Effective and Scalable Teacher Professional Development: A Report of the Formative Research and Development

E. Caroline Wylie

Christine J. Lyon

Elia Mavronikolas

December 2008

ETS

Listening. Learning. Leading.®

Effective and Scalable Teacher Professional Development:

A Report of the Formative Research and Development

E. Caroline Wylie, Christine J. Lyon, and Elia Mavronikolas

ETS, Princeton, NJ

December 2008

Abstract

This study is a qualitative analysis of data collected during a yearlong series of teacher learning community meetings and classroom observations. The participants are middle and high school mathematics teachers from 2 school districts. Teachers were introduced to the research behind formative assessment and how to apply that research to their teaching during a 3-day summer workshop. Teachers then met monthly in small school-based groups to deepen their understanding of formative assessment and to talk about their own classroom experiences. The analyses focused on how teachers' understanding changed over time, how the teacher learning communities supported the teachers, how learning translated into classroom practice, and the factors that supported or hindered the development of teachers' understanding and practice. Lessons learned during the study itself and from subsequent analyses of the data had a significant impact on the development of ETS's *Keeping Learning on Track*® program.

Key words: Formative assessment, teacher learning communities, professional development

Acknowledgments

Table of Contents

List of Tables

Introduction

There is a robust research base supporting the connection between teachers' use of formative assessment in everyday teaching and improved student learning (see for example, Black & Wiliam, 1998; Brookhart, 2005; and Nyquist, 2003, for reviews that reflect slightly different perspectives but reach similar conclusions). Formative assessment can be defined by a guiding principle: students and teachers using evidence of learning to adjust teaching and learning to meet immediate learning needs minute to minute and day by day. Another way to think about formative assessment is to define it as a process "in which information about learning is *evoked* and then *used* to modify the teaching and learning activities in which teachers and students are engaged" [emphasis in the original] (Black, Harrison, Lee, Marshall, & Wiliam, 2003).

To enact this principle, five research-based strategies have been articulated for use within the classroom:

- Clarifying and sharing learning intentions and criteria for success
- Engineering effective classroom discussions, questions, and learning tasks
- Providing feedback that moves learners forward
- Activating students as the owners of their own learning
- Activating students as instructional resources for one another

Taken together, these five strategies define the scope of formative assessment (Leahy, Lyon, Thompson, & Wiliam, 2005).[1] The five strategies along with the guiding principle are important to any classroom; however, how teachers choose to implement these ideas may be specific to their own classroom, subject area, teaching style, and/or students. For this reason, a variety of practical classroom techniques for each strategy have been developed and documented by ETS staff, by teachers with whom ETS staff have worked, and by other practitioners.[2]

British research following Black and Wiliam's seminal research synthesis showed that with professional development, teachers could integrate these practices into their everyday teaching practice and these changes resulted in substantial gains in student achievement scores on standardized tests (Black et al., 2003). However, a formal professional development model had not yet been developed.

A review of the current literature on best practices for teacher professional development revealed increasing agreement that effective professional development needs to attend to both *process* and *content* elements (Reeves, McCall, & MacGilchrist, 2001; Wilson & Berne, 1999).

Garet, Porter, Desimone, Birman, and Yoon (2001) suggest that professional development is more effective when it is local, sustained, and involves collective participation. Work by Cobb, McClain, Lamberg, and Dean (2003) supports the idea that when professional development is situated where teachers operate it can be sensitive to local constraints. A sustained effort of professional development is also more effective than 1-day workshops (Cohen & Hill, 1998; Ingvarson, Meiers, & Beavis, 2005). One model that fulfills each of these requirements is based on teacher learning communities (TLCs). TLCs attend to each of these considerations by forming a school-based, ongoing workshops where active, collective participation is a requirement for participation.

In addition to these process elements, research indicates that professional development is more effective when it has a focus on deepening teachers' knowledge of the content they are to teach, the possible responses of students, and strategies that can be utilized to build on these (Supovitz, 2001). An assessment for learning focus within established TLCs can provide a way for teachers to learn to think systematically about student thinking and how to tailor instruction to meet the immediate learning needs of students.

Additionally, while introducing new material to teachers may be of intellectual interest to them, unless it becomes embedded as a habit of mind and part of their regular practice, it is of little long-term value. The challenge of all professional development efforts is to help teachers transfer new knowledge into practice. Research on the nature and development of expertise (Berliner, 1994) provides some insight: Expert teachers develop *automaticity* for the repetitive operations that are needed to accomplish their goals, they are more opportunistic and flexible in their teaching than novice teachers, and they perceive meaningful patterns in the domain in which they are experienced (Newell & Simon, 1972). The book *How People Learn* (Bransford, Brown, & Cocking, 1999) identifies situations that support learning, such as opportunities for people to reflect on experiences in systematic ways so that they can create and build an accessible knowledge base and learn from their own mistakes.

Etienne Wenger's research (1998) focuses on the development of communities of practice that establish themselves, formally or informally, among groups of people who work in similar areas. By interacting with people who are more expert in an area, a group member can see ways in which he or she could also develop that same expertise. While the observer's future is not determined in terms of his or her ultimate level of expertise, there is a paradigmatic

trajectory. While the TLC model does not necessarily have an identified expert in formative assessment, each teacher brings different experiences and expertise to the group and some individuals may develop local expertise in a particular technique, which then serves as a model for other group members.

Overall this review of literature supported a model of teacher professional development that focused on formative assessment while using TLCs as the vehicle for teacher change.

Background

The current study, the Effective and Scalable Teacher Professional Development (ESTPD) project, is one part of an extended research-and-development initiative to explore what is necessary for teacher professional development to be both effective in changing teacher practice and scalable within and across districts. This study built directly on the literature cited in the Introduction section as well as lessons learned from a previous research project, Evidence Centered Teaching in Algebra (ECTA).

To briefly summarize the ECTA project, ETS staff were engaged in a small-scale study from August 2003 to July 2004. During that time, eight 9th-grade algebra teachers attended a 3-day workshop that provided professional development training on a version of formative assessment that was supported through the use of two ETS products: the Teacher Assistance Packages from the *Pathwise*® series and *Discourse*® system (a computer system that allows teachers to see the work of all students simultaneously). Pairs of teachers were from four schools located in four different states. As a result of this geographical spread, teachers only had contact with one other teacher in their school, and no contact with the other teachers in the project after the summer workshop. The only support offered during the school year was a webinar in February 2004 that provided an opportunity for participating teachers to discuss implementation experiences and issues.

The year of work in the ECTA study illuminated significant problems with the design of the intervention that were consistent with our review of the literature. In the summer of 2004, the data were analyzed, and it confirmed the preliminary findings: The approach to professional development was ineffective in getting teachers to routinely incorporate the materials provided or key elements of formative assessment into their practice, and as a result there was no discernable change in the classroom. It was clear that a 3-day workshop and minimal ongoing supports for teachers were inadequate for effecting teacher change. In order for the intervention

to be effective, sustained and embedded support was necessary to maintain teachers' focus and motivation. These insights led to the revision of the intervention and the conceptualization of a new set of research questions.

Current Study

The current study was designed to build on current research in the field and to overcome many of the shortcomings in the ECTA project. The belief, based on a solid research foundation, remained that a focus on formative assessment would be highly beneficial to teachers. However, in contrast to the previous study, the professional development spanned the entire year rather than relying on a one-time workshop. In the ESTPD project, teachers engaged in a professional development program that began with a 3-day summer workshop that focused on formative assessment, sometimes termed assessment *for* learning (AfL). They then continued to meet monthly throughout the school year in school-based TLCs in order to further explore the ideas presented in the summer workshop. Each aspect of the professional development will be explained in more detail in the following sections.

Research Questions

Since considerable effort was spent developing practical mechanisms for supporting teacher change, the research questions for this project were focused on the feasibility and practical utility of the intervention, rather than attempting to prove the effects on student learning before the intervention itself was fully understood. Therefore, the study was designed to be observational rather than experimental. This focus allowed us to trace the trajectory of the professional development intervention with a group of middle and high school mathematics teachers who met over the course of a year in five TLCs. The research follows case-study methodology, a deliberate choice of the researchers, since the main goal of the study was to understand the *how* and the *why* of the professional development approach.

The theory of action for the ESTPD intervention reflected a three-step model common to interventions predicated on teacher professional development. The first step in the model requires that teachers be exposed to professional development that helps them learn a better way to teach. This in turn leads to teachers' improving and/or adopting better teaching practices related to the focus of the professional development, and these improvements in teaching lead to improvements in student learning.

4

In order to understand the impact of the professional development, it is important to consider the model at all four points: to understand the components of the professional development itself, ways in which teachers' understanding or knowledge changes, how teaching practices change, and how student learning is improved because of improvements in teaching. From this theory of action, research questions were developed that primarily attended to ways in which teacher change was supported and realized. The current study sought to address the following four research questions:

1. How does teachers' understanding of formative assessment change over time?
2. How do TLCs contribute to the development of teacher understanding of formative assessment?
3. To what extent does teacher learning translate into teacher practice?
4. What factors supported or hindered teachers' development of formative assessment understanding and practice?

We began the work with a question that focused on the impact on student learning but realized that it was premature to ask such a question, given the exploratory nature of the previous questions and significant data-collection issues. In addition to the formal research questions, the project team also sought to learn in less-formal ways about the structures and materials needed to support teachers' engagement with the ideas of formative assessment. This project was seen as an early part of a longer-term development process. By the end of the year the team needed to be in a position to make a recommendation for further development and refinement of materials, even if the official analysis was not complete.

Overview of the Effective and Scalable Teacher Professional Development (ESTPD) Intervention

There were three components to the ESTPD intervention: a summer workshop, monthly TLC meetings, and in-school support provided by an ETS staff person. Each component is described in the following sections.

Summer Workshop

In the summer of 2004, participating teachers attended a 3-day summer workshop. The Tupelo and Hickory [3] summer workshops were conducted separately but had parallel agendas.

During each workshop, the five key strategies and the "one big idea" were introduced. For each of these, the research background was presented and practical techniques that other teachers had used to implement the strategy were introduced. Additionally, these practical techniques were modeled by the workshop leaders when appropriate.

Participating teachers were also introduced to the Algebra Teacher Assistance Packages in the ETS *Pathwise* series. The Algebra Packages, a series of nine teacher-targeted booklets that are rich in algebra content and rooted in the principles of formative assessment, were created to enhance any algebra curriculum with carefully designed constructed-response tasks. Each Algebra Package begins with an overview of a main student task, including a list of the National Council of Teachers of Mathematics (NCTM) algebra standards (NCTM, 2000) to which the task corresponds, a review of relevant algebra topics, and an invitation to the teacher to complete the task before assigning it to students. The Algebra Packages provide teachers with suggestions for how and when to use the task to most benefit students and the overall algebra curriculum. The Algebra Packages also include samples of student work (both correct and incorrect) that utilize various solution methods, and encourage teachers to use an accompanying rubric to assess the sample work before assessing their own students' work. Finally, each Algebra Package contains strategies for addressing common student errors and misconceptions as well as a list of discussion questions to be used during class time. A similar set of packages was developed for pre-algebra teachers.

One of the three days at the summer workshop focused on formative use of the Algebra Packages. For example, using the sample student work as a way to explicate rubrics and share success criteria with students was discussed. Participating teachers also identified places within their current curricula where the activities could replace a current lesson or sequence of lessons. Teachers were asked to plan for the formative implementation of the Algebra Packages across the school year.

At the end of the summer workshop the teachers were to develop an individual action plan following the process used by Black et al. (2003). To help teachers focus on gradual change, they were advised to choose one class—their focal class—with which to try out the ideas suggested during the summer workshop. Workshop leaders acknowledged that ideas might bleed over into other classes, particularly ideas that were proving to be successful. However, it was not a requirement that teachers implement the AfL techniques with more than one class.

In their action plans, teachers each selected the strategies that they were going to address within their focal class, the techniques they would employ, and, crucially, what they would give up doing in order to make time; it was important that the teachers make time to introduce the new ways of teaching rather than try to fit more activities into an already packed schedule. During this section of the workshop there was a strong emphasis on the notion that there is no single correct way of doing this. While part of the action planning process was to hold teachers accountable for making changes in their practice, they were never told what to change because the workshop leaders could not know enough about the teachers' schools, their curricula, and their personal teaching styles to be certain what would work and what would not.

Teacher Learning Community (TLC) Meetings

During the school year, a series of nine monthly TLC meetings were held in the schools of participating teachers. The meetings ran approximately 2 hours, and the majority of the meetings were facilitated by ETS staff, except for two of the meetings (December and January), which participating teachers facilitated. The participants used these meetings to discuss the implementation of formative assessment in their classrooms. Each meeting started with the teachers explaining what they had tried out since the previous meeting, what had worked, what had not, and why. Central to each meeting were learning activities that were designed to engage teachers in some aspect of formative assessment (e.g., starts and ends of lessons, comment-only grading, an article that focused on student perceptions of teachers). At the end of each meeting the teachers updated their action plans with the techniques they planned to try before the next meeting. In some cases, teachers decided to continue using a particular technique, while in others they decided to attempt something they had not previously tried.

In-School Support

One of the two project managers had previously worked both as a mathematics teacher and as a school principal. Furthermore, she had extensive background knowledge of formative assessment and its implementation. Her role in the original project plan was to run TLC meetings, conduct model lessons or team teach if invited by one of the teachers, and observe the participating teachers once per month and provide them with feedback. Additionally, she was to act as an advocate for the teachers within the districts with respect to obtaining necessary resources and administrative support. In other words, it was planned that her role would

approximate coaching support for the participating teachers. However, her role was modified after the first couple of months as it became clear that the majority of the teachers were not very interested in receiving coaching support.

During the first 3 months, the project manager visited every teacher at least once. The focus of these meetings was determined based on teacher requests. Nine teachers invited her to observe, and on two occasions she also team-taught. On other occasions, classroom visits entailed little more than delivering various supplies necessary for some of the formative assessment techniques. After the first 3 months, however, few teachers requested in-class support or were open to being observed. Furthermore, the project manager's roles as an advocate and a TLC meeting leader became more demanding and time-consuming than originally projected. As a result, her work during the remainder of the year focused primarily on developing the monthly materials and activities and running the series of nine monthly TLC meetings. At the start of the intervention we assumed that the coaching role would be an important part of supporting teacher change during the project, but over time we realized that the leverage for change came from the initial workshop and monthly TLCs. In terms of the broader issues of scaling, it was important to learn that the initial level of support that had been planned was neither entirely feasible nor necessary for the project to continue for the year.

Participating Teachers

One of the lessons learned from the earlier ECTA study was that working with teachers in distant districts was difficult, if not impossible. Therefore, ETS staff solicited participation from two local school districts. Working with local districts ensured that ongoing support was feasible, that observations and student data could be collected with minimal travel, and good cooperation from administrative staff could be ensured. Each district recruited teachers to attend a 3-day summer workshop and the after-school meetings during the school year. Although the principles of AfL are equally applicable to all subjects and all grade levels, both districts chose to focus on mathematics teachers. Each district and the participating teachers are described in the text that follows. The demographics, attendance, and results that are reported focus on a set of core teachers. A cutoff for inclusion in the core group was attendance at a minimum of four TLC meetings.

Hickory School District. The first school district, Hickory School District, elected to focus on their middle school mathematics teachers. This district is a lower-middle class,

suburban district that borders the urban school district that we also worked with in this project, and serves a diverse population, although the majority of students are White.

Six teachers and three district-level administrators attended the summer workshop. In September, mathematics teachers who had not been able to participate in the summer workshop were invited to attend the TLC meetings. As a result, the group grew to nine teachers, representing the three middle schools in the district. Two schools had four participating teachers each, while the third school had only one participating teacher. The majority of teachers taught pre-algebra classes, but three of them also had algebra classes. All nine teachers in this group attended a sufficient number of TLCs to be included in the core group of participants.

Within this group of teachers, eight were female and one was male. There was one Asian American teacher, and the other eight were White. The modal age range of the group was 51–60 (five teachers), two were in the 41–50 bracket, and there was one each in the 31–41 and 21–30 age-groups. All of the participating teachers were certified teachers six held mathematics certifications), and the group had an average of 20 years' teaching experience.

Tupelo School District. The second district, Tupelo School District, is a low-income urban district with a high proportion of African American students. Approximately 21 teachers and the interim high school mathematics supervisor attended the summer professional development opportunity. Most of these participants were high school mathematics teachers who came from two separate high schools within the district. Four teachers from middle schools also attended. Seven more teachers joined the project during the school year, and several dropped out after the summer workshop or in the early part of the year. As a result, there was a core group of 20 teachers who attended the Tupelo TLCs regularly.

The majority of this core group of teachers taught both algebra and pre-algebra. In addition, two were special education teachers, one taught geometry, one taught algebra II, and two teachers taught a pull-out 4-week course for students who had failed the mathematics portion of the state-mandated high school proficiency test.

Eight teachers in the core group were female and 12 were male. Four of the teachers identified as African American, 4 as Asian American (all originally from India), 1 as Hispanic, 10 as White, and 1 as Other. The modal age range of the group was 51–60 (9 teachers); 5 teachers were in the 31–41 age range, and 3 teachers were in the 61–70 age group. The years of teaching experience in this group ranged from 1 to 36 years, with an average of 19 years. Fifteen

of the participating teachers had permanent certification, 1 had a temporary or provisional certificate, and 1 noted Other. There were no data with respect to age, years of teaching experience, or certification for three teachers.

Teacher Learning Community (TLC) Structure

The Tupelo and Hickory summer workshops were conducted separately due to scheduling issues. The Tupelo middle school teachers formed a single, small TLC (Group A). Toward the middle of the school year, when it became clear that only three teachers were consistently attending the meetings, the active teachers in this group were invited to join one of the larger high school groups, but the teachers chose to continue to meet as their own group.

Because the number of Tupelo high school teachers was too large for a single group, it was necessary to form several smaller school-based TLCs. The high school teachers indicated a preferred day of the week to meet and were assigned to one of three TLCs (Groups B, C, and D). Teachers were originally assigned to groups in a way that the attended to the preferences of each teacher while providing each group with a range of experience levels and a mixture of both men and women. For the most part, the high school teachers met in the school building where they taught.

Since there were not quite enough teachers from Hickory to split into two groups, the Hickory teachers continued to meet as a single group, although they came from three separate middle schools (Group E). Meeting locations alternated among the three schools.

The resulting five TLCs each met monthly and followed a similar agenda each month. Attendance for the core members is described in Table 1.

To summarize, 62% of the core members (18 of 29) attended seven or more of the TLC meetings. Within Group A, the smallest group, 2 of the core members attended all meetings (one meeting was cancelled due to inclement weather). Groups B and D had the most variable attendance. In Group B 1 core member attended all nine meetings, 1 attended eight, 1 attended seven, and 2 attended five. In Group D 2 core members attended all nine meetings and the other members attended between four and six meetings. Group C had the most consistent attendance, with full participation at four of the nine meetings and attendance at all nine meetings for 3 participants. Finally Group E, the largest group, had 5 members who attended eight or nine meetings, 1 who attended seven, and 3 who attended four or five meetings.

Table 1

Monthly TLC Attendance for Core Members

	Group A Core $N = 3$	Group B Core $N = 5$	Group C Core $N = 6$	Group D Core $N = 6$	Group E Core $N = 9$
September	3	4	6	4	8
October	3	3	6	6	8
November	3	5	5	5	7
December	2	4	6	4	8
January	2	5	6	4	6
February	0[a]	4	5	5	7
March	3	2	4	2[b]	7
April	2	3	5	3	5
May	2	4	5	3	7

[a] The February meeting for Group A was canceled due to inclement weather. [b] The March meeting for Group B was canceled due to inclement weather; two core members attended other group meetings.

In addition to the core members of each group, each group (apart from Group C) had teachers who attended between one and three meetings during the year. Group A had only one teacher in addition to the core group, Group B had four additional teachers, Group D had two additional teachers, and Group E had one teacher, although this teacher was a long-term substitute teacher who attended while the permanent teacher was out on maternity leave. Even with the additional members, TLC meetings were generally small, with attendance ranging between two and eight members and a modal attendance of five participants.

Data Sources and Analysis

Data Collection

The research questions focused on the impact of the ESTPD project in terms of teacher understanding of formative assessment, their classroom practice, and the appropriateness and feasibility of the TLC support. Two primary sources of teacher data were collected: narratives from observations of TLC meetings and classroom observations of focal classes. TLC meetings

were documented first to provide information to help the developers understand which activities were most successful at promoting teacher understanding, and to provide evidence of how much material could be covered in a 2-hour period. In addition, the TLC observations provided opportunities to better understand how the teachers talked about formative assessment, how their understanding deepened over time, and the impact the professional development had on their classroom practice. The classroom observations were completed in order to see whether the reported changes were evident in practice. Each is described in the following sections.

TLC meeting observations. All TLC meetings were observed by one of four ETS staff persons assigned as notetaker for each TLC meeting (this role rotated among four primary researchers). The notetaker wrote a running narrative, capturing the major points of conversations, questions asked, and interactions among the participants and between the participants and the facilitator. The narrative was typed up following a standard format (in preparation for input into N6, a qualitative software package) the following week and shared with the other team members.

In addition, TLC participants were interviewed either one-on-one or in pairs by a member of the ETS research team during the second-to-last TLC meeting (April) of the year.

Classroom observations. To help focus the classroom observations on the aspects of teaching practice influenced by the formative assessment professional development, the ECTA observation protocol (Thompson et al., 2008) was revised based on the experiences of a previous project. In order to simplify the protocol, a number of fine-grained ratings were collapsed into larger categories. As a result, quantitative scores (given on a 5-point scale) with qualitative justifications were given for seven aspects of instruction that included lesson planning and design, algebra and mathematics content, classroom interactions and environment, questioning and classroom talk, assessment and analysis of student understanding, providing feedback to students, and responsive and reflective teaching. In addition, each classroom observation included a one- to two-page written narrative describing the sequence of the lesson, questions asked, and student-to-student and teacher-to-student interactions. Following the aspect score and observation narrative, each lesson was given an overall assessment of lesson quality. This provided an overall score focused on the learning that occurs during the lesson, student engagement, and the use of the question-answer-feedback (QAF) cycle. Each rating was made on

a 6-point scale that had three broadly described categories with specific descriptors at each point, as shown in Figure 1.

Overall weak -no learning -no intellectual work		Some strengths, some weaknesses - some learning attempted - teacher does most of intellectual work - mostly procedural learning		Overall strong - lots of learning - students do the intellectual work - mostly conceptual learning	
1	**2**	**3**	**4**	**5**	**6**
Extremely weak lesson either because of severe management/ environment issues or because no real math is taking place. Lesson lacks any sense of purpose and/or a clear link to conceptual development.	On the surface there may be aspects of a good lesson, and while students may be involved in hands-on activities, individual or group work, it appears to be activity for activity's sake. Overall, however, little is learned and there is low student engagement.	Limited learning takes place in the lesson, with occasional signs of student engagement and some algebra content being communicated /discussed. Teacher is exploring aspects of the QAF cycle, but implementation is weak or sporadic.	Adequate learning takes place in the lesson, with some use of the QAF cycle. Performance is characterized by unevenness, although strengths outweigh weaknesses.	Good lesson, with most students engaged in active learning of important math content. Teacher makes use of QAF cycle in order to advance the learning of students throughout the lesson.	Great lesson, hitting on all cylinders—all students are actively engaged in deep, conceptual learning of important algebra content. Teacher makes powerful use of QAF cycle in order to advance the learning of students throughout the lesson.

Figure 1. **Overall assessment of lesson quality.**

Two classroom observations, approximately 5 months apart, were conducted. Ideally, observers would have seen the same focal class on both occasions. However, in the high schools, classes followed a block schedule (e.g., students had mathematics class every day for a double period and the course was completed in one semester). As a result, the high school teachers had a different focal class for each semester. Since the focus of the observation was to document changes in teachers' practice rather than learning changes in specific students, it was more

important to observe the teacher at the end of the intervention rather than at the end of the semester; therefore, this change in focal classes was not seen as a concern.

After each lesson, the observer briefly interviewed the teacher to learn about the lesson's goals; how the lesson fit into a longer sequence of instruction; the reasons behind the teacher's choices of instructional strategies, resources, and student grouping; the success of the lesson; and any future instructional steps that were planned. In addition, teachers were asked to reflect on the impact of the ESTPD project overall.

Six classroom observers participated in training session to ensure a common understanding and interpretation of the data-collection and documentation process as well as consistent rating of the seven aspects of instruction and assignment of overall scores. There were several practice opportunities using video case studies. One staff person selected the videos and scored them. The whole group then viewed the same clips and scored them. Significant time was spent discussing the examples and rubrics and coming to an agreement on a set of scores for each example. A previously unseen example was then reviewed and, as homework, all participating staff scored it and, in cases when scores differed, participated in discussions with the lead trainer to resolve the discrepancy. Initial observations, in both the autumn and the spring, were done in pairs. Observers independently wrote up the observation, exchanged and reviewed their work, and discussed and resolved any discrepancies. After the initial calibrations, observations were conducted independently. At the conclusion of data collection, each observer exchanged write-ups with a second trained staff member for a final review. These reviews were completed, questions were answered or clarified, and revisions were made when necessary.

Data Analysis

The written narratives from both the TLC meetings and the qualitative aspects of the teacher observations were imported into N6, a qualitative data-analysis software package. The data were coded to facilitate retrieval of specific parts of the data later in the analysis process (Merriam, 1998). For this data set, codes were developed both *deductively* and *inductively*. Deductive codes are used when the data are analyzed using pre-existing categories. Inductive codes are developed in the course of the data analysis when a new category is identified from the data and then that code is applied to the rest of the data.

The data were initially coded for demographic information and the various techniques that the teachers discussed using in their classroom. A second round of coding was carried out

against a series of research questions. For example, one research question focused on how the teachers' understanding of the assessment for learning techniques changed or deepened over time. Supporting data for this question were identified by specific codes directed toward identifying conversations where the teachers talked about the logistics of using a technique, the impact that a technique had on student learning, or adaptations that they made to a technique.

Two ETS researchers independently coded the meeting narratives for one group (20% of the narratives), compared their codes, and resolved discrepancies in a face-to-face meeting. During the initial resolution meeting, additional inductive coding categories were added to the list. For example, both researchers noticed instances where one or more of the teachers focused on reasons why a technique would not work for them, so a code to track resistance to change was created, along with a second code to identify instances when other members of the TLC challenged negative comments. This additional code further supported another research question that was focused on understanding how the TLC functioned as a group.

The narratives for the remaining four TLCs were then coded using the revised set of codes. A third researcher was trained to use the revised set of codes. The remaining four sets of narratives were then double-coded by one of the two original researchers and the third person. In cases of disagreement, differences were discussed and resolved prior to entry in N6.

Quantitative scores on the seven aspects of lesson quality and the score for overall lesson effectiveness were compared across the two sets of observations, conducted in the fall and in the spring of the intervention.

Results

Before reporting on each of the research questions themselves, it is useful to develop a picture of the five TLCs. While all of the participants were mathematics teachers, there was still considerable variation in the gender, ethnicity, age, and years of experience both within and across groups. The groups themselves also differed according to school level, district, and size. Members of Groups A and E were middle school teachers, while the teachers in Groups B, C, and D taught high school. Groups A through D all were situated in an urban school district, while Group E was an urban fringe/suburban district. Group A was consistently the smallest, and Group E the largest. The location for Group E rotated each month among the three participating middle schools while, for the most part, the Tupelo groups (except for Group B) met in the same room within a particular building.

While there were differences in the composition of each group, other more subtle differences emerged over the year. For example, individual group members demonstrated both commitment and a lack of commitment to the project in verbal and nonverbal ways. These differences contributed to the working environment and support structures that developed within each group and examples of such are outlined in the following discussion.

Group members demonstrated a commitment to the project and to their group in a wide variety of ways, such as being willing to travel to another school to meet with their group, reminding other group members about the meetings, coming to meetings prepared with their materials and binders from previous meetings, sharing materials and resources with other group members, collecting materials and handouts for absent group members, and, in one instance, a long-term substitute teacher attended the meetings in place of a teacher who was on maternity leave. A lack of commitment was demonstrated mainly by negative comments that participants made, such suggesting an earlier starting time (before the official end of the school day) for the meetings "so that we can finish this thing earlier," or by nonverbal behavior, such as arriving late or packing up to leave a meeting before it was officially over. For Groups B, C, and E, the numbers of instances of negative behavior were significantly outnumbered by the examples of positive behavior. Instances of negative behavior in Group A were evenly divided between showing commitment and showing a lack of commitment. For Group D, however, the examples of lack of commitment and unwillingness to participate outnumbered the examples of commitment by a factor of 2. The majority of the examples of negative behavior were associated with two members of the group, but these comments negatively affected the environment of the entire group.

Although the groups consisted of teachers who were meeting to talk about issues of formative assessment and its impact on pedagogy, it is not surprising that the reactions of students to particular techniques or the ways in which students were impacted by techniques were often mentioned. Since comments about students were not limited to one meeting or one group, all meeting narratives were coded for comments made about students. Initially, all teacher comments related to students were coded, and then those were further coded into positive, negative, or neutral/factual comments. What is striking is the proportion of negative comments about students: almost eight negative comments for every positive comment. Each group had from one to three positive student comments across the nine meetings. These comments tended

to be empathetic statements expressing a general sensitivity toward students' social and emotional needs, showing respect for their feelings and vulnerabilities around learning, and social concerns with respect to their personal lives and the realities they faced outside of the classroom. For example, one teacher remarked, "The kids don't really want to be pushed, but when you do and they fly, then they really appreciate that teacher. They say 'Wow! I can do this!'" (Group D, January meeting).

> While another group member discussed techniques that enable students to ask for help:

> Paula said that asking to use a consultant allows a student to save face or sometimes she would tell the student behind to "whisper the answer" to the person in front. Phoebe commented that it was a nicer way of saying "do you want to ask for help?" and that "it's all about honoring the kids." Lucy shared her homework strategy: There is a space on the board for "homework help" and as students come into the room they can write the homework question number on the board if they had a problem. (Group E, November meeting)

The number of these comments is too low to make differential comments by group; however, it is worth noting that Group C had more positive comments than any other group. The negative comments that teachers made ranged from comments about student ability, motivation, or expectations, to occasionally blaming the parents or previous teachers for the problems. The following example provides an illustration:

> No matter what I use there are always certain students who are "waste products." They are just kids who don't belong in G&T [gifted and talented]. They are kids who can't even multiply. I seriously want to know who they bribed to get into G&T. They don't contribute. They don't pay attention. (Group A, March meeting)

There was some variation across the groups in terms of the number of negative comments. Considering the number of comments per group according to the core group size, Group E (Hickory middle schools) was the least negative, with only one negative comment per core group member, while Group A (Tupelo middle school) was the most negative, making more than seven negative comments per member over the course of the year. Of the three Tupelo high school groups, Group C was the least negative (under two comments per person), followed by Group D (just over two comments per person), and then Group B (three comments per person).

The majority of negative comments focused on one of two themes. First, student ability, with teachers making comments like "they are clueless" (Group A, November meeting). Second, teacher comments focused on student motivation with comments like "students don't want to do the work" (Group D, May meeting), "students don't want to read" (Group A, October meeting), "students don't understand responsibility" (Group B, April meeting), and "students don't want to think" (Group C, May meeting).

One thing worth noting is that when teachers talked about student reactions to the various formative assessment techniques, the positive comments outnumbered the negative ones. It may be that having new approaches to try in the classroom, and specifically approaches aimed in part at improving student engagement, provided an opportunity for teachers to see their students in a more positive light. This notion appears to hold true, at least for one of the teachers interviewed at the end of the year. One teacher attributed changes in his attitudes about his students directly to his participation in the project.

> Sure, one of the most important changes because of the workshops was that I started looking at teaching in a different perspective. The thing is that I started looking at the students as a whole body. You know, as a whole body, a group of students, many times they think alike. They want to learn, but often they don't want to look like they're learning, but they do learn. I've tried putting them in groups, and they do accept the concept. Sometimes it gradually falls apart in the sense that some kids may not accept their role or may become indifferent or try to show off. But others can still benefit. The grouping seems to be working. So, that is one important thing I've learned. Another thing I've learned is that every student has his own pride. I try not to put them down as much as I used to. Sometimes they act rowdy and bad, and you have to. But I try not to. It's not always easy, because they are really sensitive about it. They mostly think about how their peers will look at them. I've tried to be more aware of that, and they like it better. That is what I have learned from the workshop. (Group B, April meeting)

The working environment was also affected by the group members' attitudes toward the program itself. One of the coding categories that emerged from reviewing the TLC narratives was given the shorthand label of *resistance*. Teacher comments assigned to this category were typically reasons why they could not implement a particular technique. Reasons were related to a lack of time in general, or to time constraints related to curriculum pacing pressures in particular,

or to perceived problems that students would have with some of the techniques. For example, one teacher discussed the tensions between pedagogy and coverage:

> When I try to work in groups it doesn't help me. . . . I have to prepare them for the exam and the exam contains all the material. Time is limited, I can't use different strategies. (Group B, April meeting)

Another teacher anticipated problems with the *no hands up: Popsicle sticks* technique:

> He shared that often times [when calling on a student] the student will just say they don't know how to do it and then they won't go to the board or answer the question. He did not feel that using the Popsicle sticks would help with this, and he didn't think his kids would answer the questions if he called on them in this way. (Group B, January meeting)

In other instances, teachers did not provide an excuse for not trying something, but rather dismissed the idea outright. One teacher, for example, dismissed the idea of marking student work with a green pen by saying, "I love my red pen. It gets me through the mountain of work that needs to be graded" (Group A, January meeting).

Again, the groups varied in the degree to which this occurred. Group E was the least resistant group, with less than one comment per group member over the course of the year, and Group A was the most resistant, with more than four comments per member. Groups B, C, and D were very similar, each having more than two comments per group, with Group D being the most negative of the three.

In terms of developing a sense of collegiality, Groups C and E stood out among the five. In Group C, members talked about checking in with each other a day or two before a TLC meeting to remind each other about the meeting. In addition, the only examples of team-teaching developed in this group.

> Irwin then told the group about an experiment he and Ben ran with their classes. He talked about how they meet regularly about lessons and do a lot of collaboration so this time they did a group lesson. (Group C, April meeting)

On a couple of occasions, a group member came to a TLC meeting with materials to share with the group. Teachers in Group E also brought materials with the group on a number of occasions; on one occasion a teacher returned to her classroom to bring something to show her

colleagues, who waited after the official meeting had ended. It is worth noting that the two groups with the most variable attendance, Groups B and D, did not seem to develop as much collegiality. In addition, the sporadic attendance of a couple of people not in the core group, particularly for Group B, may have countered efforts to establish collegiality.

In general, Groups C and E developed the most positive working environments, with Group B the next most positive. The teachers in these groups behaved positively, made the least number of negative comments toward students, and developed the most collegiality. This positive working environment may have provided some of the necessary supports that enabled the teachers to make changes to their practice that were sometimes incongruent with their current teaching practice, or even incongruent with the culture of the school. The remaining subsections within the Results section describe the particular data that were examined in order to engage with each of the four research questions.

How Does Teachers' Understanding of Formative Assessment Change Over Time?

The first research question to be addressed in the analysis focused on the impact that the professional development had on the teachers' understanding of formative assessment. Teachers' understanding was examined by reviewing the kinds of comments that teachers made about what they were doing in the classroom with respect to formative assessment and the questions that they asked during the TLC meetings.

The written narratives from the series of five meetings per month over 9 months were reviewed as previously described, and teacher comments that demonstrated clear understanding or misunderstanding of formative assessment were identified. The phrase *clear understanding* refers to the comments that reflected an understanding of the ideas and concepts that had been presented in the initial workshop or in subsequent meetings, or of the application of AfL techniques. By *misunderstandings* we mean comments or actions that reflected limitations in the teachers' understanding of the intent of the project or ideas of formative assessment.

In the excerpts that were identified as representing clear understanding, teachers spoke of catching student misconceptions, eliciting deeper thinking in their students, getting hard-to-attain feedback from students of all levels, using feedback to inform their future instruction, using particular techniques that resulted in broader changes in teaching practices and types of assignments given, abandoning unproductive methods of instruction and an emphasis on summative grading, understanding the nature of learning in a deeper and more complex way,

looking for evidence of learning before moving on with a lesson, shifting away from focusing on "covering the curriculum" to ensuring time for learning and observing for "evidence of learning," understanding the need to have clearly defined objectives and goals for a lesson in order to determine what ought to be assessed, and shifting from a teacher-centered environment to a more student-centered learning environment. In many of these comments, teachers indicated that they were doing something different from before the start of the project, wished that they had started sooner, or indicated that they were thinking differently about their practice. For example, one teacher discussed how the use of comment-only marking was forcing him to examine multiple aspects of his teaching:

> Jacob spoke about the type of questions that you need to give when working in groups, and how that relates to the comment-only grading. Because he is doing comment-only grading, he has a pressure now to have a different type of assignment. He feels that it is pushing him towards a different type of teaching, and it is changing not only the grading—it is changing his teaching to a less-structured environment. He thinks that this may be the same type of shift that he needs for group work because he is having trouble figuring out how to do roles with the types of tasks that he gives students to do in their groups currently. (Group C, April meeting)

Across the excerpts that were identified as representing teacher misconceptions or misunderstanding of some aspect of formative assessment, teachers' comments suggested that they viewed the project as a collection of techniques not necessarily tied to assessment; focused on the implementation of techniques as each related to classroom management issues rather than learning; indicated confusion about specific aspects, such as the relationship between grading and self-assessment or formative assessment; or assumed that teacher-created tasks were the same as formative assessment techniques. On other occasions, a teacher would discuss what he or she perceived to be a successful use of a particular technique but further discussion revealed the teacher's significant misunderstanding of the technique. One teacher discussed his use of whiteboards to engage students in a lesson: "The whiteboards work well. I'll ask a question like, 'Who won the football championship?' just to get them going" (Group D, March meeting).

Although there is nothing wrong with this use of whiteboards, it does miss the formative nature of the technique. In this instance, the technique is used as a way to engage the class at the

start of a lesson rather than as a means to collect information that can be used to adjust future instruction.

Groups C and E had an even balance of examples of clear understandings and misconceptions, while Group A had slightly more examples of misconceptions than clear understanding. Groups B and D, however, had more examples of misconceptions than clear understanding. The fact that there were more misunderstandings of the AfL techniques and the larger ideas of the project may be related to the degree to which teachers in these two groups tried out techniques in their own classrooms.

The TLC narratives were reviewed to identify how many techniques teachers reported using during the How's It Going? section of each meeting. It was during this part of the meeting that teachers talked about what they had been trying out in their classes. At this stage in the development history, the meeting facilitators did not insist that everyone share (even if only to say that they had not been able to try something that month), so it is possible that techniques were tried out by teachers and not recorded during the meeting. Even so, these data at least suggest something about how teachers were trying out various ideas. From the narratives, a total number of techniques tried was counted for each group and then scaled according to the size of the group. Participants in Group A, the smallest group, reported on an average of 17 techniques per person in the group. Participants in Groups C and E both reported on an average of 10 techniques per person, while participants in Groups B and D reported on an average of 5 techniques per person. Considering that there were nine meetings for each group over the course of the year, on average participants in Groups A, C, and E were reporting on at least 1 technique per meeting. That the participants in Groups B and D were reporting back on many fewer techniques may in part explain why these two groups had higher rates of misconceptions or weak understandings of the project ideas. One way in which participants learn is through hearing the examples of others' implementation of ideas, but if there are fewer reports of actual practice, there are also fewer opportunities for incorrect ideas to be challenged. In subsequent training for TLC leaders and in their support materials we have also paid more attention to the notion of group members supportively challenging ideas and working together to keep the focus on formative assessment. In this initial implementation, ETS staff were not as proactive as they might have been in challenging ideas.

Across the year of implementation the number of misconceptions or misunderstandings outnumbered the examples of good understanding by about two to one. However, no strong patterns emerged across the year, although there were both fewer misconceptions and fewer good understandings in the middle part of year during the January, February, and March meetings, suggesting either teachers' interest had dwindled or facilitators were not probing as much as they could have, resulting in more superficial conversations that were not coded into either category.

Looking for changes over time proves to be more difficult than it might first appear. Previously, we used a scale for analyzing the quality of implementation of AfL techniques (Wylie, Lyon, Ellsworth, & Martinez, 2007). An implementation score was awarded based on the level of sophistication with which the technique was used for any technique that was presented with sufficient detail. These scores ranged from 0 to 3. A Level 1 score was awarded when a technique was implemented weakly in such a way thazerformance was one that tended to lack consistency in execution, or in which students were given insufficient structure or guidance when generating or acting on feedback, or for which the teacher collected and analyzed information about student learning but failed to apply it appropriately to determine next instructional steps. A Level 3 performance represented the highest-quality implementation and was one in which the teacher and students made maximal use of the information collected about student learning to move instruction forward. In the excerpt that follows, a teacher's discussion of Stop/Slow Signals (students use red and green disks to indicate that they understand [green] or that they need the teacher to slow down or provide help [red]) provides one example of a report that received a Level 3 on the implementation scale.

> Sharon asked if she could report first. She had used the stop lights —red and green cards. Everyone in the class had a card. If they weren't sure, they flipped it over and then they didn't have to be embarrassed because they didn't understand something. She had asked the kids if they thought it worked, and they said it did because they didn't have to put their hands up for the whole period. Sharon said that she was at the front of the room and could see all of the cards. After she saw where the red cards were, she was looking for the most lost faces and she went to those first. She noticed, "At one point I had a whole row of reds, so I could go there and work with that whole group of students." [The ETS staff person] asked, "If you hadn't used the lights would they have told you that they didn't get it?" Sharon indicated that they would have given up. (Group A, November meeting)

Instances in which an identified AfL technique was changed in such a way that rendered it almost impossible to make formative use of the technique were coded as a non-formative use. Often these techniques were modified in a way that affected the purpose or formative nature of their use. For example, the purpose of whiteboards was to collect information from every student in the class simultaneously. However, a second advantage of the technique was that students did not use multiple pieces of scrap paper while answering consecutive problems. When the technique was implemented with only the second purpose in mind and students did not simultaneously respond to a question by holding up a response, the technique was no longer formative in nature. The techniques that the teachers discussed in the meetings were coded using this implementation scale. Table 2 illustrates how the implemented techniques were scored.

Table 2

Frequency of Implementation Scores by Month

Quality of implementation	Sept.	Oct.	Nov.	Dec.	Jan.	Feb. [a]	March	April [a]	May [b]
Level 3	10	30	24	25	10	0	12	0	10
Level 2	0	27	32	25	11	0	4	0	10
Level 1	9	25	18	21	11	5	14	0	0
Non-formative	2	32	18	16	16	0	0	0	0
Total	21	114	92	87	48	5	30	0	20
Proportion of Levels 2 and 3	48%	50%	61%	57%	44%	0%	53%	-	100%
Proportion non-formative or Level 1	52%	50%	39%	43%	56%	100%	47%	-	0%

[a] During the February and April meetings there was no How's It Going? activity, which resulted in the very low counts. [b] During the May meeting the How's It Going? activity was in the form of a reflection across the entire year, and participants were asked to comment on one positive thing from the year—on many occasions they talked about something broader than the use of an individual technique.

In Table 2, the counts for Levels 2 and 3 were combined in order to examine the proportion of higher quality implementations. Similarly Level 1 and non-formative were also combined to examine the proportion of lower-quality implementations. In the months where there was a standard How's It Going? activity, the proportions of higher- and lower-quality implementations hovered around 50% in each category. However, earlier in the year there was a steady increase in the proportion of higher-quality implementations, while the proportion of lower-quality implementations decreased. Although there was an interruption in this pattern in January, March had among the highest proportion of high-quality implementations, with 53% of the techniques scored scoring a Level 2 or Level 3 on the implementation scale.

Although the number of reported techniques decreased in March, the proportion of Level 3 techniques increased, meaning that teachers were discussing fewer techniques but at a higher level of implementation. These results are encouraging for the first year of implementation, as they indicate that the teachers were able to successfully discuss the implementation of these formative assessment techniques.

As a result of the structure of the meetings in the second half of the year, it is difficult to determine how the teachers changed in their discussions about formative assessment over the course of the year. The numbers from the January and March meetings suggest that fewer techniques were discussed, regardless of implementation score, compared to the October through December meetings. October had the highest proportion of non-formative examples, but by the next month the proportion had dropped, although it spiked again in January. While the data are sparse for the second half of the year, there is certainly evidence of successful implementations of individual techniques.

This first research question focused on the changes in teachers' understanding of formative assessment over time. Although we did not formally assess teachers' knowledge of formative assessment at the start and end of the year, their comments during TLC meetings provided a window into this understanding. The comments made and questions asked, the teachers' reports of implementation, and the range of implementation scores indicate that the teachers thought about their practice and behaved differently as result of the intervention. This is especially evident in the implementation scores, as the quality of the examples indicated an improvement in quality over the course of the year. So although the analysis of the comments and questions asked by the participants showed that misconceptions accounted for more of the

comments than those that demonstrated a clear understanding of formative assessment, these implementation scores present a more positive picture. Additionally, some teachers themselves talked about how they thought about their practice differently as a result of the intervention.

How Do Teacher Learning Communities Contribute to the Development of Teacher Understanding of Formative Assessment?

The literature on professional development suggests that TLCs contribute to teacher learning through three mechanisms (Reeves et al., 2001; Wenger, 1998; Wilson & Berne, 1999)

1. Provision of additional content knowledge
2. A structure of supportive accountability
3. The development of a community of learners

The following sections provide descriptions of each mechanism and an outline of its implementation in this project, as well as any relevant results.

Additional AfL content knowledge. A structured agenda with a series of activities that focused on one or more aspects of formative assessment was developed for each month. These meeting protocols were designed to build on the AfL content presented at the introductory summer workshop, expand on one or more of the five key strategies, and/or examine a specific technique in more depth. (Comments on the AfL content of the TLCs and what was learned on-the-fly over the course of the year about this content is presented in a later section of this report, Formative Impact of the Work.)

A wide range of topics were discussed; some were broad topics (such as strategies for improving teaching and learning and how to find out what students know), while other topics focused on a single AfL strategy or technique (such as comment-only marking, whiteboards, and group work). Finally, general topics that were of concern to the specific districts, such as how to effectively use calculators and using pre-test results, were also discussed. The goal of the meeting protocols was to present new AfL content, to highlight techniques that may have been overlooked during the introductory workshop, to entice teachers to make additional changes to their practice, and to provide additional time and support for teachers to examine and think deeply about how information gathered could be used formatively and the impact that may have on the classroom.

To evaluate the effectiveness of the new AfL content, the How's It Going? section of each the TLC narrative was reviewed. Instances where a strategy or technique that was presented

during a previous meeting was reported on during the subsequent meeting were identified. Since techniques were never mandated, but rather the decision to implement was left up the individual teacher, we cannot assume that all teachers would implement a new technique or that all new techniques would be implemented. However, across the meetings there were several instances where new techniques were adopted by participating teachers. For example, in Meeting 4, teachers engaged in an activity designed to promote student self-assessment. The teachers were provided with eight examples of techniques designed to promote this type of activity and were asked to choose one and consider how it could work in their classrooms. Two teachers discussed an idea called Debrief Regularly and decided to revise the technique for their classroom. The following excerpt summarizes their discussion:

> Jacob and Ben each chose the same card out of their pile [Debrief Regularly] and even began designing a new learning log for use with their classes. . . . [The ETS staff person] asked the group if they would be willing to share with the others what they had chosen to use. . . . Ben has picked the strategy called Debrief Regularly, but he wants to use it as an exit ticket. He wants to give the kids three things and make the page student-friendly. [The ETS staff person] offered to take their page and run off the new learning logs/exit tickets for them. (Group C, December meeting)

In following meetings, Ben often discussed how he implemented this new technique with his classes and even provided copies of the handout for other participating teachers. This example illustrates how providing new AfL content throughout the year kept teachers involved with the professional development and encouraged them to continue making changes to their practice.

A second example was observed after Meeting 8. During this meeting, the TLC groups engaged in an activity designed to help them think about group work, more specifically, student roles that can enhance the productivity of students during group work. In a subsequent classroom observation, the observer noted:

> He [the teacher] helps students arrange desks so they form groups with students facing each other. He reminds them what the cooperative group roles are. "Nobody does all the work. The recorder writes and reporter tells the rest of the class, but you all have to think about it." . . . The teacher asks the time managers how much more time they will need. One boy says 5 minutes. (Classroom Observation, May 2005, Group C member)

This example is a clear illustration of AfL content presented during a workshop being translated into actual classroom practice.

   Structure of supportive accountability. In addition to providing a venue for new information and formative assessment content, the TLCs also established a supportive environment that held teachers accountable for making changes to their practice while providing the necessary support. Two specific activities were developed to accomplish this and were repeated at each TLC meeting. First, each meeting began with an introductory activity called How's It Going? This activity provided an opportunity for participants to share the formative assessment techniques they had implemented since the last meeting. Participants were encouraged to discuss the techniques they had tried, the benefits of those techniques, and the challenges. These reports produced several positive outcomes. First, reports of successful implementation encouraged other participants to implement the technique, especially those who had been reluctant to change their practice. For example, one teacher was initially hesitant to provide students with a rubric and sample student work prior to an assignment. After several meetings with colleagues where the successful implementation of rubrics and the impact that their use had on student learning was discussed, this teacher tried the technique and was surprised by its positive impact on his class. The previous reports had provided proof that the technique could work in his school, with students similar to the ones he taught. Had this proof not been provided, the teacher may never have experimented with the technique. In addition to encouraging teachers to make changes, this activity also allowed for the transfer of knowledge (Nonaka & Takeuchi, 1995) among the group.

   As teachers begin to implement techniques, they often adapt ideas to fit their classroom or find new techniques from other sources that align with the five key strategies and one big idea of assessment for learning. The How's It Going? activity provides a forum where these teachers can discuss these adaptations and new techniques. Sharing within the group not only benefits the teacher who is sharing by allowing that teacher to think about the technique, discuss its advantages, and identify any problems, the sharing process also benefits other members of the group because their repertoire of possible techniques increases each time a group member discusses a new technique or adaptation. One example of this is the technique Homework Help-Board. Homework Help-Board is a technique where students identify homework questions they

struggled with, put them on the board, and solve them for one another (ETS, 2007). During the second meeting, a teacher first shared this technique with the group.

> [One teacher] said that she has an area on her board for homework help that she uses with her honors class. When students first come into the class, if there was a problem that they had trouble with from their homework, they put that problem number up on the board and then any student who thinks that they have the correct answer can then put the problem and the solution on the side board. She said that she will probably start using this in the spring with her other classes. (Group E, October meeting)

Throughout the year, several teachers referenced this technique, and it became common practice among the teachers in this district. For example, another teacher reported on a successful implementation of this technique during How's It Going? in the fourth meeting:

> [A teacher] reported that she has been trying the Homework Help that was brought up at the last meeting and that it has been wonderful. She especially likes to use it with the algebra kids. At first she would forget why the kids were jumping out of their seats but in algebra, as soon as a problem is put up on the board, several kids will jump up and solve it and she rarely has to do the problems for them. In pre-algebra they are not as confident but will do some of the problems. She tends to get more problems than solutions in that class. But overall she felt that even if the students just put them up it is helpful. Before I would just give them the key and I would get silence. Now I at least know which problems they need help with and they don't have to ask for help in front of the entire class. (Group E, December meeting)

Another teacher reported implementing the technique even later in the year, during the ninth meeting: "We've been doing homework help on the side board. It's easy for me to see if the whole class is catching on" (Group E, May meeting).

These examples illustrate how knowledge shared during the How's It Going? activity can not only hold teachers accountable for making changes, but can also provide additional resources and support for ongoing changes and engagement with the professional development ideas.

Challenges or difficulties were also discussed during How's It Going? sessions. These discussions provided participants with an opportunity to problem solve with colleagues and potentially overcome any difficulties. The following example illustrates an exchange between

one teacher in Group E and the rest of her group. In this example she discusses her difficulty with finding time to integrate exit tickets into her daily instruction.

> Paula reports that she has been wanting to use the exit tickets but is afraid that she won't have time to do it correctly. She asked the rest of the group how they integrate it into the classroom. Lucy shares that she uses it as a real quick thing. "I mention it at the start of class the next day and tell them how they did, or sometimes I will hand them back. It is not easy to think of a good question, but when you do it doesn't take much class time to do. I usually put the question up on the overhead and the students have to respond during the end-of-class cleanup time. The time you don't always use anyway, the last 4 to 5 minutes of class. (Group E, December meeting)

In a second example, a teacher in Group E shared a problem she was having with the implementation of the Popsicle sticks technique:

> Katrina shared a problem with the Popsicle sticks that she was having with just one class. One or two students had been complaining bitterly about not getting picked. It was driving her crazy, since they were beginning to call out answers. [The ETS staff person] asked her if in the past these students had been getting more attention, and then asked the group if they had advice for Katrina. Katrina said she had tried letting the students pick the next name, and Lucy suggested letting them put their own sticks in the cup (so they could be sure their name was in the cup). [The ETS staff person] returned to Paula's idea of letting students "hire a consultant," and wondered if these problem students would like to be the consultants. Throughout the rest of the workshop various comments were made back to this issue, a good example of teachers working collaboratively to solve a problem in teaching. (Group E, November meeting)

The data suggest that the teachers did utilize this time to discuss the techniques they were trying out and the difficulties they were having. As presented in the previous research question, the total number of techniques reported on during the How's It Going? session was counted for each group and scaled according to the size of the group. Participants in Group A, the smallest group, reported on an average of 17 techniques per person, participants in Groups C and E reported on an average of 10 techniques per person, and participants in Groups B and D reported an average of 5 techniques per person. Since it was suggested that participants choose only one

or two new ideas to try each month, an average of 1 technique per meeting, per participant in Groups A, C, and E suggests that participants continued to use this time to share the techniques that they had been working on, the challenges they had been having, and new issues as they arose.

The second activity designed to increase accountability was the development of monthly action plans. This activity required teachers to take 5 minutes at the end of each meeting to write down what they thought they would implement by the next meeting. Although teachers were initially resistant to this process, they later acknowledged that the informal low-stakes documentation helped to keep them accountable, as is illustrated in this quote:

> I think specifically what was helpful was the ridiculous action plans. I thought that was the dumbest thing, *but* I'm sitting with my friends and on the action plan I write down what I am going to do next month. . . . It was because I wrote it down and I had it in my little packet and that idea of making improvements and sort of informally, which is much more powerful than formally committing to doing it. I was surprised at how strong an incentive that was to actually do something different . . . just the idea of sitting in a group, working out something, and making a commitment, even something as informal . . . I was impressed about how that actually made me do stuff. (Classroom observation, May 2005, Group C member)

As this teacher noted, writing down a plan and knowing he would be asked to report on it motivated him to follow through.

A learning community. Although opportunities for teacher learning were explicitly designed into the TLCs through the additional AfL content and the structures designed to promote supportive accountability, learning was also expected to take place by virtue of there being a community of learners focused on the same topic. The literature suggests that teachers involved in a community of practice that examines teaching and learning should learn from each other as they share problems that they experience in their classrooms and challenge one another to further their own learning (Wenger, 1998).

In order to examine problem solving, the TLC narratives were reviewed. Evidence of problem solving was typically seen in three forms: direct requests for help, general statements of a problem, and challenges to a peer's perception or attitude. We will examine these three forms of problem solving as they were manifested in each group.

Direct requests for help were typically posed as a question (e.g., "How do you keep students from cheating while using the whiteboards?" [Group B, January meeting]). The majority of these requests focused on formative assessment issues, with only three about discipline or student behavior issues. The requests that did focus on formative assessment issues centered on concerns or issues that the participants had about particular techniques. No single technique dominated, and questions ranged from asking for an explanation about how to implement a technique, such as "How do you use exit tickets?" (Group A, January meeting), to more specific questions, such as "When students use the traffic lights to signal their understanding, what you do if you have a whole class of green and one red?" (Group C, March meeting). While a few of the questions focused on the procedural aspects of using a particular technique, most focused on substantive issues of teaching. Approximately two-thirds of the time, the ETS facilitator answered the question, and the remainder of the time a group member answered. This was consistent across all groups. Given how the groups were established (that is, with an outside facilitator), it is not surprising that the groups tended to look primarily to this person for answers. (It is worth noting that in subsequent iterations of this work, TLCs have been led primarily by a teacher from the school, and that, particularly in their second year, TLC members have been encouraged to learn collectively from each other rather relying on the TLC leader for all the answers.)

In general, problem statements were not so much direct requests for help as they were comments that helped to identify shortcomings or challenges within teachers' practices. This is illustrated in the following quote:

> [A teacher] pointed out that the discussion questions were good because he was just thinking that he has trouble giving comments sometimes, and he thinks that it is because they are not good questions. He felt that the discussion questions will give him some good questions to use with his students. (Group C, September meeting)

Although the teacher in this quote does not specifically ask for help, he does identify a problem within his practice—challenges in developing and using good questions. Across the full set of meetings, the majority of these statements focused on formative assessment issues, and only 7 out of 27 instances were focused on discipline or student behavior issues that were not connected to formative assessment. Comments that were focused on formative assessment ranged from issues tightly connected to a particular technique (e.g., "[O]ne teacher said he was

also trying [wait time] but that it was very difficult for him. . . . [H] said that he ended up slipping right back to questions that he knew they [the students] could answer and that the strategy was harder than he thought" [Group C, October meeting]), to broader topics of teaching and learning that were often connected to a technique or strategy (e.g., the relationship between comment-only grading and the depth of thinking required by the types of tasks that the teacher normally selected [Group C, April meeting], to student behavioral issues related to the use of the techniques (e.g., "one teacher shared that students don't want to work in pairs, since they only wanted to have their questions answered by the teacher" [Group B, December meeting], and another teacher expressed concern about the noise level when students worked in small groups [Group A, April meeting]). These differences provided evidence that for some of these teachers, changing practice was leading to fundamental shifts in classroom norms, while for other teachers, even when discussing formative assessment techniques, discipline issues were still a significant problem.

Not surprisingly, since these comments were not framed as direct questions or requests for help, about a third of the time there was no follow-up to the comment. About one-third of the time the ETS facilitator responded with a comment, and the remainder of the time a group member answered. Again, as noted before with respect to direct requests for help, because of the nature of the intervention the ETS facilitator was regarded as the primary source for answerers to questions. However, the fact that one-third of the time a group member responded even when an ETS facilitator was present provides evidence that the teachers within the group could provide support and advice themselves.

The evidence of problem solving was evenly divided between direct questions and general statements of a problem. However, the number of questions posed and problem statements made differed by group, with Groups B, E, and C evidencing 20, 13, and 11 examples, respectively. Members of Groups A and D both asked questions directly or commented on problems on five or fewer occasions during the year. It may be that the positive working environment created in Groups C and E, which created an atmosphere where it was OK to ask questions or admit a problem, contributed to this difference. There were not enough instances to examine by month when questions were asked or problems identified at the group level, but across the five groups the greatest number of questions being asked and/or statements of problems occurred during the November, December, and January meetings.

In addition to examining the TLC meeting narratives to identify the types of questions the teachers asked and problems they shared, we also examined the data for instances of the participants challenging their peers' perceptions and attitudes. We viewed the TLC dynamic as illustrative of the thinking and learning that was going on within the group. The willingness of the group to challenge ideas about formative assessment and attitudes toward students showed that the teachers were not passive listeners but rather active participants in the learning process. The challenges occurred most frequently during Group C's TLC meetings, with five occasions of challenges, followed by four occasions in Group B, and only once in Group D and once in Group E. No instances of challenging among peers were recorded in Group A, the smallest TLC group. None of the challenging instances occurred before the groups' November meetings, and the majority took place during or after the February meeting (their sixth meeting), suggesting that the participants required a certain comfort level with their peers before they were willing to engage in these kinds of discussions could. The ETS facilitator for these meetings often asked questions to help the teachers think about what they were doing, modeling the kinds of challenging questions that we wanted the participants to ask. The following two examples illustrate how the facilitator used questions to deepen the teachers' understanding:

> [A teacher] used the white side of the whiteboards for pre-algebra with her SRA [Special Review Assessment, for students who had previously failed the High School Proficiency Assessment] students. It went over well because it was novel. . . . [The ETS facilitator] asked the teacher, "Why are we using whiteboards? What is the point?" (Group B, October meeting)

> [A teacher] used exit cards with the problem $1/x + 1/y = ?$, and only a couple of students were able to answer it, even though every student in the class at the start of the period had been able to answer the question $1/3 + 1/5 = ?$. When [the ETS facilitator] asked [the teacher] what he did in the next lesson, [the teacher] responded with "used another starter." [The ETS facilitator] probed to see if [the teacher] had done anything with the results and found out that he had not. (Group C, October meeting)

Within Group C, three of the five instances of a colleague challenging a peer's attitude or opinion came from the same person. None of the challenges were negatively received, and the challenge often served to move the conversation along. One example occurred in the context of a

discussion on the value of comment-only grading. One participant was advocating for summative grades as the way to "tell whether the kid is improving or not." The TLC leader tried to help the group member think of other ways to find out this information without grades per se, and one of the other group members said, "Sometimes you're kicking the students when they're down with the grades." The teacher who was advocating for the use of grades replied, "They don't care about comments; they want grades!" (Group C, February meeting). The conversation continued for several more exchanges. Although initially the dialogue had been primarily between the one participant and the ETS facilitator, after the challenge was made the conversation broadened to include several more group members. Another example of how a challenge from a peer served to deepen the conversation occurred when the TLC meeting topic was focused on group work. One member commented that group work "teaches kids to socialize rather than do math" to which a peer responded, "I learned more about math during the mock group work exercise because I felt comfortable and learned new information from others in the group" (Group C, April meeting).

In Group B there were four instances of a member challenging another participant's attitudes or beliefs. For this group, the challenger was always the same teacher, and the four documented instances occurred during group's the sixth meeting. This particular teacher offered suggestions when another teacher reported that he was struggling with discipline issues, was supportive of students when other group members conveyed negative attitudes about students, and urged a colleague to "just try to it for a day" when the colleague was listing reasons why comment-only grading would not work for his students.

Groups D and E each had only one instance of a challenge. Given some of the earlier descriptions of the culture that developed in each group, the lack of peer-to-peer challenges in Group D is in keeping with the general working environment of this group. Members of Group D displayed less commitment to the project than the other groups showed and were more likely to make excuses for not trying some of the ideas. Given the environment, is it not surprising that members of the group were unlikely to challenge each other. However, the lack of challenges among the Group E members is more surprising. This group included teachers from the middle schools in the suburban/urban fringe district. In some ways, this group seemed to function better than the other groups. There was more support for the TLC from administrative levels, teachers were less negative about their students, and teachers were less resistant to trying out new ideas. However, as illustrated by the lack of participant challenging ideas, it was quite a passive group.

The following excerpt, taken from the meeting narrative, illustrates the only time a participant challenged the ideas of another teacher. In this example a teacher is explaining an approach for holding students accountable for completing homework.

> Frank shares, "For the whole year, my class has been set up in groups. . . . Each group starts out the week with 5 points. If someone doesn't have their homework, that's a point taken away. If someone is acting up in class, that's a point taken away. When they get down to 3 points, the whole group has to copy out a page from the student handbook, like copy out page 18. They really hate page 21 because it's got all kinds of tiny print."

> The rest of the teachers are pretty interested in this—some seem to wish they could do it at their school. One teacher seems pretty bothered by it. Apparently, this punishment is standard at one school, but teachers at other schools have been told not to use punishments that involve copying things over and over. There was a discussion of whether the memo that banned this sort of punishment applied only to the repetitive copying of the same phrase or to copying of all text. [The ETS staff person] has to do a little facilitation to get folks to refocus on learning instead of behavior management.

> Frank proclaims: "There's a positive aspect to this too!" And Paula replies, "Thank god!" Frank continues, "At the end of the week, the group that has the most points wins Jolly Ranchers." (Group E, April meeting)

Notably, even though Group E had established one of the more positive working environments, the level of challenge in this example and in the group as a whole was quite low. It is not possible to know why was the case. It was the largest of the three groups, which may have impacted collegiality. Compounding this issue was the fact that the teachers were spread across three schools. Furthermore, many of the participating teachers in this group viewed themselves as already accomplished teachers. Their perception was that the purpose of the project was to share with other less-experienced or weaker teachers rather than to improve their own pedagogy. This is illustrated by the following observation narrative, which was made after the facilitator's second observation:

> The workshops were helpful but they were way too long. Sharing ideas with other teachers is always helpful, but I think it is better for some of the newer teachers. I got a few good ideas, like whiteboards, ABCD cards, but it is all about time. I don't have time

to do a lot of this stuff when I have to get classes ready for the GEPA [the New Jersey Grade Eight Proficiency Assessment]. I have five different classes and I have to prepare five different lessons, so there really isn't time in class or out of class. There really is not anything that I want to change that I haven't. I think I do a pretty good job. (Classroom observation, May 2005, Group E member)

It is evident from this quote that Martha did not understand or agree with the idea that when pedagogy is improved by incorporating the principles of assessment for learning, students will be prepared for state tests, since the focus will be on checking for level of understanding and taking instructional action in the light of evidence. It is clear that Martha thinks she is a strong teacher and does not need to make additional changes. Her attitude toward the project, and professional development in general, may have limited the group's ability to challenge one another, since that was not a part of this particular group's working norms. This suggests that TLC leaders and professional development providers will need to model this type of discussion for some TLCs and will need to create an atmosphere where it is not only acceptable to challenge another member, regardless of experience, it is expected.

The second research question focused on the contribution of TLCs to the development of teacher understanding of formative assessment. ETS staff structured these meetings to provide opportunities for teachers to discuss new AfL content each time they met and to ensure that each meeting conducted both the How's It Going? and the Action Planning activities. Analysis of the TLC meeting narratives showed that these activities did support the learning processes by providing new ideas, holding teachers accountable for making changes, and providing support for the TLC leaders and other group members, allowing participants to continue to engage with the professional development ideas throughout the year.

While a structure can be provided and TLC meetings scheduled, it was still an open question whether the TLC participants would learn only from the AfL content provided, or additionally from one another forming true learning communities. We examined problem solving from three perspectives and found that there were variations across the groups that were closely tied to the working environments and norms that were established early on. This has led to the development of several activities designed to promote healthy working norms and to examine roles and responsibilities within TLCs for future versions of the intervention.

To What Extent Does Teacher Learning Translate Into Teacher Practice?

In order for professional development to impact student learning, not only must teachers learn about new teaching practices, but this new learning must result in changes to instruction. For that reason, it is not enough to examine teacher understanding of formative assessment, we must also investigate whether this new understanding had an impact on the classroom. To do this, classroom observations were conducted at two points during the year, approximately 5 months apart. Within the core group of teachers, we have two observations for 17 teachers. The classroom observations provided us with information at three levels:

1. An overall assessment of lesson quality
2. Quantitative scores with qualitative justifications for seven aspects of instruction
3. Examples of techniques in use within these classrooms

The overall assessment of lesson quality provides a score focused on the learning that occurs during the lesson, student engagement, and the use of the question-answer-feedback cycle. It was hypothesized that teachers who practiced formative assessment would have higher quality lessons and that as teacher learning translated into practice that these scores would increase. For the 17 core teachers who had two observations, the average score at Time 1 was a 2.6 (out of a possible 6 points), while the average score at Time 2 was a 2.9. A paired-samples t-test showed no significant differences between the two sets of observations for overall score or the other seven dimensions. Even though the observation protocol was revised, a more sensitive instrument may be needed in order to record observable changes in practice. The protocol focused on best teaching practices and the question-answer-feedback loop. Subsequent protocols have been revised to focus more specifically on the overall use of assessment for learning. Hopefully, this change will provide a more sensitive measure of the types of changes for which we would like the professional development to contribute.

Seven aspects of instruction were also rated on a 5-point scale. These aspects included lesson planning and design; algebra/mathematics content; classroom interactions/environment; questioning and classroom talk; assessment and analysis of student understanding; providing feedback to students; and responsive and reflective teaching. Since each of these aspects should be improved by use of formative assessment, it was hypothesized that ratings on each aspect would improve as teachers' understanding of formative assessment improved, although as noted no significant differences were seen. In subsequent protocols, these general aspects of instruction

have been replaced with descriptions of specific formative assessment techniques used within the lesson. While the quantitative scores were not as informative as we had hoped, narratives from each observation were used to identify the use of formative assessment techniques within the lesson. If teacher learning was in fact translating into practice, we would expect to see implementation of assessment for learning techniques during the lesson observations, and as the year progressed we would expect to see more techniques implemented. Across the 17 observations at Time 1, the implementation of an assessment for learning technique was observed on 32 occasions. On average, teachers implemented approximately two techniques per lesson. At the second observation, the frequency with which assessment for learning techniques were observed was similar. Across the 17 observations at Time 2, implementation of an assessment for learning technique was observed on 30 occasions. Again, this is an average of approximately two techniques per lesson. Table 3 shows the techniques and the frequency with which they were observed at both Time 1 and Time 2. The appendix provides short descriptions of each technique.

Although the average number of techniques per lesson was consistent across both observations, the techniques observed did change. At Time 1, 11 separate techniques were observed. The most popular techniques included learning intentions and whiteboards, which were observed in eight and six lessons, respectively. Both of these techniques were discussed during the introductory workshops. Several techniques were observed on only one occasion, including higher-order questioning, learning logs, rubrics, peer assessment with rubrics, exit tickets, and keep the question going. At Time 2, 11 separate techniques were again observed. The most popular techniques remained the same; however, several of the less-popular techniques were not seen at the second observation. Therefore, although the range of techniques remained consistent, four new techniques were observed. These new techniques included wait time, hot seat questioning, think/pair/share, and systematic variation in problems. These techniques were all discussed during the new AfL content presented at later TLC meetings, which may indicate that the new AfL content presented throughout the year does in fact translate into changes in teacher practice.

Even though the average number of techniques per lesson and the range of implemented techniques did not increase from Time 1 to Time 2, the fact that these numbers stayed consistent is a positive result. Researchers such as Cohen and Hill (1998) have discussed the limited value

39

of one-time workshops with respect to sustaining changes in teacher practice. The finding that 5 months after the initial workshops, teachers who participated in this professional development program were still using AfL techniques is encouraging. Not only were there observed changes to teacher practice but these changes were sustained and/or different techniques were implemented after 5 months. This may provide additional support for the use of ongoing TLC meetings and the extension of the program to cover a full 2 years of material.

Table 3

Frequency of Techniques Observed at Time 1 and Time 2 for 17 Teachers

Technique	Time 1	Time 2	Total
Learning intentions	8	7	15
Whiteboards		6	12
Group work	5	3	8
No hands up	4	2	6
Class poll	2	3	5
Higher-order questioning	1	2	3
Learning logs	1	1	2
Rubrics	1	0	1
Peer assessment with rubrics	1	0	1
Exit tickets	1	0	1
Keep the question going	1	0	1
Wait time	0	1	1
Hot seat questioning	0	2	2
Think/pair/share	0	1	1
Systematic variation in problems	0	1	1
Total	32	30	62

The number of observations per group is too low to conduct a comparison by group; however, some interesting patterns were seen in the two groups with the most observations. All five members of Group C were observed at both Time 1 and Time 2. Across the 5 observations at Time 1, the implementation of an AfL technique was seen on 9 occasions. However, the same

three techniques were observed in all 5 observations (learning intentions, whiteboards, and group work). At the second observation, the frequency of AfL techniques increased by 1, from 9 occasions to 10 occasions, but the range of techniques increased from 3 occasions to 6 occasions.

Six of the 10 members of Group E were observed on both occasions. At Time 1, three techniques (learning intentions, whiteboards, and no hands up) were observed in at least two lessons, while five techniques were seen only once. At Time 2, the frequency of AfL techniques decreased from 12 observed techniques to 8. The range of techniques also decreased from 8 different techniques to 5.

These differences between these two groups suggest how different TLCs may support one another in the learning process. Although members of both groups were still implementing assessment for learning techniques at both observations, there were different patterns in the frequency of techniques observed and the range of techniques implemented. The teachers in Group C began with more shared techniques, which diverged over time, while in the teachers in Group E began with a broader list that coalesced around a subset of these techniques. Any conclusions about these patterns must remain tentative, given the low numbers being considered. However, these results do suggest an avenue for future research focused on understanding the different learning pathways of for both individuals and groups, and from there to ensure that a professional development program is flexible enough to support all pathways.

In addition to identifying the frequency with which each technique was observed, the observed techniques were also scored with the same implementation scale used to evaluate the teachers' reports during the How's It Going? activity. Each observed technique was scored based on the level of sophistication with which it was used. A score of 1 was awarded when a technique was implemented weakly in such a way that the neither the teacher nor the students were able to make formative use of the information. A score of 2 was assigned when performances tended to lack consistency in execution, where students were given insufficient structure or guidance to get the most benefit from a particular technique, or where the teacher collected and analyzed information about student learning but failed to apply it appropriately to determine next instructional steps. A 3 was given to performances that represented the highest-quality implementation and was one in which the teacher and students made maximal use of the information collected about student learning to move instruction forward. Finally, instances when an identified assessment for learning technique was changed in such a way that rendered it

almost impossible to make formative use of the technique were coded as a non-formative use. The average implementation score for all techniques observed was 2.0 at both Time 1 and Time 2. Although at first glance this may indicate that there was little improvement in the implementation of assessment for learning techniques, it is important to remember that the techniques implemented varied from Time 1 to Time 2. Participating teachers were encouraged to not only refine and adapt existing techniques but also to implement new techniques throughout the school year. So although the average score did not increase for Time 2, Time 2 also includes techniques that may have only been implemented for only a short period of time.

While the overall distribution of scores was similar for the two observations, there were multiple examples of teachers showing improvements in their use of the same technique. For example, one of the middle school teachers used whiteboards with her students in both of the lessons in which she was observed. During the first lesson, she asked the students a question that they were to respond to on their individual whiteboards. The observer noted:

> While students hold up their whiteboards showing their answers, only half the class gets it right. The teacher moves right along by announcing: "10 points for the correct answer to the homework problem!" (Classroom observation, October 2004, Group A member)

A little later in the same lesson, in response to another question from the teacher, the observer noted, "Students hold up whiteboards, but again, there is no elaboration of answers shown." This pattern of students using whiteboards to write their answers but receiving little response or comment back from the teacher regarding their answers was repeated on multiple occasions during this lesson. In a post-observation interview, the teacher was asked to comment on how she selected the formative assessment strategies she used in the lesson. She responded that, "Whiteboards are great and the kids really love using them as you can tell." At this stage, the teacher was unable to elaborate on the broader purpose, and her overall implementation of this technique was judged to be low.

However, the picture changed 6 months later, when this teacher was observed a second time. On this occasion, the directions she gave her students were much clearer. The teacher reminded the students that they would have time to solve each problem and, on the count of three, they would hold up their whiteboards for her to see. For one problem, several students were confused and the teacher asked another student to work with the students who didn't understand how to solve the problem until they indicated their understanding. The teacher

repeated this strategy at multiple points during the lesson. In the post-observation interview for this second occasion, the teacher discussed her use of the whiteboards in a much more sophisticated way, indicating that she had used the whiteboards to get a quick assessment of her students' understanding and to see where they might have had difficulty with the steps needed to solve the problems. The overall implementation of this technique was judged to be of higher quality than the earlier implementation. She attributed this improvement, in part, to monthly TLC meetings and the reporting that was routinely expected of each participant:

> The monthly workshops were very helpful. Although I try things it kept me on target and made me think because I knew that I would be asked what I had done over the month. (Classroom observation, May 2005, Group A member)

In another example, one of the high school teachers had students work in small groups to complete a task during both observations. During the first classroom observation he allowed the students to divide into two groups so that one group could work on the computer for part of the class period and then switch with the other group. Within the groups the distribution of effort was uneven, with one student doing most of the work while the others observed. The teacher encouraged those watching to make sure that the person who was working was correct, but this effort was quite ineffective.

The setup of the group work started off in a similar manner during the second observation, with the teacher allowing students to select their own groups. However, when this teacher saw how his students had arranged themselves, he intervened. In his post-observation interview he noted the following:

> I realized that all the strong students were together. So I moved the groups around because all the strong students were in one group, which would have made the other groups too weak. (Classroom observation, May 2005, Group C member)

On this occasion he reminded the students of the specific cooperative group roles (recorder, reporter, timekeeper, etc.) but also reminded them that they were all expected to participate in the group's thinking process. During this observation, students stayed on task more, and the observer noted that conversations were focused around solving the mathematics problem at hand. What is particularly noteworthy about the second observation of small-group work is that this observation was conducted in May. In early April the TLC meeting focused on

various aspects of group work and how group work can help students learn collaboratively. At that meeting, this teacher had advocated assigning roles to students within the group, an approach that had not been evident during the first observation but was in the second.

The third example of a teacher's practice changing from one implementation to the next differs from the previous two in that on both occasions the teacher was seen implementing whiteboards in an accomplished way. However, by the second observation the teacher had broadened the ways in which he used them. During the first observation the lesson focused on finding the area of regular shapes, and the teacher had students complete a series of problems on their individual whiteboards and hold them up so he could see their responses. Students used both sides of the whiteboard: On one side they drew diagrams to show the teacher, and on the other side they completed the problem and wrote the final answer for the teacher to see. For some questions, the teacher encouraged his students not only to share their answers with him but with the rest of the class so that they could see the variety of approaches used to solve the problem. In response to one question there was a range of responses, and the teacher's initial feedback was to say only that he saw a range of responses. All of the students looked at their whiteboards again to check their responses, and several students made corrections. Overall, he used the immediate feedback from the students' whiteboards to direct the class, to capitalize on various problem-solving approaches, and to select the next question to move learning along.

During the second observation, the students were again using whiteboards while working on the topic of exponents. During this class the teacher did not ask students to hold up their whiteboards. Instead, he moved around the classroom to look at student work. As students completed each problem he selected one or more student responses. The whiteboards were made by inserting a white card into a clear document sleeve. Removing the card allowed the sleeve to act as a transparency, and the teacher put the selected response on the overhead projector to share with the whole class. Given that students were expanding exponents, this approach made it easier to share students' work with the rest of the class. It also gave the teacher an opportunity to demonstrate different approaches to the problems. Using this approach, the teacher was able to move the students along to expanding algebraic expressions with exponents, since he was able to monitor student understanding during the class. This example illustrates how refinements and adaptations to techniques also translated into practice throughout the year and provides evidence that even accomplished teachers can make changes and improvements to their practice.

A second interesting aspect of this observation is the transfer of knowledge between the TLCs. Although ETS staff had advocated the use of whiteboards throughout the project, the idea of removing the white card from the clear sleeve in order to share responses on an overhead projector came from a teacher in the Hickory school. ETS staff members who had attended the TLC meetings shared the whiteboard idea across districts. So although the idea originated in one school district, it was shared with teachers in the other district and was later seen being implemented quite successfully. There is an important lesson here for future use of this model: It is important to provide TLC leaders with opportunities to meet and share ideas across TLC groups so that everyone in a district can benefit from examples of good teaching practice and successful implementations. This is an important means of keeping things fresh and growing expertise beyond an immediate TLC group and into the district as a whole.

To recap, the third research question focused on the extent to which teacher learning translates into teacher practice. To examine this, teacher observations at the start and end of the year were examined. Scores for overall lesson quality and the seven aspects of instruction were weak in general, and improvements were too small to attribute to the intervention. However, analysis of the techniques implemented during the observed lessons revealed some interesting findings. The implementation of AfL techniques was observed at a similar rate at both Time 1 and Time 2, indicating that participating teachers were employing ideas from the professional development. In addition, new techniques were observed at Time 2, indicating that not only were teachers still using new learning from the summer workshops, they were also translating learning from the TLC meetings into their practice. Finally, although the average quality of implementation did not improve, there were several examples that indicate improvement in the implementation of assessment for learning techniques across the year. Given that the most useful source of information for this analysis was the specific identification and analysis of assessment for learning techniques, the overall protocol has been revised to focus more specifically on those teaching practices that are important to the implementation of assessment for learning.

What Factors Supported or Hindered Teachers' Development of Formative Assessment Understanding and Practice?

The previous questions focused on teachers' understanding of AfL, the ways in which the TLCs contributed to the development of that understanding, and how this learning translated into changes in teacher practice. This question focuses on the broader contexts in which the TLCs

were situated, looking at factors within the schools and districts that supported or hindered the development of TLCs, the development of the teachers' understanding of formative assessment, and the ability of teachers to make changes to their classroom practice. In addition, the impact of ETS support is examined.

One striking difference between the two districts was the level of support offered to the teachers by school- and district-level administration. Examining first the Hickory school district, three district staff attended the summer workshop. In addition, district staff attended two of the nine meetings: the assistant superintendent and the K–8 mathematics supervisor at one, and the K–12 professional development supervisor at another. Throughout the year various teachers reported being observed by a principal, an assistant principal, or the K–8 mathematics supervisor, and on each occasion they received positive feedback about the new ideas they were implementing in their classrooms. The school district purchased sets of whiteboards for every teacher in the district as a result of the feedback they had received from the teachers in the pilot implementation. Toward the end of the year, an assistant principal asked one teacher to share experiences with others in the school, not just the mathematics teachers. Overall, the Hickory teachers knew that the project had the support of the district- and school-level administrators, and in various ways received feedback that indicated that their efforts were noticed and valued.

By way of contrast, district support was not evident in the Tupelo school district. Unlike the Hickory teachers, all of the Tupelo participants were paid a stipend for the time they spent in TLC meetings. However, that was one of the few tangible ways in which the teachers' efforts were recognized. Even though the interim mathematics chair attended the introductory workshop, no district- or school-level administrative staff attended any of the subsequent 36 TLC meetings (three TLCs for the high school teachers and one for the middle school teachers, with 9 meetings each). In the meetings, the only positive examples of district support were comments from ETS staff. These comments informed the teachers that the district had agreed to continue to support the project by offering a second year of stipends. The only other gesture of support came at the start of the school year, when the district agreed to provide overhead projectors for every teacher in the project. Initially, this gesture was viewed very positively by the teachers, but given that the teachers were still lobbying for the projectors and bulbs by the December meeting—4 months into the project—this support fell short of what it could have been.

46

The TLC narratives were also examined for examples of school or district hindrances to the teachers' development of formative assessment practices. Nothing was noted in this category for the Hickory TLC. Unfortunately in the Tupelo high school there was not a positive attitude among the teachers toward the administration. Beyond attitudes, communication and information seemed lacking within the school. As an example, when asked, teachers in one of the Tupelo groups had no idea who was in charge of professional development within the school. There was no permanent mathematics chair in the school, and as a result no leadership of the mathematics department. The teachers viewed many of the behavior and tardiness issues in the school as problems that were a result of the administration failing to set up and enforce rules in the school. Consequently, students were allowed to walk the halls long after a bell indicated the start of class time, and few of the observed lessons began on time.

Some of the barriers to the teachers' deepening their understanding of formative assessment were somewhat indirect, and related to the policies and practices at the Tupelo high school. These practices were not specifically hindering formative assessment practice but rather teaching in general, so they often served as a distraction to the teachers, who wanted to talk about them. Some of these issues, such as a packed curriculum and hard-to-meet pacing guide, are not limited to the schools in our study—such complaints are almost universal. Other decisions made by school administration were difficult to understand. For example, in October some students in the pre-algebra classes were moved to newly formed algebra classes, and some teachers were given new assignments to cover these classes. Other pre-algebra teachers gained additional students from classes that were divided up between several teachers in order to allow the original teacher to take on the new algebra class. Moving both the teachers and the students resulted in a general level of discontent among the teachers. Considerable time was spent in the second meetings of Groups B, C, and D, and the third meeting for Group C discussing the administration's decision and its outcome. The somewhat capricious nature of this decision only added to the teachers' sense that the administration was not supportive of them as teachers. The teachers were not happy with the block scheduling: Pre-algebra and algebra classes met daily so that the course only took one semester. The movement of students from pre-algebra to algebra and between pre-algebra courses only exacerbated the feeling that they did not have enough time with their students. Other issues included meetings scheduled to coincide with the TLC meetings, lax detention and tardiness policies, and a lack of connection with administrative staff.

As a result, guidelines were implemented in a very ritualistic and procedural way, without much thought to the purpose or impact they would have on learning. For example, objectives were written on the board, since, as one teacher noted, "If you're being evaluated, and the objective is not on the board, you're dinged" (Group C, November meeting). However, these objectives were rarely if ever referred to by the teachers. Another example was the high school principal's policy on group work. One teacher noted, "The principal said she'd fire us if she didn't see students in groups" (Group C, May meeting). However, during observations, it was noted that teachers who had students working in groups were more often visited by administrative staff to ensure that the noise was related to learning.

ETS staff offered support in a variety of ways, such as bringing materials (index cards, whiteboards, markers, tickets), photocopying materials from the Algebra and Pre-Algebra Packages, and providing in-class support (e.g., offering to team-teach). Very few teachers asked for in-class support, and the majority of the requested support was for ETS to provide materials. As noted earlier, the in-class support was offered during early months of the project, but as requests became few and far between, support became more focused on the TLC meetings. However, requests for materials continued throughout the year.

Occasionally the support offered by ETS may in fact have made the teachers overly reliant on us. In Group E during the April meeting, the ETS staff person mentioned that the district was going to buy a set of whiteboards for every teacher because of the feedback from the group. One teacher asked if they were going to buy dry erase markers for everyone too, but the ETS person did not know. The teacher responded, "Why don't you ask—you seem to get a good response" (Group E, April meeting).

A somewhat similar comment occurred during the last meeting of one of the high school groups. The teachers were discussing the schedule for the algebra classes and were unhappy with the proposal. Although this topic was quite beyond the scope of the project, one group member said, "We need ETS to support our position in front of the superintendent" (Group C, May meeting).

Reflecting on the role that ETS had in this project, for the most part the teachers were very grateful for the material support they received from ETS but were not interested in receiving in-class support. This may have negatively affected the program, as teachers viewed ETS as a provider of manipulatives instead of a provider of professional development.

Additionally, because of its outsider status, ETS was also viewed as a potential advocate in arenas beyond formative assessment, further diluting ETS's role within the district.

Overall, if the ESTPD intervention is to be effective at scale, the factors that support and hinder implementation within any district must be considered. This project highlighted several aspects of district support that should be maintained for a professional development project to be effective, including an administrative presence at trainings, acknowledgment of the types of changes that participating teachers are being asked to make and any conflicts there may be with existing or new policies, and the need for a defined role for the TLCs and professional development providers. Additionally, if the intervention is to be effective, ETS staff cannot hold teachers' hands, so to speak, by providing materials, acting as an outside advocate, and/or discussing problems that are not directly related to assessment for learning. ETS's role must be to provide new and important AfL content knowledge and teaching practices to improve instruction and learning.

Formative Impact of the Work

The purpose of this section is to describe lessons that were learned as LTRC staff worked with the Tupelo and Hickory school districts. Long before the writing of this report was started, the next iteration of the project began. In the school year that followed the effort described here, we began work with the Cleveland Municipal School District. Rather than working directly with teachers, TLC leaders ran the TLC meetings. As a result we needed to develop training for those leaders and materials for them to use. This development leap would not have been possible had we not had the experience of being teacher leaders the previous year.

The TLC narratives from the 2004–2005 school year were reviewed multiple times before being formally analyzed for this report. After each round of monthly meetings was complete, the narratives were written and shared among members of the project team. At that stage, each one was read by most of the LTRC staff. Within a few days it was time to focus on the agenda and structure for the next meeting, so the notes from the previous series of meetings were discussed in terms of what the teachers were interested in and struggled with, along with our impressions of how they were implementing the ideas in their classrooms. Most of the workshop narratives ended with an overall comments section, which represented the first analysis of each workshop as the notetaker switched from documenting the meeting to commenting on it. Concerns about misunderstandings that were becoming apparent in some of

the groups were first voiced here. Subsequent learning opportunities in later meetings were structured to address these concerns.

The opportunity to train teachers and TLC leaders in Cleveland required the development of materials that we had not needed in the first year of the work. The introductory workshop was modified from what was used with the Tupelo and Hickory teachers, and new materials were created for the TLC leader training. Materials were also developed for the TLC leaders to use in the monthly TLC meetings. As a result of the work with these teachers over the course of 1 year, we also became convinced that teachers would be better served by having 2 years in which to engage with the ideas surrounding formative assessment.

The following list represents lessons that we learned from the Tupelo/Hickory experience, lessons informed by the process of conducting and observing the TLC meetings, and subsequent discussions among LTRC staff of the participants, their learning, their misunderstandings, and their successes over the course of that year. All of these lessons significantly informed the shape of our immediate future work.

1. Early in the school year we began to realize that the participants did not always understand the purpose behind the practical ideas and tools we were providing. In fact, by the second meeting, observers were noting that the teachers seemed to be treating the ideas as "a bag of tricks" to be used almost as treats for good student behavior. Because of this concern, we clarified the main research aspects of formative assessment and called them *formative assessment strategies*, which could be implemented in a variety of ways that we called *formative assessment techniques*. We also put in writing the one big idea of formative assessment, which tied the strategies together into a coherent whole and provided a useful touchstone against which to test various techniques by asking the question, Does implementing this classroom technique support some aspect of the big idea? By creating a specific vocabulary we could be much clearer both in our presentation of the material in the introductory workshop and in the materials that we developed for the monthly meetings.

2. After working with the various groups of teachers over the course of the year, we came to appreciate their struggles processing the volume of material presented. The strategy/technique distinction provided a tighter framework. Seeing how teachers had misunderstood aspects of formative assessment led us to significantly increase the

amount of modeling of formative assessment techniques used in the workshop and also to explicitly discuss how and why they had been incorporated into the workshop.

3. During the first year of TLC meetings the groups met once a month. Meetings generally were 2 hours long. Groups ranged in size from as few as two people (Group A) to as many as nine (Group E). The teacher feedback we received indicated that the frequency and duration of the meetings were appropriate, so in subsequent iterations we have continued with monthly meetings and the 2-hour time frame. In the light of our experience of varying group sizes, we modified our recommendation for the ideal group size to a range of between four and eight members. A group of two or three participants seemed too small to allow for the depth of conversations we would like the teachers to engage in, but groups that are too big make it difficult to ensure that everyone has an opportunity to participate.

4. Reflecting on the structure and focus of the original Tupelo/Hickory TLCs led us to realize that the amount of material we intended to cover in each meeting was too ambitious. There were often more topics or activities than could meaningfully be covered in a single 2-hour meeting. Ideas got jumbled and the overall focus on formative assessment at times even got lost. As a result, the meeting materials for the following year were much more tightly focused on a single topic per meeting, with a much clearer articulation of which strategy was to be the focus of the meeting and what techniques could be employed to support that strategy. Several of the activities or discussions that were used in for the Tupelo/Hickory meetings were not incorporated into subsequent iterations because the discipline of having to connect the topic to a strategy made us realize that some of the topics covered during the first year were not actually relevant.

5. As we reviewed the TLC narratives with the benefit of hindsight, we realized that on occasion we not only allowed off-topic discussions to continue, we actually promoted them. A key example is the tickets technique. This technique was first mentioned by a teacher in Group E during the November meeting. The discussion on the starts and ends of lessons concentrated on how teachers can use learning intentions at the start of the lesson to focus students on the particular learning for that class period, and the

value of returning to those learning intentions at the end of the lesson to review progress. One teacher told the group about her ticket system for rewarding students "for coming to class on time, starting the warm-up, and for having a pencil, notebook and calculator." Little was said in this meeting about the approach, since it did not have a formative assessment component per se. However, at the next round of meetings, ETS staff mentioned this approach during the Group B meeting. Over the next couple of months this approach grew in popularity, almost exclusively among a number of the Tupelo high school teachers who were struggling with discipline issues. We had been providing these teachers with materials (such as index cards for entrance/exit tickets and page protectors for whiteboards), so we also provided them with rolls of raffle tickets. Enthusiasm for the ticket system grew, with some teachers sharing the idea with others who were not directly involved in the project. However, the ticket system became problematic for two reasons: First, the system may not have actually improved student behavior, and second, the approach diluted the focus on formative assessment. Although some teachers reported positive improvements, another teacher said, "Well, I started using the tickets. At first, it was for their work, but now it's for almost anything. But it's still like pulling teeth to get them to turn in their work. My prizes were mainly candy, gum, or 5 points on their grade—most take the points. One of my classes doesn't turn in work, so I widened the ticket use." Research on motivation (Elliot & Dweck, 1988) suggests that external rewards may be problematic, and this particular example seems to support that finding. The second issue was that discussion about the system distracted teachers from focusing on learning more about how to use formative assessment to increase student motivation and engagement. The ticket system was the first technique several teachers mentioned when asked about techniques they thought were particularly successful, suggesting that some of the teachers had lost sight of the underlying focus on formative assessment. From this we learned that there is a fine line between respecting the expertise, suggestions, and enthusiasm of the teachers in areas that may not be directly related to formative assessment and allowing those ideas and suggestions to shift the focus away from the goals of the TLCs. The conversation should have been redirected back to formative assessment, and the teachers should have been referred

to other resources for help with their management and engagement issues. In the work the following year, we coined the phrase *So what's formative about that?* to help teachers focus more clearly on the big idea.

6. Another lesson learned during this year of work had to do with the issue of maintaining a clear focus on the professional development. The intervention focused in part on supporting the teachers' use of the Algebra and Pre-Algebra Packages. These materials provide rich mathematical problems for students, along with pre-assessment and discussion questions for the teachers. The Algebra and Pre-Algebra Packages complimented the multiple formative assessment approaches that we had presented to the participating teachers. However, we struggled in this project to help the teachers' efforts to embed these tasks into their existing curriculum. They tended to treat the tasks as interesting but something extra to be done for the project, rather than as a solution for finding good questions, rich tasks, and sample student work. Only a small proportion of the teachers used the materials at the start of the year, although we set time aside in each meeting to talk about the next task that they would be trying. After the January meeting, we stopped including this discussion as part of the meeting agendas, and without the constant reminders, teachers stopped using the materials, or at least stopped talking about them. On one hand, we had been overly ambitious when we decided to include these materials as part of the intervention. On the other hand, we may have seen better results if we had returned to the activity of the summer workshop and set time aside for the teachers to figure out how to meaningfully incorporate the materials into their instruction and to learn from the experiences of other teachers in their group.

7. We also learned that the leadership role could be given to group members—it did not need to be held by an ETS staff person. For three of the monthly meetings (December, January, and February), a member of each group facilitated the meetings. In almost every instance, the designated leader was provided with the materials before the meeting, came to the meeting obviously prepared to lead it, and ensured that everyone had an opportunity to participate. The one aspect that was a struggle for the leaders was dealing with an unanticipated question, and they often looked immediately to the ETS staff member for an answer. In addition, the materials

provided to the meeting leader were little more than an annotated agenda and the various handouts that were needed for the class. It was only in subsequent revisions that the leader notes contained enough information to provide teachers with the supportive content that they needed, such as learning intentions for each meeting, key talking points, and handouts summarizing main points. This additional support helped the leaders support the rest of the group as they engaged with the ideas of formative assessment. As a first step, however, it was important to see that leadership could be shared with group members and to understand the kinds of support that we would need to develop. The other encouraging aspect of shared leadership was that although the ETS staff person provided the majority of responses to questions asked by group members, on occasion other teachers had ideas to share or suggestions for problems. Even though the teachers were still grappling with the ideas of formative assessment, they were prepared to help each other think through ideas and solve problems.

8. The final lesson we learned over the course of the year related to our own collective understanding of formative assessment. As we worked together each month to create meeting agendas and respond to the teachers' questions and struggles, we began to understand the nuances of formative assessment more deeply ourselves. During the first 6 months of the project we were collectively operating with an under-theorized conception of the intervention's AfL content. With hindsight, we can see this in two distinct ways. First, although we conducted classroom observations, after examining those protocols we realized that the observation ratings focused primarily on the question-answer-feedback cycle and student engagement. Other aspects of formative assessment, such as self-assessment and peer assessment, were not accounted for in this first year's observation protocol. Second, before we had articulated the notion of strategies and techniques, we had earlier versions of those ideas that included such things as review games that were popular with the Hickory group, and group work, which was more about instruction than formative assessment. It is important to recognize that there was an increasing level of sophistication in our own thinking.

Discussion

The overall goal of the implementation of the ESTPD project was to understand the *how* and *why* of this particular approach to teacher professional development. From previous work, it was clear that a minimal-support approach was insufficient. In this project, teacher learning was supported through a 3-day workshop and monthly TLC meetings. Although we initially assumed that coaching would also be part of the support, coaching did not play as central a role as initially envisioned.

There were various cycles of analysis for this particular research-and-development project. The most frequent forms of data collection were the observations of the TLCs and subsequent writing of the narratives. The group facilitators wrote these narratives the week after the meeting and shared them with the rest of the research team. Twice during the 1-year project, we conducted a series of classroom observations to see whether the teachers were incorporating formative assessment approaches into their instruction. Short-cycle analysis was also important so that we could answer the question, What do we need to do next month? In terms of a medium-cycle analysis, we needed to be able to answer the question, What will the next iteration look like? The final analysis was the most formal one and involved our working from a set of research questions to develop a formal coding structure:

1. How does teachers' understanding of formative assessment change over time?

2. How do TLCs contribute to the development of teacher understanding of formative assessment?

3. To what extent does teacher learning translate into teacher practice?

4. What factors supported or hindered teachers' development of formative assessment understanding and practice?

Having used the data collected to respond to both the short-cycle and medium-cycle analyses, we had intuitive responses to these questions; however, formally examining the data for each question was valuable in order to clarify and, in some cases, revise our memories. It was enlightening to have five TLC groups functioning within the same study and using the same materials. We learned that not all groups are created equal. The different working environments and norms created within the groups both positively and negatively shaped the potential for

improving teacher understanding, changing teacher practice, and providing new learning for participants.

The first research question focused on the changes in teachers' understanding of formative assessment over time. Although we did not formally assess teachers' knowledge of formative assessment at the start and end of the year, their comments during TLC meetings provided a window into this understanding. The comments made and questions asked, the teachers' reports of implementation, and the range of implementation scores indicate that the teachers thought about their practice and behaved differently as result of the intervention. This is especially evident in the implementation scores, as the quality of the examples indicated an improvement in quality over the course of the year. So, although analysis of the comments and questions asked by the participants showed a high proportion of misconceptions and suggested some struggles to understand what we meant by formative assessment, the implementation scores present a more positive picture.

The second research question focused on the contribution of TLCs to the development of teacher understanding of formative assessment. ETS staff structured these meetings to provide opportunities for teachers to discuss new AfL content each time they met and to ensure that each meeting had a How's It Going? session and Action Planning activities. Analysis of the TLC meeting narratives showed that these activities supported the learning processes by providing new ideas, holding teachers accountable for making changes, and facilitating collegial support between TLC leaders and other group members, thus allowing the participants to continue engaging with the professional development ideas throughout the year.

While a structure can be provided and TLC meetings scheduled, it was still an open question whether the TLC participants would learn only from the materials provided, or also from one another and thus form true TLCs. We examined problem-solving approaches from three perspectives and found that there were variations across the groups that were closely tied to the working environments and norms that had been established early on. This has led to the development of several activities designed to promote healthy working norms and to an examination of the roles and responsibilities within TLCs for future versions of the intervention.

The third research question focused on the extent to which teacher learning translates into teacher practice. To examine this, teacher observations at the start and end of the year were examined. Scores for overall lesson quality and the seven aspects of instruction were weak in

general, and improvements were too small to attribute to the intervention. However, analysis of the techniques implemented during the observed lessons revealed some interesting findings. The implementation of AfL techniques was observed at similar rates at both Time 1 and Time 2, indicating that participating teachers were employing ideas from the professional development. In addition, new techniques were observed at Time 2, indicating that not only were teachers still using new learning from the summer workshops, they were also translating learning from the TLC meetings into their practice. Finally, although the average quality of implementation did not improve, there were several examples that indicated improvement in the implementation of assessment for learning techniques across the year. Given that the most useful source of information for this analysis was the specific identification and analysis of assessment for learning techniques, the overall protocol has been revised to focus more specifically on those teaching practices that are important to the implementation of assessment for learning.

Overall, if the ESTPD intervention is to be effective at scale, the factors that support and hinder implementation within any district must be considered. This project highlighted several aspects of district support that should be maintained for a professional development project to be effective, including an administrative presence at trainings, acknowledgement of the types of changes that participating teachers are being asked to make and any conflicts there may be with existing or new policies, and the need for a defined role for the TLCs and professional development providers. Additionally, if the intervention is to be scalable, districts cannot rely on ETS staff to provide materials or act as outside advocates. ETS's role must be to provide new and important formative assessment knowledge and teaching practices to improve instruction and learning.

Epilogue

Out of the research-and-development work conducted in these two districts, ETS's *Keeping Learning on Track*® (KLT) program emerged. KLT is a sustained, interactive professional development program that helps teachers adopt minute-to-minute and day-by-day assessment for learning strategies that have been shown by research to increase student learning. Cleveland Metropolitan School District agreed to pilot the first year of this program, which necessitated the development of support materials, since LTRC staff would not be as directly involved with the participants. Many of the lessons learned in Tupelo and Hickory were applied in the development of these materials, as outlined in the section Formative Impact of the Work. For further information about the KLT program, see http://www.ets.org/klt.

References

Berliner, D. C. (1994). Expertise: The wonder of exemplary performances. In J. N. Mangieri & C. C. Block (Eds.), *Creating powerful thinking in teachers and students: Diverse perspectives* (pp. 161–186). Fort Worth, TX: Harcourt Brace College.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead, UK: Open University Press.

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

Bransford, J., Brown, A., & Cocking, R. (1999). *How people learn: Brain, mind, experience, school*. Washington, DC: National Academy of Sciences.

Brookhart, S. M. (2005). *Research on formative classroom assessment*. Montreal, Canada: Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Cobb, P., McClain, K., Lamberg, T. d. S., & Dean, C. (2003). Situating teachers' instructional practices in the institutional setting of the school and district. *Educational Researcher, 32*(6), 13–24.

Cohen, D. K., & Hill, H. C. (1998). *State policy and classroom performance*. Philadelphia: University of Pennsylvania Consortium for Policy Research in Education.

Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, *54*, 5-12.

ETS. (2007). *Keeping Learning on Track*[TM]: *TLC Leader Kits*. Princeton, NJ: Author.

Garet, M. S., Porter, A., Desimone, L., Birman, B., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Education Research Journal, 38*(4), 914–945.

Ingvarson, L., Meiers, M., & Beavis, A. (2005). Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes and efficacy. *Education Policy Analysis Archives, 13*(10). Retrieved November 13, 2008, from http://epaa.asu.edu/epaa/v13n10/.

Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment that keeps learning on track minute-by-minute, day-by-day. *Educational Leadership 63*(3), 18–24.

Merriam, S. (1998). *Qualitative research and case study applications in education.* San
Francisco: Jossey-Bass.

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

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice
Hall.

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies
create the dynamics of innovation*. New York: Oxford University Press.

Nyquist, J. B. (2003). *The benefits of reconstructing feedback as a larger system of formative
assessment: A meta-analysis*. Unpublished master's thesis, Vanderbilt University,
Nashville, TN.

Reeves, J., McCall, J., & MacGilchrist, B. (2001). Change leadership: Planning,
conceptualization, and and perception. In J. Macbeath & P. Mortimer (Eds.), *Improving
school effectiveness* (pp. 122–137). Buckingham, UK: Open University Press.

Supovitz, J. A. (2001). Translating teaching practice into improved student achievement. In S. H.
Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the States*
(vol. 2; pp. 81–98). Chicago: University of Chicago Press.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge, UK,
Cambridge University Press.

Wilson, S. M., & Berne, J. (1999). Teacher learning and the acquisition of professional
knowledge: An examination of research on contemporary professional development. In
A. Iran-Nejad & P. D. Pearson (Eds.), *Review of research in education* (pp. 173–209).
Washington, DC: American Educational Research Association.

Wylie, E. C., Lyon, C., Martinez, C., & Ellsworth, J. (2007, April). *How formative assessment
ideas translate from talk to practice.* Paper presented at the annual meeting of the
American Educational Research Association, Chicago, IL.

Notes

[1] For the Effective and Scalable Teacher Professional Development (ESTPD) project, the formative assessment strategies were presented as four strategies that resulted from the initial meta-analysis conducted by Black and Wiliam (1998).

[2] For the ESTPD project, a clear distinction between strategies and techniques had not yet been articulated. The presentation of each of the four research-based strategies was followed by the specific ways that participating teachers put those strategies into practice. For this report, the current classification of strategies and techniques is used, but quotations from teachers in this project will not make such a distinction.

[3] The names of the schools and all teachers within the schools have been changed to pseudonyms.

Appendix

Definitions of Observed Techniques

Class poll: The teacher surveys the class for their attitude about a certain topic. The teacher quickly and efficiently asks the whole class what their opinion(s) or gut feeling(s) is about a specific topic or idea. The teacher then performs a short assessment of the results and incorporates the resulting information into the lesson in a way that contributes to student learning.

Exit tickets: The teacher asks a question at the end of a lesson, and students write responses on index cards and hand them in before they can leave the classroom. The teacher uses the cards to assess student learning or understanding of key concepts or ideas from the lesson.

Group work: The teacher structures a task so that students can meaningfully work together to deepen their own understanding and that of the rest of their group.

Higher-order questions: The teacher deliberately uses higher-order thinking questions to promote thinking during class discussion.

Hot-seat questioning: The teacher asks one student a series of questions in order to probe more deeply into a topic or an idea. The teacher uses hot-seat questioning so that the one student who is the focus of the questioning, as well as the rest of the class, benefits. The teacher ensures that all students are engaged in the process, perhaps by asking other students to summarize or react to what the focus student has said.

Keeping the question going: The teacher asks one student a question and then asks another student if the first student's answer seemed reasonable or correct. Then, the teacher asks a third student for an explanation of why or why not there is an agreement between the first two students. This helps keep all students engaged, because they must be prepared to either agree or disagree with the answers given.

Learning intentions: The teacher writes clear, accessible, and valuable learning intentions on the classroom board or on a handout, makes purposive reference to them at the start of the lesson, and refers back to them during or at the end of the lesson in a way that supports student learning.

Learning logs: Near the end of a lesson, the students write summaries or reflections explaining what they learned during the lesson (what they liked best, what they did not understand, what they want to know more about, etc.). The students usually hand these in for review and response, either periodically or at the end of each lesson.

No hands up: The teacher calls on students to answer questions. The teacher may call on any student or students, often using a random approach such as selecting from a collection of Popsicle sticks that each has a student's name written on it. For their part, the students only raise their hands if they want to ask a question.

Peer assessment with rubrics: The students trade papers and check each other's work against a familiar rubric.

Rubrics: The teacher uses rubrics to illustrate expectations for performance on extended tasks.

Systematic variation in problems: A group-work strategy in which all members of the group work on a similar problem but with slight variations. This strategy promotes group work with individual accountability, since each student has a unique problem to solve.

Think/pair/share: The teacher asks a question, and then the students think on their own, share ideas in pairs (perhaps receiving feedback from their partner), and share their best thinking with the whole class.

Wait time: The teacher waits after asking a question before calling on a student or students for an answer to give them time to think before responding.

Whiteboards: The teacher asks or presents a question and waits an appropriate amount of time while students write responses on whiteboards, and then the students individually and simultaneously hold up their whiteboards for the teacher to see.