (2004), three years of professional development is generally necessary to bring about broad-based change in teaching practices. We would have liked to collect data on these teachers for another year, but funding was not available.

A second limitation was the sample size. Only twenty teachers participated in the professional development experience and forty teachers in study. Although they taught in a variety of schools and school districts, we would have liked to have had a larger and more diverse sample. Although our study met most assumptions in order to conduct the analyses, a larger sample would have increased our confidence in the results.

Finally, because of limited resources, we collected classroom data through only two observations of each teacher. While the two "snapshots" of teaching provided some data about their use of formative assessment practices, more observations would have enhanced the validity of the results. Furthermore, teachers knew when observers were coming, so there is the possibility that the lessons were not representative of their typical teaching approaches.

Discussion

The analyses yielded several noteworthy findings. First, the use of DOK level 3 questions during instruction seemed prevalent with experimental teachers, particularly the seventh-grade teachers. Leaders of the professional development experience worked hard throughout the sessions to help teachers not only understand the differences among tasks and questions with respect to DOK levels, but they also provided teachers many opportunities to create and write DOK level 2 and 3 tasks. Second, experimental teachers used the funneling interaction pattern significantly more often than control teachers. Although funneling has not been deemed as effective as the focusing interaction pattern, it is better than the IRT interaction pattern (Herbal-Eisenmann & Breyfogle, 2005). Focusing is a difficult interaction pattern to master, and we believe that teachers needed much longer than a year to master this technique. Although we discussed and practiced the focusing interaction pattern during sessions, it is much more of a challenge to actually implement with classes of middle-grades students. Third, experimental teachers used student peer assessments more often than control teachers. This strategy seemed intuitively valuable to experimental teachers, and many were willing to experiment with it during instruction. Fourth, participation in the experience seemed to have a mild affect on self confidence and enjoyment in teaching mathematics.

Several nonsignificant findings of this study are worth mentioning. There were no significant differences in experimental and control teachers' growth in knowledge of rational numbers during the experience. This result was not surprising because building teachers' knowledge of rational numbers was a secondary goal. We used many rational number examples during our presentations, but we did not seek to build their knowledge of rational numbers purposely, and obviously we did not. Also, there were no significant differences among growth in the rational number knowledge in students of experimental teachers and control teachers except at the seventh-grade level. Again, we were not surprised by this result. The students spanned all three grade levels; therefore their opportunities to learn rational number concepts and skills varied greatly across grades and classrooms. The assessment that we administered was not directed specifically at the mathematics content that students might have learned. Furthermore, teachers were just beginning to incorporate formative assessments in their classroom during the first year, and it is unlikely that their actions would have an immediate impact on students' knowledge of mathematics.

Implications for Professional Development on Formative Assessment

Related to the first limitation above, we believe that additional professional development beyond this first year would have yielded stronger evidence. Some experimental teachers embraced the concept of formative assessment and made significant changes in their classroom practices almost immediately. Some experimented with particular strategies, like using different questioning interaction patterns, asking higher level questions, and improving the nature of their feedback. Some teachers tinkered with some of the ideas and tried them on a superficial level. And, one or two teachers did not seem influenced by the experience. For the last two groups, it would have been beneficial to have at least another year, with perhaps more intensive coaching, to help them with the process.

In particular, we felt that more time was needed to help teachers with the questioning interaction patterns. The results showed that experimental teachers used funneling questions more often. While this represents a better interaction pattern than the less effective IRF questioning pattern often used by teachers, it is not as conducive to quality formative assessment as the focusing interaction pattern. Focusing is more difficult to implement, especially with middle school students, and another year of work on it might have helped teachers.

Finally, experimental teachers asked more questions at all levels and used student peer assessment more often than the control teachers. Of the array of formative assessment strategies, these two seemed to be easiest to implement early for the experimental teachers. Therefore, professional development providers might consider these as important topics on which to focus early in the experience to give teachers some easier strategies to incorporate.

Future Research

Because of the length of this study, it raised more questions than it answered. As mentioned earlier, a subsequent study should be replicated over a longer period of time, with more teachers and more classroom observations. These changes would yield more reliable and valid results. Also, researchers studying the impact of professional development on formative assessment practices might consider focusing on a few critical variables. This study revealed that a more focused investigation on the impact of professional development on questioning patterns, depth of knowledge of questions, or any of the components of formative assessment might yield more informative results.

References

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139-148.

Black, P., & Wiliam, D. (2005). Developing a theory of formative assessment. In J. Gardner (Ed), *Assessment and learning* (pp. 81-100). London: Sage.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan, 86*(1), 9-21.

Bush, W. (2006). Diagnostic teacher assessments in mathematics and science. Paper presented at the *Colloquium Series at the University of Maryland.* College Park, MD.

Bush, W., & Leinwand, S. (2000). *Mathematics assessment: A practical handbook for grades 6-8*. Reston, VA: National Council of Teachers of Mathematics.

Guskey, T. R. (2003). What makes professional development effective? *Phi Delta Kappan, 84*(10), 748-750.

Heck, D. J., Banilower, E. R., Weiss, I. R., & Rosenberg, S. L. (2008). Studying the effects of professional development: The case of NSF's local systemic change through teacher enhancement initiative. *Journal for Research in Mathematics Education, 39*(2), 113-152.

Herbal-Eisenmann, B. A., & Breyfogle, M. L. (2005). Questioning our patterns of questioning. *Mathematics Teaching in the Middle School, 10*(9), 484-489.

Howe, A. C., & Stubbs, H. S. (1997). Empowering science teachers: A model for professional development. *Journal of Science Teacher Education, 8*(3), 167-182.

Loucks-Horsley, S., Love, N., Stiles, K. E., Mundry, S., & Hewson, P. W. (2003). *Designing professional development for teachers of science and mathematics*. Thousand Oaks, CA: Corwin Press.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author..

National Staff Development Council. (2001). *Standards for staff development, revised: Advancing student learning through staff development*. Oxford, OH: Author.

Saderholm, J., Ronau, R., Brown, E. T., & Collins, G. (in press). Validation of the diagnostic teacher assessment of mathematics and science (DTAMS) instrument. *School Science and Mathematics*.

Sparks, D. (2004). The looming danger of a two-tiered professional development system. *Phi Delta Kappan 86*(4), 304-306.

Stiggins, R. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice, 18*(1), 23-27.

Stiggins, R. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappan, 86*(1), 22-27.

Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan, 85*(4), 324-328.

Tapia, M. (1996). *The attitudes toward mathematics inventory*. ERIC Reproduction Service No. ED404165.

Tapia, M., & Marsh, G. E. (2002, November). *Confirmatory factor analysis of the attitudes toward mathematics inventory*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Chattanooga, TN.

Tapia, M., & Marsh, G. E. (2004). An instrument to measure mathematics attitudes. *Academic Exchange Quarterly, 8*(2), 16-21.

Tschannen-Moran, M., & Woolfolk-Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*(7), 783-805.

Van de Walle, J. (2007). *Elementary and middle school mathematics: Teaching developmentally*. Boston: Allyn and Bacon.

Webb, N. L. (1997). Criteria for alignment of expectations and assessments in mathematics and science education. *Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6*. Madison: University of Wisconsin.

Wiliam, D., Lee, C., Harrison, C., & Black, P. J. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy, and Practice, 11*(1), 49-65.

Wiliam, D., & Thompson, M. (2008). Integrating assessment with instruction: What will it take to make it work? In Dwyer, C. A. (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53-82). Mahwah, NJ: Lawrence Erlbaum Associates.