

*Tight but Loose:
Scaling Up Teacher
Professional Development
in Diverse Contexts*

E. Caroline Wylie, Editor

June 2008

ETS RR-08-29



Tight but Loose: Scaling Up Teacher Professional Development in Diverse Contexts

Edited by E. Caroline Wylie
ETS, Princeton, NJ

Papers by

Laura Goe, Dawn Marie Leusner, Christine J. Lyon, Cynthia Tocci, and E. Caroline Wylie
ETS, Princeton, NJ

Donna Cleland and Maureen Gannon
Math Science Partnership of Greater Philadelphia, PA

Judith Ellsworth
University of Wyoming at Laramie

Margaret Heritage
CRESST, University of California, Los Angeles

Jeff Maher
St. Mary's County Public Schools, MD

Diane Mardy
Ho-Ho-Kus School District, NJ

W. James Popham
University of California, Los Angeles

Donna Snodgrass
Cleveland State University, OH

Gayle Taylor
Vermont Department of Education

Marnie Thompson
Fund for Democratic Communities

Dylan Wiliam
Institute for Education, University of London

June 2008

As part of its educational and social mission and in fulfilling the organization's nonprofit charter and bylaws, ETS has and continues to learn from and also to lead research that furthers educational and measurement research to advance quality and equity in education and assessment for all users of the organization's products and services.

ETS Research Reports provide preliminary and limited dissemination of ETS research prior to publication. To obtain a PDF or a print copy of a report, please visit:

<http://www.ets.org/research/contact.html>

Copyright © 2008 by Educational Testing Service. All rights reserved.

ETS, the ETS logo, KEEPING LEARNING ON TRACK, and LISTENING. LEARNING. LEADING. are registered trademarks of Educational Testing Service (ETS).



Abstract

This series of papers was originally presented as a symposium at the annual meetings of the American Educational Research Association (AERA) and the National Council on Measurement in Education (NCME) held between April 9, 2007, and April 13, 2007, in Chicago, IL. The authors represent school districts and departments of education across the United States, as well as researchers at Cleveland State University, Educational Testing Service (ETS), the Institute for Education in London, and the University of Wyoming at Laramie. All of the current ETS staff, along with Dylan Wiliam and Marnie Thompson, worked at ETS for several years on an iterative research and development program, out of which grew the *Keeping Learning on Track*[®] (KLT) program.

These papers represents the thinking about the theory behind the KLT program, describes the range of contexts used to implement the program, and illustrates the inherent tensions between the desire to maintain fidelity to a theory of action and the need to demonstrate flexibility in order to accommodate local situations. Papers 2 through 6 present descriptions of five implementations in chronological order.

Key words: *Keeping Learning on Track*, KLT, teacher learning communities, TLCs, assessment for learning, AfL, formative assessment

Table of Contents

	Page
Tight but Loose: A Conceptual Framework for Scaling Up School Reforms by Marnie Thompson and Dylan Wiliam	1
A Teacher-Driven Implementation of Assessment <i>for</i> Learning in New Jersey by Laura Goe and Diane Mardy.....	45
The <i>Keeping Learning on Track</i> [®] Program in New Teacher Induction by Jeff Maher and Dylan Wiliam.....	55
Letting Go of the Reins: Learning About Scalability in the Context of One District-Wide Implementation by Christine J. Lyon, Donna Cleland, and Maureen Gannon	67
<i>Keeping Learning on Track</i> [®] in an Urban District’s Low Performing Schools by E. Caroline Wylie, Marnie Thompson, Christine J. Lyon, and Donna Snodgrass	79
A State-Sponsored Pilot Project in Selected Schools in Vermont by Cynthia Tocci and Gayle Taylor.....	93
Scaling Up Across Diverse Contexts: Lessons Learned From Five Implementations of the <i>Keeping Learning on Track</i> [®] Program by Dawn Marie Leusner, Judith Ellsworth, and Laura Goe	105
Tight but Loose: Through the Looking Glass by Margaret Heritage and W. James Popham.....	125
References.....	135

Tight but Loose: A Conceptual Framework for Scaling Up School Reforms

Marnie Thompson (*Fund for Democratic Communities*)
and Dylan Wiliam (*Institute for Education, University of London*)

Introduction

Teaching and learning aren't working very well in the United States. Considerable effort and resources, not to mention good intentions, are going into the formal enterprise of education, theoretically focused on teaching and learning. To say the least, the results are disappointing. Looking at graduation rates (one aggregate measure of the effectiveness of current practice) is sobering. Nationally, graduation rates hover below 70% (Barton, 2005), a figure that is certainly not the hallmark of an educated society. Worse, for the students who are most likely to land in low performing schools—poor students and students of color—graduation rates are even more appalling. The Schott Foundation (Holzman, 2006) reported a national graduation rate for African American boys of 41%, with some states and many large cities showing rates around 30%. Balfanz and Legters (2004) even went so far as to call the many schools that produce such abysmal graduation rates by a term that reflects what they are good at: *dropout factories*. The implications of such educational outcomes for the sustainability of any society, much less a democratic society, are staggering.

Learning—at least the learning that is the focus of the formal educational enterprise—does not take place in schools. It takes place in classrooms, as a result of daily, minute-to-minute interactions between teachers and students and the subjects they study. So it seems logical that if we are going to improve the outcomes of the educational enterprise—that is, improve learning—we have to intervene directly in this black box of daily classroom instruction (Black & Wiliam, 1998b; Elmore, 2002, 2004; Fullan, Hill, & Crevola, 2006). And we have to figure out how to do this at scale if we are serious about improving the educational outcomes of all students, especially students now stuck in chronically low performing schools.

Scaling up a classroom-based intervention is not like gearing up factory machinery to produce more or better cars. Scaling up an intervention in a million classrooms (roughly the number of teachers in the United States) is a different kind of challenge. Not only is the sheer number of classrooms daunting, but the complexity of the systems in which classrooms exist, the separateness of these classrooms, and the private nature of the activity of teaching mean that each teacher has to independently get it and do it right. No one else can do teachers' teaching for them, just as no one else can do students' learning for them. No matter how good an intervention's theory of action or how well designed its components, the design and implementation effort will be wasted if the intervention does not actually improve teachers' practices—in all the diverse contexts in which they work. This is the challenge of scaling up.

This report discusses one promising intervention into the black box—a minute-to-minute and day-by-day approach to formative assessment called the *Keeping Learning on Track*[®] (KLT) program, which deliberately blurs the boundaries between assessment and instruction. The report also describes our attempts to build this intervention in a way that tackles the scalability issue head on. While KLT is in many ways quite highly developed, we are in midstream in understanding and developing a theory and infrastructure for scaling up at the levels required to meet the intense need for improvement described above.

So, in addition to describing the theory of action and components of the KLT intervention, this paper also offers a theoretical framework that we call Tight but Loose as a tool

to assist in designing and implementing classroom-based interventions at scale. The Tight but Loose framework focuses on the tension between two opposing factors inherent in any scalable school reform. On the one hand, a reform will have limited effectiveness and no sustainability if it is not flexible enough to take advantage of local opportunities while accommodating certain unmovable local constraints. On the other hand, a reform needs to maintain fidelity to its core principles, or theory of action, if there is to be any hope of achieving its desired outcomes. The Tight but Loose formulation combines an obsessive adherence to central design principles (the tight part) with accommodations to the needs, resources, constraints, and particularities that occur in any school or district (the loose part), but only where these do not conflict with the theory of action of the intervention.

This tension between flexibility and fidelity can be seen in five place-based stories that are presented in the next papers in this report. By comparing context-based differences in program implementation and examining the outcomes achieved, it is possible to discern rules for implementing KLT as well as more general lessons about scaling up classroom-based interventions. These ideas are taken up in a concluding paper, which examines the convergent and divergent themes of the five place-based stories, illustrating the ways in which the Tight but Loose formulation applies in real implementations.

How This Paper is Organized

Because the Tight but Loose framework draws so heavily from an intervention's theory of action and the details of its implementation, this paper begins with a detailed examination of the components of KLT, including a thorough discussion of its empirical research base and theory of action. We will then present our thinking about the Tight but Loose framework and how it relates to the challenges of scaling up an intervention in diverse and complex contexts, incorporating some ideas from the discipline of systems thinking. Finally, we will discuss the Tight but Loose framework as it might be applied to the scaling up of KLT across diverse contexts.

Keeping Learning on Track: What It Is and How It Works

KLT is fundamentally a sustained professional development program for teachers, and as such, it has deep roots in the notion of capacity building described by Elmore (2002, 2004). We were led to teacher professional development as the fundamental lever for improving student learning by a growing body of research on the influences on student learning, which shows that teacher quality trumps virtually all other influences on student achievement (e.g., Darling-Hammond, 1999; Hamre & Pianta, 2005; Hanushek, Kain, O'Brien, & Rivken, 2005; Wright, Horn, & Sanders, 1997). Through this logic, we join Elmore and others—notably Fullan (2001) and Fullan et al. (2006)—in pointing to teacher professional development focused on the black box of day-to-day instruction as the central axis of capacity building efforts.

KLT is built on three chief components:

1. A content component (what we would like teachers to learn about and adopt as a central feature of their teaching practice): minute-to-minute and day-by-day assessment for learning (A_fL);
2. A process component (how we support teachers to learn about and adopt A_fL as a central part of their everyday practice): an ongoing program of school-based collaborative professional learning; and

3. An empirical/theoretical component (why we expect teachers to adopt AfL as a central part of their everyday practice, and the outcomes we expect to see if they do): the intervention's theory of action buttressed by empirical research.

Attention to the first two components (content and process) has been identified as essential to the success of any program of professional development (Reeves, McCall, & MacGilchrist, 2001; Wilson & Berne, 1999). Often, the third component is inferred as the basis for the first two, but as we will show in this paper, the empirical and theoretical basis for an intervention should be explicitly woven into the intervention at all phases of development and implementation. That is, not only must the developers understand their own theory of action and the empirical basis on which it rests, the end users—the teachers and even the students—must have a reasonably good idea of the why as well. Otherwise, we believe there is little chance of maintaining quality at scale.

The interplay of these three components (the what, the how, and the why) is constant, but it pays to discuss them separately to build a solid understanding of the way KLT works. In the next sections of the paper, then, we outline these three components in some detail. We find that there are so many programs and products waving the flag of AfL (or formative assessment) and professional learning communities that it is necessary to describe exactly what we mean and hope to do in the first two components. Not only does this help to differentiate KLT from the welter of similar-sounding programs, it legitimizes the claims we make to the empirical research base and the theoretical basis described in the third component.

The What: Minute-to-Minute and Day-by-Day AfL

Knowing that teachers make a difference is not the same as knowing how teachers make a difference. From the research summarized above, we know that it matters much less which school you go to than which teachers you get in the school. One response to this is to seek to increase teacher quality by replacing less effective teachers with more effective teachers—a process that is likely to be slow (Hanushek, 2004) and have marginal impact (Darling-Hammond, Holtzman, Gatlin, & Heilig, 2005). The alternative is to improve the quality of the existing teaching force. For this alternative strategy to be viable, three conditions need to be met.

First, we need to be able to identify causes rather than correlates of effective teaching. This is effectively a counterfactual claim. We need to identify features of practice such that, when teachers engage in these practices, more learning takes place, and when they do not, less learning takes place. Second, we must identify features of teaching that are malleable—in other words, we need to identify things that we can change. For example, to be an effective center in basketball, you need to be tall, but as one basketball coach famously remarked, “You can't teach height.” Third, the benefits must be proportionate to the cost, which involves the strict cost-benefit ratio and also issues of affordability. The issue of strict cost-benefit evaluation turns out to be relatively undemanding. In the United States, it costs around \$25,000 to produce one standard deviation increase in one student's achievement. This estimate is based on the fact that 1 year's growth on tests used in international comparisons, such as Trends in International Mathematics and Science Study (TIMSS) and Programme for International Student Assessment (PISA), is around one third of a standard deviation (Rodriguez, 2004) and the average annual education expenditure is around \$8,000 per student. Although crude, this estimate provides a framework for evaluating reform efforts in education.

Class size reduction programs look only moderately effective by these standards, since they fail on the third criterion of affordability. A 30% reduction in class size appears to be

associated with an increase of 0.1 standard deviations per student (Jepsen & Rivkin, 2002). So for a group of 60 students, providing three teachers instead of two would increase annual salary costs by 50%. Assuming costs of around \$60,000 per teacher (to simplify the calculation, we do not consider facilities costs), this works out to \$1,000 per student for a 0.1 standard deviation improvement. This example illustrates the way that one-off costs, such as investing in teacher professional development, can show a significant advantage over recurrent costs such as reducing class size.

Even here, however, caution is necessary. We need to make sure that our investments in teacher professional development are focused on those aspects of teacher competence that make a difference to student learning, and here, the research data are instructive. Hill, Rowan, and Ball (2005) found that a one standard deviation increase in what they called teachers' mathematical knowledge for teaching was associated with a 4% increase in the rate of student learning. Although this was a significant effect, and greater than the impact of demographic factors such as socioeconomic status, it is a small effect—equivalent to an effect size of less than 0.02 standard deviations per student. It is against this backdrop that the research on formative assessment, or AfL, provides such a compelling guide for action.

Research on Formative Assessment

The term *formative assessment* appears to have been coined by Bloom (1969) who applied Michael Scriven's distinction between formative and summative program evaluation (Scriven, 1967) to the assessment of individual students. Throughout the 1980s in the United Kingdom, a number of innovations explored the use of assessment during, rather than at the end of instruction, in order to adjust teaching to meet student needs (Black, 1986; Brown, 1983). Within 2 years, two important reviews of the research about the impact of assessment practices on students had appeared. The first, by Gary Natriello (1987), used a model of the assessment cycle beginning with purposes and moving on to the setting of tasks, criteria, and standards; evaluating performance; and providing feedback. His main conclusion was that most of the research he cited conflated key distinctions (e.g., the quality and quantity of feedback), and was thus largely irrelevant. The second review, by Terry Crooks (1988), focused exclusively on the impact of assessment practices on students and concluded that the summative function of assessment had been dominant, which meant that the potential of classroom assessments to assist learning had been inadequately explored. Black and Wiliam (1998a) updated the reviews by Natriello and Crooks and concluded that effective use of classroom assessment could yield improvements in student achievement between 0.4 and 0.7 standard deviations, although that review did not explore in any depth the issue of the sensitivity to instruction of different tests (see Black & Wiliam, 2007, for more on this point).

A subsequent intervention study (Black, Harrison, Lee, Marshall, & Wiliam, 2003) involved 24 math and science teachers who were given professional development training designed to get them to use more formative assessment in their everyday teaching. With student outcomes measured on externally mandated standardized tests, this study found a mean impact of around 0.34 standard deviations sustained over a year, at a cost of around \$8,000 per teacher (Wiliam, Lee, Harrison, & Black, 2004). Other small-scale replications (Clymer & Wiliam, 2006/2007; Hayes, 2003) have found smaller, but still appreciable effects, in the range of 0.2 to 0.3 standard deviations, but even these suggest that the cost-benefit ratio for formative assessment is several times greater than for other interventions.

It is important to clarify that the vision of formative assessment utilized in these studies involved more than adding extra assessment events to the flow of teaching and learning. In a classroom where assessment is used with the primary function of supporting learning, the divide between instruction and assessment becomes blurred. Everything students do, such as conversing in groups, completing seatwork, answering questions, asking questions, working on projects, handing in homework assignments—even sitting silently and looking confused—is a potential source of information about what they do and do not understand. The teacher who is consciously using assessment to support learning takes in this information, analyzes it, and makes instructional decisions that address the understandings and misunderstandings that are revealed. In this approach, assessment is no longer understood to be a thing or an event (such as a test or a quiz); rather, it becomes an ongoing, cyclical process that is woven into the minute-to-minute and day-by-day life of the classroom.

The effects of the intervention were also much more than the addition of a few new routines to existing practices. In many ways, the changes amounted to a complete re-negotiation of what Guy Brousseau (1984) termed the *didactic contract* (what we have come to call the *classroom contract* in our work with teachers)—the complex network of shared understandings and agreed ways of working that teachers and students arrive at in classrooms. A detailed description of the changes that occurred can be found in Black and Wiliam (2006). For the purposes of this report, the most important are summarized briefly below.

The first change is a change in the teacher's role from a focus on teaching to a focus on learning. As one teacher said, "There was a definite transition at some point, from focusing on what I was putting into the process, to what the pupils were contributing. It became obvious that one way to make a significant sustainable change was to get the pupils doing more of the thinking" (Black & Wiliam, 2006, p. 86). The key realization here is that teachers cannot create learning—only learners can do that. What teachers can do is to create the situations in which students learn. The teacher's task therefore moves away from delivering learning to the student and towards the creation of situations in which students learn; in other words, engineering learning environments, similar to Perrenoud's (1998) notion of regulation of the learning environment. For a fuller discussion on the teacher's role in engineering and regulation, see Wiliam (2007) and Wiliam and Thompson (2007).

The second change is a change in the student's role from receptivity to activity. A common theme in teachers' reflections on the changes in their students was the increase in student responsibility: "They feel that the pressure to succeed in tests is being replaced by the need to understand the work that has been covered and the test is just an assessment along the way of what needs more work and what seems to be fine" (Black & Wiliam, 2006, p. 91).

The final change is a change in the student-teacher relationship from adversaries to collaborators. Many of the teachers commented that their relationship with the students changed. Whereas previously the teacher had been seen as an adversary who might or might not award a good grade, increasingly classrooms focused on mutual endeavor centered on helping the student achieve the highest possible standard.

The changes described above were achieved through having the teachers work directly with the original developers of the intervention. In order to take any idea to scale, it is necessary to be much more explicit about the important elements of the intervention, and this makes clear communication paramount. In the United States, reform efforts around formative assessment face a severe problem, due to the use of the term *formative assessment* (and, more recently, *AfL*) to denote any use of assessment to support instruction in any way. In order to clarify the

meanings, we have expended much effort, over a considerable period of time, in simplifying, clarifying, and communicating what, exactly, we mean by AfL or formative assessment. In this process, our original views about what kinds of practices do and do not constitute formative assessment have not changed much at all, but our ways of describing them have.

The central idea of formative assessment, or AfL, is that evidence of student learning is used to adjust instruction to better meet student learning needs. However, this definition would also include the use of tests at the end of learning that are scored, with students gaining low scores being required to attend additional instruction (for example on Saturday mornings). While such usages may, technically, conform to the definition of the term *formative*, the evidence that supports such practices is very limited. For that reason, within KLT, the key principle—known as the Big Idea—is expressed as follows:

Students and teachers
Using evidence of learning
To adapt teaching and learning
To meet immediate learning needs
Minute-to-minute and day-by-day

Of course, while such a formulation helps clarify what is not intended, it provides little guidance to the teacher. In unpacking this notion, we have found it helpful to focus on three key questions, derived from Ramaprasad (1983):

- Where the learner is going
- Where the learner is right now
- How to get there

There is nothing original in such a formulation of course, but by considering separately the roles of the teacher, peers, and the learner, it is possible to unpack the big idea of formative assessment into five key strategies, as shown in Table 1.

The empirical research base behind each of these five strategies is extensive and beyond the scope of this paper. See Wiliam (2007) for a fairly exhaustive treatment.

The five strategies certainly bring the ideas of AfL closer to the point where they are of practical use, but through our work with U.S. teachers, we came to understand that these generic strategies offer a necessary but still insufficient framework. The reasons for this are complex, and relate to the difference between *know-how* (craft knowledge, or technique) and *know-why* (knowledge of universal truths). For a fuller discussion of this contrast, see Wiliam (2003). We argue here that the scalability of a complex intervention requires both, because helping teachers achieve know-why empowers them to make implementation decisions that enhance, rather than detract from, the theory of action. However, exclusive attention to the know-why does not answer teachers' need for know-how. As one of us (Wiliam, 2003, p. 482) has written earlier:

The kinds of prescriptions given by educational research to practice have been in the form of generalized principles that may often, even usually, be right, but in some circumstances are just plain wrong. ... But more often research findings also run afoul of the opposite problem: that of insufficient specificity. Many teachers complain that the findings from research produce only bland platitudes and are insufficiently contextualized to be used in guiding action in practice. Put simply, research findings underdetermine

action. For example, the research on feedback suggests that task-involving feedback is to be preferred to ego-involving feedback (Kluger & DeNisi, 1996), but what the teacher needs to know is, “Can I say, ‘Well done’ to this student, now?” Moving from the generalized principles produced by educational research to action in the classroom is not a simple process of translation.

Table 1
Deriving the Five Key Strategies of AfL

	Where the learner is going	Where the learner is right now	How to get there
Teacher	Clarifying learning intentions and criteria for success	Engineering effective class-room discussions, questions, and learning tasks that elicit evidence of learning	Providing feedback that moves learners forward
Peer	Understanding learning intentions and criteria for success	Activating students as instructional resources for one another	
Learner	Understanding learning intentions and criteria for success	Activating students as the owners of their own learning	

So, in addition to the theoretical framework provided by the five strategies, teachers also need exposure to a wide range of teaching techniques that manifest the strategies. The techniques represent specific, concrete ways that a teacher might choose to implement one or more of the AfL strategies. Working with researchers and teachers in dozens of schools, we have developed or documented a growing list of techniques that teachers have used to accomplish one or more of the strategies named above. We do not claim to have invented all of these techniques; rather, we have gathered them together within the larger framework of minute-to-minute and day-by day formative assessment. At this point, we have catalogued over 100 techniques, roughly evenly distributed across the five strategies. We expect the list to continue to grow, as teachers and researchers develop additional ones. To give the flavor of the techniques, we describe here just two techniques for each of the five strategies.

Strategy: Clarifying Learning Intentions and Sharing Criteria for Success

Example Technique 1: Sharing exemplars. The teacher shares student work from another class or uses a teacher-made mock-up. The selected exemplars are chosen to represent the qualities that differentiate stronger from weaker work. There is often a discussion of the strengths and weaknesses that can be seen in each sample, to help students internalize the characteristics of high quality work.

Example Technique 2: Thirty-second share. At the end of a class period, several students take a turn to report something they learned in the just-completed lesson. When this is a well

established and valued routine for the class, what students share is usually on target and connected to the learning intentions stated at the start of the lesson. If the sharing is off target, that is a signal to the teacher that the main point of the lesson hasn't been learned or it has been obscured by the lesson activities and needs further work. In classrooms where this technique has become part of the classroom culture, if a student misstates something during the thirty-second share, other students will often correct him or her in a nonthreatening way.

Strategy: Engineering Effective Classroom Discussions, Questions, and Learning Tasks That Elicit Evidence of Learning

Example Technique 1: ABCDE cards. The teacher asks or presents a multiple-choice question and then asks students to simultaneously (on the count of three) hold up one or more cards, labeled A, B, C, D, or E, as their individual response. ABCDE cards can be cheaply made on 4-inch by 6-inch white cardstock printed with one black, bold-print letter per card. A full set might include the letters A–H plus T. This format allows all students to select not only one correct answer, but multiple correct answers or to answer true/false questions. This is an example of an all-class response system that helps the teacher to quickly get a sense of what students know or understand while engaging all students in the class. The teacher may choose to ask the question orally or to present it to the class on an overhead. The teacher then uses the information in the student responses to adapt and organize the ensuing discussion or lesson.

Example Technique 2: Colleague-generated questions. Fellow teachers share and/or write better questions—questions that stimulate higher order thinking and/or reveal misconception—to be used in ordinary classroom discussions or activities. Formulating good questions takes time and thought. It makes sense, then, to share good questions and the responsibility for developing them among a group of colleagues. Once developed, good questions can be reused year after year. Questions may have been previously tried out in one teacher's classroom, or they may be brand new to all, with teachers reporting back on how well they worked. Time to develop questions is sometimes built into a regular schedule (such as team or grade level meetings), or it may have to be specially scheduled from time to time.

Strategy: Providing Feedback That Moves Learners Forward

Example Technique 1: Comment-only marking. The teacher provides only comments—no grades—on student work, in order to get students to focus on how to improve, instead of on their grade or rank in the class. This will more likely pay off if the comments are specific to the qualities of the work, designed to promote thinking, and intended to provide clear guidance on what to do to improve. Consistently writing good comments that make students think is not easy to do, so it is a good idea to practice this technique with other teachers for ideas and feedback. Furthermore, the chance of student follow-through is greatly enhanced if there are established routines and time provided in class for students to revise and improve their work.

Example Technique 2: Plus, minus, equals. The teacher marks student work with a plus, minus, or equals sign to indicate how this performance compares with previous assignments. If the latest assignment is of the same quality as the last, the teacher gives it an equals sign; if the assignment is better than the last one, he or she gives it a plus sign; and if the assignment is not as good as the last one, he or she gives it a minus sign. This technique can be modified for younger students by using up and down arrows. There should be well-established routines around this kind of marking, so that students can use it formatively to think about and improve their progress.

Strategy: Activating Students as the Owners of Their Own Learning

Example Technique 1: Traffic lighting. Students mark their own work, notes, or teacher-provided concept lists to identify their level of understanding (green = *I understand*; yellow = *I'm not sure*; red = *I do not understand*). Younger students can simply draw a smiling or frowning face to indicate their level of understanding. The teacher makes colored markers or pencils available, provides instruction on their purpose, and provides practice time, so students know how to use them to code their levels of understanding. It is important that time and structure be allotted for students to get help with the things they do not understand or this technique will simply result in frustration.

Example Technique 2: Learning logs. Near the end of a lesson, students write summaries or reflections explaining what they just learned during the lesson (what they liked best, what they did not understand, what they want to know more about, etc.). Students can periodically hand these in for review, or hand them in at the end of selected lessons. These summaries or reflections may be kept in a notebook, journal, online, or on individual sheets. The teacher, in turn, periodically takes time to analyze them, respond, and, based on the information in them, perhaps modify or adapt future instruction. Students may also review their own learning logs to take stock of what they have learned over time and also to note areas of continuing interest or difficulty.

Strategy: Activating Students as Instructional Resources for One Another

Example Technique 1: Peer assessment with a preflight checklist or rubric. Students trade papers and check each other's work against a preflight checklist or rubric to improve the quality of the work they submit to the teacher. To close the feedback loop, there should be clear structures for when and how students are to take this feedback on board to improve their work. A preflight checklist is a list of the required, basic components for an assignment, such as "title page, introduction, five-paragraph explanation, conclusion." The preflight checklist differs from a full-fledged rubric in that it is used primarily to check that all the required components are present, whereas a rubric is more likely to get into the quality of those components. Some checklists and rubrics will be generic—applicable to many assignments. Others may be specific to a particular assignment. Whether a checklist or rubric is used, peers should be taught to provide accurate feedback. Note that students should not provide grades of any kind, just feedback.

Example Technique 2: Homework helpboard. Students identify homework questions they struggled with, put them on the board, and solve them for one another. As students enter the classroom, they write the problem number or other identifier for homework questions that they could not figure out in a predesignated section of the blackboard. At the same time, they and classmates who succeeded at any of the identified problems show their solutions on the board, with minimal involvement from the teacher. This technique results in an efficient review of homework that is targeted to the areas of difficulty. The teacher need only assist on those problems that no one else can solve, and even then, this may only require the teacher to ask an appropriate question, offer a suggestion, or begin a solution—then the students can take over.

These 10 techniques represent less than 10% of the techniques now catalogued in the KLT program. See Leahy, Lyon, Thompson, and Wiliam (2005) for descriptions of additional techniques used by U.S. teachers in enacting A/JL. Wiliam (2007) listed several more and went into a great deal of detail on the empirical basis behind many specific techniques.

All the techniques are decidedly low technology, low cost, and usually within the capabilities of individual teachers to implement. In this way, they differ dramatically from large-scale interventions such as class-size reduction or curriculum overhauls, which can be quite expensive and difficult to implement because they require school or system level changes. Most of the techniques do not, in themselves, require massive changes in practice. Nevertheless, the research shows that these small changes in the flow of instruction can lead to big changes in student learning (Black et al., 2003; Leahy et al., 2005).

Finally, it is important to point out that the instantiation of any one of the strategies in a particular set of teaching techniques can vary substantially from grade to grade, subject to subject, and even teacher to teacher. For example, a self-assessment technique that works for middle grade math teachers may not work well at all in a second grade writing lesson. It is even true that what works for seventh grade pre-algebra in one classroom may not work for seventh grade pre-algebra in another classroom, even if it is right down the hall—because of student or teacher differences. Given this variation, it is important to provide teachers with multiple techniques and to give them scope to customize these techniques to meet the needs of their students, subject matter, and teaching style.

To make it easier for teachers to see the relationships among the Big Idea, Five Key Strategies, and 100+ practical techniques, see Figure 1.

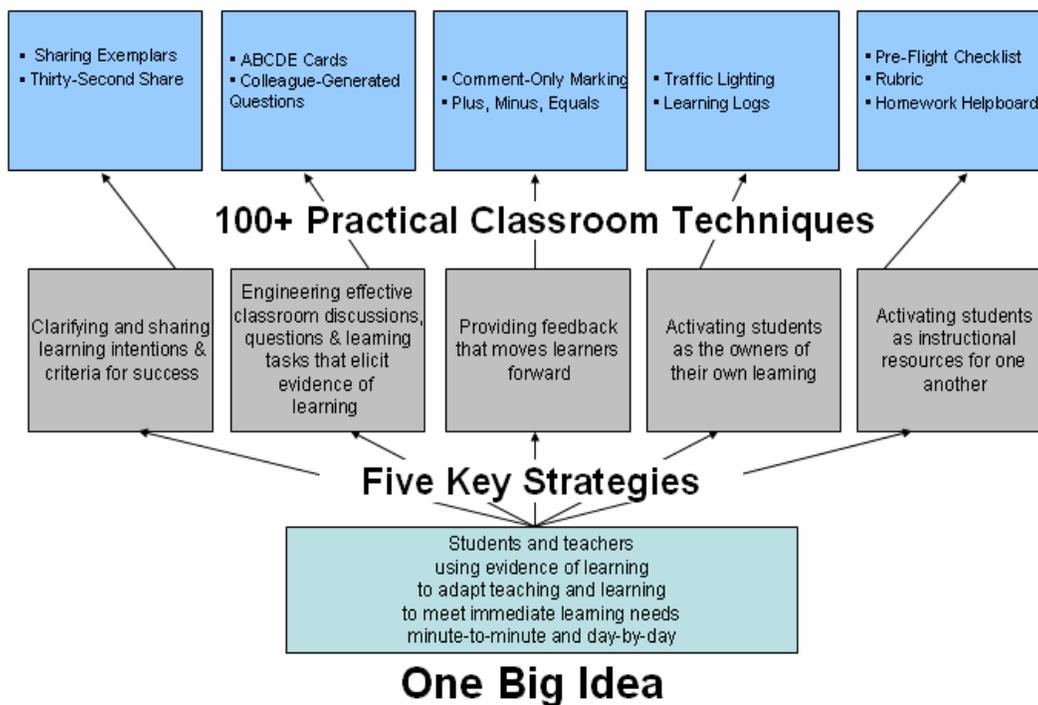


Figure 1. How the Big Idea, Five Key Strategies, and practical techniques of KLT are related.

The How: Sustained, School-Based, Collaborative Professional Learning

For teachers to take on wholeheartedly the new roles and new paradigms that minute-to-minute and day-by-day A/L requires of them, they need more than just a quick exposure to its principles and methods. Many teachers have a great deal of the required knowledge and skills to

understand and implement the AfL strategies once they are exposed to these ideas, but they need sustained opportunities to consciously develop, practice, reflect upon, and refine this skill set so that it works in the context of their own classrooms. Mandated state standards, testing, pacing guides, and scripted curricula have left many teachers feeling divorced from goal setting and assessment—core practices in AfL. These skills have atrophied in teachers who feel their role in establishing goals and measuring progress toward learning has been pre-empted. Besides opportunities for learning, practice, and reflection, these teachers also need experiences that explicitly counteract the isolation, frustration, and deprofessionalization that have occurred in many school faculties.

Without effective professional development systems to teach teachers to do AfL, the potential of the intervention will never be realized. By effective, we mean that the professional development leads to observable, measurable improvements in teaching practice, a requisite step toward improving student learning. The sad truth is that most professional development is not effective, by this definition (Garet, Porter, Desimone, Birman, & Yoon, 2001). The challenge is to develop models of professional development and scalable systems of delivery that faithfully disseminate the content of AfL, while also providing sustained, meaningful assistance to teachers who are attempting to replace long-standing habituated practices with more effective ones.

Two Phases of Professional Learning

In response to these challenges, KLT supports two distinct phases of professional learning: (a) initial exposure and motivation, and (b) ongoing guided learning, practice, reflection, and adjustment. These phases and the research and development process that led to their current structure within KLT are explained in detail in an earlier paper by Thompson and Goe (2006), from which much of the following discussion is drawn.

In Phase 1, teachers and school leaders are exposed to an overview and basic information about AfL through an interactive, two-day workshop or seminar, presented within a motivational framework so that they can see the advantage of making a longer-term commitment to changing practice. Topics covered in the introductory workshop include the following:

- The Big Idea that unifies and drives the five strategies of KLT.
- The five AfL strategies of KLT.
- A sample of the 100-plus teaching techniques, each associated with one or more of the strategies, that teachers can select from and customize to make AfL come alive in their classrooms. In the course of the introductory workshop, participants get direct experience with dozens of the KLT techniques, which are used by the workshop leaders to facilitate the teachers' own learning.
- The nature of teacher expertise: why one-day workshops, or even sequences of workshops, cannot effectively change teaching practice and an introduction to the nuts and bolts of KLT TLCs.

Woven throughout the workshop are presentations on the research base for KLT: how we know AfL and sustained TLCs work to change teacher practices and improve student learning. To motivate interest in AfL as a central component of daily practice, we rely on compelling research that shows the student learning gains that can be obtained by becoming expert at it: primarily the Black and Wiliam 1998 and 2003 studies, but increasingly adding in evidence that is accumulating in the United States (Clymer & Wiliam, 2006/2007; Wylie, Thompson, Lyon, &

Snodgrass, 2007). To motivate the commitment to a years-long learning process, we cite both research and teachers' own experiences with the limited effects of one-off workshops, and then make a logical argument for sustained collegial learning. Participants also develop a personal action plan for taking the first steps in implementing AfL in their own classrooms and are expressly invited to begin Phase 2 by joining an ongoing teacher learning community focused on developing further expertise.

Without Phase 1, most teachers would not know where to begin or even see that they needed to begin. But in fact, it is in Phase 2 that the learning has the potential to actually change teaching, learning, teachers, and schools. Phase 2 represents a guided learning-by-doing stage, the stage where the knowledge learned at an explicit level is translated into tacit knowledge that is accessible and applicable in practice in increasingly transparent ways. Opportunities for practice in real settings followed by reflection have to be structured. Otherwise, the pace of teaching and the daily life in schools do not naturally allow teachers and school leaders to develop expertise in complex interventions such as AfL. Thus, a primary process by which KLT attempts to effect these changes in teachers' practice is via school-embedded teacher learning communities (TLCs). These have the potential to provide teachers with the information and support they need to develop their practice in deep and lasting ways, and they are designed to build school capacity to support individual and institutional change over time.

Developing Teacher Expertise Through Teacher Learning Communities

TLCs embody critical process elements that are needed in order for professional development to result in actual changes in teacher practice. Specifically, effective professional development is related to the local circumstances in which the teachers operate (Cobb, McClain, Lamberg, & Dean, 2003), takes place over a period of time rather than in the form of one-day workshops (Cohen & Hill, 1998), and involves teachers in active, collective participation (Garet et al., 2001; Ingvarson, Meiers, & Beavis, 2005).

There are, of course, many professional development structures that would be consistent with this research base. But we believe that TLCs as advocated in the *Standards for Staff Development* of the National Staff Development Council (NSDC, 2001) under the name *professional learning communities* provide the most appropriate vehicle for helping teachers become skilled practitioners of AfL.

There is a growing body of evidence on ways to build and sustain TLCs (Borko, 1997, 2004; Borko, Mayfield, Marion, Flexer, & Cumbo, 1997; Elmore, 2002; Garmston & Wellman, 1999; Kazemi & Franke, 2003; McLaughlin & Talbert, 1993; 2006; Putnam & Borko, 2000; Sandoval, Deneroff, & Franke, 2002). We note a pattern in this literature—if the practices you are hoping to get teachers to change are recurrent, central, and entrenched within everyday teaching and school culture, then teachers will need sustained support to change them. Not only must the support be sustained over time (at least a year and often much longer in many of the studies cited above), that support must embed teachers' learning within the realities of day-to-day teaching in their own schools and classrooms and allow for repeated cycles of learning, practice, reflection, and adjustment within their native context.

To some extent, these cycles map onto the cyclical depiction of knowledge creation and knowledge transmission of Nonaka and Takeuchi (1995), wherein the tacit knowledge of a person or group is turned into explicit knowledge so that it can be taught to another person or group. Until the new knowledge is practiced and made operational (through a process labeled *learning by doing* in Nonaka and Takeuchi's framework), the knowledge remains explicit. It is

only through sufficient learning by doing (or practice) that the knowledge can be combined with existing knowledge structures, internalized, and made accessible and useful in relatively seamless ways (essentially making the knowledge tacit again). Nonaka & Takeuchi's framework seems particularly apt in this application, because it treats learning as a phenomenon that is situated in a social milieu, which is certainly characteristic of teacher learning.

The seamless, transparent, and highly accessible quality of internalized tacit knowledge is one of the distinguishing features of expertise in any field (Ross, 2006), and AfL is no different. An expert in AfL is able to rapidly note essential details of the complex social and psychological situation of a lesson (especially the state of student learning), while disregarding distracting yet nonessential details. The expert teacher is then able to swiftly compare that situation with the intended goals for the lesson, the teacher's knowledge of the content being taught, the teacher's developmental knowledge of students in general and of these students in particular, and other relevant schema. Guided by the results of these comparisons, the teacher then selects the next instructional moves from a wide array of options—most well-rehearsed, some less familiar, and some invented on the spot—such that these next steps address the students' immediate learning needs in real time.

Such expertise is certainly marked by the speed of cognition, but there is more to it than speed alone. Expert teachers don't just think faster than non-expert teachers; they think and behave in qualitatively different ways. This has been borne out in the work of Berliner (1994), who documented eight ways that expert teachers function like experts in other fields. For example, Berliner notes that expert teachers perceive meaningful patterns where nonexperts cannot, in the domain of their expertise.

The story of how Berliner came to understand this particular feature of expert teaching is instructive and directly related to the need for teachers to practice and reflect upon teaching in real contexts. In the early 1990s, he produced a series of videotapes depicting common teaching problems in staged classroom settings. When he showed these tapes to novice teachers, experienced but nonexpert teachers, and expert teachers, he expected the experts to be able to describe the videotaped interactions in rich detail and provide plausible, nuanced solutions to the problems revealed. Instead, he found that the experts were completely stymied by the videotapes, whereas the novices and other nonexperts were able to converse at length about what they had seen (though not necessarily cogently or plausibly).

Through later conversations with the teachers, he discovered that the staged depictions felt realistic only to nonexperts. There were subtle but essential details of real classroom life that were either absent from the staged depictions or out-of-sync. The nonexperts did not miss or notice these. Without these subtle details in their proper place, however, the experts were thrown off in their search for meaningful patterns—they couldn't even begin to make sense of what they were seeing, because it did not map onto their relatively dense knowledge webs concerning what goes on in teaching and learning in a real classroom. When the staged videos were replaced with videos filmed in real classrooms, the experts were easily able to respond with detailed, nuanced, cogent, and plausible descriptions and prescriptions—even though the technical quality of these spontaneous videos sometimes made it difficult to hear and see all the relevant action.

This same kind of observation and pattern-matching is an integral part of the teacher's role in AfL, as it is for most complex teaching behaviors. Learning to do AfL requires the development of expertise, not the rote application of declarative or procedural knowledge.

To feed the development of teacher expertise in general and expertise in AfL in particular, we sought a learning vehicle that would support the kind of socially supported

knowledge creation and transfer described by Nonaka and Takeuchi (1995) and provide support for the sustained, reflective practice that marks the learning of experts (Ross, 2006). TLCs have potential to accomplish this and to represent a learning model that could be scaled to reach many, many teachers in all kinds of schools.

Though the idea of TLCs is usually warmly greeted by teachers—who generally wish for more collegiality in their professional lives—we did not elect to employ these communities just to be nice to teachers; we chose this vehicle because it is the only one we’ve found that works to change teachers’ practices in the ways that we need them to change. When the practices in question are recurrent, central, and entrenched within school culture, a sustained and school-embedded learning vehicle is needed to counteract the force of old habit. Furthermore, because the kind of teaching we were trying to develop in teachers has all the hallmarks of expertise, we needed a vehicle that could provide support for extended practice, where here we mean the word practice in the sense of piano practice or of chess players playing literally thousands of practice games. As Ross (2006) pointed out, experts practice differently from nonexperts, going well beyond simple repetition of the thing to be learned. Instead, they approach practice systematically and apply critical analysis and reflection to the results of their practice efforts.

We were also attracted to TLCs for KLT because their grassroots character lends itself to scaling up the intervention. As Black, Harrison, Lee, Marshall, and William (2002) noted, few teachers make use of formative assessment in day-to-day teaching. If you accept the notion that it would be good for teachers to do more of this kind of teaching, then the need for professional development is tremendous, and the issue of scalability rises to crucial importance. It is not enough to devise a program of professional development that works effectively when it is delivered by its original developers and their hand-picked expert trainers. Where would we find the army of experts needed in the 100,000-plus U.S. schools that could benefit from AfL? There simply are not enough qualified coaches and workshop leaders to be found, and the mechanisms for disseminating learning through such top-down models are dauntingly complex and expensive. This is not to say that there aren’t serious challenges involved in bringing TLCs to scale with fidelity to the original AfL content, given that we must assume that there will be no experts (at least not at the outset) in any given learning community. This is a design issue that must be faced squarely, by building bootstrapping strategies into the professional learning portion of the intervention (these are described later in this paper).

There are several other ways that TLCs seem to be particularly functional vehicles to support teacher learning about AfL. First, the practice of AfL depends upon a high level of professional judgment on the part of teachers, so it is consistent to build professional development around a teacher-as-local-expert model. Second, school-embedded TLCs are sustained over time, allowing change to occur developmentally, which in turn increases the likelihood of the change sticking at both the individual and school level. Third, TLCs provide a non-threatening venue, allowing teachers to notice weaknesses in their content knowledge and get help with these deficiencies from peers. For example, in discussing an AfL practice that revolves around specific content (e.g., by examining student work that reveals student misconceptions), teachers often confront gaps in their own subject-matter knowledge, which can be remedied in conversations with their colleagues.

In a related vein, TLCs redress a fundamental limitation of AfL, which is its (perhaps paradoxical) generality and specificity. The five AfL strategies are quite general—we have seen each of them in use in pre-K, in graduate-level studies, at every level in between, and across all subjects—and yet implementing them effectively makes significant demands on subject-matter

knowledge. Teachers need strong content knowledge to ask good questions, to interpret the responses of their students, to provide appropriate feedback that focuses on what to do to improve, and to adjust their teaching on the fly, based on the information they are gathering about their students' understanding of the content. A less obvious need for subject-matter knowledge is that teachers need a good overview of the subject matter in order to be clear about the big ideas in a particular domain, so that these are given greater emphasis. TLCs provide a forum for supporting teachers in converting the broad AfL strategies into lived practices within their specific subjects and classrooms.

This bonus feature of TLCs focused on AfL—attention to the development of teacher content knowledge—is certainly a good thing, given well-documented deficiencies in U.S. teachers' preparation and content knowledge for teaching the subjects they teach (Fennema & Franke, 1992; Gitomer, Latham, & Ziomek, 1999; Kilpatrick, 2003; Ma, 1999; National Commission on Mathematics and Science Teaching for the 21st Century, 2000). But it is important to note that the learning communities we describe here are not expressly designed to redress these deficiencies, even as we see evidence of teachers using them to advance their subject matter knowledge when we observe learning community meetings.

This issue of deficiencies in teachers' subject matter knowledge raises a question for the model we describe: are there limits to the effectiveness of TLCs focused on AfL in transforming teacher practice, given pre-existing limits on teachers' subject matter knowledge? We do not have a definitive answer to this question, though we can report that we have repeatedly observed groups of teachers improve their pedagogical practice, even when no teacher in the group has had strong content knowledge. Furthermore, evidence from one school district suggests that the students of these teachers are learning better and faster (Wylie, Thompson, et al., 2007) despite weaknesses in their teachers' content knowledge—because changes in teachers' practices have led to students' changing their own relationship to their learning and the content they are learning about. These results suggest that simply improving teachers' pedagogy works to boost student learning, even in the absence of strong content knowledge on the part of the teacher. Whether or when this effect will top out remains to be seen in later research. (This is not to say that further gains could not be achieved by deliberately focusing on improving current teachers' content knowledge. However, the policy infrastructure and institutional capacity for achieving this goal at scale are not yet in place.)

Finally, TLCs are embedded in the day-to-day realities of teachers' classrooms and schools, and as such provide a time and place where teachers can hear real-life stories from colleagues that show the benefits of adopting these techniques in situations similar to their own. These stories provide *existence proof* that these kinds of changes are feasible with the exact kinds of students that a teacher has in his or her classroom. This contradicts the common lament, "Well, that's all well and good for teachers at *those* schools, but that won't work here with the kinds of students we get at this school." Without that kind of local reassurance, there is little chance that teachers will risk upsetting the prevailing classroom contract. While limiting, the old contract at least allows teachers to maintain some form of order and matches the expectations of most principals and colleagues. As teachers adjust their practice, they are risking both disorder and less-than-accomplished performance on the part of their students and themselves. Being a member of a community of teacher-learners engaged together in a change process provides the support teachers need to take such risks.

Because learning by doing is integral to the development of expertise in the complex realm of AfL, expertise cannot be developed quickly. Furthermore, it can only be developed in

those who have ample opportunity for practice, reflection, and adjustment—teachers. Just as a chess master needed to play a lot of chess to become an expert at chess, a teacher also needs to practice AfL extensively to become expert at it.

While we have not had a chance to make a formal study of learning community leadership, a review of the results we have seen to date suggests that teachers themselves can provide effective leadership for their peers. Because they are going through the same learning and change process, they have essential insights into the pace of change, the kinds of dilemmas faced, and the types of support that make sense, all within the context of the classroom. When we place teachers in the leadership role, we caution them not to assume any extra expertise just because they are facilitators and advocates for their respective TLCs. We find that this is often a relief to them—they do not want to hold themselves above their peers or feel pressured to act as if they know more than they do. They want the breathing room to ask for help in the places where they are struggling.

Coaching by curriculum coaches or building principals is a common model for school-based professional development. However, it is limited in its utility for advancing AfL, unless the coach has previously developed expertise in AfL through using and refining it in his or her own classroom. Since we know that such expertise is rare in classroom teachers, we have reason to believe that few coaches and principals could have developed this expertise before they left the classroom. This is not to say that coaches and principals can play no useful role in supporting TLCs focused on AfL, but we do think they should refrain from holding themselves up as experts, unless they have “walked the walk” (and we are also aware that many coaches and principles believe that they were implementing AfL effectively in their own classrooms when they were teaching, even though it is clear that they were not).

We think the notion of legitimate peripheral participant developed by Lave and Wenger (1991) is useful here. In describing the idea of a community of practice, Lave and Wenger described the role of apprentices as legitimate peripheral participants. While Lave and Wenger resisted the decomposition of this term into its constituent elements and the idea that the term should be understood in terms of its antonyms (e.g., illegitimate, central, non-participants), they saw the term *peripheral* as a positive term that “places the emphasis on what the partial participant is not” (p. 37). In many, if not most, of Lave and Wenger’s examples, the implication is that peripheral participation will eventually lead to full participation, but this is not necessarily the case. As they suggested, “legitimate peripherality can be a position at the articulation of related communities of practice ... affording ... articulation and interchange among communities of practice” (p. 36). Within TLCs, those who are not attempting to make changes in their own practice can never be full participants in the community—not least because they do not share the same goals. However, provided they recognize and accept their peripherality, they can be of substantial help to the community, brokering ideas, acting as advocates, and facilitating the community’s learning.

TLCs are certainly catching on, with federal and state education policy now moving to acknowledge that these kinds of embedded, teacher-driven, drip feed approaches can be an effective way to shift teacher practice. (See, for example, (Division of Abbott Implementation, 2005; Librera, 2004; North Carolina Department of Public Instruction, 2005; U.S. Department of Education, 2005). We note, however, that implementing TLCs consistently and effectively is not as simple as changing federal and state regulations and funding frameworks for professional development. Significant structural barriers to TLCs exist in many schools, including daily and weekly schedules that provide little or no time with colleagues during the normal school day,

personnel policies and practices that do not recognize or value teacher expertise, local bargaining agreements that discourage teachers from meeting outside scheduled hours, inadequate resources to support teacher time away from the classroom, competing demands on teacher time, and school cultures that do not easily align with the needs of sustained, school-embedded, collegial work with colleagues.

Given the steep institutional challenge associated with mounting TLCs, it is important to say that we are not endorsing TLCs as a one-size-fits-all solution for all teacher learning. Rather, we are endorsing a more flexible concept, one of matching the nature of the content to be learned with learning processes that are appropriate for that particular content. Where the content to be learned draws on the kinds of complex cognition and behaviors that are typical of experts, and perhaps contradicts habituated or encultured practices, then we would argue that TLCs provide a suitable, and perhaps a necessary, learning modality.

There are other kinds of teacher learning, however, that are probably best dealt with in other ways. For example, if the goal is to boost teachers' subject matter knowledge, then there are learning vehicles that put teachers in close contact with expert sources (e.g., professors, texts) that may be more efficient than exploring that subject matter with colleagues who are not experts in that realm. If the knowledge to be learned is procedural and highly standardized (for example, learning to use new grading and record-keeping software), then a workshop learning experience will be faster, cheaper, and more likely to result in uniform compliance.

Using Modules to Guide Teacher Learning About AfL

The most popular notions of professional learning communities assert the need for teachers to meet and plan collaboratively, and they generally insist that data occupy a central place in the discussion. DuFour (2004) and McLaughlin and Talbert (2006) do not agree on all aspects of how such learning communities should function, but they both leave it up to the teachers to collectively select the topics that they will focus on and the data they will consider in that discussion. The problem of bootstrapping expertise led us to a decision to provide significantly more guidance on the content and processes of KLT learning communities than is typically the case in the professional learning community literature.

The primary means of providing this guidance is through a set of modules for use by the facilitators of the learning communities. These modules, comprising directions and materials for 90-120 minutes worth of group study, represent an attempt to incorporate sufficient content about the program to make it reasonably likely that fidelity can be maintained, one of the challenges of using TLCs to scale up any intervention. Recognizing that there are no formative assessment experts in most TLCs, each module provides explicit guidance for the conduct of a monthly learning community meeting. Each module contains an agenda, detailed leader notes with guidance for timing and discussion points, and informational and activity handouts that are to be photocopied for the use of participants (to give an idea of the detail in the leader notes, the material for one module typically runs to over 30 pages). KLT offers enough modules to cover 2 years worth of monthly learning community meetings.

Every module begins and ends the same way, with what we have come to call the bookend activities. In order to model and gain the benefit of the assessment for the learning strategy "Clarifying learning intentions and sharing criteria for success," each module begins with a clear statement of the meeting's learning intentions. To model closing the loop on learning intentions, each module ends with a quick look back at these. The group as a whole decides whether the learning intentions were achieved, and if not, plans ways to redress that problem.

Bracketed within the two reviews of the learning intentions are two other recurring, counter-balanced activities. First, every module includes the How's It Going? segment: time for every teacher to report on and ask for feedback and help on his or her most recent experiences trying out AfL techniques. Black et al. (2003) reported that teachers' awareness that they would be asked to report on their most recent efforts was a helpful, even necessary, spur to action. This is balanced by an activity near the end of the module: a segment known as Personal Action Planning. This is a time for teachers to describe on paper their next steps in trying out and refining AfL techniques in their own classrooms. The personal action planning segment includes time to make arrangements to exchange observation time in colleagues' classrooms, or to collaborate with colleagues in other ways, perhaps to generate hinge questions for key concepts or to practice writing formative comments on student work. The expectation conveyed in this segment is that between meetings, teachers will practice AfL techniques in their classrooms. With this kind of between-meeting effort and the support of colleagues, teachers can gain progressively more skill and insight into ways to improve student learning through AfL.

Our goal in repeating these particular opening and closing activities in every meeting is to create a climate and expectation of both support and accountability, which we explicitly refer to as *supportive accountability*. By emphasizing these two concepts together, we hope to convey that ongoing teacher learning is worthy and necessary, that teachers are expected to work on improving their practice on an ongoing basis, and that they will be supported to do so. We believe that accountability is an important and useful tool in any organization. However, many teachers feel alienated from the concept of accountability, due to pervasive, test-heavy accountability measures that are often out of balance with capacity building measures. To recapture the concept of accountability and put it in service of improving teacher effectiveness, we have made the capacity-building component (Elmore, 2002) explicit in the way we structure TLC meetings. This is most apparent in the required sharing and feedback segment (How's It Going?) and the explicit statement of what each teacher is going to commit to practicing next (Personal Action Planning).

In addition to the repeated bookend activities, each module also includes a teacher learning activity that is designed to deepen teachers' knowledge of a particular AfL strategy and introduce one or more associated techniques, illustrated by stories of how real teachers have made this strategy come alive in their classrooms. These learning activities address such topics as planning lessons so that the learning intentions are well understood by students, developing quality questions, using techniques for providing formative feedback, and so on. As a general rule, each module's new learning segment addresses only one of the five strategies or the big idea of KLT, to keep a clear focus for the meeting. This information is usually embedded in group activities that require teachers to reflect on their current practices and figure out how they might adapt newly learned techniques to their own classrooms.

Though the printed bulk of any given module (that is, the agendas, leader notes, and handouts) is mostly taken up by the pages associated with the new learning activity and we have put considerable effort into designing, pilot-testing, and refining these activities, their inclusion in each module is actually of secondary importance. Consistent with the research on effective professional development, our theory of teacher learning prioritizes activities that require teachers to reflect on the details of recent practice and outcomes in their own classrooms, not activities in which they simply hear about and speculate about how it might work in their classrooms. We include the new learning segments to lure teachers to the next meeting, since many teachers don't start out seeing the utility of simply talking with their colleagues about the

details of their practice, much less enjoying this kind of self-exposure. We also advise community leaders to drop the new learning activity in favor of longer, more in-depth attention to the supportive accountability activities of How's It Going and Personal Action Planning if time is limited. At first, many leaders cannot believe that their main job is not coverage of the material (an extension of the pressure they are under in their own classrooms).

Providing Extended Support and Guidance for Learning Community Leaders

Experience in the first districts we worked in taught us about the importance of institutionalizing ongoing support for teacher learning, and the TLC modules were our first step in this direction. These proved to be necessary but insufficient—as it turns out, sustaining TLCs has its own complexity and context-driven peculiarities, much like AfL. The development, pilot testing, and refinement of the AfL modules goes a long way toward ensuring that when TLCs meet, they maintain a strong and faithful focus on helping teachers adopt AfL strategies and techniques. However, we learned that in most districts, the modules by themselves are not enough, because their availability does not ensure that the TLCs will, in fact, convene or survive, in the face of numerous structural and cultural barriers that must be overcome.

For just as there is an implicit classroom contract, there is an implicit school contract (only partially represented in the text of any bargaining agreement), and that contract has not, historically, supported teachers to be learners. We therefore began looking for ways to provide ongoing support to develop the expertise of TLC leaders and the school leaders who must help them carve out space and time. Eventually we developed a set of supports for the leaders of the learning communities that parallels the two-phase approach taken with regard to AfL, where the content focus was the institutionalization of TLCs. That is, we developed an initial two-day immersion workshop and ongoing embedded support for learning community leaders.

A critical problem facing those who would like to establish TLCs is the lack of time within regular school hours when teachers can meet to discuss teaching and learning or to observe in each other's classrooms. Without this time during the regular paid day, learning communities can never hope to be attractive to the vast majority of teachers. Learning community leaders and allies thus have to work together to communicate and demonstrate that the learning community is a priority, and to do this they need two things. First, they need some level of knowledge of the research base supporting TLCs, to use as leverage when arguing for time and resources. Much of this knowledge base exists in an explicit form that can be conveyed in the initial exposure workshop, backed up by printed reference materials.

What learning community leaders need on top of that are ongoing, structured opportunities for new learning, practice, reflection, and adjustment—about leadership of learning communities. Without this ongoing support, they will not be able to facilitate meetings that deliver high value to the participants, and motivation to participate will decline. In one district's implementation (described in Wylie, Thompson, et al., 2007), there is a required, monthly, day-long workshop for all the TLC leaders in the district. In other districts, the support meetings for TLC leaders are less frequent or not as long, but the gist is the same. In these meetings, the leaders function in two roles successively: first, as teachers learning about applying AfL in their own classrooms, and second as advocates and facilitators for TLCs.

For each role, there is time set aside to reflect on successes and challenges and to plan for next steps. The general framework employs the same principle of supportive accountability that is used in the modules, which is often expressed as push back from colleagues or the leader of the meeting: gentle challenges to explain further or make a direct connection to the theory of

action of AfL. A series of activities to facilitate this level of critical analysis has been developed (though not yet codified to the degree that the modules are), including school action planning forms and protocols for getting a learning community off the ground or for sustaining it once it is up and running; a protocol for reviewing the next module in a sequence; exercises to develop coaching skills for facilitators; and exercises for evaluating the quality of the How's It Going and Personal Action Planning segments of the meetings.

But the toughest challenge lies in developing the critical or analytic abilities of the leaders. Thus, a good deal of any meeting with TLC leaders is spent in a kind of whole-group How's It Going session, in which the leaders report on their own experiences using AfL. A KLT expert leads the process, taking pains to model a level of critical push back that demands that teachers connect their stories directly to the Big Idea and the Five Key Strategies and reflect critically about what is or is not working, from within that framework.

One of the first problems encountered in these sessions is a kind of vagueness that pervades teachers' first expressions of what it is they have been doing in the name of AfL. This vagueness often masks shaky understanding—not only of the overall theory, but even of the specifics of how and why one would choose to implement a particular technique. It is not uncommon, for example, to hear teachers state that they are using a technique, and on further questioning it turns out they are doing nothing of the kind. (They may, in fact, be using a different AfL technique fairly ably or weakly, not doing anything at all, or implementing a technique that stems from a different reform altogether, one that has surface features that reminded them of AfL.) So the KLT participant who is leading the session will gently but firmly probe as much as is needed to fully understand in detail exactly how the teacher is implementing the technique in question, with questions like: “Which types of lessons do you use it in? When in the lesson do you do this? How exactly do you select the students involved? When exactly is it that you take time to parse the students' responses? What kinds of changes to your instruction do you make when you get that kind of response?” The tone and duration of this questioning is reminiscent of the AfL technique called *hot seat questioning*.

Once the exact nature of what is being done is made clear, the leader can then move on to questions concerning the effectiveness of the technique, beginning with clarifying why, exactly, the teacher is using this particular technique: “What exactly are you trying to achieve, and which of the five strategies of AfL does it pertain to?” (Often followed by, “Really? I don't see the connection—tell me why it applies to that strategy more clearly.”) Once the reason for selecting a particular technique has been established: “Do you think it's working as well as you'd like? If yes, what's your evidence? If not, why not?” Where this line of questioning has shown up a problem of practice, it is then appropriate to query the teacher and colleagues for concrete ideas to remedy the problem.

The purpose of all this questioning is certainly not to embarrass the teacher. The purpose is to find out what is really going on, so that the full power of that particular technique's theory of action can be brought to bear on analyzing what's working, what's not working, and what can be adjusted to yield better results. We find that once a learning community leader has experienced this type of interested push back one time (and seen it applied to his or her colleagues), that person comes to the next meeting having already sorted through these preliminaries, able to present a much more concrete and grounded narrative of AfL efforts. In fact, at the next meeting, many arrive with questions or problems about the specifics of their practice—a hallmark of the reflective practitioner.

We also find that the burden of push back starts to be distributed among all the learning community leaders in the room—teachers saying to teachers things like, “Wait a minute, you said you were working on the strategy providing feedback that moves learners forward, but I don’t see feedback in what you just described.” This is one way that the collective expertise of the group starts to develop. It also develops simply from hearing the stories of each teacher’s practice. It is not at all unusual to see teachers madly scribbling down ideas for their own classrooms while listening to another teacher on the hot seat. If that teacher had not been queried deeply about what he or she was doing, all kinds of good practice would have remained hidden.

Not only does this level of assertive questioning help the learning community leaders tighten up their own thinking about and practice of AfL, it also teaches them that this is what supportive accountability looks like, and that every teacher needs it. By seeing this kind of assertive coaching modeled (and by surviving it themselves), the leaders have a better idea of how to facilitate the crucial How’s It Going? segment of their next meeting.

Of course, the provision of ongoing expert support for teacher learning community leaders begs the question of who is going to lead this kind of work, and that goes directly back to the problem of bootstrapping expertise. But we are now faced with a smaller problem: how to develop a large enough cadre of skilled KLT leaders who are capable of modeling and teaching this kind of expertise to learning community leaders. This is a smaller problem to solve than the problem of placing an expert—right now—into every TLC.

The Why: The Theory of Action of *Keeping Learning on Track*

In the preceding sections, we have described the content and process components (the what and the how) of KLT in considerable detail, interwoven with numerous references to the empirical and theoretical studies that led us to design the intervention in the specific ways that we did. This level of detail is needed to fully understand the intervention, but such detail can also have the effect of obscuring the flow of the intervention’s complete theory of action, or the why. In this section, then, we attempt to state the theory of action in a way that makes its flow transparent and accessible, while highlighting certain essential implications for practice. For, as we mentioned at the outset, understanding the why is not only important to legitimizing our claims to the empirical research base, it is also part of the intervention itself.

As we said earlier, KLT is fundamentally a teacher professional development program. Thus, our overall theory of action reflects the three-step model common to all interventions predicated on teacher professional development: (a) teachers learn about a better way to teach through professional development, (b) teachers adopt the better approach to teaching, (c) student learning is improved because of these improvements in teaching. In the case of KLT, the three-step process looks like this, with three supersteps: (a) teachers learn extensively and deeply about minute-to-minute and day-by-day AfL via an initial workshop and sustained engagement in teacher learning communities (TLCs), (b) teachers make minute-to-minute and day-by-day AfL a central part of their everyday teaching practice, implementing the Big Idea and five strategies of AfL through judiciously chosen practical techniques, and (c) student learning improves as a result of the particular ways in which the teaching is made more responsive to the immediate learning needs of students and the changed classroom contract.

Each of these three supersteps are explicated in turn, expanding on the theory of action of each step.

Superstep A: Teachers Learn Extensively and Deeply About Minute-to-Minute and Day-by-Day AfL Via an Initial Workshop and Sustained Engagement in TLCs

There are three important underlying aspects to this superstep: (a) teachers are learners; (b) there has to be a complete and correct transmission of the knowledge base for minute-to-minute and day-by-day AfL; (c) The nature of the learning required to become proficient at AfL is akin to the development of expertise, and this takes time and structures that support extended, systematic, reflective practice. We argue that deep attention must be paid to these three aspects, or the intervention will fail.

Teachers are learners. Consider first the idea that teachers are learners. Though the phrase is often thrown around as a platitude in educational reform circles, the daily lives of most teachers do not bear many signs that this is so. Teachers are simply not provided the time and structures to be learners. Furthermore, under pressure to cover all the material that might be on the end-of-course or end-of-grade test, in many schools the task of teaching has been reduced solely to the function of curriculum delivery. A teacher's understanding of that curriculum or how to teach it is presumed to be handled by the pacing guides and scripted lessons that are becoming increasingly common. Learning? That's for kids. Teachers just stand and deliver.

But think about it—if we want teachers to make major changes in the way they are teaching, then teachers have to learn about these new ways of teaching. Hence, teachers must be treated as learners, and they must see themselves as learners. If they don't see themselves this way, then there is no hope of them even being open to change, much less doing the hard work of learning anything as complex as minute-to-minute and day-by-day AfL. This notion of teachers as learners has a parallel in the emphasis placed on professional learning as one of the three core criteria for a successful educational breakthrough in Fullan et al. (2006). We note that Fullan et al. developed their three core criteria—precision, personalization, and professional learning—primarily in relation to interventions directed at student learning, but once we accept that teachers are learners too, then these three criteria apply just as well to the portion of our intervention that is directed to adults.

Complete and correct transmission. Fullan et al.'s (2006) notion of precision comes into play in relation to the next key aspect: the need for a complete and correct transmission of the knowledge base for minute-to-minute and day-by-day AfL. This second aspect may seem self-evident, but the development, refinement, extension, and documentation of the knowledge base of KLT has been a non-trivial task, spanning more than a decade of research and development involving dozens of researchers and multiple iterations of design research on two continents—and it is still under development as the intervention is worked in yet more kinds of classrooms, subjects, and so on. Nor was the development of reliable methods of transmission non-trivial; this task is still in process 8 years from the earliest efforts in Britain, primarily because we have taken the issue of scalability to heart.

And even where the knowledge base and the transmission methods are fully developed, there are still threats to complete and correct transmission—static, if you will—in the form of disruption to workshops or learning communities, problems with attendance at these events, weaknesses in the performance of trainers and leaders as they are learning their jobs, etc. Minimizing these forms of static has to be a key concern of implementers, or the essential ingredient of precision is lost.

In practical terms, we find that one of the greatest threats to the complete and accurate transmission of the knowledge base is that teachers' time and attention is split across too many reforms at a time. It is not unusual for teachers to tell us that they would like to attend the entire

introductory workshop or the learning community meetings, but they cannot because they are required to attend an event related to another initiative. Even when they find the time to attend formal learning events, teachers' ability to focus on the ideas of AfL or the details of its implementation is often fractured by the pull of too many new reforms jostling for attention.

A key issue in a complete and correct transmission of any knowledge base has to be its coherence and manageability—its digestibility, so to speak. Minute-to-minute and day-by-day AfL may hinge on one big idea and five simple-sounding strategies, but it has over 100 techniques to choose from, draws on research and theory from multiple fields, and, when working, is daily brought to bear on many different kinds of classroom transactions. The organizational schema of the big idea, the five strategies, and the practical techniques is our attempt to provide a memorable, manageable framework onto which the teacher's growing understanding of AfL can be pinned and referenced. Nevertheless, we have to admit that the knowledge base is large, complex, and at first unwieldy, especially for a profession in which many have been treated as if they are unable to handle anything more complex than a script. It takes time and practice for teachers to internalize the AfL framework and organize the details of their thinking and experience along the new lines, which leads us directly to the third key idea of this superstep.

Time and structures to support the development of expertise. We have discussed the notion of expertise and its development at some length in earlier sections, so here we want to focus on the mechanisms of two specific components we believe lead directly to teachers' acquisition of expertise in AfL. First, let's look more closely at the How's It Going segment of a KLT learning community meeting. For this segment to have its intended impact, every teacher has to know—going into the meeting—that they are expected to take a turn, with no exceptions. While there is no actual enforcement mechanism (that is, no punishments or external rewards), the idea is to create a climate in the meetings that makes the expectation clear. The requirement that every teacher participate does not always go down easily in the United States, where it simply is not customary for teachers' practice to be exposed to others, unless it is done voluntarily. But, as Black et al. (2003) noticed, and has been confirmed in U.S. research on this model of professional development (Lyon, Wylie, & Goe, 2006; Thompson & Goe, 2006), for teachers, knowing they will be expected to report on their most recent efforts is often what motivates them to follow through on trying out some aspect of AfL. So one function of the How's It Going? segment is to provide a spur to practice.

Another spur to practice comes in the Personal Action Planning segment at the end of every learning community meeting. In this segment, teachers do a brief think-aloud with a partner about their next steps in instituting AfL in their teaching, and then write down what it is they are going to do in the coming month. Interviews with participants confirm that this simple act of writing down a statement of what they will do in the coming month serves as a significant impetus in actually doing it, even when no one else is following up with them about it. So again, this is a mechanism that leads teachers to practice what they are learning about. And, as the literature on expertise makes clear, practice is an essential ingredient in the development of expertise.

But that same literature also makes clear that practice without purpose or thought does not lead to expertise. When teachers get to practice in a purposeful, reflective way, the result is the development of the kind of seamless expertise described earlier. Each teacher's expert cognitive structures are unique, and each teacher follows his or her unique practice pathway to develop these structures. This is where Fullan et al.'s (2006) notion of personalization seems

relevant. But without purpose or reflectivity, what results is neither personalized nor likely to lead to expertise.

So another purpose of the How's It Going segment and to some extent the questions teachers answer in their personal action plans is to provide tools and support for critical reflection upon practice. There is a sizable literature on reflective teaching, but there is little evidence that most teachers engage in it very often or very deeply (Rodgers, 2002). So, teaching teachers how to reflect critically about their own or others' practice, then, has to be part of what goes on in a TLC meeting. Otherwise, we tend to see cycles of polite, serial turn-taking in the How's It Going? segments, an approach that will not promote deep learning, and in fact debases the content, such that it becomes just a collection of random tricks.

We cannot emphasize how important this aspect of the program is, nor how difficult it is to achieve it, due to the bootstrapping expertise problem discussed earlier. Here we are talking about two distinct aspects of expertise: a critical habit of mind coupled with a clear enough understanding of the theory of action. In many places, there are very few people who come to the process already equipped with these. This is not a problem for KLT alone—it's a central problem in institutionalizing any worthwhile reform. All the simple educational reforms have been instituted, their gains harvested, their cumulative effects found to be insufficient. We must now grapple with more durable organizational and learning problems that evade simple solutions. For these we need interventions that attend to the highly variable complexity of classrooms, schools, and districts (Hendry, 1996; Schein, 1996). We can't solve these complex problems with dumbed-down solutions. We have to create expertise—in every school and classroom—that can handle this complexity. This takes time and targeted capacity building.

Implications for practice. The most important implications for practice of this superstep in the theory of action are as follows:

- Schools must treat teachers as learners (by providing suitable learning environments), so that teachers will begin to see themselves that way.
- Threats to the complete and correct transmission of the KLT knowledge base, including the problem of competition from other initiatives, must be detected and minimized.
- It takes time to develop expertise in AfL, so teachers and the intervention itself must be given time to for this to happen. A school or district that is not willing to make a 2-year commitment to this process is doomed to fail, so they should not even bother.
- That commitment should include structures that give teachers time to meet where they focus on nothing but minute-to-minute and day-by-day AfL.
- Meetings need to last long enough that every teacher gets focused, individual attention for the specific problems he or she is facing in the classroom.
- Attendance at these meetings is necessary but insufficient—merely showing up does not cut it. Rather, active and full participation is a requirement. Everyone must report on his or her latest AfL efforts. Reports and feedback must be detailed and bend toward a rigorous application of the theory of action of AfL. And all participants must commit—in writing—to their next steps in instituting AfL in their classrooms.
- There is an unavoidable upper limit to the number of teachers in a TLC, because if there are too many members, there simply will not be enough time for each teacher to get individual attention for his or her unique problems in instituting AfL.

- Since the essential parts of a TLC meeting are the reflective self-reports, the supportive accountability shown by colleagues, and a written commitment to take specific next steps, any structure that supplies these will do. In other words, it is not necessary to meet in a KLT-sanctioned learning community that follows a KLT module. Other formats will work as well, as long as these essential ingredients are present.

Superstep B: Teachers Make Minute-to-Minute and Day-by-Day AfL a Central Part of Their Everyday Teaching Practice, Implementing the Big Idea and Five Strategies of AfL Through Judiciously Chosen Practical Techniques

It is our contention that if all the aspects of Superstep A are implemented with fidelity, then teachers will indeed make minute-to-minute and day-by-day AfL central in their teaching practice. We would have to concede that the KLT intervention would be a failure if that didn't happen. So at first it may seem silly to include this as an explicit step. Unfortunately, however, this is the step that is most frequently missed when professional development reforms fail to lead to hoped-for improvements in student learning. As Fullan (1991) noted, "Nothing has promised so much and has been so frustratingly wasteful as the thousands of workshops and conferences that led to no significant change in practice when teachers returned to their classrooms" (p. 315). So it is worth noting it as a critical link, and examining it more closely.

There are two aspects to this step in our theory of action: (a) The paired notions that AfL becomes central to practice (centrality) and is used minute-to-minute and day-by-day (frequency), and (b) the idea that teachers judiciously choose the techniques they use. As earlier, we argue that deep attention must be paid to these aspects, or the intervention will fail.

Frequency and centrality. Let's examine the first part of the second step in our theory of action: "Teachers make minute-to-minute and day-by-day AfL a central part of their everyday teaching practice." The notions of frequency and centrality loom large here. In essence, this phrase says that teachers do AfL a lot. Not just every once in a while—say, when their principal comes in to observe them or for special lessons. It says they do it every day, minute-to-minute even. That is asking a lot of teachers, and the only way teachers can achieve it is if their entire approach to teaching has been changed so that AfL is second nature and interwoven into everything they do. And this can only be achieved through a sustained program of support such as that outlined in the first step in the theory of action.

While it is difficult to achieve a change of this magnitude, it is still worth doing, precisely because of these twin notions of frequency and centrality. Simply, the potential for impact is magnified hugely if the lever for change is used often. This is one of the criticisms we have of the use of quarterly benchmark assessments as a form of formative assessment. Even if all the structures for making sensible, formative use of these data are in place, it only happens three or four times a year! This kind of long-cycle formative assessment (William & Thompson, 2007) just does not provide enough information on enough days to have much of a chance to make a difference in students' learning. It is reasonably good at helping teachers make larger decisions about things such as curricular emphases, but it is useless when it comes to knowing what to do in the next day, hour, or minute. So, it may not be a bad investment, but it certainly is not one with the big payoffs that are so often quoted from Black and William's 1998b meta-analysis on the effects of formative assessment.

Minute-to-minute and day-by-day AfL, in contrast, works its magic partly by dint of sheer frequency. By making the learning environment more responsive to students' learning needs on a more frequent basis, tangible support for students' learning increases as well.

Furthermore, when implemented in a dense way, such that it really does become central to practice, AfL is an intervention that changes the entire classroom culture to one focused on learning (Black et al., 2003). This in turn enhances the speed and depth with which students learn.

Judicious choice of practical techniques. Let's talk now about the second part of this step in the theory of action: "implementing the big idea and five strategies of AfL through judiciously chosen practical techniques," with special attention to the final words, "judiciously chosen practical techniques." The operative notion here is that teachers must choose the specific techniques that will work for them, given their teaching style, students, and curriculum. One of the breakthroughs, we think, of the framework of AfL expressed in KLT is that it recognizes the variability and complexity of classrooms, and leaves it to teachers to apply professional judgment about what will work. But, though we never tell teachers what to do, we do two complementary things that increase the likelihood that the intervention will actually be put into practice: (a) we require teachers to take professional responsibility for their classrooms in regard to each of the five strategies; and (b) we provide teachers with a large set of classroom-tested techniques that they can use to make those strategies come alive. This is a delicate balance that moves abstract ideas derived from research—like feedback that moves learners forward—into practical, day-to-day techniques for teachers.

Implications for practice. The most important implications for practice in this step of the theory of action are as follows:

- If a sincere effort has been made to implement the key aspects of the first step of the theory of action, and teachers are not yet showing changes in practice, then program leaders should stand back and examine what is getting in the way. It just does not pay to keep doing the same thing if the professional development is not leading to actual changes in teaching practice.
- Anything that gets in the way of teachers' daily and deep use of AfL will undercut the effectiveness of the intervention. (A frequent source of conflict is the imposition of mandatory or perceived-as-mandatory pacing guides or scripted lessons.)
- Any attempt to dictate which techniques teachers should use undercuts the delicate balance between professional discretion and professional accountability for improving practice so that it is more responsive to students' learning needs.

Superstep C: Student Learning Improves as a Result of the Particular Ways in Which the Teaching Is Made More Responsive to the Immediate Learning Needs of Students and a Resulting Shift in the Classroom Contract

There are two key aspects to the third step in the theory of action: (a) the particular ways in which teaching is made more responsive to the immediate learning needs of students, and (b) the changed classroom contract. As before, we argue that deep attention must be paid to both these aspects, or the intervention will fail.

The particular ways in which teaching is made more responsive to the learning needs of students. There is not scope in this paper to delve into the specific mechanisms of each of the 100-plus AfL techniques represented within KLT, though Wiliam (2007) goes quite a way in this direction. But since we are making the case that theory of action matters deeply, we will explicate the general mechanisms for each of the five strategies and illustrate each with an explication of one or more techniques within that strategy.

The strategy, “Clarifying learning intentions and sharing criteria for success” works via two distinct paths. First, by making the learning intentions of a lesson known to students, students become more engaged and are more apt to bring the right intellectual skills to bear on the learning activities in which they will be participating. That is, once students know the point of a lesson, they are able to deploy their own meta-cognitive abilities in the instructional mix, as well as the value of their personal interest in the topic. When students don’t know where they are headed, they have to be slowly and painfully guided through each step of the learning activities, a task that can feel Herculean to any teacher faced with 25 students staring blankly when asked “What do you think comes next?” Second, when students understand success criteria (that is, what a successful performance/paper/ response looks like), they can begin to internalize what quality looks like. This in turn, can lead students to produce first efforts that are more on target and second efforts that show meaningful refinements.

One Sharing Exemplars technique discussed earlier involves displaying samples of student work—say, for example, a lab report— from previous years (names removed of course) that show different levels of accomplishment, and then having students discuss in small groups the differences among the papers. Another class of techniques, which includes the Thirty-Second Share technique, makes sure that each lesson closes out with a look back at the stated learning intentions, so that students have an opportunity to draw meaning from the learning activities they have just completed.

The strategy, “Engineering effective classroom discussions, questions, and learning tasks that elicit evidence of learning” is fundamental to the idea of using information about student learning to adjust instruction in real time. In simpler terms, the point of this strategy is to make as many students as possible think as deeply as possible as often as possible, and to get evidence of that thinking as often as possible. There are several mechanisms at work here. The first is engaging more students, and several techniques are geared to doing exactly this. One technique in this category involves the use of wipe-on, wipe off markers on homemade whiteboards. Every student has such a whiteboard, and holds it up to show answers to questions asked by the teacher. This requires every student in the class to show his or her thinking, not just the students in the front row who always raise their hands. Using Popsicle sticks or another randomization tool to select individual responders puts everyone in the pool of possible responders, and thus also boosts engagement.

The second mechanism involves pushing students to think deeply more of the time, since learning does not happen except when there is thinking going on (Bransford, Brown, & Cocking, 1999). The techniques that attempt to effect this mechanism take longer to learn than some of the others. Writing better questions by working with colleagues is a long-term collaborative process that begins with distilling the key concepts to be learned and then figuring out the best formats and phrasings to stimulate higher order thinking. Using Wait Time and Think Time can yield deeper thinking from students, but only if the questions asked really have a higher cognitive demand. Hot Seat Questioning, where a single student is queried at some length and in some detail to get at higher order thinking, brings with it the pedagogical problem of ensuring that the other students don’t tune out; we have seen teachers join this with other techniques to make sure everyone has reason to think they may be called on next to expand on or explicate the thinking of the student on the hot seat.

The third mechanism within this strategy involves getting evidence of student learning as often as possible, which goes right to the heart of the Big Idea. The theory of action here is rather an obvious one: without that evidence, the teacher has no idea when or how to adjust

instruction to better meet students' learning needs. His or her next decisions will be taken in the dark, with the result that the teacher may move on before students get it or dwell too long on a topic that everyone already understands. Even if the teacher does not move on at the wrong time, he or she will be in the dark about the exact nature of any misconceptions students harbor if the teacher does not seek out information about what students are thinking. The technique that best illustrates this mechanism is the Hinge Point Question technique, where the teacher develops a single, well-designed question, to be asked at a critical moment in the lesson. The question is deliberately designed ahead of the lesson so that every student's answer can be provided and parsed within no more than two or three minutes. The teacher has also—ahead of time—prepared two or more alternative pathways through the remainder of the lesson. The choice of pathway is conditioned on the nature of the responses to the hinge point question.

The strategy, “Providing feedback that moves learners forward” operates through two different mechanisms. The first is to give students feedback that makes them think, and thus stimulate learning. A technique that exemplifies this mechanism is called, Find and Fix Your Errors. In this technique, teachers do not explicitly identify problems that are incorrect or the grammar errors in a passage. Instead they simply indicate that a certain number of errors are present in the document, page, section, or line of the student's work, and it is up to the student to find them and fix them. To find the errors, students have to think anew, perhaps by applying grammar rules or employing a checking algorithm. Of course, this technique will not work if the students have no clue about what they did wrong, so another feature of effective feedback has to be in force: it either explicitly or implicitly tells the student what it is they need to do to improve. It's a judgment call on the teacher's part, deciding how much scaffolding to provide within any instance of feedback—there is a balance between prompting students to think on their own and giving them an idea where to start. And none of this makes sense if there is not a clear target toward which the student is striving.

The second mechanism by which the feedback strategy works is perhaps more direct, and that is providing time and structures to take feedback on board. While this seems like an obvious requirement for feedback to have any impact, it is often overlooked in the rush to cover the curriculum. All the effort that teachers may put into writing the most thoughtful comments in the world will go for naught if there is no time for students to consider the feedback and do something with it. A technique that some teachers use is a mastery grading system that requires students to submit work as many times as needed to show they have mastered the concepts involved. This clearly shows that students are expected to respond to feedback, and teaches them to make the most of the feedback they receive—otherwise they will spend an awful lot of time revisiting past assignments. Other teachers build draft and revision time into every major assignment and set aside class time for this purpose. To help students make sense of feedback, teachers may use individual conferencing or whole class instruction in which they discuss the standards for quality that the feedback references.

The strategies, “Activating students as instructional resources for one another” and “Activating students as the owners of their own learning” share an important mechanism in common: both strategies redistribute the work of the classroom in such a way that students take greater responsibility for their learning. This activation means that there are many more minds applying themselves to the challenges of teaching and learning, thus increasing the resources applied and the chance of success. When these strategies are working in an A/L -focused classroom, the students are doing the intellectual heavy lifting in grappling with the concepts under study, which leads to deeper learning than if they act as passive recipients in the process.

This is not to say that the teacher has no intellectual responsibilities, but rather that his or her intelligence is freed up to attend to the business of engineering the learning environment so that student learning is enhanced.

In addition to redistributing the cognitive work of the classroom, the strategy, “Activating peers as instructional resources for one another” operates through several mechanisms unique to it. In some techniques, such as Think-Pair-Share, students serve as sounding boards for one another, whereby thinking can be developed and deepened. In techniques such as Best Composite Paper or the various Jig-Saw techniques, student thinking is pooled to make a joint work product that is better than what a single student could come up with on his or her own (exemplifying the Chinese proverb, “When you have three people in a room, you have a genius”). In the various peer assessment techniques, students provide feedback to other students, which increases the rate at which feedback can be given in the classroom and takes some of the load off the teacher so he or she can concentrate efforts in other areas of planning or feedback. Furthermore, providing feedback on another student’s work requires that a student grapple again with the nature of quality with regard to a particular assignment, which advances the student’s own understanding of the concepts at hand. And finally, some techniques within this strategy rely on the principle of peer tutoring (e.g., Homework Helpboard and Question Strips in Groups), which has been shown to benefit the teaching student as much or more than the student receiving the help. Peers have a way of speaking their understandings that is often more easily taken in by students, and students will ask their peers to explain it again more readily than they will ask the teacher.

The strategy, “Activating students as the owners of their own learning” has its own unique mechanisms as well. One class of techniques involves students signaling to the teacher that they need him or her to slow down (e.g., the Traffic Signal technique or the Thumbs-Up/Thumbs-Down technique). Such signals provide critical information to the teacher that he or she can use to adjust instruction so that it is more responsive to students’ learning needs. Traffic lighting can also be applied as a form of meta-cognition that is never shared with the teacher: the teacher can structure review sessions, for instance, where students traffic light lists of concepts that are about to be tested. Students then focus their studying on the areas marked yellow or red. Self-assessment using rubrics or Pre-flight Checklists can also help students internalize the notion of quality.

The changed classroom contract. The mechanisms by which each of the five strategies of minute-to-minute and day-by-day AfL impact student learning imply changed roles for both students and teachers. Not every teacher begins to implement the strategies and techniques with this big picture in mind; rather, teachers tend to begin by focusing on just getting the technique to work reasonably well. But after a while, implementing one technique leads to a cascade effect, wherein the teacher sees the necessity of making another, and another, change in practice. For example, Black et al. (2003) described what happened to a teacher who decided to implement the technique Wait Time, which involves providing adequate time (roughly three seconds) after each question for students to formulate a thoughtful response. When he first began using the technique, students commented that he seemed, well, “slow.” This led him to see that the quality of the questions he was asking was weak: he was asking recall questions with little cognitive demand. The awkward three-second delay between questions made plain just how superficial his questions were—first to his students, and only later to him. That led him to begin a deliberate process of developing better questions that promoted student thinking. That proved to be much more difficult and important, in the end, than simply waiting three seconds after each question before calling on a

student to answer. In this way, the teacher came to see himself as the engineer of the larger learning environment, not just the clever implementer of a technique or two.

As the teacher starts to see him- or herself as the engineer or regulator of student learning, he or she begins to let go of the idea that it is one's teaching performance in the classroom that is paramount, and instead starts to focus on students' learning as the central goal. This change of focus, coupled with the use of various techniques that explicitly require students to take more responsibility for their own learning, leads to students' abandonment of heretofore passive roles for the more active role of being in charge of their own learning. These changing roles eventually lead to an overall rewriting of the implicit classroom contract.

The new classroom contract reflects a far more efficient and effective deployment of the personnel within the classroom. First of all, under the old roles, the teacher was struggling to jam some learning into students' heads, which flies in the face of how people learn (Bransford et al., 1999). Second, with the teacher acting as engineer and regulator of the learning environment, positive conditions for learning are more likely to be present. And third, with students behaving as the active directors of their own learning, there are simply more minds applied to solving learning problems. This is how the changed classroom contract advances student learning.

Implications for practice. The most important implications for practice in this step of the theory of action are as follows:

- The professional learning portion of the intervention needs to support teachers to have a reasonably good idea of how to implement the individual techniques and help them clearly understand why they work, so that the mechanism of each technique can be brought to bear.
- Teachers' persistent practice of selected AfL techniques will lead to even more important changes in the classroom contract, and this, in turn, strengthens the effects of the intervention. Therefore, every effort should be made to help teachers persist, even when initial efforts do not pay off right away.
- Shifts in the classroom contract may represent substantial deviations from accepted practice in some schools. For example, when students start acting as if they are in charge of their own learning, some adults in the school may react negatively. Or, some conceptions of teaching may not have room for the teacher playing the often more subtle role of engineer or regulator. School principals, for instance, may want to see more rote practice, or at least see the teacher in more traditional teaching roles. Sometimes, even students will push back against the new roles. Because there might be resistance from any number of quarters, the lone wolf teacher who implements AfL on his or her own is at a distinct disadvantage from the teacher who is doing this as part of a larger school-wide effort.

The Motivation for the Tight but Loose Framework

For KLT, or any intervention, for that matter, to be both effective and scalable, three inter-related factors must be satisfied. First is a very clear idea of what it is you're trying to enact and why you think this is a worthwhile thing to do—that is, a clear idea of all the program components, including its theory of action. Second is a comprehensive notion of what it means to scale up an intervention across diverse contexts. Third is consideration for the particularities of the actual contexts into which the intervention is to be scaled. We discuss each of these factors in turn.

Factor 1: A Clear Idea of What It Is You're Trying to Enact and Why You Think This Is a Worthwhile Thing To Do

This factor goes to both the effectiveness and scalability of an intervention. First, clarity on the components and theory of action for a reform is needed to achieve reasonable implementation fidelity, which is a powerful factor in the success or failure of an intervention at scale. If this clarity is lacking or spotty, then there is little hope that the intervention can be successfully communicated in one locale, much less in multiple sites. This factor also goes to the issue of effectiveness, for if an intervention's developers cannot articulate a complete and logical theory of action—the mechanism by which the intervention is supposed to attain its advertised outcomes—then any claims of effectiveness are, at best, merely hopes. We are familiar with a number of interventions (benchmark testing systems come to mind) that are highly detailed in describing their component parts but lack the same kind of attention to their theory of action: How, exactly, is it that testing students more often leads to improvements in student learning?

We are also aware of interventions that succeed on the theory of action count and are buttressed by deep empirical work, but fail to provide sufficient detail on the exact components or steps to be undertaken that turn the theory into practice. Some of the most promising research-driven interventions, such as Dweck's (2000), call for changing the kinds of praise we provide students, Slavin's (1995) work in cooperative learning, or Kluger and DeNisi's (1996) work on feedback are clear about the mechanisms through which these impact student learning, but provide insufficient specification or supports for exactly how, in the flow of a lesson, a teacher is to implement these. We note that in all these examples the reforms cannot be handled prescriptively—they require a high degree of skill and judgment on the part of the implementing teacher. But it is doubtful that any of these reforms will become scalable unless and until specifications and tools that match the theory of action are created to support teachers' development of expertise in these areas.

Lewis, Perry, and Murata (2006) did a good job of explaining the reasons why a strong theory of action (what they called “explication of the innovation mechanism;” p. 5) must be coupled with clarity on the intervention itself (the “descriptive knowledge base;” p. 4). They pointed to the problem of an innovation's surface features obscuring its true underlying mechanism. For example, they noted, “a focus on the surface features of ‘reform mathematics,’ such as hands-on activities and discussion, may provide a lethal substitute for attention to the underlying mechanism of developing students' mathematical reasoning through problem solving” (p. 5).

Lewis et al. then applied this idea to the distortions they have seen in the way lesson study has been translated into practice in the United States, arguing that the current state of understanding of lesson study is still largely caught in the surface feature stage, with many believing that its effects on student learning stem solely from the refinement of lesson plans. In fact, the underlying mechanism (the theory of action) for lesson study is far more complex and includes changes along three inter-related pathways: strengthening teachers' knowledge, developing teachers' commitment and community, and using powerful learning resources for both students and teachers (p. 5). To the degree that lesson study proponents in the United States fail to appreciate its true underlying mechanism, they will short-circuit its development, and almost certainly do damage to its potential as a lever for change.

The same fate awaits any intervention that is not clear about its theory of action or that fails to successfully communicate that theory of action. This is why we have been so insistent in

weaving the empirical and theoretical research base into the KLT intervention. The end users—the teachers—have to understand the why just as much as they have to understand the what or the how. Otherwise, they will be susceptible to the same kinds of distortions identified by Lewis et al., mistaking seductive surface features for the underlying mechanisms of change, and in the process undermining their own hard work. Even with our efforts to make the theory of action plain to end users, we have seen some of these same kinds of distortions in certain implementations of KLT—a testament to the pervasiveness of this problem (more about these later).

Factor 2: A Comprehensive Notion of What It Means to Scale Up an Intervention Across Diverse Contexts

As Coburn (2003) pointed out, “definitions of scale have traditionally been restricted in scope, focusing on the expanding number of schools reached by a reform” (p. 3). While recognizing the simplicity, intuitiveness, and ease of measurement associated with this commonly held definition, Coburn went on to raise questions about its meaning:

But what does it really mean to say that a reform program is scaled up in these terms? It says nothing about the nature of the change envisaged or enacted or the degree to which it is sustained, or the degree to which schools and teachers have knowledge and authority to continue to grow the reform over time. (p. 4)

These are questions that get to heart of successful scaling of complex, classroom-focused reforms such as A/L. Coburn proposed a needed expansion of the notion of scaling up that is rooted in the notion of consequential change, by which she meant change that makes a difference for teaching and learning. She discussed four inter-related dimensions to scale: depth, sustainability, spread, and shift in reform ownership. Her characterizations of these dimensions show a great deal of alignment with the demands placed on teachers and schools by interventions as complex as KLT. For example, she discussed depth as, “change that goes beyond surface features or procedures... to alter teachers’ beliefs, norms of social interaction in the classroom, and underlying pedagogical principles” (p.4). Among other points of connection, this sounds like the level of change required to alter the implicit classroom contract, which is part of the theory of action for KLT.

Coburn then turned to the dimension of sustainability, focusing on the idea of consequential change sustained over time. She pointed to the long, discouraging history of reforms that began with “a short-term influx of resources, professional development, and other forms of assistance” only to fall into disuse “in the face of competing priorities, changing demands, and teacher and administrator turnover” (p. 6). As developers of an intervention, we are trying to build in hedges against this kind of ebb and flow, in the form of sustained TLCs, but there is no guarantee that the resources to support these will be continued in any particular district or school. So this dimension seems to be beyond the scope of the intervention by itself, and is really more of an issue with regard to the larger systems into which the intervention is placed. It is tempting to then expand the scope of the intervention to handle these kinds of issues as well, which moves it toward a kind of systemic change or systemic reform initiative. The problem with this approach is that it can dilute the focus on the classroom, since more loci of change now have to be kept front and center. We speak about this topic later.

Coburn next took up the idea of spread, which she defined as encompassing both outward spread from schools and classrooms and spread within schools and classrooms. The notion of outward spread is more or less in line with the traditional ways of conceptualizing scale—more

teachers and more schools are drawn into the reform. Spread within is subtler, and edges into the areas of depth and sustainability. An example would be the way a reform might eventually be worked into the day-to-day policies and practices of the school or district. In the case of KLT, spread within might show itself in a classroom when a teacher who has been applying A/L to one subject begins applying it across all the subjects he or she teaches. At the school level, spread within might show itself when the school's grading policy is changed to make room for comment only marking or mastery grading. Another example would be when the school rearranges its schedule to allow teachers time to meet in their learning communities during normal school hours.

The last dimension Coburn took up is shift in reform ownership from external reformers (e.g., developers, researchers, or vendors) to internal players, with “authority for the reform held by districts, schools, and teachers who have the capacity to sustain, spread, and deepen reform principles themselves” (p. 7). She cautioned against an under-conceptualization of this dimension as being only about buy-in or acceptance of the reform. Rather, she emphasized a real shift in knowledge about and authority/capacity for extending the reform. The KLT theory of action partially addresses this dimension in its concern for complete and accurate transmission of the knowledge base, and its insistence on weaving the theory of action directly into that knowledge base. However, it is clear that practice still lags theory in this regard—otherwise, we would not be seeing distortions of implementation or so many teachers or schools falling away after the initial blast of investment in the reform (see, for example, Wylie, Thompson, Lyon, & Snodgrass, 2008 and other papers in this report).

The expanded notion of scale that Coburn proposed certainly complicates things for the developers and researchers who set out to improve teaching and learning through a clever, research-based intervention. It is tempting to rule most of these dimensions out of bounds or at least outside the responsibility of program developers, because of the cognitive and emotional overload they push on us. On the other hand, if we are serious about making the difference in teaching and learning that is needed (consequential change, in Coburn's language), we have to consider these dimensions at all points in the process, from the earliest stages of development through to full-scale delivery into the thousands of schools and tens of thousands of classrooms where the innovation is needed. This path seems to be pushing us to adopt a systemic reform approach.

There are two problems here. First is the previously mentioned issue of the focused intervention—in this case, minute-to-minute and day-by-day A/L —getting lost within the larger systemic reform effort. The other problem is more serious: a review of the various systemic reform programs shows no consistent track record of effectiveness, and the track record is even spottier when we look at these programs' effects across diverse contexts. Getting something to work across diverse contexts is the very problem we started out with! Apparently, systemic reform programs have a hard time in the scale up process too, partly because they also suffer from bootstrapping problems. So moving from a smallish intervention to a larger program of systemic reform (in which the intervention is embedded) does not seem to be the answer either.

There is a body of theory that may be of some assistance in dealing with this dilemma: the discipline of systems thinking or systems design, as articulated by C. West Churchman. Churchman's basic mission in life was to struggle with the question of whether or not it was possible to “secure improvement in the human condition by means of the human intellect” (Churchman, 1982, p. 19). In answering this question, Churchman (and others in the operations research/management world in which he worked) first tried to become as comprehensive as

possible, striving for a degree of understanding and control of the whole system that bordered on the omniscient and the all-powerful. Where Churchman began to diverge from his colleagues was in admitting the inherent impossibility of that task, while recognizing that morally, he was still bound to try to fix serious social problems. As his student Werner Ulrich summarized, Churchman came to understand that “what matters is not ‘knowing everything’ about the system in question but understanding the reasons and possible implications of our inevitable *lack* of comprehensive knowledge” (Ulrich, 2002, para. 3).

Churchman came to think that reformers who were not aware of this dilemma were prone to commit what he called the *environmental fallacy* (Churchman, 1979, pp. 4–7), which would inevitably lead to solutions that were more destructive than helpful. Churchman quite rightly pointed out that every problem exists in an environment and that each environment has an environment, and so on. When you try to fix a problem, you will always impact the environment in which it exists, and its larger-still environment, and so on—in ways you cannot possibly predict or control. Human cognition is just not able to see through to every connection and consequence in the system. What to do?

The first thing Churchman thought we should do is acknowledge the problem: the limits of human cognition and control will never allow us to see or control the whole system. The second thing we should do—because you are still morally bound to “secure improvements in the human condition”—is to move ahead, but as you go, engage in a constant sweep-in process: a self-conscious effort to sweep in a wider arc of information, experiences, and values to your understanding of the system. This approach suggests a distinct posture of inquiry tempered by humility.

What this means for reformers like us—researchers with a focused intervention that has to be scaled up within a diverse array of contexts (read: systems)—is that we have to think systemically, with Churchman’s environmental fallacy looming in the back of our minds to force us into awareness of the larger system. We should stay focused on the specifics of the focused intervention we are espousing, but as we go, we should find ways to sweep in the place-based particularities of the systems in which it will be operating. The “think globally, act locally” mantra of the environmental movement comes to mind.

This is quite different from the general notion of systemic reform. The various systemic reform models each take different approaches to helping a system change at multiple levels. While on the surface these models look as though they are sensitive to systemic particularities (if only because they have tended to include the word *systemic* in their names), in fact there is actually a wide range of context sensitivity about them. Some are very prescriptive about the methods and content of the changes to be made, where others deliberately keep a very open stance about the methods and goals. But what they all have in common is the driving assumption—from the outset—that the whole system must change.

This is not our starting point. Our starting point is that we want to see teachers adopt minute-to-minute and day-by-day AfL as a central part of their teaching practice. Where the system works to support that, leave it alone; where the system is in the way, change it. We agree wholeheartedly with Elmore (2002, 2004) and Fullan et al. (2006) that ultimately, all levels of the system have to be aligned toward the goal of improving teaching and learning. But instead of starting with whole system reform (as the system reformers do—not Elmore and Fullan), we begin with the focused intervention. Its very use within the system shines a light on the environmental problems that have to be taken care of for the original intervention to work. To

the teachers and administrators charged with implementing reforms, this is much more clarifying and motivating than an abstract goal such as systems alignment.

Factor 3: Consideration for the Particularities of the Actual Contexts Into Which the Intervention Is To Be Scaled

Systemic thinking, or the use of the sweep-in approach leads us directly to the third factor: Consideration for the particularities of the actual contexts into which the intervention is to be scaled. In large measure, this factor’s meaning turns on the meaning we give to the word *consideration*. On the one hand, consideration might be limited to something as minimal as maintaining awareness of the conditions of a particular context. At the other extreme, consideration might mean making frequent, visible, concrete adjustments in response to the conditions of that context. For leaders of educational reforms, the negotiation between these two extremes is exactly the place where guidance would be helpful. Because on the one hand, a reform will have limited effectiveness and no sustainability if it is not flexible enough to take advantage of local opportunities, while accommodating certain unmovable local constraints. But on the other hand, a reform needs to maintain fidelity to its core principles, or theory of action, if there is to be any hope of achieving its desired outcomes. Negotiating this tension is where the Tight but Loose framework can be of value.

Finally, the Tight but Loose Framework

The Tight but Loose formulation combines an obsessive adherence to central design principles (the tight part) with accommodations to the needs, resources, constraints, and particularities that occur in any school or district (the loose part), but only where these do not conflict with the theory of action of the intervention. We hope it is now apparent why we have been so deliberate in the development and articulation of the theory of action of KLT. Not only do teachers need to understand it to make it work, they (and the systems that surround them) have to understand what is not part of the theory of action, so they can make good decisions about which pieces of the intervention they must hold onto in the face of contextual challenges, and which pieces they can be flexible about. Clarity on the theory of action allows for rigor without rigidity.

We contend that many interventions fail to maintain effectiveness at scale because they err on the too-loose side of this formulation. Further, this often happens because the developers of the reform either do not have clarity on the exact mechanisms and targets of the reform—that is, they do not thoroughly understand their own theory of action—or they have not sufficiently articulated these such that adopters can make good decisions in support of the reform’s implementation. We also contend that reforms can fail on the too-tight end when a theory of action requires such a tight specification of conditions that scaling is impossible beyond a small number of idealized settings.

Larry Cuban’s (1998) brief history of the Effective Schools reform effort illustrates a failure on the too-loose side. In an aptly titled article, “How Schools Change Reforms: Redefining Reform Success and Failure,” Cuban pointed to the reform’s origins in late 70s and early 80s empirical studies, which identified urban schools that were “beating the odds.” This was followed by further empirical work that attempted to draw out the features that these schools held in common. The theory behind this work was that whatever these schools were doing could and should be replicated in other urban schools. Research on the beating-the-odds schools led to different lists of factors, but all the lists prioritized these four: “All children, regardless of

background, can learn and achieve results that mirror ability, not socioeconomic status; top-down decisions wedded to scientifically derived expertise can improve individual schools; measurable results count; and the school is the basic unit of reform” (p. 462).

Cuban went on to detail the way that this research-based notion catalyzed a reform movement that took the formal name of Effective Schools. In the wake of *A Nation at Risk* and other educational polemics, along with grave concern about urban schools in particular, Effective Schools took off in popularity, even to the degree that it was specifically named in 1988 amendments to ESEA, and state agencies were advised to set aside federal funds to help schools establish programs based on the Effective Schools factors. But by the late 80s and early 90s, the movement toward national goals, curriculum, and tests, as well as other policy waves (including such diverse methods as school vouchers and systemic reform) had started to crowd the ideas and methods of Effective Schools—even as they used the Effective Schools research to justify their own approaches. By the end of the 90s, the (capital E) Effective Schools movement had given way to the more generic concept of effective schools, which meant pretty much any reform that purported to improve schools, as long as test scores, top-down reforms, and at least the idea of research figured in its justification or method.

Cuban examined this history while outlining five competing, seldom explicit, criteria that are used for judging a reform’s success or failure. He stated that policy elites tend to use the standards of effectiveness, popularity, and fidelity, whereas practitioners (teachers and administrators) tend to use the standards of adaptability and longevity. By the practitioners’ standards, the Effective Schools reform has worked beautifully: it has adapted across thousands of schools, albeit in a highly reductionist form, and it is for this reason that it has achieved a certain measure of longevity. Just look at the number of schools that employ top-down accountability reforms and prioritize test scores above all else! (We note the irony of this result: the practitioners’ criteria show that the (small e) effective schools reform was quite successful, even though many practitioners are not at all happy with this fact—school-based administrators and teachers are now forming a significant block of resistance to top-down accountability programs.)

Applying the policy elites’ criteria of effectiveness, popularity, and fidelity leads to quite a different judgment as to the reform’s success, however. As Cuban said:

There is some evidence of partial success (e.g., individual schools that have performed consistently above expectations; test-score evidence of gains in basic skills for urban children) but no clear long-term trend of student improvement in academic performance. For popularity and adaptiveness, there is no question that both have been in full display. Effective Schools programs have been tailored to meet school settings different from those for which they were originally conceived. If some Effective Schools reformers disliked the constant modifications and dilution of their correlates of effectiveness, other administrators and practitioners enjoyed the reform’s flexibility. Its resiliency and popularity have given the ideology and program a remarkable reach. However, such plasticity and popularity—a reform for all seasons—mean that whatever ideological and programmatic bite it contained softened considerably as it spread to small towns, suburbs, states, and the embrace of the federal government.

Hence, as Effective Schools became a generic program of improvement, even losing its brand name, its potential to meet the standard of effectiveness lessened considerably. (p. 469-470)

This is the too-loose problem in a nutshell! The very plasticity that allowed the reform to move into so many diverse settings ensured that it lost its meaning and effectiveness.

In addition, the story of the Effective Schools movement illustrates another key point of our Tight but Loose theory: An innovation's empirical basis is important but ultimately not sufficient; rather, that empirical basis has to be stitched into a larger theory of action. Empirical work should sow the seeds for a promising intervention and give a boost to the development of its theory of action. It should be used to resolve problematic discontinuities in that theory of action, which are likely to emerge as the innovation is under development and pilot testing across diverse contexts. (See Lewis et al., 2006 for an excellent discussion on the uses of design research for this exact purpose.) And, of course, empirical studies should be used ultimately to prove or disprove an intervention's effectiveness. But empirical work should not be mistaken for the actual understanding and articulation of why an intervention works. The Effective Schools story illustrates this perfectly. For, even though the reform was predicated on extensive empirical research of high quality, its empirical origins did not in themselves provide a well-reasoned and complete theory of action that could stand up to the pressures that ultimately bent the reform into a thousand weakened and distorted forms.

There are any number of small and large educational initiatives that have failed on the too-loose side of the formulation. But there are also initiatives that fail, or at least fail to scale up, because of problems on the too-tight end of the equation. A recent commentary (Cossentino, 2007) on the publication of a study that looked at the effectiveness of Montessori schools illustrates this point. Angeline Lillard and Nicole Else-Quest (2006) conducted a randomized experiment made possible by over-subscription to a lottery for entry into a public Montessori school in Milwaukee. The experiment showed that Montessori works, finding statistically significant learning advantages for both 5- and 12-year-olds who got into the program by lottery, compared to students who were not admitted in the lottery.

Cossentino's (2007) commentary on the study and on Montessori was quite enthusiastic. She began by highlighting the deep empirical and theoretical work that stands behind the Montessori method:

Contemporary psychology has caught up to Montessori's revolutionary insights (insights gained from close and ongoing child study), and many of the elements of Montessori thought to be "quaint" and "unscientific" not only have been validated by experimental psychology, but also have been absorbed into the educational mainstream. It is now common, for instance, to find child-size furniture, manipulative materials, mixed-age grouping, and differentiated instruction in all manner of American classrooms. Likewise, new research on brain development, embodied cognition, and motivation provides striking confirmation of Montessori's claims regarding sensorial learning, attention, and intrinsic vs. extrinsic rewards. (p. 32)

Transmitting the deep theory and knowledge base behind Montessori is not an easy task. Its proponents have relied for years on a form of teacher training that immerses teachers in an "all-Montessori, all-the-time" educational environment that is somewhat unique—at least in the United States—in its commitment to the theory of action of a single approach. Cossentino recognized this in her commentary and then spoke about it in a way that could be an advertisement for the tight part of the Tight but Loose framework:

As researchers such as Harvard University's Richard Elmore and his colleagues in the Consortium for Policy Research in Education have argued, building capacity takes deep

and system-wide understanding of the core technologies of teaching and learning. In Montessori schools, this means deep knowledge of what Montessori is (and is not). And that knowledge comes first and foremost from the training centers that prepare teachers to work in these schools.

Montessori teaching practice is among the most technically complex approaches to instruction ever invented. Doing it well requires teachers to have mastered both the details of developmental theory and the carefully orchestrated sequences and activities that make up the Montessori curriculum. Deploying this vast knowledge base is further supported by ongoing clinical observation, which forms the basis for all interactions with children.

In Milwaukee, public Montessori schools are supported by a rigorous training program that adheres to strict standards based on an interpretation of Montessori education that is both complex and stable. While in most schools the knowledge base for teaching is a moving target—contested, contingent, contextual—in most Montessori schools, and especially in the Milwaukee schools studied, that knowledge base has changed little in the hundred years since it was first developed by Maria Montessori. Critics may charge that such stability amounts to a “stale” or “dogmatic” approach to pedagogy, but the results suggest otherwise. These results should prompt us to look much more closely at the “what” as well as the “how” of capacity.

Coherent reform means improvement efforts that hang together in a systematic and consistent manner. The how, why, and what of education must make sense in practical as well as theoretical ways, which means that improvement plans cannot be grafted together in a random or piecemeal fashion. When the reform involves Montessori, achieving coherence takes leadership that appreciates both the complexity of the Montessori knowledge base and the totality of Montessori as a system. (p.32)

Clearly, Montessori proponents get the tight part. But here is the problem: Montessori has been around for a very long time—almost 100 years—and has been in place in a small number of (mostly private) schools in the United States for almost as long. Yet it is still not used at any kind of scale. In the past few decades, there have been notable attempts to bring it into use in public schools; the Milwaukee experiment is the most successful and well-known of these. But other such efforts have foundered. The pressures to conform to more conventional notions of schooling have led to one of two outcomes in most locales: either the Montessori method has been watered down to be almost meaningless (and therefore ineffective), or the conflicts between Montessori’s theory of action and conventional notions of schooling have led to the removal of Montessori from the schools that tried it.

It is important to state that we are not suggesting that Montessorians should just loosen up. It may be that their obsessive adherence to their theory of action will ultimately lead to a steady gain in popularity, as the good effects of their approach slowly become known (and as the deleterious effects of wandering, unprincipled approaches become more obvious). And certainly as proponents of a different complex intervention with a deep theory of action and knowledge base, we are in exactly the same boat. We are hoping that by holding tight to our theory of action—and convincing others to do so as well—we can reap the consequential change that is clearly needed to make schools into places of learning instead of the dropout factories that so many schools currently are.

But our Tight but Loose framework—as well as our starting point for KLT—professional development within the schools as they currently exist—gives us a slightly different take on the

notions of context and scale than the Montessorians seem to have. We are not aware of any attention to the issue of scaling within the Montessori literature, whereas scale has been built into our thinking almost from the beginning of our development process. And that's because of the moral imperative we feel to not abandon the 49 million students in the nation's schools (National Center for Education Statistics, 2005). Schools aren't going to close down and start from scratch anytime soon. It is not that we believe we have an intervention that can—overnight—fix everything that's broken in schools. That perspective would reek of the arrogance that Churchman spent the latter part of his life trying to counteract. We believe that we have an intervention that can usefully be put to work in lots of different contexts by the people who teach and go to school there, in ways that make sense for them, while still holding onto the essentials of the theory of action, so that it has a decent chance of success. And for that reason, we continually bother ourselves with the problem of negotiating the Tight but Loose boundary.

This means that we have to concern ourselves with the ecological validity of KLT. That is, we have to include in the design of the intervention guidance, support, and tools that increase the likelihood that it will succeed within the thousands of school ecologies in which we hope it will come to reside. Our reading of Cobb et al. (2003) certainly spurred us to be more mindful of these ecologies, and, in particular to the importance of the brokers in school communities—the teachers and school and district leaders who play pivotal roles in bringing reforms to life—and the boundaries that they traverse in the process. But ultimately, we steered in a different direction, because we thought that adopting Cobb et al.'s focus on these players would leave the intervention too dependent on the question of whether there were an adequate number of really smart, well-placed people in a school or district. This is not to denigrate the role or influence of the people in the schools and districts we work in—there is a substantial body of research detailing their capacity to make or break a reform. In fact, this will be seen in some of the later papers in this report. Keeping awareness of this aspect of context—which is completely outside our control—is still necessary if we take Churchman seriously.

Thinking systemically leads us to believe, however, that solving the problem of “not enough smart people” or “not enough people in the right positions to make a difference” is a problem to be solved by local implementers, with help from us. And currently, we conceive of that help as being in the form of explicit guidance on what is essential to hang onto and what can be jettisoned as the intervention is transmitted across boundaries. This approach not only saves us from becoming overly prescriptive (too tight, not to mention offensive to people's intelligence); it also allows us to take advantage of the times when there are already really smart people in place. Or the times when a system (a grade level team, a department, a school, a district) is in just-good-enough shape to begin the process of capacity building required by KLT. Just-good-enough shape is all that is needed to get started, and explicit attention to capacity building is what increases the likelihood that there will be enough smart people ready for the next crisis of implementation.

Tight but Loose Applied to *Keeping Learning on Track*

Having applied the notion of Tight but Loose to two other reforms, it is only fair to turn the lens on our own. What does Tight but Loose look like when applied to KLT? That is the text, or at least the sub-text, of the papers that follow this one, which will relate a set of place-based stories of implementing KLT in five diverse settings. In anticipation of these stories, let's briefly discuss some areas of implementation that have or could have benefited from thinking in a tight but loose fashion.

A good example is the range of practice we see in the ways that teachers use the whiteboard technology. If you want to use whiteboards to boost student engagement and to get information on what student thinking looks like, then you have to regularly expect every student to hold his or hers up. Unfortunately, that is not always what we see in classrooms. In a classic example of confusing the surface features of a reform with its underlying mechanism, a few teachers have eagerly brought whiteboards into the classroom, and then use them as a glorified form of scratch paper, “because the kids really like using the wipe-on, wipe off markers.” Needless to say, the whiteboards in these classrooms are not leading to any noticeable improvements in engagement or learning, and are certainly not working a change in the classroom contract. This is an example where we need to be more explicit about the theory of action behind a technique.

This one example is illustrative of the kinds of things that we (and teachers) have to be tight about. The tight list is actually quite long, as can be seen by our lengthy disquisition on the components and theory of action of KLT. Coming to know and understand everything on this list is what is involved in becoming an expert at minute-to-minute and day-by-day AfL. The list is not static or exactly the same for every teacher, which is why it requires expertise to master it, instead of brute memorization.

We also need to be tight about the essential elements of the professional learning portion of the intervention. It is pretty well proven that a bunch of well-meaning researchers at ETS or a university coming up with a clever intervention with a strong theory of action and empirical support is not sufficient to produce change in the black box of day-to-day instruction. So another part of the theory of action has to address the process by which teachers learn about, practice, reflect upon, and adjust their instruction so that they eventually become expert at AfL. That is why we build in the explicit expectation that teachers participate in learning communities focused on AfL. And it’s why we provide such explicit guidance for the content and tone of the learning community meetings, and provide ongoing support to learning community leaders so they can get tighter about their own understanding of AfL.

However, at this stage of development, we would have to say that we are far less sure of the things we believe we must be tight about with regard to growing teacher expertise than we are with regard to the practice of AfL itself. A few things are coming clear, and these have been noted in this paper: things like teachers needing to have a regular time and place where they are required (by custom or rule) to tell a story about their most recent efforts at AfL in their own classrooms, get feedback, and come up with plans for their next steps. They do not necessarily have to operate within one of our learning communities or follow one of our modules, but we are sure that they need the personal story-telling/feedback/planning cycle.

If we are tight about this, then learning communities that have 20 people in them can’t be allowed (unless they split up into smaller groups for the How’s it Going? segment), simply to give each person adequate time to tell his or her story and receive critical feedback. We ran into the problem of over-large learning communities in a school district that had recently reorganized itself into K-8 schools. The teachers wanted to stay all together so they could get to know one another, as they had just been thrown together from a number of schools. We argued about this with the initiative’s leaders at both the district and school level, but ultimately we did not prevail. (The level of growth shown by these teachers was not great, though there were other problems that could have led to this result as well.)

Another example of where we are tight is the idea of never telling teachers which techniques they should employ in their classrooms. A few sites we have worked in have

attempted to meld KLT with other reforms, and they looked for techniques that mapped nicely onto these. In essence, they wanted to use the fact that KLT included these particular techniques as a basis for requiring these techniques in every classroom. Knowing we have to be tight about this issue, we have argued and prevailed, explaining that leaving the choice of techniques up to each teacher is consistent with two intersecting points in our theory of action: teachers are accountable for both taking charge of their own learning and for making steady improvements in their practice. Selecting and practicing the techniques that make sense to you, as the person in charge of your classroom, is part of the learning process. If administrators fail to treat teachers as accountable professionals, the learning is short-circuited, and the expertise never develops.

For the administrators who worry that teachers need to be held more accountable, we remind them that we do hold out very clear expectations for teachers. A teacher who is learning to become expert at A/L needs to learn how it applies in all five of the strategies, not just the one or two that hold immediate appeal. They don't have to use every technique, not at once, and not ever. But they do have to, over time (the span of 1 to 2 years, we would say), work on techniques from each of the five strategies. This is a non-negotiable, another thing we are tight about.

Our development of the theory of action for KLT and the Tight but Loose theory has occasionally led us to identify an area that we can be decidedly loose about. An example has to do with the question of whether teachers who join the program must be volunteers as opposed to conscripts, forced to participate by school or district mandate. There is no question that we see many advantages to at least beginning the process in a school with volunteers. Not only does this make the bumpy first steps of a new program go a little easier, it also leads to the creation of local existence proofs that can be used to disarm the doubting late adopters. But there is nothing in our theory of action that would strictly rule out the possibility of entering a school or district under a top-down mandate—as long as that mandate was backed up by adequate resources and true support for the teachers, who are the ones taking the biggest risk.

In general, we would say that anything the theory of action does not require us to be tight about is something we can be loose about. This approach allows us to explicitly carve out areas of flexibility, and being flexible enables the intervention to adapt to different locales. But never forget that being tight is what ensures that it will work. Under this definition of looseness—where the loose list is defined as everything we are not tight about—ensures that the two lists will never come into conflict (except for the cases where we should be tight about something but we haven't yet learned that lesson).

It appears that the loose list will include many things that are outside the realm of the classroom, things that have a more system feel to them—like where the funding comes from, exactly how often teachers must meet together, how KLT is to relate to certain system policies and practices such as parent communications, report cards, and the like. Because the loose list is likely to include a lot of things outside the classroom, it is easy to think of these as systemic issues, and then to jump to the idea that we are loose about systemic issues. But that does not make sense, given that we know that systemic conditions exert positive or negative pressure on classroom activities. This is where “thinking globally, acting locally” will come in handy—it can guide us in figuring out which parts of the environment we have to attend to.

There are also a number of places where we have to develop a very nuanced statement of tight but loose—it's okay to be loose about X, but only in Y circumstances. For example:

- Yes, we can schedule the TLC meetings after school, but only if we no longer require these teachers to attend the literacy sessions that we had previously scheduled for alternate weeks. Otherwise, the time demands will be too great, teachers won't attend

regularly, or they'll resent the program instead of embracing it.

Or,

- No, we cannot require every teacher in the school to “choose” to adopt the KLT “Find and Fix” technique, even if it does seem perfectly in line with our math curriculum. That approach would violate the “rule” of never telling teachers what to do. Once that rule has been violated, teachers will lose the sense that they have to take charge of their own learning, and worse, it really might not be appropriate for some teachers and students. We'll have to look for other ways to make the connections to the math curriculum apparent.

Conclusion

In this paper, we have tried to set out some of our preliminary ideas for a framework for thinking about school reform at scale. Our starting point has been the need to accept and embrace the bewildering diversity of schools and school systems. We do so, not out of some noble desire to honor the individuality and idiosyncrasies of our schools, but rather because we see the differences between our schools as inevitable reactions to the diversity of contexts in which they operate, the variety of problems they face, and the variety of resources at their disposal—it might be possible to try to make all schools the same, but this would inevitably make them worse. This diversity means that one size fits all interventions cannot succeed. The natural response to this need to allow reform efforts to be adapted to local circumstances is to allow flexibility in implementation and operation.

However, allowing flexibility requires a much deeper understanding of the theory of action of the intervention than is necessary for rigid replication. Even the simplest intervention is in reality extraordinarily complex, with many components, some of which will be more effective than others. Without a strong theory of action for the intervention, there is a real danger that modifications of the intervention leave out or neutralize the effects of the most powerful components (even with a strong theory of action, this risk is substantial in the absence of empirical evidence about the relative effectiveness of the components). Thus if we are to design complex interventions that can be implemented successfully in diverse settings, then we must find ways of ensuring that the changes that are made to allow this (the intervention has to be loose) are made in such a way as to minimize the likelihood that the most important components—the active ingredients if you like—are not compromised (the intervention has to be tight). This leads us to the central idea that an intervention has to be both tight and loose. The Tight but Loose formulation combines an obsessive adherence to central design principles (the tight part) with accommodations to the needs, resources, constraints, and particularities that occur in any school or district (the loose part), but only where these do not conflict with the theory of action of the intervention.

With such a formulation, there is a danger that the loose components are seen as not important—rather like the protective outer core of beliefs that Imre Lakatos proposed for the methodology of scientific research programs (Lakatos, 1970): components that can be discarded without damage to the main theory. However, we believe that the loose components play a much more significant role. They are much more like the delivery mechanism for a drug. While the drug is the active ingredient, the drug is effective only when it can be delivered to the right place, in the right dosage, and at the right time. For some applications it might be delivered by injection, in others by inhaler, and in others, orally with a timed-release coating. Without the

delivery mechanism, the drug is useless, but conversely, without the drug, the delivery mechanism on its own is also useless.

We do not claim that the need for interventions to be both tight and loose is original. Indeed, it seems to us that all interventions that have been successful at scale in the past have been both tight and loose. What we do claim is that conceptualizing interventions explicitly in terms of the Tight but Loose formulation forces attention onto important aspects of the design of the intervention, and increase the likelihood of successful implementation at scale. In particular, we suggest that the adoption of the Tight but Loose formulation forces attention to three processes: what we want to change, how we propose to effect such changes, and why these changes are important.

In addition to the general points about school reform at scale, we have discussed in detail one particular intervention—a professional development program entitled KLT. We have described its origin in the well-established research base on the effects of classroom assessment practices on student achievement, and also some of the steps we have taken in designing interventions to bring these practices to scale. While our basic thinking about what classrooms implementing effective assessment should look like have changed little in the last 10 years, we have developed radically, and continue to develop, the ways we communicate about these practices, and the structures that will support their adoption. As a result of extensive development work in over a hundred districts, we are convinced that the development of minute-to-minute and day-by-day assessment practices offer the possibility of unprecedented improvements in student achievement, that TLCs offer the most appropriate mechanism for supporting teachers in making the necessary changes in their practice, and that the Tight but Loose formulation provides a design narrative that optimizes the chances for taking these changes to scale.

A Teacher-Driven Implementation of Assessment for Learning (AfL) in New Jersey

Laura Goe (ETS, Princeton, NJ) and Diane Mardy (Ho-Ho-Kus School District, NJ)

Introduction

The story of the Ho-Ho-Kus, New Jersey school district and how it came to embrace the principles of assessment for learning (AfL) will be told from two perspectives: that of the researcher who was assigned by ETS to document the implementation and institutionalization of AfL, and that of the teacher who played a key role in ensuring that AfL was adopted with a great degree of enthusiasm in this small K-8 school district. The story of AfL in Ho-Ho-Kus will unfold in alternating voices, as the researcher, Laura Goe, describes events from her perspective and the teacher/researcher, Diane Mardy, does the same. This format for describing what transpired in Ho-Ho-Kus is a form of narrative inquiry, a type of qualitative investigation that Chase (2005) has described as, “an amalgam of interdisciplinary analytic lenses, diverse disciplinary approaches, and both traditional and innovative methods—all revolving around an interest in biographical particulars as narrated by the one who lives them”(p. 651).

The purpose of describing the implementation and institutionalization of AfL (which later became the *Keeping Learning on Track*[®] [KLT] program) is fourfold. First, this case study provides an interesting description of how a professional development program was introduced and spread throughout this school district over a 2-year period. Second, it serves to illustrate one version of a particularly high level of implementation where the strategies and techniques introduced as part of the professional development spread to other classrooms, even the classrooms of those who never attended the professional development sessions. Third, it provides an interesting view of how teachers adapted the AfL strategies and techniques to suit their own purposes in their own unique contexts. Fourth, it serves to illustrate how the Tight but Loose framework provides a way to examine the introduction and implementation of the program in a particular environment.

Tight but Loose is a way of thinking about how to scale up reforms, programs, and interventions (Thompson & Wiliam, 2007). Some components must be implemented exactly as intended (tight), while others may be adapted to work best in local contexts (loose). The key is to create professional development, interventions, or other reforms with enough structure to ensure that users adhere to the parts that must be tight (programmatic elements) while allowing for adjustments at the local level—the loose part—where appropriate. The Tight but Loose framework applies to the implementation and institutionalization of an intervention at the district, school, and classroom level, since these are all places that have local differences. AfL was the focus of the professional development developed by Dylan Wiliam and others to support teachers in implementing formative assessment practices in their classrooms. The ideas that were shared and tested with teachers were later incorporated into what came to be known as the KLT program. During the development of a set of theories, strategies, and techniques about scaling up KLT, the Tight but Loose framework was conceived.

In the next section, we will relate the story of how each of us observed and/or participated in the introduction of AfL in the Ho-Ho-Kus school district. Interestingly, Diane also conducted research of her own related to AfL strategies and techniques during this time period. Thus, she is labeled a *teacher/researcher*, rather than solely a teacher. Details about her research and what she learned from it will be woven into the dialogue.

What is the *Keeping Learning on Track* Program?

Before moving into the narrative, it may be helpful to provide some general information about the approach to formative assessment that was used in this district. Staff at ETS were interested in the work around formative assessment – or AfL that had been recently conducted by Paul Black and Dylan Wiliam, education researchers in England, as a way to bring minute-by-minute formative assessment to the classroom (Black et al., 2003). When Dylan Wiliam relocated to ETS, he began a course of development with ETS researchers that would allow the transmission of the ideas of AfL to be scaled up in a way that would make it possible to introduce the ideas to entire districts rather than to just a few classrooms or schools.

The professional development approach used in this district to support teachers' adoption of minute-to-minute and day-by-day AfL strategies has been shown by research to powerfully increase student learning (Black et al., 2002). Formalized into five key strategies, it focused on using evidence of learning to adapt instruction. The five strategies are as follows:

1. Sharing learning expectations: Clarifying and sharing learning intentions and criteria for success,
2. Questioning: Engineering effective classroom discussions, questions, and learning tasks that elicit evidence of learning,
3. Feedback: Providing feedback that moves learners forward,
4. Activating self: Activating students as the owners of their own learning, and,
5. Activating peers: Activating students as instructional resources for one another.

These strategies each are associated with specific techniques that allow teachers to operationalize the strategies.

A key component of this approach is the idea that teachers should have complete freedom to choose which strategies and techniques to use. This freedom ensures that teachers' professional judgment is a major component of the successful implementation of strategies and techniques. Each teacher knows his or her students, learning goals, and classroom context better than anyone else and thus should know best how to choose appropriate strategies and techniques to further his or her needs.

AfL Comes to Ho-Ho-Kus

Ho-Ho-Kus is a small K-8 school/district in New Jersey. The state has many small districts such as this one, where the school and district are contiguous and housed in one building. There is only one school, which serves about 600 high-achieving students. The district benefits from strong parental involvement and the presence of experienced, highly qualified teachers. Key features of the district include strong support for professional development from the administration and trust in teachers as professionals.

Laura (the ETS researcher): In the spring of 2004, a curriculum coordinator in charge of several small school districts contacted Dr. Dylan Wiliam about AfL after reading a newspaper article about his work.

Diane (the teacher/researcher): The curriculum coordinator contacted all educators to apprise them of a study group that would be meeting the following year. This group would meet once in the summer and then periodically during the school year to learn about and implement formative assessment with Dylan.

Laura: Dylan, then employed by ETS, subsequently agreed to work with the district consortium, and the curriculum coordinator sent out a request for volunteers with a brief description of the professional development approach.

Diane: Intrigued, but hesitant too, I contacted this coordinator and told her that if formative assessment had anything to do with giving more standardized tests, I was not interested. But if it involved looking at learning in new ways, I was definitely on board.

Laura: Eight teachers volunteered—two from the high school, three from a K-8 district (Ho-Ho-Kus), and three from two other local K-8 districts. The teachers agreed to meet with Dylan six times during the 2004-05 school year for about 2 hours each meeting. They were not paid for attending but did receive release time from their classrooms to meet during the school day. Dylan asked me to observe and take field notes during these meetings with this small group of teachers, and I subsequently attended Dylan's sessions with these and other teachers over a 2-year period.

Diane: I have taught for over 25 years in Ho-Ho-Kus School (a pre-K to 8th grade school), but in many different elementary and middle school grades. I am presently acting in a more official leadership role as Coordinator of Student Services and Staff Development. When I first began my work with Dylan Wiliam, I was teaching third grade. There were also two colleagues from my school who participated in this professional development, three other teachers from different schools, and the curriculum coordinator. The size of the group was ideal for encouraging real dialogue, and we all benefited immensely from working closely in round table discussions with Dylan Wiliam. We were consistently challenged to reexamine our methods of teaching and theories of how students learn, as well as to think "outside of the box." Having other colleagues from my school to share AfL strategies, discuss ideas that worked and didn't, and engage in productive dialogue about formative assessment was invaluable.

Laura: As part of this professional development, teachers also agreed to keep learning logs in which they recorded their attempts to implement AfL. All but one of the teachers completed the six sessions during the 2004-05 school year. The learning logs became important tools in understanding how teachers grew from knowing very little about formative assessment and AfL to implementing a variety of strategies and techniques in their classrooms.

Diane: During this time I was also pursuing a degree in educational leadership. One of the requirements for successful completion of the program was to carry out an action research project that would span more than 1 year (2004-05 and 2005-06 school years). I decided to focus my efforts on AfL, specifically its impact upon mathematics instruction. In order to increase the value of this experience, the action research was also collaborative. I shared what I had learned in discussions with Dylan during weekly team planning meetings with the three other third grade teachers. We devised activities to use in our classrooms and then discussed what worked and didn't and new ideas to try the following week.

Laura: Diane's experience of sharing with the other teachers in her school was true for others in this group, and so teachers participating in the AfL professional development began going back to their schools and discussing what they were learning with other teachers and administrators in both formal and informal settings. I learned this from listening to the teachers describe sharing their new knowledge about formative assessment. Later, during interviews with the other two Ho-Ho-Kus teachers, one high school teacher, and two teachers in the other K-8 districts, I confirmed that there was a second-generation effect; the teachers who learned about formative assessment directly from Dylan shared it in their schools. Thus, for many teachers in the four participating schools, their first encounter with AfL was through presentations by

participants in the small group that met the 1st year. Diane, in particular, seemed to be a key person in this transmittal of AfL to teachers in her school, though the other two teachers also talked enthusiastically with their colleagues about AfL, according to what they told me in interviews. The teachers also used their weekly grade-level meetings to share AfL strategies and report how they were working in their classrooms. The teachers also worked with Diane to present professional development around what they were learning from Dylan's group meetings. Their obvious enthusiasm for AfL and their apparent success in implementing strategies in their classrooms created a strong interest among their colleagues. All three of the teachers were experienced and respected by their colleagues, so their excitement about AfL aroused interest among the other teachers. Diane's leadership role in the school was likely a factor in her becoming the main proponent and unofficial spokesperson for AfL at Ho-Ho-Kus.

Diane: Going beyond just the collaborative sharing that I was doing in weekly team meetings, during an in-service professional development day in January 2005, I presented an AfL Workshop for the entire teaching and administrative staff in the school. Interest was piqued immediately, and many teachers began to slowly implement some of these methods in their classrooms.

Laura: The positive experiences the initial group of teachers had with AfL led to ramping up the program for the following year, with Dylan delivering a one-day summer workshop for all interested teachers followed by four half-day workshops during the 2005-06 school year. I attended these workshops, collecting learning logs, recording teachers' comments, observing, and interviewing teachers at all four schools during the 1st year. I interviewed Ho-Ho-Kus teachers more intensively in the 2nd year and observed in classrooms where teachers were using AfL strategies and techniques. I also attended two informal sharing sessions organized by Diane in which teachers met after school and shared AfL strategies and techniques they had tried and asked each other questions about implementing specific practices.

Diane: Ho-Ho-Kus was ready for the implementation of formative assessment. The atmosphere in the school is extremely collegial, with teachers, administrators, students, and parents working together to create a community of learners. Teachers and administrators work together to develop new initiatives and then focus efforts toward this collective goal for the school. Prior to its introduction to AfL, Ho-Ho-Kus School was involved in a 3-year implementation of a learner active technology-based instruction initiative. The stage was set for another layer to be added to this doctrine which would give teachers some further methods for differentiating learning and focusing assessment to drive instruction. In addition, students were ready to take the next step in having more control over their learning outcomes, and parents were primed for greater challenges and independence for their children.

Laura: The introduction and implementation of AfL in Ho-Ho-Kus was quite different than in any of the other sites we worked with in that 2-year period. While most teachers were introduced to AfL or the KLT program through a workshop lasting several days, the Ho-Ho-Kus teachers had already heard about and developed positive expectations for formative assessment through the small group of volunteers who worked closely with Dylan in the 1st year. These volunteer teachers generated a high level of enthusiasm for AfL among their colleagues who were not participating. This positive buzz meant that there was an interest in and openness to the AfL strategies and techniques before the larger-scale professional development was begun in the 3rd year. This led me to speculate about models for delivering professional development. It seemed as though this model was very effective, but it was also highly dependent on having both

very fertile ground and the right growing conditions to support the rapid spread of the ideas that had been introduced.

Diane: During the 04-05 school year in which I was introduced to AfL, I began designing and conducting an action research project at Ho-Ho-Kus which affected approximately 70 third grade students and their four classroom teachers. The students involved were those who participated in the regular education mathematics class. These are children who were performing on grade level or above in this subject. I believed that daily use of formative assessment techniques in mathematics instruction, including those I had learned through participating in sessions with Dylan, would help students to become more confident in their ability to solve open-ended math problems. I hoped that this would result in greater student performance and confidence in this area and an increase in standardized test scores as well.

Laura: Because Diane decided to focus her action research project on AfL strategies and techniques, it was necessary that she share specific strategies and techniques with participating teachers. In addition, Diane was asked by her administration to share what she had learned about AfL at a professional development session for her entire school. The other teachers trusted her when she told them about the benefits to student learning that could be gained from using AfL strategies and techniques, because Diane was a highly-respected teacher and leader at the school. This is an important point in this school's implementation story; further expansion of AfL in this school was precipitated by the interest generated through learning about AfL directly from teachers—Diane and her two colleagues—not from Dylan, the expert on formative assessment.

Diane: There were a number of challenges facing both students and teachers participating in my action research project. These revolved around time, skill level, and interest. The teachers involved were operating with differing backgrounds in formative assessment awareness and experience, whereas I had taken part in in-service training in this area for all of the 2004-05 school year and already incorporated many of these techniques into the classroom.

Laura: Two other teachers in Diane's school also attended Dylan's small group sessions in the 1st year. Thus, Diane was not the only teacher in the school who was championing AfL strategies and techniques. This factor probably contributed to the enthusiasm for AfL among teachers who participated during the 2005-06 school year. Rather than having only one teacher state that AfL worked well in their classroom, there were two other teachers with very different teaching contexts who were also successfully employing AfL strategies and techniques in their classrooms. The fact that these other teachers had quite different contexts (one had no classroom but worked as a specialist in teachers' classrooms) may have helped teachers from a variety of subjects and grade levels to imagine that these strategies and techniques were likely to work in their unique contexts.

Diane: When the opportunity arose to study with Dylan during the 2005-06 year, more than two-thirds of the staff, kindergarten through eighth grade, signed up. In fact, Dylan Wiliam, along with ETS staff studied the Ho-Ho-Kus School to try to ascertain what it is about this school that has enabled formative assessment to take hold so readily and whether it can be pinpointed and replicated elsewhere.

Laura: Towards the end of the 1st year of working with Dylan, Diane began her action research project at Ho-Ho-Kus. She worked with three teachers and started by determining (through interviews) how much these teachers already knew about strategies and techniques for formative assessment.

Diane: As I undertook the action research project at Ho-Ho-Kus, I learned that the amount of formative assessment that occurred in the classrooms I was studying ranged from a

small amount to none, as reported to me by the classroom teachers. This created a different comfort level for the students and teachers across the grade level, as some had familiarity with these methods and some were novices. The innate skill level of students was also a consideration when evaluating the impact of formative assessment on student learning in my project.

Laura: During the 2nd year of implementation of AfL in Ho-Ho-Kus, Dylan conducted half-day professional development workshops with teachers from three K-8 schools. The first meeting took place before schools opened for the fall, and Dylan spent most of the time introducing AfL strategies and techniques. All of the subsequent meetings took place at Ho-Ho-Kus, and all of the participating teachers were given classroom coverage to allow them to attend the meetings. During these sessions, Dylan discussed AfL strategies and techniques briefly and spent a large part of the time encouraging discussion in small- and whole-group sessions where teachers shared the various ways that they were implementing the strategies and techniques, and reported the results they obtained due to their efforts. Meanwhile, Diane carried out her action research project, focused on mathematics teaching and learning.

Diane: As more and more teachers began to utilize some of these techniques, interest grew about the prospect and possibility of learning more about AfL. As the original small group had the fortune to work directly with Dylan Wiliam the previous year, progress that teachers were making in these areas became known, and an opportunity for continued professional development became a reality. As many of these AfL techniques were being tried for the first time, teachers shared some of these practices with colleagues in team planning meetings and more informal gatherings such as in the faculty room. The development of a common language was facilitated by discussing what worked and what did not, and encouraging others to expand their teaching strategies by incorporating some of these new ideas. Traffic Lighting¹ and Two Stars and a Wish² are two examples of formative assessment techniques that are recognizable in almost all classrooms.

Laura: During the 2nd year of the Ho-Ho-Kus implementation, I observed in the classrooms of six teachers and interviewed them afterwards. The observations were designed to document the AfL strategies and techniques that teachers were using and learn how they were using them. Afterwards, I interviewed the teachers about how AfL had affected their teaching, and also asked them why they believed AfL had been received so enthusiastically at Ho-Ho-Kus. In these interviews, the teachers focused on two aspects of their school that they felt made it possible for AfL to gain a foothold among classrooms. The first factor was a supportive, progressive administration that was open to new ideas and trusted teachers to do their jobs. The second factor was the collegiality among the staff at the school, particularly among grade level teams. Some of the teachers indicated that their grade level teams frequently shared with each other which techniques they had tried and how they had worked in their classrooms.

Diane: The overwhelmingly positive results from using AfL prompted us to begin a formative assessment study group, independent of Dylan's sessions with us, where ideas could be shared and teachers could support one another in trying new methods and techniques. The inaugural meeting of the group took place in October 2005, and we then met a number of other times during the year. One exciting result was the way the group took the ideas of AfL and adapted it to help address some classroom management issues that some grade levels were experiencing. The group was able to utilize the ideas of setting measurable goals and sharing expected outcomes with students in a way that could affect classroom behavior while still calling upon the self-monitoring abilities of the students.

Laura: While Dylan introduced AfL to Ho-Ho-Kus, he had limited opportunities to work directly with the teachers. But the teachers took ownership of the strategies and techniques soon after they were introduced. After that, Dylan focused on helping teachers keep tight what needed to be kept tight, constantly challenging them to think “What’s formative about that [technique]?” Meanwhile, teachers adapted and experimented with the strategies and techniques within their classrooms and discussed with each other what they had tried and how it had worked. The primary way the teachers were able to be loose was in choosing which strategies and techniques were of most interest to them and best suited to the subject, grade level, and needs of their students. They were also able to extend their thinking about AfL strategies to assist with behavior and classroom management. Many of the teachers described adapting the self-monitoring aspects of AfL to help students monitor their behavior more effectively. While the KLT program explicitly addresses academic learning, the teachers felt that aspects of the strategies could be applied to students’ self-monitoring efforts.³ They also used methods of randomly selecting students (such as Popsicle sticks) to choose students for classroom activities such as taking attendance, passing out papers, or heading up the line to go to the next class, thus ensuring that all students had an equal opportunity to participate in the social and functional aspects of their time in the classroom. There was some discussion among ETS staff about whether these adaptations of AfL strategies and techniques were actually lethal mutations and indicated that teachers did not fully understand formative assessment. The adaptations I observed could be characterized as encompassing other purposes of education—beyond the academic—focusing on behavior and self-monitoring strategies to encourage students to learn to monitor their activities and actions. While AfL does not directly address applications of these strategies for non-academic purposes, the teachers used logical extensions of what they were learning about formative assessment, and many of them were both innovative and successful.

Conclusion

Laura: This paper combines two perspectives to reveal how AfL garnered a devoted following within one K-8 school, leading to the incorporation of AfL practices in many grades and subjects. This case study is important in that it illustrates one set of optimal conditions for professional development implementation and institutionalization, including a high level of institutional and administrative support—encouraging teachers to present about AfL during professional development days, providing release time for attending half-day sessions with Dylan, and attending sessions Dylan conducted to introduce administrators to AfL. Further, an exceptionally collegial, experienced, teacher workforce, a respected teacher leader whose enthusiasm for AfL was embraced readily by other teachers, ownership by teachers of the AfL strategies and techniques, and students and parents who were open to innovation in teaching and learning all contributed to high implementation. Moreover, it illustrates how the teachers in Ho-Ho-Kus were able to negotiate the tight and loose components of AfL. The teachers maintained a tight grip on the basic tenets of AfL while at the same time they experimented with applications of the techniques in furtherance of non-academic (classroom management and behavioral) goals. The teachers understood the academic focus of AfL and the appropriate strategies and techniques that they could employ to improve student learning and felt that similar strategies and techniques might improve student self-monitoring as well. They created these adaptations with considerable input and support from other teachers.

Diane: AfL is a critical and crucial part of instruction in Ho-Ho-Kus School. Taking small steps can make great strides in incorporating some of these methods. Seeing students’

confidence in mathematics grow in such measure is always a positive. Using that confidence to go further and embrace greater challenges is the goal. In visiting classrooms, I see student ownership of learning that is exciting and continually adapting to the needs of students. I recently observed a kindergarten mathematics lesson where students were using beans that were glued to a wooden stick to measure the lengths of various pictures of objects. This teacher uses green, yellow and red cups for students to traffic light their understanding of concepts and activities. One boy had his red cup out, indicating that he had a question or problem, and I approached him. When I asked him what made this activity red for him, he responded that he just didn't see how this could work. He indicated that when he measured with one side of his stick the object was three beans long, but with the other side of the stick it was four beans long because the beans were different sizes. This child had internalized the concept of the need for consistent units of measure and was able to verbalize it as well. What followed from this encounter was a meaningful class discussion of how measurement works and the idea of recognized units of measure. Giving students opportunities such as these can only enhance the learning experience.

Laura: I, too, observed many interesting incidents that resulted from application of AfL strategies and techniques. In the music classroom, I saw the students' use of the self-assessment technique known as traffic lighting whereby students drew red, yellow, and green circles next to lines of music, to mark parts they were having trouble with and parts they had mastered. In a math classroom, I observed a student teaching two students using whiteboards to write their answers—in effect giving his own mini-lesson while the teacher was giving a mini-lesson of her own to other students. These incidents reflect how teachers were able to claim these strategies and techniques and incorporate them in their classrooms in interesting and sometimes unexpected ways, thus providing an illustration of how the looseness of AfL implementation should work. But the tightness of AfL implementation is reflected in the continued sharing of practice and discussing how the strategies and techniques they had learned could be adapted to new situations.

Diane: Through my years in education, I have observed that students are adept in figuring out what a teacher wants and what they must do to achieve within that classroom environment. Traditionally, much of instruction has remained teacher-centered and students have become more and more dependent upon their teachers to tell them what to do next and how to do it. I believe for learning to be truly meaningful, students must be involved and engaged in their own learning. The onus must be put upon the student if learning is to become transferable to everyday life and future endeavors. AfL provided a means to do that.

Laura: This paper has focused on showing one high-level implementation of AfL (the KLT program) in one school district, but the lessons learned through observing this process have wider applications. First, the implementation was tight in that the basics of the AfL theory were introduced directly by Dylan, thus ensuring that the first generation of teachers got a good, solid foundation to build upon. Second, the voluntary nature of the professional development opportunity worked well in this situation, since teachers who participated in the 1st year were motivated by a desire to learn more about formative assessment, while the teachers in the 2nd year had additional motivation because they had heard positive things about AfL. Third, the implementation took advantage of the collegial nature of the school, allowing the teachers to develop their own study groups (TLCs) and meet on their own terms. Since ongoing sharing and discussion of AfL strategies and techniques is a tight component of successful implementation, the fact that they were motivated and disciplined enough to form and conduct these groups on their own was key to the success these teachers had. Finally, the resources available to support

the implementation were limited to providing teachers with coverage so they could attend professional development with Dylan. This suggests that for a school or district that is rich in certain kinds of resources (motivated, experienced, collegial staff and a supportive administration), fewer monetary resources are needed to ensure a successful implementation. Thus, resources fall into the loose category, since flexibility about the types and quantities of resources is acceptable.

Why would this implementation work, or not work, for other districts and schools? In other words, what can we learn about scalability from this implementation? This implementation combined a strong dose of AfL by Dylan Wiliam with a particularly open and willing set of teachers, enabling the strategies and techniques discussed to grow over time, with one generation of teachers passing on their AfL strategies and techniques to the next, supported by regular, ongoing attention from Dylan. Having a combination of message and openness to the message is certainly a basis for a strong implementation, but there was also, in the case of Ho-Ho-Kus, a messenger impact—the teachers were engaged by the messenger, not just the message. In addition, a strong teacher leader, Diane Mardy, was central to the growth of AfL and its continued presence in the school. While having a well-respected leader is relatively common in schools, there is no guarantee that the leader will get behind AfL (or KLT). Thus, there was almost a perfect storm in terms of having the right pieces in place to insure a successful implementation. However, like perfect storms, this is a rare event. And while you know such an event is possible and can even be predicted, it is impossible to rely upon its occurrence. What we learned about scalability from this implementation is that this particular implementation—highly dependent on Dylan, Diane Mardy, and a very receptive school and group of teachers—is unlikely to be scalable.

Epilogue

It remains to be seen whether Ho-Ho-Kus will be able to sustain and build from the strong foundation they had in the strategies and techniques of AfL. Dylan returned to England, and Diane took an administrative path and is no longer a classroom teacher. Thus, it will be up to the second- and third-generation AfL teachers to carry the techniques and strategies forward. That is likely to happen only if some of the teachers have thoroughly incorporated what they learned from Dylan and Diane into their everyday teaching and into their conceptualization of teaching. Moreover, support for continued professional development around AfL was not negotiated, so the teachers are truly on their own. Thus, the issue of sustainability remains to be investigated.

Notes

¹ Traffic Lighting is where students use the colors of a traffic light to indicate whether they are having difficulty or feel that they are ready to go on. Using colored plastic cups or laminated circles, students can indicate at their desks that they need help (red), that they are not yet secure with the material (yellow), or that they are confident they know what to do (green). Similarly, they can use colored pencils or markers to draw circles next to questions on homework, classwork, or tests to indicate whether they found the problem easy, slightly difficult, or challenging. With this information, teachers are better able to assist students and assess when some students are ready to move on to other work while others require remediation.

- ² Two Stars and a Wish provides both teachers and peers with a way to acknowledge a job well done by listing two things that the student did well, usually in written work or in a presentation; these are the stars. The wish allows the teacher or peers to indicate one thing they wish they had seen in the work or would like to see in future iterations of the work.
- ³ Note that adaptations of strategies and techniques for behavioral management are not part of the framework of KLT (which focuses solely on academic learning) and should be considered cautiously. In particular, teachers attempting to use such adaptations in isolation from their academic component may have limited success with regards to improving student achievement.

The *Keeping Learning on Track*[®] Program in New Teacher Induction

*Jeff Maher (St. Mary's County Public Schools, MD) and
Dylan Wiliam (Institute of Education, University of London)*

Introduction, Background, and Context

While there is ample and increasing evidence that assessment for learning (AfL) can increase student achievement (Black & Wiliam, 1998b; Crooks, 1988; Kluger & DeNisi, 1996; Natriello, 1987; Nyquist, 2003) much less is known about the best ways to support teachers in developing these practices. We were aware that a significant contributory factor in the success of the initial AfL projects (e.g., Black et al., 2003) had been the choice and flexibility given to teachers in choosing how to implement AfL, but equally important was that choice had been constrained, so that the focus stayed on AfL—the Tight but Loose metaphor that is the organizing theme for this collection. If teachers benefited from having a menu of options from which to select, then presumably, the same would be true for districts. For that reason, we were concerned to develop a range of models that districts could use in implementing the *Keeping Learning on Track*[®] program (KLT)—the professional development package that we were developing at ETS—and a number of these models are described in this collection. The impetus for the project described in this paper came from a chance meeting between the two authors at the annual Pathwise[™] professional development conference organized by ETS in Philadelphia in June 2004.

St. Mary's County is one of 24 local jurisdictions (23 counties and the City of Baltimore) in the State of Maryland. St. Mary's County is a peninsula with an area of 372 square miles and 500 miles of shoreline on the Patuxent River to the northeast, and the Potomac to the southwest. The county has a population of approximately 100,000, and a population density approximately half the Maryland average. The economic activity of the county is diverse, including agriculture, fishing, tourism, and a number of defense related industries, although the per capita income is within a few dollars of the U.S. average. In Maryland, school districts are organized on a county basis. St. Mary's County's 24 public schools (16 elementary, four middle, and four high) enroll approximately 17,000 students, with another 3,000 students attending one of the 31 private schools within the county. At the time, St. Mary's County Public Schools (SMCPS) had recently completed the development of a mission statement, which signaled its intention to “establish, maintain, and communicate high expectations for teaching and learning while supporting a tailored approach to system initiatives, based on the needs of individual schools through instructional leadership; standards-based curriculum; analysis of data; systematic and focused staff development; and allocation of resources” (SMCPS, 2004).

In addition, the mission statement committed SMCPS to ensuring that “All instructional decisions will be guided by proven practices and ongoing assessment” and work on the development of an action plan for using assessment to guide instruction was well underway. Since then, the mission statement has evolved to the following: “Know the learner and the learning, expecting excellence in both. Accept no excuses, educating all with rigor, relevance, respect, and positive relationships.” This mission succinctly emphasizes the point that learning above all is the target. An important starting point for the development of the action plan was the work of Wiggins and McTighe (2000) and Marzano (2003) in establishing the clear connection between the goals of instruction, designing instruction backwards from there, and monitoring the results to provide clear and specific feedback on how to reach the goals. This was combined with

the work of Richard Stiggins (Stiggins, 2001; Stiggins, Arter, Chappuis, & Chappuis, 2004) on the design of student-involved assessments to yield two priority areas in the action plan:

1. Fine-tuning the long-range and daily lesson planning structure to refine practice
2. Creating and using assessments to guide and inform instructional decision-making

The district's original focus on "assessments to guide and inform instructional decision-making" had been on what might be termed medium-cycle (William & Thompson, 2007) formative assessment—using more or less formal assessment episodes that would yield evidence of student achievement. However, in discussion, the advantages of a focus on short-cycle formative assessment became clear, and a formal decision to begin the project was taken.

One clear advantage of the Tight but Loose formulation is that it allows a rapid evaluation of whether a particular context is likely to provide a fruitful environment for a particular intervention. Within SMCPs, the existence of a new teacher induction program provided a ready-made forum for the introduction of the KLT program, and this seemed a particularly useful context to explore the implementation of KLT.

At some level, the key ideas of AfL would appear to be relevant to all teachers, whether they are just beginning to teach in the context of a preservice teacher preparation program, or whether they are veterans with 20 or more years' experience. Certainly we have found that with experienced veterans, breaking habits developed over many years is difficult, but for novice teachers, it may be that the basic classroom processes are so novel, so far from being routinized, that AfL is at least as difficult to develop. Through the pairing of a teacher new to the district with a more experienced mentor, the new teacher induction program offered the possibility of exploring whether the same professional development would be equally appropriate for those at the beginning of their teaching careers and for experienced teachers.

Initially, to participate in the scheme, teachers had to volunteer as a pair consisting either of a teacher in the 2nd year of teaching in the district and her or his mentor. However, there was also interest among more experienced teachers in participating, even though they were not mentoring novice teachers, so coteaching pairs (typically, a regular education teacher paired with special education teacher in a coteaching inclusion setting) were also allowed to join. In this paper, we describe the experiences of the teacher pairs in the 1st year of the project, drawing on evidence collected at the monthly meetings of the participating teachers and their reflective accounts completed over the pilot year.

Beginning the Project

Having agreed to a way forward, the next step was to secure support from administrators by outlining the research base for KLT to a meeting of school and district administrators in December 2005. The project was formally launched in March 2006 with a presentation to 600 SMCPs teachers followed by workshops with the 16 pairs of teachers who had volunteered to be a part of the pilot project. The assignments of the teachers and their teaching experience (in years) are shown in Table 1 (all names of teachers and schools are, of course, pseudonyms).

Through two further meetings in May 2005, teachers learned more about the main ideas of KLT and refined their action plans, ready for the new school year in September 2005. These meetings were supplemented by a series of readings (e.g., Black & Wiliam, 1998b; Leahy et al., 2005) and all participating teachers were given a copy of a book describing the experiences of a cohort of British teachers attempting to make similar changes in their practice (Black et al., 2003).

Table 1*Teaching Assignments of Participating Teachers*

Name	School	Assignment	Experience
Janice	Birchtree ES	Kindergarten/Grade 1	14
Mandy	Birchtree ES	Kindergarten/Grade 1	1
Penny	Hazeltree ES	Grade 3	16
Christopher	Elmtree MS	Grade 6 mathematics	8
Mark	Elmtree MS	Grade 6 mathematics	31
Bill	Goldentree HS	Grades 9-12 mathematics	7
Marcus	Goldentree HS	Grades 9-12 mathematics	36
Sophie	Goldentree HS	Grades 9-12 art	1
Darlene	Goldentree HS	Grades 9-12 art	8
Diana	Goldentree HS	Grades 9-12 media	30
Jennifer	Goldentree HS	Grades 11-12 English	20
Tricia	Goldentree HS	Teacher for the deaf	4
Marian	Gingko ES	Teacher for the deaf	3
Claire	Gumtree ES	Grade 3	10
Juliet	Gumtree ES	Grade 3	6
Margaret	Linden MS	Grades 6-8 IRT	7
Pamela	Linden MS	Grade 6 mathematics	3
Joe	Maple MS	Grade 8 language arts	20
Lacey	Maple MS	Grade 8 language arts	3
Margarita	Poplar ES	Kindergarten	1
Rebecca	Poplar ES	Kindergarten	4
Sally	Poplar ES	Grade 2	25
Kirstie	Poplar ES	Grade 3	30
Marcy	Poplar ES	Grade 3	26
Leanne	Poplar ES	Grade 3	3
Linda	Poplar ES	Grade 5	
Moira	Poplar ES	Grade 5	
Derek	Peachtree ES	Grade 5	7
Julie	Peachtree ES	Grade 5	1
Kelly	Cedar MS	Grade 7 social studies	7
Susan	Cedar MS	Grade 7 CTE	6
Megan	Willow ES	Grades K-5 IRT	18

Note. ES = elementary school, HS = high school, MS = middle school, CTE = career and technology education, IRT = instructional resource teacher.

Teachers also were encouraged to keep reflective logs, although these varied in detail substantially, with some teachers contributing only a few lines each month, while others generated two or three pages every two or three weeks. These journal entries, the final evaluation reports contributed by the participants, and the action plans together provide the main evidence base for this paper.

Although all the teachers completed action plans during the introductory sessions, only 27 of the 32 action plans were formally submitted and available for analysis. The five teachers

who did not submit action plans appeared to have participated little in the project beyond the introductory workshops—certainly no journal entries or reflective writing was received from them and attempts to get copies of their action plans were unsuccessful. Some form of reflective writing was received from all the 27 participants who submitted action plans. Below, we describe the main outcomes of the pilot in terms of the action planning process, the impact on teachers, and the impact on students.

Action Planning

Initially, the teachers' action plans tended to be in the form of long-term intentions rather than clear priorities for implementation at the beginning of the school year and tended to be lists of all the interesting ideas that the teachers had picked up during the introductory sessions. For example, Moira's initial action plan, compiled in May 2005 identified 18 different priorities for the coming school year. By September, this had been reduced to just five areas, all of which were related to classroom questioning (no hands up policy, use of Popsicle sticks to choose students, increased wait time, allowing students to pass or "phone a friend," and the use of some sort of all-student response system).

In her reflections at the end of the year, Tricia, a high school teacher for the deaf and hard of hearing, referred to the same process of reducing the number of priorities.

Even though my action plan started out bigger at the beginning of the year, it was good for me that way. It gave me the opportunity to see which ones were really where I wanted to focus and let me be flexible in the way I went about using them. I eventually reduced them, but stuck with the core idea that I had going into it.

As the teachers began to implement their plans, one of us (JM) supported the teachers with monthly meetings during which teachers talked about their experiences, did some new learning about AfL, and revised action plans, similar to the more formalized procedures now included in the KLT training materials (Thompson & Wiliam, 2007). Between meetings, the teacher pairs met more or less regularly, although these meetings were almost invariably outside the school day, and teachers found it difficult, if not impossible, to arrange to observe each other teaching. The log compiled by Christopher, a sixth-grade mathematics teacher at Elmtree Middle School, chronicles the frustrations one pair of teachers encountered in undertaking even a single peer observation.

September 22: Mark and I plan to observe each other once a week in the next couple of weeks. Mark will be observing my classroom during sixth period and I will be observing Mark's classroom during third period.

October 6: Mark and I have had a difficult time getting into each other's classrooms, but we talk on a regular basis.

October 20: Mark and I have talked regularly, but have not got in to observe each other's classes yet.

November 3: Mark and I still haven't been able to observe one another for one reason or another, but plan to observe each other a number of times in November, December, and January.

November 17: The other issue Mark and I have right now is when will be able to observe one another. Hopefully we will be able to start after Thanksgiving break and observe each other a number of times before our January meeting.

Changes in the Teachers

In their reflective accounts, almost every teacher mentioned time as an issue in the implementation of their action plans. Penny, teaching third grade in Hazeltree Elementary School said,

I wish administrators really understood the real nuts and bolts of good teaching. Many tend to be driven by a score mentality approach. It seems teacher discretion is very limited in school systems today. Many practices that really work seem to be forgotten.

However, echoing the remarks of participants in an earlier study (Black, Harrison, Lee, Marshall & Wiliam, 2004), most of the teachers tended to regard time as a resource to be optimized rather than as a constraint that rendered the implementation of their action plans impossible.

Also echoing the findings of earlier studies (Black et al., 2003), the freedom given to teachers to develop their own plans led to some uncertainty. One teacher wrote “I wish I knew what Dylan really wants us to do” and another wrote,

I wish we could share with others what they are doing. I wish I knew how the others feel about the process. I wish we could work together to generate ideas. (Susan, seventh grade information technology, Cedar MS)

Many of the teachers spoke, and wrote, about the fact that their work on A/L cut across some well-established practices in their schools, and in some cases even appeared to contradict school district policy.

At first, I was nervous because this is so different from the way that many classrooms are set up, but with time, practice, and most of all collaboration I feel that this was a smooth transition that could be successful in any classroom. (Tricia, Goldentree HS)

Some of the teachers also found considerable resistance from the students themselves: “This was a challenging year for me. I was truly pushing myself to think outside of the box and not give out. The students did not embrace many of my ideas at first (second, ...)” (Kelly, seventh grade social studies, Cedar MS). She went on: “I realized early in the year that the students were not used to independent work. They wanted their hands held during all lessons. They resisted all attempts at first. I had to regroup and introduce my ideas on a slower scale.” One student in her class wrote, “Why did we do so much work today? We are only 12.”

Many of the teachers reported feelings that some quite profound changes in their practice had taken place. Juliet, a teacher who ran a support group for third grade students, realized how much she had changed through a conversation with a colleague.

Towards the middle of the school year this particular teacher asked me if I would give her students a grade from their time spent with me. I explained to her that grades were not a component of our time together. I went on to tell her that our group was successful because students knew that tasks we were completing, and the assessments we were conducting, were used as tools for learning. I went on to explain that students knew that they were to learn from our time together and that they would not be judged by a grade. I

also told her that students were always expected to improve their work to exhibit growth in their learning.

I remember walking away from the teacher feeling confident about my position but believing that she did not really “understand” what we were doing. I remember feeling that, in her, I was looking at a reflection of myself from the past. In the past I had been a teacher that relied heavily on grades. I remember believing that if I assigned an assignment, and students spent time completing it then it should be graded so students understand that I value the time and effort they put forth. I never demonstrated through this process that I valued their learning. Well, a couple of months went by, and the discussion I had had with the teacher was forgotten until she approached me one day. On this particular day, she approached me, and thanked me for working with her students. She went on to comment about how well they were doing as a result of time spent in our group. I remember thinking, even though she didn’t say it, that she finally believed that student progress did not need to be linked to a grade. In the future, I would like to sit down with this teacher and explain that this is not my success; it is the students’ success. Without having pressure of grades looming over their heads, these students learned, like I did, how ongoing assessment should be used to improve their own learning.

Every teacher who contributed reflections felt the process to be beneficial for their practice. Some just reported a feeling of being more effective in their practice, while others had much clearer insights into the benefits, such as the ability to judge the optimum pace of the instruction more effectively:

I think using the [popsicle] sticks has provided me with immediate direction in my lesson. I can pace the lesson more efficiently. In math I have noticed that I thought they were getting the lesson, but when questioned with the sticks, the level of understanding that I wanted wasn’t there. I slowed my lesson down, so that I was able to achieve the understanding of the objective. Sometimes the opposite is true. I have had to change my lesson in mid-stream because they understand more than I thought they would when I was planning out the lesson. This has worked out great for all concerned. I have less behavior problems. I can keep my students engaged and on task easier. (Kirstie, Poplar ES)

While all the teachers who contributed reflective journals spoke of the benefits of AfL techniques, and most teachers managed to sustain and to build on their successes, it was clear that the developments were fragile. One of the most reflective of the participants, who had contributed regular monthly logs, described in her final reflection in May 2006 how her attempts at implanting AfL, while initially successful, had foundered halfway through the year.

Before writing this final reflection, I reread the compilation of reflections that are in my binder so as to “reflect” on the whole year of participation in AfL. As I read the reflections, tears welled in my eyes as to the deflation of my excitement of implementing these learning tools. I had such high expectations, hopes and dreams for my students at the beginning of this year. As the year progressed, I was not able to see total implementation of my Action Plan, but rather disintegration of my Action Plan as well as the lost hopes and dreams. Even though I try daily to return to that “different” classroom, a classroom with think time, a “no hands” policy, Phone a Friend and Pass Cards and the use of Popsicle sticks to encourage student participation, it is within the first few minutes of the day that all is lost to behavior and discipline issues. ... The end of year summary of action research reflects last year’s results as well as the beginning of this year, but is not

indicative of the last half of this school year. I saw great things occurring when able to implement the AfL techniques, and I would begin a new year with the same expectations, hopes, dreams and AfL tools in my toolbox. I believe in what I had implemented, and I know that they make a difference in student learning. (Moira, fifth grade, Poplar ES)

While AfL can clearly produce substantial benefits, it is clear that its successful implementation is far from straightforward, even for experienced practitioners.

Changes in the Students

Almost all of the teachers reported positive changes in their students, either in comparison with the same students previously or with previous students. One of the changes most commonly mentioned by teachers was with the degree of attention students paid in class:

I have been using sticks for student assessment. It is interesting how everyone sits up straighter and pays more attention when I pull them out. (Kirstie, third grade, Poplar ES)

Susan, teaching seventh grade information technology, found that students were much more motivated during the AfL pilot than previously. In the previous year, only 56% of students had submitted all required assignments, whereas during the AfL pilot, 96% of her students had submitted all required assignments, and the average scores were also higher.

All of the teachers who had included questioning in their action plans mentioned benefits from the use of increased wait time.

I have noticed that the extra effort on my part in allowing extra wait time better motivates the students to produce questions of their own.... They produced higher order questions such as, "Why do you skip the ones place on the decimal side of the place value chart?" When students begin to evaluate and produce questions, they also begin to construct their own knowledge. Students are now empowered because of their active participation in their own learning. (Pamela, sixth grade mathematics, Linden MS).

Sticks are working well for me. I have less noise in my class since students have finally figured out that I am not going to call on them when they raise their hands unless I ask for hands. Using sticks has also helped wait time. Students have no choice but to sit quietly and wait until their classmate has answered the question. I have definitely seen the benefit of wait time lately. One particular student in my class will just shrug his shoulders and try to wait me out. I have not given in, and just this last week, he has answered several questions correctly. (Marcy, third grade, Poplar ES).

Another strong theme in the teachers' reflections was the development of increased student autonomy in their learning through self-assessment. Kelly (seventh grade social studies, Cedar MS) described the difference she noticed in her students' self-awareness over the course of the year. She gave the following examples of self-assessments to the question "What did you learn today" given by her students at the beginning of the year:

I did nothing today.

I did something about Social Studies.

I did worksheets today.

Today I learned that I do not like the way my group members look.

Towards the end of the year, she reported that the same students were making far more sensible, reflective comments such as the following:

In class today, we learned about the human messing up the geography of the area. These people are stupid because they are hurting the future.

Today we worked in group to see what is happening to the Aral Sea. This was hard because a group member did not want to work. We ignored him. Today's lesson was very science. I thought this was Social Studies. [The result of Kelly's attempts to introduce a degree of cross-curricular work!]

In her final evaluation, Juliet, a teacher who ran an intervention group for third grade students, wrote,

During the time I am providing feedback to a student, others are either listening to help provide peer feedback, or working to improve their own work. In this environment, students are very honest about their work. Knowing that there is no report card grade linked to our group work, they are more willing to honestly reflect on their own progress. Before this situation, I would have expected students to be less willing to revise and edit their work since there is no grade linked to it. However, the environment we have created together encourages all students to do their best for themselves, not for others. In the two months that I have worked with my comprehension groups, there has only been one time where a student was unable to completely revise their work to their satisfaction due to time constraints. Despite being forced to stop working, this student was calm and willing to move on because she knew that the point was to understand how to improve her work and even though she was no longer going to have time in class to correct her work she knew how she could have revised her work to make it better and she was thus comfortable with moving on.

By using assessment *only* for learning, I got to see how students took ownership of their learning. They no longer relied on me to tell them how they were doing. They knew the criteria and goal that was set for them. They were able to monitor their own progress and they could tell me when they were ready to exit the group.

She went on to describe an incident where she was discussing with a student how much progress he had made:

He was telling me how proud he was of himself. He was pointing out the progress he had made since joining our group and how happy he was that he had accomplished his goal. As I was walking him back to his class, he asked if he could come back occasionally so that I could check his progress. I had said of course he could come back if he felt he needed. After a minute, he told me, "Never mind." He went on to explain, in his own words, that he would be able to evaluate his own progress and instead of me checking his progress for him, he would know from his ongoing self-assessment if he needed to come back and join our group.... I couldn't have received a better compliment. This student left our group with the understanding of how he could assess his own learning. Though he didn't verbalize it, he grasped the idea of using A/L, and not for "a grade."

Other teachers noted that students were also able to sustain their own learning without constant support from the teacher:

The discussions that we had every week, whether they were based on questions I asked or ones they had thought of, were amazing. Sometimes I would forget that they were nine year olds. It was like listening to adults discuss and debate over their interpretation of what was happening in the story. They came up with issues and points of view that I would never have thought of. (Megan, K-5 resource teacher, Willow ES).

The students were held accountable for their own learning and it appeared in front of them each week with their individual graphs. It made the students more aware of what they know and what they did not know. The test wasn't just tucked away after it was taken. They knew what they had to do to improve the graph the next week. (Sally, second grade, Poplar ES).

Anecdotal evidence from the students themselves, as reported by the teachers, indicated that these changes were on the whole, appreciated by students. Students in Pamela's class (sixth grade math, Linden MS) made the following comments.

I think the Popsicle sticks are a great reliable resource in order to pick a student. In the classroom we each had an equal chance of being picked. We have a 1/24 chance of being picked.

I believe that these Popsicle sticks are a good idea. I think it is a good idea because many people raise their hand and people may get called on more than once. I like them because everybody gets a chance. Also it is fair to everyone. I like the Popsicle sticks.

Of all the A/L techniques to which the teachers had been introduced, the use of comments, rather than grades, presented perhaps the most significant challenge. However, where teachers had moved, however cautiously, from grades to comments, the experience had been positive.

I have not issued a letter grade on any work so far this year [October 13th] and grades are not even mentioned by the students or the parents. I find writing comments that provide real insight to be laborious but I have observed a small but significant improvement in the written work of my students. I also have students reflect upon and evaluate their own work habits and quality of participation and written assignments. When I return written work students read my comments and can revise and resubmit. We look at and critique anonymous examples of good and poor student work so that they may see a range and compare their own to it. I think their focus is on what they are learning rather than what grade they received as it had been in past year. (Mark, sixth grade mathematics teacher, Elmtree MS)

Overall, there was a suggestion that younger students took to these changes more easily than older students, although it appears that even the high school students could be won round to the changes in classroom practice in time:

Written critiques are often not well received at the beginning of the school year as students do not expect to write much in the art classroom. But they become second nature after the first few projects. (Sophie & Darlene, ninth to twelfth grade art, Goldentree HS)

Impact Within the District

Since the implementation of KLT in SMCPS was designed primarily to explore the development of different models of teacher professional development, specific studies on the impact on student achievement were not designed into the study. Nevertheless, it has been possible to shed some light on the likely impact on student achievement by examining changes in student scores on state mandated tests. Although the specific work of the pilot was undertaken with a small number of teachers, as a result of the meetings with administrators and the presentation to the 600 teachers in March 2005, work on AfL has begun to impact the work of SMCPS systemically, and it seems possible that this is a factor in the rises in test scores seen across the district between 2005 and 2006, as shown in Tables 2, 3, and maybe 4.

Table 2
Percent Increase in Students Scoring Proficient/Advanced in Mathematics 2005-2006

Grade	All students	White	African American	FARMS	Special education
3	6.0	5.8	6.5	2.9	18.5
4	9.3	9.3	12.1	9.0	11.2
5	1.5	-0.2	6.5	2.9	6.8
6	9.3	9.3	9.5	9.7	17.7
7	9.7	10.7	9.9	9.1	1.0
8	7.4	7.5	7.2	5.5	7.9

Note: FARMS = free and reduced meal students.

Table 3
Percent Increase in Students Scoring Proficient/Advanced in Reading 2005-2006

Grade	All students	White	African American	FARMS	Special education
3	5.0	4.7	6.8	4.8	12.4
4	0.4	3.3	-9.7	-8.5	-2.2
5	2.1	0.6	8.4	6.3	8.7
6	0.9	0.5	2.0	-4.0	5.9
7	8.7	9.2	11.8	12.7	3.8
8	-3.4	-4.3	-0.2	-1.0	0.3

Note. FARMS = free and reduced meal students.

Table 4
Percent Increase in Students Scoring Proficient/Advanced on the HSA 2005-2006

	Algebra			Biology			Government		
	2005	2006	Gain	2005	2006	Gain	2005	2006	Gain
Cypress	57.9	74.1	16.2	73.0	81.1	8.1	67.2	80.1	12.9
Goldentree	33.1	58.3	25.2	43.3	75.2	31.9	56.7	72.8	16.1
Laburnum	49.5	62.0	12.5	75.3	83.4	8.1	76.3	85.4	9.1
SMCPS	58.3	70.2	11.9	66.1	80.1	14.0	67.2	79.8	12.6

Note. HSA = Maryland State High School Assessment, SMCPSS = St. Mary's County Public Schools.

On the Maryland State High School Assessment (HSA), the proportion of SMCPs students scoring proficient or advanced in Algebra/Data Analysis increased from 58.3% passing in 2005 to 70.2% passing in 2006. This surpasses the Maryland state average of 66.6% passing and is the first time the SMCPs score has been above the state average. Furthermore, all disaggregated student groups saw significant gains in achievement. Of special note, the proportion of special education students scoring proficient or advanced rose 13.5% and that of African American students rose 12.3%.

For the Government HSA, the proportion of SMCPs students scoring at proficient or advanced rose from 67.2% in 2005 to 79.8% in 2006, again taking SMCPs above the state average of 74.2%, with all minority groups showing even greater gains. The proportion of African American students scoring at proficient or advanced showed the largest increase—21.2%—and that for special education students rose 12.8%.

The highest pass rate was 80.1% on the Biology HSA, up from 66.1% in 2005. This bested the state's pass rate of 67.8% by 12.9%. On this test as well, student groups showed remarkable increases, with African American students leaping forward by 26.5% and special education students gaining 20.2% over their 2005 scores.

Of course, it is impossible to attribute these increases solely to KLT—KLT was only one of many initiatives being implemented within SMCPs during 2005. Nevertheless, these substantial improvements in student achievement show that KLT can play an important role in systemic school improvement. Perhaps most importantly, these improvements have been built through sustained developments of teacher competence that provide a basis for continued improvements. As Juliet put it in her end-of-year reflection:

In the end, I learned more from this process than I had ever imagined. I am a better instructor now that I have refined my thinking about assessment. This process has not only improved my thinking inside the classroom but also outside of it. As teachers, we all participate in the Teacher Performance Assessment System. I participated in this assessment system for 2 years without ever fully understanding exactly what is meant when it identified formative and summative assessments. After 3 years I can honestly say I finally get it! Thanks!

In May 2006, a second cohort was established. This second cohort extended to other teachers in schools across the county and again included coaching partners. Some of the practices emphasized in KLT have begun to spread more widely within the schools in SMCPs and have begun to be integrated into more general school improvement processes. For example, Team Action Plans are developed and implemented on a quarterly basis for each department or grade-level team. These plans identify specific goals, assessments of those goals, professional development needs, and extension/remediation approaches as a result of the assessments

Conclusion

From previous studies (e.g., Black et al., 2003), we know that the development of AfL can have powerful impacts on teachers' classroom practices and on student achievement, although those studies involved relatively close involvement from researchers in the field, and therefore would not, by themselves, provide models that could be taken to scale. The current study has shown that more limited interventions, when appropriately supported by school district staff, can lead to the kinds of profound changes in teacher practice found in earlier studies. Moreover, the current study has shown that the findings of the earlier studies, conducted in the United Kingdom, appear to be applicable in the United States and can be successfully integrated

with other reform efforts at district level. Because the current study was not designed to investigate the impact on student achievement, it is too soon to say whether the impact on student achievement is comparable to that found in earlier studies, but the success of the current project suggests that such investigations are worth undertaking. It will also be interesting to see over the coming years whether the initial interest in, and impact of, AfL in St. Mary's County Public Schools can be sustained.

The work in SMCPS is also useful in illustrating that effective interventions in a district may not come in through the front door in the sense of a mandated, top-down, district-driven approach to reform.¹ The resources provided for the mentoring program, in terms of staffing, a venue for the meetings, and legitimacy for the content, created a policy space in which the KLT program could be introduced in the county. The work of the participating teachers provided an existence proof that demonstrates at least the relevance and feasibility of the practices of AfL within SMCPS, if not their efficacy. The tight but loose formulation allowed the program to build on the affordances available within the district while maintaining its adherence to the research that motivated the intervention in the first place.

Epilogue

Two and a half years after the start of the project, in October 2007, Jeff Maher had this to say, “We continue to make major strides in our achievement, especially in regards to closing achievement gaps. As one example, *all* of our elementary schools made AYP, and those elementary schools who were on the ‘school improvement’ list published by the state of Maryland are now off, as a result of the progress they have made. Our high school assessments have gone up as well—with pass rates near, and sometimes exceeding 90%, placing the county first in the state in biology, and in the top five in other subjects. I cannot express how effective AfL has been for us. We may not have the hard evidence to prove that AfL is what has made the difference, but the soft evidence is in the teaching we see every day.”

Notes

¹ We are grateful to an anonymous reviewer for suggesting this metaphor.

Letting Go of the Reins: Learning About Scalability in the Context of One District-Wide Implementation

*Christine J. Lyon (ETS, Princeton, NJ) and Donna Cleland and Maureen Gannon
(Math Science Partnership of Greater Philadelphia, PA)*

Introduction

The *Keeping Learning on Track*[®] (KLT) program is a school-based teacher professional development program that supports teachers to use assessment for learning (A/L) in their everyday teaching, via sustained, school-based teacher learning communities (TLCs). There is strong evidence that assessment used with the primary goal of supporting learning can result in significant learning gains (Black & Wiliam, 1998b; Brookhart, 2005). Through professional development opportunities, the KLT program exposes teachers to a wide range of classroom techniques, all unified by a central idea, using evidence of student learning to adapt instruction in real time to meet students' immediate learning needs.

During the development of the KLT program, scalability was deliberately designed into the program by ensuring that components could be delivered by people who were not the original developers themselves, but instead could take advantage of districts' internal and external capabilities. To test this, ETS researchers and facilitators from the Math Science Partnership of Greater Philadelphia (MSPGP) collaborated in a district-wide implementation in the School District of Huntington Township. The MSPGP acted as a third-party provider and took responsibility for collaborating with the School District of Huntington Township to plan the implementation and provide ongoing guidance and support, while ETS provided the content and expertise in A/L. The purpose of the pilot was to move responsibility for the program further away from the original developers by relying on MSPGP staff to be the primary district contact and method of support. To accomplish this, expertise within both A/L and facilitation of the KLT program had to be developed within ETS, the MSPGP, and the district.

The paper will describe the results of this pilot beginning with the events that led up to the district-wide implementation. The district-wide implementation took place over 2 years and resulted in two distinct phases. The first phase, which illuminated several problems, also provided knowledge that was used to shape a second more successful model. Both of these models, as well as an evaluation of their strengths and weaknesses will be described. Finally, a description of how each version differed from the standard implementation of the KLT program will be provided with a discussion of how those differences supported or hindered each implementation.

It is first necessary to describe what is now viewed as the standard implementation of the KLT program in order to explain how this pilot differed. Participating teachers are introduced to the Five Key Strategies and One Big Idea of A/L through a 2-day introductory workshop. This workshop explains each A/L strategy, its empirical research base, and practical classroom techniques that can be used to implement the strategy within a classroom. Teachers are provided with opportunities to deepen their understanding of each strategy and a few specific techniques by engaging in hands-on, interactive learning activities. A subset of teachers are then identified to serve as TLC leaders. During a second 2-day workshop, these TLC leaders are introduced to the research supporting the value of school-based TLCs, provided with materials designed to ensure effective leadership, and provided with opportunities to engage in several activities designed to support the deepening of conversations during TLC meetings. Finally, all

participating teachers are provided with ongoing support through monthly TLC meetings. These meetings are facilitated by the trained TLC leaders and use a set of structured AfL activities developed by ETS—the KLT program modules (ETS, 2007). These modules provide directions and materials for a series of 16 meetings, each lasting 2 hours. To support the TLC leaders throughout the program, leaders meet two to three times a year to reflect on their own implementation of AfL, discuss the facilitation of their TLCs, and plan for upcoming TLC meetings. Although this represents the standard implementation that has been advocated by ETS for several years, the components of this implementation and the theory of action for the KLT program were not adequately communicated to the third-party provider at the start of this partnership. In addition, although there is empirical support for the use of AfL and the standard implementation was research based, there was no empirical evidence supporting the use of the standard implementation to impact teacher practice with regard to AfL.

Background

ETS and the MSPGP began their partnership in the 2003-2004 school year. Building on previous work (Black & Wiliam, 1998b), ETS staff had begun developing a professional development program designed to enhance teachers' use of AfL in the classroom. The goal of initial development was to design an effective and scalable professional development program for training teachers in the principles of AfL through an initial workshop and ongoing and sustained support.

The MSPGP is supported by a National Science Foundation funded research grant. The grant has three broad goals:

1. To improve student access to and achievement in mathematics and science
2. To improve the quality, quantity, and diversity of the teacher workforce through professional development
3. To understand, through research, how to effectively perform those tasks

The MSPGP facilitates collaboration between 46 school districts in Pennsylvania and New Jersey, 13 colleges and universities, and a number of nonprofit organizations. The organization primarily provides free professional development and administrative consultations to schools and districts within the consortium.

In the spring of 2004, the MSPGP expressed interest in offering professional development focused on AfL, and at the same time, ETS was interested in expanding the research base of formative assessment to include American teachers (at that time the research and development had been conducted primarily in the United Kingdom). ETS and the MSPGP jointly offered a professional development opportunity to the MSPGP's 46 districts. Two middle school teachers from the School District of Huntington Township attended these initial trainings¹. Both teachers were enthusiastic and began implementing formative assessment within their classrooms. As a result of this enthusiasm, the district became interested in offering this professional development to all its teachers and approached the MSPGP in the spring of 2005.

The school district of Huntington Township is a suburban, middle class district with low minority student enrollment. The administration is deeply committed to student achievement and tries to stay abreast of current trends in professional development, so there was strong central support for the initiation of the project. The district includes five elementary schools, one middle school, and one high school and serves a total of 5,475 students. In 1986, 1987, and 1989, all the schools won recognition as Schools of Excellence at either the national or state level. Test scores

are well above state and national norms, and 90% of students go on for further education after high school. The school district uses inquiry based curricular materials and are pioneers in the use of some of these materials. The district was viewed as a good choice for district-wide scaling, since it was performing well and had strong central support for the project.

The Initial Huntington Implementation

The initial district-wide implementation was planned, organized, and implemented by the MSPGP with facilitation and content assistance from ETS staff. The implementation included a one-day introductory workshop for participating teachers, a half-day administrator and principal workshop, and three follow-up workshops. Teachers were to be supported throughout the year by regular professional learning community (PLC) meetings². Each of the workshops and PLC meetings are described in detail in the following sections.

Introductory Workshop

The project began in June of 2005 with a 1-day workshop focused on AfL. The workshop was held on a district-wide professional development day, and all teachers in the district were required to attend. The morning presentation was facilitated by ETS staff and focused on the Five Key Strategies and one Big Idea of AfL, the research supporting these strategies, and practical techniques for classroom implementation. Afternoon breakout sessions organized by grade level or discipline were led by administrators from the district. Each group's task was to choose one of the five strategies that the group would like to pursue in the fall.

Administrator and Principal Workshop

ETS staff held a second half-day session for administrators and building principals over the summer. This session was designed to garner support from the administration. An overview of AfL and the empirical support for its implementation were presented. Although all district level administrative staff members were present, the conversation remained at a surface level. As a result, this staff was generally supportive and amenable to the implementation, but little enthusiasm for or dedication to the effort was demonstrated.

Follow-Up Workshops

There were three follow-up workshops in the initial implementation. The first was facilitated by ETS staff on October 31, 2005. A 4-hour session focused on the strategy "Providing feedback that moves learners forward" was presented. Teachers were provided with a brief review of the research and were introduced to additional AfL techniques that support this strategy's implementation during this time. Teachers were then divided by discipline and practiced one technique—Comment Only Marking—by working collaboratively to write formative comments on student work. In general the workshop was not well received. The workshop was held in a school auditorium, and poor acoustics made it difficult for staff to hear the facilitator and for the facilitator to interact with small groups. When teachers dispersed into break out groups to work on comments, some groups were disappointed that they were not learning more about the key strategy of interest that they had identified in June.

The second follow-up workshop was facilitated by MSPGP staff in November of 2005. This workshop was a full-day, interactive session offered as part of the district's catalogue of professional development offerings. Teachers chose workshops from the catalogue to participate in. Fewer than 20 teachers registered for the AfL workshop, and as a result the administration

conscripted several teachers into the workshop. Although this session did serve as an opportunity to identify teachers who were genuinely interested in AfL, in general, there was a lack of teacher buy-in and enthusiasm.

The final workshop was also facilitated by MSPGP staff in February of 2006. Participants broke out according to content area groups to discuss the problem, “How to find out what students know” during this hour and a half workshop. They then participated in a matching activity to familiarize themselves with several new AfL techniques. Each of these groups was facilitated by a curriculum supervisor who was provided with guidance and materials by the MSPGP staff.

Professional Learning Communities

Follow-up was to be done through professional learning communities (PLCs) organized by grade level or discipline and led by curriculum supervisors within the district. Each PLC had previously chosen one of the five strategies to focus on, and time during each of the curriculum supervisor’s monthly meetings was to be devoted to discussions about the implementation of this strategy. All participants were provided with a copy of the book, *Assessment for Learning: Putting It Into Practice* (Black et al., 2003), and curriculum supervisors were provided with a list of questions for each paper that were intended to guide the discussions. Despite access to these materials, supervisors reported that they were uncomfortable facilitating the meetings, and they felt that they did not know any more about AfL than the teachers. In fact, some of the teachers who were genuinely interested knew more than the supervisors. Finally, these conversations did not occur with any regularity. The curriculum supervisors had difficulty finding time to include AfL discussions in their meetings, and many of the participating teachers felt limited by the mandate to focus on one particular strategy.

Evaluation of the Initial Implementation

The initial implementation differed from the standard implementation of the KLT program in four ways that limited its effectiveness. First, it is clear that changing teachers’ practice in ways consistent with AfL requires structured, ongoing support. However, the mechanism for providing this support was ineffective on several levels. There was no structure to ensure time for teachers to meet in a setting where the sole focus was on AfL. The curriculum supervisors were asked by the assistant superintendent to add these discussions into an already full administrative agenda, and often time was not reserved for these discussions.

Second, the supervisors did not have sufficient materials, training, or support to lead these meetings. Although the supervisors were provided with copies of the book, *Assessment for Learning: Putting It Into Practice* and a list of guiding questions, they were still uncomfortable with their role. Since the KLT program assumes that teacher leaders are developing their AfL expertise along with members of the TLC, structured guidance is provided through a set of modules for use by the facilitators of the learning communities. These modules provide activities that facilitate conversations focused on deepening and expanding teachers’ understanding of AfL. Although the book and guiding questions may have provided a starting point for the discussions, there may not have been sufficient structure or activities to guide the types of conversations that are necessary to deepen teachers’ understanding and enhance the implementation of AfL.

Third, the PLCs were facilitated by curriculum supervisors who in many instances did not have classrooms. One of the mechanisms by which teachers develop expertise is by sharing their

experiences with a group, discussing their successes and challenges, and refining a plan for the next month. Therefore, the KLT program is run on the assumption that expertise is developed in AfL by the implementation of a technique within a classroom. For this reason, the follow-up discussions are most beneficial when facilitated by teachers who are learning about AfL with their colleagues.

The limited follow-up provided by the PLCs resulted in low levels of teacher buy-in and follow-up support from the administration. Traditional professional development efforts are typically one-shot efforts that move by quickly and have little impact. As a result, most teachers do not expect professional development initiatives to be sustained over time. This leads to a general mistrust and feeling that “this too shall pass” which fuels teacher reticence to buy in. Although at the onset of the project the MSPGP understood the need for sustained professional development, the specific method of implementation (e.g., PLCs with no time solely devoted to AfL) was not sustainable. In later conversations with the district, the need for a commitment to deep, ongoing professional development rather than a one- or two-shot whole-district effort was emphasized.

Finally, this initial implementation required groups of teachers to come to a consensus on one strategy that the PLC would focus on. The goal of the KLT program is for teachers to focus on a range of techniques that will support all five strategies. However, teacher change is slow and often initially focuses on one strategy. Over the course of time, additional strategies and techniques may be incorporated. Although all five strategies should be incorporated into a teacher’s classroom, not all techniques are appropriate in every context, and, therefore, the techniques chosen will vary by teacher. As a result, neither the order in which a teacher approaches the strategies, nor the specific techniques used should be dictated. The structure for the initial PLCs did not reflect this thinking, but rather limited the teachers’ implementation of AfL.

It is clear that the initial implementation failed to impact teachers’ AfL practice and student achievement. The theory of action for KLT and most teacher professional development is based on a three-part model: teachers learn about a better way to teach through professional development, then teachers adopt the better approach to teaching, and finally student learning is improved because of these improvements in teaching (Thompson & Wiliam, 2007). The first Huntington implementation failed to teach teachers a better way to teach, since the primary vehicle for teacher learning, the PLCs, either stopped meeting or did not focus on AfL. Since the first step in the theory of action was not achieved, we can therefore assume that there was no impact on practice or student achievement.

A Second Huntington Implementation

Although the first implementation was less than successful, staff from both the MSPGP and the districts strongly believed that AfL could produce positive changes in teacher practice, student engagement, and student achievement. As a result, ongoing evaluative conversations with district administration led to a second implementation, which was planned by MSPGP staff. This implementation included more hands-on support from the MSPGP and spanned 2 years—the remainder of the 2005-2006 school year and the 2006-2007 school year. Since the district and MSPGP recognized the difficulty of adding yet another topic to the curriculum supervisor meetings, a decision was made to switch from PLCs to TLCs. The TLCs would be run during district professional development days, and the time would be reserved solely for discussing AfL and its implementation. The 1st year of the implementation focused on building capacity within

the district by training a core team of TLC leaders. The 2nd year focused on rolling out the district-wide implementation, with TLC leaders facilitating TLCs for all teachers while receiving ongoing support from MSPGP staff. Each aspect of the intervention is described in the following sections.

TLC Leader Training and Support

Although the initial implementation included all teachers in the district, for the second implementation, the central administration decided to initially focus on a subset of highly motivated teachers. The administration invited approximately 35 teachers to participate in the TLC leader training. Teachers with the capacity for leadership or with an interest in AfL were chosen. In order to have representation from the five elementary schools, the middle school and the high school, four TLCs with approximately eight teachers per TLC were formed. Elementary teachers formed two of the TLCs, and middle school and high school teachers formed the other two TLCs.

Each TLC meeting was facilitated by two MSPGP facilitators and used one of the KLT modules. The modules provide a detailed agenda, discussion points, and blackline masters for each activity. Every TLC module includes the following five activities:

1. *Introduction and Housekeeping.* This activity provides time to address any housekeeping concerns, announcements, or group norms to ensure that the group is ready to focus on AfL. In addition, the learning intentions for the meeting are shared with participants. These statements describe what participants should understand, know, or be able to do by the end of the meeting. We know from the research that it is important to explain to students what they will be learning and this is modeled through the sharing of learning intentions at the start of each meeting.
2. *How's It Going?* This activity provides time and structure for every participant to give a brief, informal report on his or her latest AfL efforts. During each report, participants listen attentively and use a structured protocol to give feedback and support to their colleagues.
3. *New Learning.* This activity provides time for participants to consider a particular aspect of AfL and build on the material presented at the initial workshop. The aim of this section is to deepen teachers' understanding and broaden the repertoire of techniques that they can use in their classrooms.
4. *Action Planning.* This activity provides time for participants to plan out what AfL techniques they will try out in the next month and identify any support or help that they will need to do so.
5. *Summary of Learning.* This activity provides time for participants to revisit the learning intentions to see whether or not they were achieved. Just as presenting the learning intentions models good classroom practice, revisiting these during the last activity in the TLC meeting also models one of the practical classroom techniques.

MSPGP facilitators reviewed the materials and evaluated and modified each module as it was used with each of the four TLCs.

The How's It Going? session during each TLC emphasized the expectation that teachers were implementing the AfL techniques in their classrooms and encouraged rich, reflective

conversations about practice. Teachers shared techniques that they had developed themselves, which often became ideas that other teachers adopted for use. Although the new learning in the modules is often most valued by teachers, the collective expertise and creativity shared by teachers during How's It Going? activities often has the greatest impact. Teachers deeply enjoyed the opportunity to share with their colleagues, and their reports often act as existence proofs, which encouraged teachers in the same context to try something new. Since teachers share the same context, they often feel that another teacher's suggestion is more valuable and attractive than a suggestion from an outside facilitator or written materials.

Each TLC leader participated in four 3-hour TLC meetings, which covered the new learning from seven KLT modules. Teacher leaders were allowed the opportunity to discuss and practice the skills in each module by modeling them in the way that they would use to facilitate that same module the following year. This provided additional support for the learning of the leaders and allowed the leaders to develop their own AfL practices before leading a TLC of their own.

After the 2005-2006 school year, TLC leaders reconvened for a 2-day summer session designed to support them as leaders. The 2 days were spent reviewing the modules and planning for turn-around training during the 2006-2007 school year. During the 1st day, the five-part structure of the modules was presented, and participants were given time to become familiar with the structure and format of the materials. During the 2nd day, the two modules for the first two TLC meetings were selected. Each TLC was assigned one leader who worked to become familiar with the content of the two modules and plan for their presentation. Many of the TLC leaders worked in pairs to either prepare or present the modules. In cases where TLC leaders teamed up for the presentation, TLCs were combined into larger groups of up to 30 teachers.

Once the school-based TLCs got started, the TLC leaders continued to meet with MSPGP staff during the 2006-2007 school year. Three meetings were planned—one in October, one in January, and one in April. The purpose of these meetings was two-fold:

1. To continue to deepen the leaders' understanding of AfL
2. To reflect on the leaders' experiences as facilitators and build expertise in the facilitation of TLCs

A three part agenda was planned to accomplish both goals. Meetings began with a How's It Going? activity focused on the teachers' implementation of AfL. Next, participants experienced the new learning from a Year 2 module, and finally, the meeting ended with a How's It Going? segment focused on their experiences as leaders. The final activity provided time for participants to review the module that they would facilitate with their TLC and assign tasks and responsibilities to ensure a smooth flowing meeting.

The four TLCs were collapsed into two with approximately 15-20 participants per TLC, due to substitute shortages and scheduling constraints. The new organization resulted in TLCs that included participants from all grade levels and content areas. Participants were asked to attend one of the two meetings, but depending on substitute coverage, participants could switch between meetings each month. In addition, the leader TLC meetings for the 2006-2007 school year were shortened from 3 hours to 2 hours.

An AfL How's It Going? activity was successfully run during the first two meetings; however, many teachers continued to discuss their most successful strategies from the previous year, and there was less talk of new techniques and fewer discussions regarding techniques with which they needed help. In general, there was less enthusiasm for sharing their experiences. This

activity also took much longer than usual because the group was larger. Both meetings also included a How's It Going? activity focused on their experiences as leaders. In the October meeting, this time was spent discussing the needs and plans for the initial TLC meetings, while in January, participants discussed their roles as TLC leaders, more specifically discussing concerns that surfaced as a result of their first two TLC meetings. Due to time constraints, the lengthened AfL How's It Going? activity, and the additional How's It Going? activity focused on their experiences as leaders, there was no new learning presented during either of these leader TLC meetings.

TLC leaders presented some interesting concerns during the How's It Going? activity focused on their experiences as leaders. Responses varied but centered on four themes:

1. The TLC leaders' perception of the support they were receiving from their principals
2. The TLC leaders' perceptions of district expectations for them
3. The TLC leaders' personal comfort level with the role of facilitating their groups
4. The TLC leaders' perception of their own expertise with AfL

Although these concerns were focused at a higher level and consumed more time than was anticipated, the activity itself was deemed valuable, since the teachers could not be successful leaders unless the concerns were addressed. Unfortunately, district staff was not pleased that time was spent encouraging teachers to express their concerns, preferring instead that the How's It Going? activity was minimized and the leaders were trained on the new learning in additional modules.

Teacher Learning Community Meetings

TLCs were organized by grade and subject area. There were approximately 15 participants per TLC, and each meeting was facilitated by one or two TLC leaders following a KLT module. Meetings were scheduled for district professional development days and were mandatory for all staff. TLCs were identified as the sole focus for the morning or afternoon sessions, providing approximately three meetings, each 3 to 4 hours in length. During longer meetings, TLC groups combined similar new learning topics and ran one long TLC meeting. During shorter meetings, the groups were instructed to choose one new learning topic. Every meeting included a How's It Going? activity and an Action Planning activity. Leaders were generally positive about the first two meetings, although many leaders found it necessary to split the group for the How's It Going? activity due to the large numbers and time constraints. District staff visited each building after the initial TLCs to debrief with teachers. Teachers were asked to complete a one-page evaluation of the initial meeting, and the collated responses were all very positive.

Since the TLCs were only scheduled to meet four times during the school year, building principals agreed to devote time to the KLT program during at least two faculty meetings. During this time, teachers were to split into groups of 10. Each teacher was to share his or her experiences with at least one AfL technique, while group members listened attentively and provided feedback or suggestions for improvement. Each group would then choose one report to share with the larger group. Due to competing priorities and initiatives, the principals have struggled to find time to facilitate these discussions within the already full faculty meeting agenda; thus, the activity has not occurred at any of the participating schools.

Evaluation of the Second Implementation

The second implementation was far more successful than the initial district-wide attempts. Given the failure of the first implementation to cause teacher learning, the existence of ongoing TLCs with a focus on AfL is a success, because it is within these structures that teacher learning occurs. During these meetings, teachers reported on their use of AfL techniques in the classroom. This indicates that teachers did learn about a new way of teaching, but also that they had begun to adopt this new way of teaching. Although we have no measure of the frequency with which these techniques were used or the quality of the practice, the reports provide an initial indication of changes to teacher practice. Future research should focus on those issues of frequency and quality as well as on an examination of the impact on student achievement.

This second implementation took into account many of the features of the KLT program that have now been acknowledged as necessary standard features. First, ongoing support for the implementation of AfL was ensured by providing separate meeting times for both the TLC leaders and participating teachers. TLC leaders were provided with release time to reflect on their own experiences with AfL, learn the new modules, and plan for facilitation. Every teacher in the district was assigned to a TLC that met during contract hours on a district-wide professional development day. Since this time was reserved for AfL, the focus could be maintained. Second, the TLCs were facilitated by teacher leaders. These leaders were provided with additional training in AfL, written materials and activities, and ongoing support for facilitators. Finally, although the focus of the new learning for each module is oriented around one of the Five Key Strategies, teachers were not confined to the implementation of a single strategy or set of techniques.

Although many aspects of this implementation were brought into alignment with the standard model of implementation, other aspects differed. First, the 3-year implementation provided every TLC leader with a year's worth of experience—both as a member of a TLC and as an implementer of AfL—prior to becoming a facilitator of a TLC. This additional training and experience has allowed leaders to be more comfortable and confident with the concept of AfL and the materials they are presenting to their colleagues. A second difference is the use of district professional development days for TLC meetings. Each TLC was scheduled to meet for approximately 3 to 4 hours, three times during the school year. Although the meetings have been more successful than they were in the initial implementation, the professional development days are scheduled intermittently, with either too much time or too little time between TLC meetings. In order to maintain the momentum during long lapses between meetings, *How's It Going?* activities during faculty meetings were planned. Unfortunately, these activities have not materialized. This was identified as a shortcoming of the current implementation, and the district has agreed to hold a monthly AfL TLC during the 2007-2008 school year.

Finally, the district has developed a professional development plan that combines AfL and differentiated instruction. Both topics have been allocated time on district-wide professional development days. Even though the two initiatives are presented as separate topics, the leaders have reported that they felt there was a competition between the two initiatives and the information delivered was too much to absorb at once time. Although with careful planning it may be possible to combine two complimentary initiatives, the way in which time is allocated should be thought through. It may be more beneficial to have entire days focused on one topic rather than allocating a half day to each initiative.

A complicating factor for the 2006-2007 school year was a large shift in central administrative staff. The school year began with a new superintendent, a new assistant

superintendent, and only three of the original five curriculum supervisors, all continuing in positions of increased responsibility. Also, one elementary school was redistricted, and the teaching staff had been redistributed as well. All the positions were filled from within the district, so these individuals had existing reputations and were somewhat familiar to the teachers. Although the new assistant superintendent has definitely committed to supporting AfL, there was a period of transition where MSPGP facilitators had to work closely with the central administration to ensure the future of the program.

Conclusions

The work in Huntington has provided a number of key insights for the development of a standard implementation of the KLT program. In particular, the experience has highlighted the structure and communication necessary to successfully involve a third-party provider and the support and advocacy required to address the struggles and challenges inherent in implementing a new professional development program.

An intervention must maintain fidelity to its core principle or theory of action while remaining flexible enough to respond to the local context to be effective. This can be described as the Tight but Loose framework that was explicated in the opening paper. The standard implementation described in this paper provides an outline of one model that can be used to implement the KLT program. An analysis of the strengths and weaknesses of the two implementations in Huntington provides support for keeping several aspects of this standard implementation tight. First, providing ongoing support with the sole focus of AfL is essential. The initial implementation in Huntington lacked this ongoing, focused support. Although the original implementation plan recognized the importance of this aspect, the combination of curriculum supervisor meetings with AfL resulted in insufficient time to accomplish both agendas. The second implementation utilized TLCs during district-wide professional development days and resulted in a stronger implementation and more regular follow-up.

Based on evidence from the Huntington implementation, the second aspect which must remain tight is that TLC leaders should have a classroom to practice in. There are many arguments to support the role of teachers as leaders. In addition to having the opportunities to learn about AfL by implementing the techniques in their classrooms, involving teachers as leaders can help to increase the participation of teachers in school decision making, can assist in recruitment and retention of the best teachers, and has the potential to improve the implementation of new programs (Ellsworth, Martinez, Lyon, & Wylie, 2007). The PLCs in the first implementation were led by curriculum supervisors. These supervisors did not feel comfortable in their roles as leaders and did not have the opportunity to deepen their own knowledge of AfL because they did not have their own classrooms. The second implementation allowed time to train teacher leaders and provided those leaders with opportunities to learn the content, experiment within their own classrooms, and support one another.

The third tight aspect identified through studying the implementation in this district is that if at all possible, TLCs should contain four to eight participants. During the second implementation, the four leader TLCs were combined to create two larger TLCs with approximately 15 participants. Facilitators found that the How's It Going? discussions during these meetings were not as productive. The groups became too large and included too many grade levels and participants who did not know one another. This made it difficult for every participant to share and for participants to learn meaningful techniques from one another. It is

important for the groups to remain small to allow for the exchange of knowledge to occur and to allow the groups to develop an identity and establish a culture of safety and support.

Finally, a core principle of the KLT program, and one that must remain tight is that teachers must choose the specific techniques that will work for them, given their teaching styles, students, and curriculum. This principle requires teachers to understand the conceptual framework of AfL in order to determine what will and will not work in their specific contexts. In the first Huntington implementation, PLCs were required to choose one strategy for the entire PLC to focus on. This limited the teachers' ability to use their professional judgment and also limited the group's exposure to the variety of techniques that teachers may have adopted.

Although analysis of the two implementations provided support for many of the tight aspects of the KLT program, it also highlighted several areas where the program can respond to the local situation. The standard implementation provides a 2-day workshop for TLC leaders. The TLC leaders then learn the new content along with their colleagues. Previous implementations have identified several problems associated with TLC leaders learning while they lead (Ellsworth et al., 2007). The second Huntington implementation provided TLC leaders with 4 months of training, which gave leaders time to not only learn the new content but also deepen their own understanding of AfL, internalize the information, and prepare to present the materials before they were responsible for the understanding of an entire group. Due to the problems previously identified, the Huntington implementation model may, in fact, be more favorable than the standard implementation. Unfortunately this model does require that the district-wide roll out is delayed for 1 year, and that 3 to 4 months is invested in training TLC leaders. In addition, the standard implementation requires monthly 2-hour TLC meetings. The second Huntington implementation utilized district-wide professional development days for the TLC meetings. This use of time seems appropriate as long as the sole focus of the meetings is AfL and scheduling is done to ensure that there is enough time between meetings (3 to 5 weeks) for the teachers to try new AfL techniques, but not so much time (more than 6 weeks) that the focus on AfL is lost.

In addition to supporting the tight but loose framework for the KLT program, this case study also demonstrates the necessity of sharing a program's theory of action. ETS developers have always made the theory of action for implementing AfL explicit. However, the structures and support mechanisms necessary for teachers and teacher leaders to succeed in this program were not adequately described. For that reason, the MSPGP did not have access to the knowledge that ETS developers had about the structures that may be necessary to support districts, teachers, and TLC leaders in this initiative. The standard implementation provides this structure through the use of initial workshops, TLCs focused on AfL and led by teacher leaders, and ongoing follow-up support for TLC leaders. Following this implementation requires a district to commit time, resources, and support for the program. This commitment implies a certain level of district support.

To provide this type of knowledge to third-party providers, it is necessary for ETS to more closely examine how knowledge is shared between and among organizations. To some extent, the knowledge that must be shared can be equated to the cycle of knowledge creation and knowledge transmission described by Nonaka and Taekuchi (1995). This cycle requires that the internalized knowledge of one person or group, or the tacit knowledge, be made explicit so that it can be shared with another person or group. Since ETS had failed to make explicit the standard implementation and thereby provide the knowledge of the structures necessary to support teachers, leaders, and districts in their adoption of the KLT program, it was impossible for this

knowledge to be transferred to the MSPGP and/or the district. The dissemination of the standard implementation, the theory of action, the Tight but Loose framework, and case studies such as Huntington all provide one step in the knowledge sharing process. However, not only does this new knowledge need to be made explicit for new organizations, but in order for those new organizations to fully understand the importance of this information, they must make it operational. Once operational, the new knowledge can be combined with existing knowledge structures, internalized, and made accessible and useful in relatively seamless ways. To accomplish this, it is necessary for an organization with tacit knowledge of the KLT program to work relatively closely with any organization that is new to the program. The reach of the program can be expanded as additional organizations become familiar with the program.

In the case of the MSPGP partnership, the MSPGP was the primary district contact. It is therefore not only important to consider how knowledge was shared between ETS and the MSPGP, but also how knowledge was shared between the MSPGP and the district. For example, during the second Huntington implementation, district staff requested that the new learning in the modules be emphasized and all How's It Going? activities be minimized. The KLT program emphasizes that the new learning is the least important component of the TLC modules. The How's It Going? and Action Planning activities allow the knowledge of the group to be conveyed to other participants and utilize the valuable expertise of all participants. These components and their rationale should have been communicated not only to TLC leaders but also to district staff so that they understood why the How's It Going? activities were prioritized. This information has since been conveyed appropriately, and district staff currently understand that the conversations facilitated by the How's It Going? activity were useful and helped the leaders to be more effective. In future implementations, resources will be developed to support third-party providers as they share the theory of action and as they discuss those parts of the program that must remain tight and those parts of the program that can be loosened to take advantage of existing structures or district resources. These resources can help guide initial conversations and ensure strong implementation plans.

This case study provides a description of one model of implementation involving a third-party provider. The implementation was taken one step further away from the original developers by collaborating with MSPGP staff as the primary district contacts and mechanisms of support. The MSPGP staff took on the main responsibilities for holding conversations with district staff, planning the implementations, and ensuring that the initiative, teachers, and leaders were supported. As with any district-wide initiative, the program encountered struggles and challenges as evidenced by the two-pronged implementation. It is a credit to the MSPGP and the district that they continued to work to develop a successful, scalable model. This indicates that for a third-party model to be successful, the facilitators must have a strong relationship with the district, the ability to secure support from the district and teachers, and a certain tenacity and advocacy to see the program through to the end. The history of Huntington as a whole provides additional knowledge that will assist the development of similar implementation models in the future.

Notes

¹ Actual names of schools and the district have been changed to preserve anonymity.

² The term *professional learning community* (PLC) is used to describe a community of professionals within a school that is similar to a teacher learning community but includes and/or is led by administrative staff other than teachers.

***Keeping Learning on Track*[®] in an Urban District's Low Performing Schools**

E. Caroline Wylie (ETS, Princeton, NJ), Marnie Thompson (Fund for Democratic Communities), Christine J. Lyon (ETS, Princeton, NJ), and Donna Snodgrass (Cleveland State University, OH)

Introduction

The *Keeping Learning on Track*[®] (KLT) program has the ultimate goal of improving student learning by strengthening day-to-day classroom teaching. However, classrooms are typically situated within schools, which are generally situated within school districts. In this hierarchy, the district has responsibility for providing professional development that aligns with district goals, and the district has an important role to play in creating the conditions under which professional development can lead to actual classroom improvements.

This paper will connect the initial context and motivations that led a high poverty urban district to bring the KLT program to 10 of the district's lowest performing schools to the specific ways in which the program was implemented. In addition, the paper will illustrate ways in which good intentions were sometimes undermined by insufficient support throughout the system and describe how the local leaders and program developers responded to this situation.

In the first section, we will describe the implementation of the KLT program, focusing primarily on what happened during the 2005-2006 school year. The second part will be a reflection on the tensions between maintaining fidelity to the core principles of the program and the local constraints that must be dealt with in flexible ways. It will also investigate the various levels of support necessary for a program to be sustained and not fall prey to other forces that cause a loosening or unraveling of original intentions.

Implementation of *Keeping Learning on Track*

The Cleveland Municipal School District (CMSD) is a large rust belt school district with a declining tax base, more than 100 schools (many in a poor state of repair), a student population that comes largely from working class or poor households, and a long-standing record of weak academic performance. Under directed leadership in the late 90s and the early part of the current decade, the district made some significant improvements in attendance, graduation, and other indicators of performance. These improvements renewed hope for some staff, parents, and students within the district. However, recent funding cutbacks at the state level and the defeat of a tax levy led to sizable and demoralizing staffing cuts, increases in class size, building closures, and the eventual resignation of the CEO who had overseen many of the improvements. The announcement of her pending resignation at the start of the 2005-2006 school year had a profound impact on the district for the rest of that year.

A member of the central office administration had met Dylan Wiliam at an AERA session in the Spring of 2005. A few weeks later, an ETS staff member approached her to suggest a research partnership between ETS and CMSD. The project was seen as an opportunity for ETS to address issues of how to scale the KLT program within a district, and given that the district is hundreds of miles from ETS, it meant that ETS staff would not be able to provide any form of regular hand-holding. Instead, after the initial exposure phase, it was going to be important to develop capacity at the district level to support the teacher learning communities (TLCs) and help them keep a strong, clear focus on assessment for learning (AfL).

ETS staff were optimistic that they would be able to support the work in the district, albeit from a distance, because the relationship with the school district was brokered by the capable, highly placed central office administrator who had made the original contact with Dylan. It was clear that she understood the principles of AfL and its potential. She was prepared to prioritize this work in every way she could. Without such an informed, committed advocate, ETS would not have taken on this challenge—having learned the previous summer the critical nature of such internal buy-in (Lyon, Wylie, & Goe, 2006).

Prior to beginning any training in the district, ETS staff worked closely with the district leadership to ensure that district leaders had a clear understanding of the time, resources, and support required for successful implementation of the program. Specifically, expectations surrounding ongoing teacher participation and support throughout the year were outlined. These expectations were detailed in a seven-page memorandum of understanding, signed by the CEO of the District and ETS, which captured the roles and responsibilities of both parties. According to the terms of the memorandum of understanding, the level of support for the initiative from the district was to be substantial, and as will be seen later in this paper, much of this support did materialize. The process of drafting, redrafting and agreeing to the scope of work in a written document was an important part of the process of garnering the support that was gained.

Completely separate from the KLT program, the central office administrator had been working with a team of district staff, spending a considerable amount of time focused on what it would take to support teaching and learning in CMSD. This group had developed seven principles that codified the ideas that emerged from these conversations. The seven principles evolved as the team incorporated the ideas they learned from the KLT program and other research into their thinking. The central office administrator felt that, while this work had not been exclusively focused on formative assessment, it had laid a strong foundation preparing district staff to support teachers and schools who would engage with the KLT program. The intention was that this core group of district staff would provide support for the schools that would engage in the project. Each member of this core team was assigned a specific school, and the role of this person was to be that of an advocate for the teachers and TLCs with the school principals. District staff identified the schools that they wished to be involved: 10 Title I schools, each school in its 3rd or 4th year of failing to make Adequate Yearly Progress (AYP), as defined by the No Child Left Behind (NCLB) Act.

Initial training for the KLT program was conducted by ETS staff and included a 1-day training session for school and district leaders, a 2-day training session for the teacher learning community leaders, and three separate 1-day training sessions for the teachers in the 10 schools. These teachers were then instructed to form smaller, school-based groups of six to eight teachers, and meet on a monthly basis after school. We note that the choice of a 1-day session to expose the vast majority of teachers to the program was not our usual way of working (the initial workshop is normally 2 days long), but the district was unable to provide more time out of the classroom than that. Each training session, along with the subsequent teacher learning community meetings, is described below.

The professional development effort began with a 1-day workshop designed for the core team of central office administrators and the principals and other school administrators who were charged with supporting the initiative in the 10 schools. The workshop was intended to motivate these leaders by presenting both the research on AfL and the research on growing teacher expertise through sustained, school embedded TLCs. Unfortunately, on the day the workshop was scheduled, the district also scheduled a district-wide student head-count, which required all

principals to stay at their schools (what this cross-scheduling implied about actual versus implied support for the program will be discussed later). Thus, only the central office-based staff (about 15 people) participated. The workshop was well-received despite low attendance. The central office-based staff recognized the importance of developing expertise in both A/L and TLCs for supporting the initiative and therefore attended the 2-day workshop for teacher learning community leaders.

ETS staff conducted a 2-day initial workshop for teacher learning community leaders, with the first day focusing on A/L, and the second day focusing on their roles as leaders of learning communities following the workshop for district and school building leaders. Each of the 10 schools was asked to send seven teachers to the workshop. However, actual participation included 47 teachers, with two to seven teachers representing each school

Following the workshop for teacher learning community leaders, ETS provided a series of three 1-day workshops for the general population of teachers in the 10 schools. A memo sent out by the CEO of the district invited every teacher in the 10 schools to attend one of these workshops, and it was expected that across the three workshops a total of 200-300 teachers would participate. However, a district-wide shortage of substitute teachers made it difficult for many to attend, even though funds for substitutes had been obtained and made available for this purpose—there simply were not enough substitutes available. This meant that many teachers were forced to ask their colleagues to cover their classes or forego the workshop entirely. Another problem stemmed from the fact that even though clear, written descriptions of the purpose and focus of the workshops were provided to each school's principal for distribution to teachers, these were not always relayed to the teachers. The lack of information may have further influenced some teachers' decisions to forego the workshop. We also found that getting teachers to the workshop was one thing; ensuring that they could stay all day was another. On occasion at these initial workshops (and occasionally in the monthly leader workshops) teachers would receive messages from their principals to come back to their schools after they had arrived at the workshops. Sometimes this was due to a true emergency that required that particular teacher's attention, but in some cases, a principal just decided to revoke permission for the teacher to attend a workshop that they knew little or nothing about. The principals did not understand the purpose of the teacher workshops, since they had missed the workshop that had been designed for them. In spite of these challenges, approximately 180 teachers participated across the three 1-day workshops.

Once there, the teachers responded very positively (learning logs collected at the workshop contained predominantly positive comments with respect participants wanting learn more about formative assessment and expressing excitement about trying new ideas out in their classrooms), similar to the response seen in other schools and districts, to the KLT workshops. They were, however, skeptical of the ongoing support that was promised, and immediately worried that this was just the latest "reform du jour," to be replaced by the next education fad taken up by the district. ETS staff worked with district staff to assure them otherwise, but truthfully, it was not clear what support would materialize as the months went by. A significant concern raised by the teachers at these workshops was the pressure they felt to adhere to the pacing guides and scripted literacy curricula used in the district; they felt that this limited their ability to find time to try new ideas. The central office administrator assured teachers she would shield them from such mandates and support teachers who decided to emphasize student understanding over coverage. In actuality, this pressure was coming day by day at the school level, and the central office administrator could not always provide the promised level of support

or shielding. Given that the majority of principals had not attended the initial workshop, they were not in a position to understand how these mandates conflicted with the KLT program.

Once all of the workshops were completed, the monthly school-based TLCs began to meet. The monthly school-based teacher learning community meetings followed the protocols developed by ETS. These materials provided a consistent structure across meetings, with time allotted at the beginning of each meeting for each teacher to report on his or her efforts to engage with formative assessment ideas in the classroom and to get feedback and support from other group members. The middle section of the meeting was a time to focus on a particular aspect of formative assessment, building on the initial material presented at the workshop, and aiming to deepen teachers' understanding and broaden the repertoire of techniques that they could use in their classrooms. Finally, at the end of the meeting, teachers were given time for planning out what formative assessment techniques they would try out in the next month, along with identifying support or help that they would need to do so.

TLCs met with varying degrees of regularity in all 10 schools. Most of the participating schools had formerly been K-6 schools or 7-8 schools, and they had been reorganized into K-8 schools just that school year. Due to this conversion, many of the staff members from lower grades (K-6) and upper grades (7 and 8) did not know each other. Some TLCs had as many as 18 members (far more than was advised, but they did not want to break into smaller groups, this being their "only chance to get to know one another!") and some had as few as three (fewer than was advised, and these groups included only those teachers who were originally trained as teacher learning community leaders). A phenomenon that had not been expected was that in some schools, recruitment extended beyond the pool of teachers who attended the initial workshop. In several schools, the TLCs had more teachers who had not attended the initial workshop than had attended. During the year, most learning communities worked through five to seven modules. Since the series of introductory workshops finished in October and the first learning community meetings were held in November, this seemed to be a reasonable number of meetings.

Throughout the school year, the leaders of the TLCs also received ongoing support in the form of a monthly day-long workshop, usually under the leadership of the internal district champion, and three times under ETS leadership. These meetings were focused on deepening individual members' understanding of A/L, and developing the structures and support needed to effectively lead sustained school-based learning communities. This form of ongoing support for the leaders came about as a result of the central office administrator's insistence that they would need this kind of support. Though ETS had not previously developed such regular support structures for learning community leaders, ETS staff worked with the central office administrator to develop a structure that was used throughout the year, regardless of whether ETS staff or CMSD staff was running the workshops.

Leaders had an opportunity to both explore and deepen their own understanding of formative assessment, as well as to reflect on their experiences as TLC leaders at these learning community leader meetings. There was also time for school-based teams to plan for the next TLC meeting, review the materials that would be used, and assign tasks and responsibilities to ensure a smooth-flowing meeting. In addition, the central office administrator often provided various supplies that would support the teachers in their formative assessment work (e.g. index cards, materials to make white boards, paper for photocopying TLC materials). Another way in which the central office administrator demonstrated her commitment to these leaders was by

ensuring that TLC leaders received graduate credit for taking on leadership roles and attending the additional leader meetings.

Throughout the school year, the central office administrator struggled to push the initiative back to the top of the priority lists of high, middle, and lower level leaders, as other imperatives kept arising that pushed the KLT program further down the stack. She often succeeded. Because of her constant advocacy, TLCs were established in all 10 schools, and the strategies and techniques of AfL can be seen in many classrooms in these schools. Positive evidence of the impact of the KLT program on teaching and teachers' thinking is reported in Ellsworth et al. (2007) and Wylie et al. (2007). The Wylie et al. paper used three data sources to consider teacher learning: teacher action plans written towards the end of the introductory workshop, detailing the changes they planned to make in their actual practice; conversations about the implementation of these changes during teacher learning community meetings; and changes in actual classroom practice. These data allowed an exploration of questions focused around understanding which techniques most appeal to teachers after the initial workshop, which of those ideas continued to be talked about during the year, which ones were evident in practice, and what was the relationship among these three stages of professional development: initial exposure, discussion in learning communities, and actual practice. Overall, the data were very encouraging about changes in teacher practice and also pointed to the value of continuing the 2-year model for the program. Importantly, the teachers who began to implement formative assessment techniques served as existence proofs for their colleagues, demonstrating that the techniques that promoted AfL could be implemented in their own schools by their peers in this inner-city district. The Ellsworth et al. (2007) paper focused on how the KLT program supported the teacher leaders who facilitated the school-based TLCs during the first of a 2-year implementation, investigating how the TLC leaders deepened their understanding of AfL over time and how their roles evolved. The paper described the support structures that the district staff provided for the TLC leaders and examined how the leaders viewed their roles and what they did as leaders, and what they deemed to be significant in supporting the TLCs. More than two thirds of the designated leaders were active in their leadership roles, which is significant given some of the systemic struggles that the initiative faced.

On the other hand, there were also a number of disappointments along the way, in particular, the inability to get building principals on board in any significant way. The level of commitment across the 10 principals varied, ranging from one or two who were deeply involved to most remaining disinterested. In many cases, it was unclear what the principal knew about the initiative, and in some schools, it was not uncommon for the principal to schedule other activities that directly conflicted with the learning community meetings.

In many ways the support at the district and school level was quite strong. At the district level, day-to-day support came mostly from staff who reported directly to the central office administrator, whereas staff who reported to others seemed to let the initiative slide from their priority lists. The soon-to-depart district CEO made her support known by making a brief address at one of the workshops and by sponsoring a day-long seminar in December 2005 on the KLT program, which all principals were required to attend. Most principals were joined by their building-level union representatives, which was an effort to garner grassroots and union support for the initiative.

Funds were found to pay a stipend to teachers for staying after school to participate in TLCs, and there were funds available for substitutes so that teachers could visit in each other's classrooms. However, such visits were hampered by the severe shortage of substitutes, and so

they never got off the ground. Funding was also set aside to deepen and spread the work in the coming school year, during which the KLT program spread to an additional five schools, a sign that this effort was seen as more than just the latest fad. Given the tight finances under which the district was operating, the level of funds applied to the effort signal a very strong commitment to the effort, even though the funding was not always joined by a commensurate organizational resolve to prioritize this particular initiative.

How Tight but Loose Evolved in the Cleveland Municipal School District

One way to think about the site-based enactment of the KLT program in CMSD is to consider the framework of Tight but Loose (Thompson & Wiliam, 2007). It is instructive to review this specific story to identify ways in which the developers of the program, along with district staff responsible for its support, had a clear idea of what it was that they were trying to enact and why. This is codified in the program's theory of action, which is what we consider to be the tight part. However, holding tightly to certain principles is different from implementing an entire project with strict rigidity, since there are ways in which the program must be able to take account of, build on, and support specific local contexts. It is important, however, that any looseness be played out in a very conscious way. Conscious loosening involves a thoughtful discussion of the local context, proposed changes to the standard implementation of a program, and the effect that those proposed changes may have on the theory of action. Changes made as a result of these discussions imply that a principle of the standard implementation is not necessary for effective implementation of the theory of action. On the other hand, unconscious loosening may lead the program to lose integrity. Unconscious looseness is often a result of district pressures or constraints that are outside the control of the reform leaders. The process of periodic reviews, looking for evidence of both conscious and unconscious looseness, can support a retightening or revision of a specific implementation or of the theory of action for the entire program.

The remainder of this section is structured by identifying both the tight and loose aspects of the implementation. The loose aspects are further subdivided into those that were unconsciously and consciously loose.

Consciously Holding Tight

During the 1st year of implementation, ETS staff occasionally received information about ways in which the KLT program was unfolding in the district and felt a need to intervene in order to maintain fidelity to the original intent of the program. One example of this was when some teachers thought that a technique called Two Stars and A Wish was being mandated for all kindergarten teachers. Two Stars and A Wish is one of dozens of techniques related to the strategy of "Providing feedback that moves learners forward." It helps teachers structure feedback to students by selecting two aspects of the work on which to comment positively (the stars) and then selecting one aspect for the student to improve on (the wish). ETS had trained a number of observers to conduct TLC observations for research purposes and reported back from a TLC meeting that Two Stars and a Wish had been mandated, and there was significant consternation among the group.

ETS staff were concerned when they learned that some teachers thought that the technique was being mandated, since an important tenet of the program was that it was always an individual teacher's decision regarding which techniques to try. Furthermore, during the KLT program workshop, techniques are introduced as practical ways to enact a key strategy that has

been shown by research to improve student learning. The research provides guidance as to why a strategy should be implemented, not just the ways to do it. Given that the majority of the teachers in the district did not have the broader context for this technique or its larger strategy, ETS staff were concerned that the technique might be applied in a way that was inconsistent with the theory of action for the program. As a result, an ETS staff member called the central office administrator to ensure that district staff understood these concerns. In the course of this conversation it became clear that district staff, who could see the power of the technique, had incorporated it into other professional development work that they were leading, which focused on the use of rubrics. The intent on the part of the district staff person had never been to suggest that this technique was mandatory, but this was how it was perceived by the kindergarten teachers who had heard about it in the workshop. As a result of the conversation, the central office administrator said she would stress to her staff the importance of allowing teachers to choose which techniques they would use.

One place where ETS staff received some push-back from learning community leaders surrounded teachers' strong desire for a technique glossary. During the workshops, many techniques are both modeled and discussed, and a list of techniques is provided as part of the training materials. However, this list includes only the titles of the techniques, with no larger explanation. The leaders were continually pushing to be given a full glossary that explained the techniques in detail. ETS staff resisted this, because from narratives written by researchers observing some of the learning community meetings, it was clear that discussions of techniques used were largely superficial listings of techniques rather than detailed or reflective accounts of what was done and why. The resistance to provide the full glossary stemmed from a desire to see the teachers deepen their understanding and skill at a few techniques, rather than perpetuate a superficial attitude of trying out the complete list of all possible techniques. So, rather than providing the glossary, ETS staff worked at several of the leader meetings to demonstrate the quality of discussion expected as teachers debriefed on their progress. They also provided support to the leaders as they strengthened their facilitation skills, so that they could deepen the conversations at the learning community meetings. Even though the leaders probably never fully gave up wanting to have the glossary, over the course of several meetings, many of them also came to understand why ETS was reluctant to just hand it over. Towards the end of the year, as teachers were becoming more secure in their understanding of formative assessment, ETS did make the glossary available; by that stage, however, we anticipated that it would be useful as an *aide memoire*, given that it was now over 6 months since teachers had attended the initial workshops, and their memories of certain techniques was less clear. This story exemplifies a more nuanced case of tight but loose, where there is an interplay between the theory of action and the way local conditions changed over time.

Conscious Loosening

The KLT program was also able to be flexible in order to accommodate complimentary district efforts that enhanced the goals and methods of the program. As mentioned earlier, the central office administrator had done some work with her core team that resulted in the definition of seven principles that defined their work with teachers. The seven principles they developed were as follows:

1. Focus on learning; teaching is only of value if it results in learning.
2. Learning moves forward when teachers use information from ongoing classroom assessments to continuously adjust instruction in real time.

3. All learning targets must be clear and manageable stepping stones that create a path to achieving the Academic Content Standards.
4. Students can reach any learning target that is close enough for them to see.
5. The teacher's role is to systematically help students see each target that lies ahead of them, and to provide strategies that help students to move their learning forward.
6. Intelligence is learned and socially constructed when children have the opportunity to model their thinking after the intelligent thinking of others.
7. Evaluative assessments do not promote learning; however they are useful for measuring how much learning has taken place at the student, classroom, school, or District level at a given point in time for the purposes of progress monitoring and program evaluation.

We did not have materials developed within the KLT program to support district-level staff engaging with the process, and so the central office administrator developed her own materials. In some instances, the research and work that she drew upon was not related directly to formative assessment, but it did not conflict with the central tenets, and, therefore, created a sense of ownership of the KLT program among district staff. This is an example of being loose in order to build on local strengths.

Another way in which ETS staff accommodated the needs of the school district was through the development, along with the district staff, of the structure for the learning community leaders meetings. This leader support had not yet been developed prior to the start of work in Cleveland. Ironically, as ETS staff had the opportunity to interact with the leaders on an ongoing basis, it became clear that what started as something done to accommodate the district was actually a critical component of the program. As a result, the KLT program now emphasizes the need for ongoing leader support in other venues. This example illustrates the ways that a theory of action may be under construction and thus learn from local efforts. This is consistent with the idea of design research, as articulated by Lewis et al. (2006).

Unconscious Loosening

The project started in CMSD with district staff selecting the schools who would participate, and so the KLT program began with the 10 lowest performing schools. Given the current climate in school districts (especially under the No Child Left Behind Act), there is incredible pressure, not only to improve low performing schools, but to do it quickly. There was funding available for an intervention because of the status of these schools (which was not available to other schools in the district), and thus these schools were selected. Given a choice, ETS might not have selected these schools, but rather might have started with a set of schools that were in slightly better shape, perhaps recruiting schools on the basis of their ability to demonstrate school-level administrative support and commitment to the project. This is because an important component of KLT is that teachers have time to meet in the TLCs and that this time be protected (e.g., no other important meetings are scheduled at that time, those teachers are not asked to coach students at that time, etc.). Support from school administrators is required to free up this time.

However, in order to get district buy-in and funding, it was important for the project to begin with these 10 troubled schools. Therefore, in an effort to ensure buy-in from the principals

of these 10 schools and secure their support for the teachers who would participate, the 1-day workshop for principals and district support personnel was written into the memorandum of understanding. However, this training was not effective, since, as noted earlier, there was a student head count day across the district, which required principals to be in their buildings, thus limiting principals' understanding of the project and, therefore, their ability to support teachers in their buildings. The fact that the student head count day was scheduled on the same day as the principals' workshop may simply indicate serious communication problems, or it may signal a more fundamental problem, that there really was not serious support for the initiative. Either reason has ongoing implications for the potential for success.

Although not fully recognized at the time, the lack of involvement of principals resonated across the year. School level support was dependent on the principal, and was reflected in the smaller than anticipated number of learning community leaders who were able to attend the leader meetings. Given all the responsibilities and demand on school principals' time, not having the opportunity to be immersed in the program's supporting research and to understand the role and importance of the TLCs was going to make it much more difficult to ensure their support for their teachers. In hindsight, ETS staff should have worked with the district staff to try a second time to bring the principals together for a full day's training, knowing the importance of their roles. However, even without an adequate educational forum for the principals, one principal was educated by her own staff about the KLT program, to the extent that she began attending the leader meetings with her team. Another principal also appreciated the program enough to qualify learning community leadership as an official committee assignment (in this school, every teacher must serve on at least two committees). The CEO-sponsored seminar for principals and union representatives went some way towards bridging the knowledge gap, but the information given was at a more general level than it would have been had the meeting only been for the principals of the 10 focus schools.

Another bit of unconscious loosening occurred because of the difficulty in locating substitutes. According to the terms of the memorandum of understanding, teachers were also supposed to meet in between the formal learning community meetings in groups of two or three, to write quality questions together or to explore student-friendly learning intentions, or whatever reflective activity they selected as being of greatest value. However, due to other pressures within schools and the lack of substitutes, these meetings never really took place on a routine basis. The lack of substitutes in the district was not something that could be rectified by the central office administrator. Although initial implementation called for teachers to have 2 hours per month to work collaboratively on A/L, these additional 2 hours per month were abandoned, not because it was decided that the time was unnecessary, but because it just happened that way.

During the initial training for the TLC leaders, ETS staff recommended establishing TLC groups of six to eight, at most 10 teachers. It was assumed that each school would end up with two or three TLCs per building, and interested teachers would group themselves according to grade or content area, whichever made most sense, depending on who was interested in being part of a teacher learning community. At the first TLC leader meeting, it was clear that in some schools the groups were significantly larger. ETS staff reiterated the suggested group size, along with the rationale that in larger groups not everyone would get an opportunity to share what they had been doing, nor would there be sufficient time for them to get meaningful feedback and support. However, teachers from those larger groups pushed back, arguing that all of the upheaval in the district during the previous summer had resulted in teachers moving schools and grade levels, and that the TLCs provided an opportunity for people who did not know each other

to come together in a collegial, focused, and productive environment. The groups were adamant that they could ensure that everyone had a voice. In the end, ETS staff could not sway their decisions, and decided that given the leaders' promise to ensure that everyone would get adequate time, and the particular situation with newly configured schools, there was a plausible rationale for maintaining the larger groups. However, with hindsight, and informed by evidence from the TLC leader meetings, it is clear that, at least in the early school-based TLC meetings, not everyone in the group was taking time for personal sharing and getting feedback from the group about how they were implementing formative assessment ideas in their classrooms, even though this is a critical aspect of the theory of action. It was also clear from the leaders that they had not fully understood the importance of everyone sharing and that this was an important aspect of accountability. The difficulty in supporting everyone sharing during the meetings stems from the fact that this level of openness is not part of the cultural norms of most schools, and colleagues can be reluctant to expose less-than-perfect implementation efforts to the group. In more recent iterations of support for TLC leaders, efforts have been made to make the importance and the benefits of everyone sharing their practice more clear.

An additional concern about having larger groups was that an individual's sense of personal responsibility and commitment to the group would be diminished. An individual may think that it did not matter whether they showed up to some meetings if there were 15 others in the group. While ETS staff had not been able to get a consistent record of TLC attendance, based on TLC leader reports at the monthly leaders' meetings, together with TLC observations, there appears to have been a downward trend in attendance at TLC meetings as the year progressed.

Another way in which some TLCs deviated from the general guidance provided by ETS was with regard to the length and frequency of meetings. The 2-hour block once a month seemed like a non-negotiable aspect of the program when the project started in Cleveland. The materials were designed to be used during a 2-hour meeting, but in most schools, teachers had to vacate buildings within 90 minutes of the end of the students' school day, due to a bargaining agreement with the custodians. These groups could only meet for an hour or so after school. Some of these groups handled the problem by meeting twice a month for an hour at a time, and some met for 45 to 60 minutes every week. ETS staff did not object to the ways that groups adjusted, although they did suggest that groups try for two 90-minute meetings, since they would probably lose more time at the start and end of the meeting on two occasions than at a single monthly meeting. The teachers showed creativity and resourcefulness in finding time and locales for meetings, which suggests that looseness here might be appropriate, although there is no available evidence of what worked and what did not work. Furthermore, the fact that the 2-hour block was unachievable indicates that something was seriously limited about the schools' or district's commitment to the program in the face of conflicting programs or ways of doing business. Even when the district office volunteered funds (left over in the substitute teacher kitty) to pay custodians' overtime rates to fix the problem, no school took up the offer.

One sign of the positive impact that the KLT program was having at the building level was that in some cases teachers had joined a learning community even though they had not attended the initial introductory workshop. ETS staff had not anticipated this in the original design for scaling within the district. However, the teachers had obviously heard something of interest from those who had attended the initial workshops, or perhaps were just interested in the stipend, but regardless, no one wanted to discourage them from exploring formative assessment within the TLCs. This was another example of an unconscious loosening of the implementation plan. We do not have evidence of the impact that having TLC members who missed the initial

workshop had on the TLC meetings. However, it appears that these individuals were able to glean enough about the main concepts of the project, since they remained involved over the course of the year, rather than dropping out due to feeling lost in the process. At the school level, having more teachers involved meant the project was more visible, which may impact its longer term viability.

Beyond the 2005-2006 School Year

In spite of these struggles in the 2005-2006 school year, there was still sufficient success that the program was expanded to five additional schools in the following year. Of the original 10 schools which had started in September, at least seven had managed to establish TLCs that met on a regular basis, and the other three met at least once or twice. The central office administrator organized a showcase in April, which provided an opportunity for schools to display ways in which the KLT program had impacted their instruction to a wider audience. Furthermore, an initial analysis of the state-wide assessment results showed that the schools that had taken part in the program had improved math and reading results over other schools in the district, and later analyses confirmed these results (Wylie, Thompson, & Wiliam, 2007). The analysis indicated a general pattern in both reading and mathematics across multiple grade levels that the KLT program schools showed greater gains (although they did not approach significance) than district schools that did not participate in the program. It was these combined results that drew in an additional five schools for the 2006-2007 school year. ETS staff were able to present these district results to the new teachers at their introductory workshop as motivating existence proofs that changes to student learning could be made.

In August of 2006, the district leadership changed, and a new CEO was hired. Although the new leadership expressed strong initial interest for the KLT program, concrete support did not materialize. The new leadership brought many of its own projects to the district, and these competed with the KLT program for funds, school attention, and teacher time. The new leadership team was very focused on the use of interim assessments as an early warning system for low performing students.

Nevertheless, the central office administrator was able to preserve the funding for the expansion of the program to the five new schools. In the 2006-2007 school year, then, she coordinated monthly leader meetings for the leaders in the original schools who were now using the 2nd year of the KLT materials, in addition to separate monthly meetings for the learning community leaders in the new schools, who were using the 1st year materials. While the five additional schools are for the most part enthusiastic about their involvement, the first cohort is struggling and losing some of its initial momentum. One of the original 10 schools was closed over the summer (for reasons having nothing to do with their participation in the program), and leaders from two other schools rarely attend the monthly leader meetings. Of the remaining seven schools, most can send only one or two of leaders to these meetings. Furthermore, while the leaders from one school attend the leader meetings, they in fact are not holding school-based TLC meetings due to so many competing school priorities. Although they maintain that conversations about formative assessment are continuing on an informal basis, the impact of these is bound to be significantly less than it would be in extended, dedicated meetings.

In some sense, the 2005-2006 school year was a rudderless year for the district. The KLT program was able to function to some degree in that chaos, but since everyone knew the superintendent was leaving, it was virtually impossible to build real district momentum and support. Everyone knew that the new superintendent's preferences would determine the future

direction, so there was a pervasive attitude of waiting to see what would happen next. Hence there was just enough room for the KLT program, but no strong commitment to it. This figures greatly into the way things played out, though it may not have been apparent at the time.

District pressure on lower performing schools is significant and often results in these schools being asked to get involved in multiple new initiatives. As a result, the teachers are struggling, first to identify which of all the “most important” things are in fact critical, and second, having decided the initiatives with which they can, or have to, be involved, to find the time for everything. It is becoming increasingly difficult for the TLC leaders to find time to meet at the district level and for teachers to attend in-school TLC meetings. There is slightly less pressure on the five new schools, perhaps because they are not quite as low performing. Time is viewed as the most critical resource needed for teachers to engage in the KLT program; time to reflect on the ideas of formative assessment as they relate to their own context, time to share about successes and challenges and receive feedback and support from peers; and time to plan upcoming changes for their classrooms. Time, then, is a critical aspect of the theory of action and is often not in good supply.

In spite of the high quality ambassadorial efforts on the part of the district administrator who championed this program, she has been unable to protect the teachers from the additional programs and efforts in which they are required to engage. While she has been able to promote the achievements of last year, these results have been insufficient to shield the teachers. She has been unable to create special dispensations that would allow them to focus on the KLT program solely rather than the battery of programs and initiatives that have come their way.

Conclusion

The work in CMSD followed what has become a fairly standard implementation of the KLT program, with introductory workshops to give initial training to both teachers and leaders (a subset of the teachers), followed by the monthly school-based learning community meetings led by a teacher from that school who had participated in the leader training. In addition, the monthly learning community leader meetings provide additional support for those leaders.

There are several aspects of this work in Cleveland that can be considered successful. The majority of the participating schools completed their 1st year of TLC meetings and progressed to the 2nd year. The TLC meetings and the materials used were of sufficient focus and quality that participating teachers were able to transform learning about formative assessment into actual classroom practice. Outside of the immediate arena of the TLCs is where the difficulties presented themselves in this implementation.

Considering the work done in Cleveland along with similar projects in other locations, it is clear that there are three levels of personnel within a district who need to be involved and committed to the project for a district-level implementation to succeed. This is not exclusive to the KLT program, but to all reform efforts. First, there needs to be a district-level sponsor, someone who has sufficient authority to work with schools across the district. Second, there needs to be higher level support, not necessarily at the superintendent level but at a sufficiently senior level to ensure that the program is given time and resources to succeed. Third, principals within each school need to understand the program and be sufficiently bought into it to support and motivate their teachers as they engage in ongoing professional development.

Part of the need to have multiple levels of support is so the program does not become just one of the many that principals and teachers are asked to attend. The old adage of “more is less” may well have some virtue when thinking about the number of initiatives within a school or

district. Urban schools in particular are under significant pressure to improve student results, and thus, systems often operate with a “more is more” mentality. Multiple levels of support within a district may help administrators at various levels find ways of focusing on fewer initiatives. This, in turn, should make it easier for participants at all levels to attend to the theory of action of the program, by reducing the noise of other competing theories and approaches. This is the tight part of Tight but Loose. On the loose side, focusing on fewer initiatives will enable the district to draw in more local resources and do it in ways that complement the theory of action.

If this kind of aligned, focused support is available, local leaders can expand their approach to implementation from one of “just do what it takes to get it done” to a more reflective stance that allows them to consider the pros and cons of different approaches, in relation to the theory of action of the intervention. The program developers were always the ones who played the part of the tight police. This was not because of the district staff’s lack of understanding of the theory of action. Rather, it was because of the difficulty that the local implementers were having in simply keeping the program running under conflicting pressures from a jumble of programs and confused lines of leadership. If the local implementers were working in a milieu where there was deeper support for the program, then they might have the luxury of concerning themselves with the theory of action on a more regular basis.

The central office administrator with whom ETS worked closely was the reason that the KLT program succeeded to the extent that it did. She ensured that meetings were scheduled, rooms reserved, reminders sent out, supplies were provided, and teachers were paid for attending meetings. In addition she led the majority of the TLC leader meetings and coordinated a showcase at the end of the year to demonstrate and celebrate what had been achieved by the participating schools and teachers. However, she did not have sufficient authority to shield the schools from competing district initiatives. She did not have the buy-in or support of the third level of support, the building administrators, who answered to a different line of authority than she did, and who had competing program demands put on them.

In the early planning of the work, the administrator’s core team of district staff were each assigned a school, and their role was to be that of an advocate, keeping principals engaged and ensuring that the time promised to the participating teachers was honored. However, the roles of many members of this team seemed to diminish as the year progressed, due to other demands on their time, usually stemming from competing demands placed on them through other lines of authority. As a result, in spite of the continued enthusiasm and determination of the project’s champion, she was stuck in the middle without sufficient support at levels above or below her. Thus, while the original plan for implementation was tightly tied to the fundamental tenets of the program and many areas of deviation (size or frequency of meetings) were done in a way so as to be responsive to the local context while considering those fundamental tenets, there was still considerable unconscious loosening, or unraveling of the rollout within the district.

The different perspectives of developers and local implementers became clear in the writing of this paper. As the paper took shape, the author with the school district affiliation was concerned that it did not portray a sympathetic enough view of the struggles and accomplishments of the many school and district staff who worked so hard to keep the program afloat. She was concerned that any school administrators reading this paper would need to see themselves in the story, and in a positive light, or they would not be able to get on board with taking responsibility for such a challenging reform effort. The program developers, however, kept insisting that it was important to write about the places where the reform diverged from the theory of action, as the first step in seeing that it got put back on track. The resolution of this

tension is exactly what tight but loose is all about, and it is necessary for program developers and local implementers to share both halves of the tension. If not, the program will not get implemented, or it will be implemented so weakly that it will lose its power.

Despite this unresolved division of labor, many important lessons have been learned from this enactment of the KLT program. Part of the process of establishing the program in a new location was the drive to once again examine a local context and determine which aspects must be adhered to as described in the theory of action, and which aspects can be modified to take advantage of particular local circumstances. In the case of CMSD, things that were undertaken to accommodate local needs, such as the monthly learning community leader support meetings, may later become an integral part of the program. Moreover, the need for multiple levels of support for the program throughout the entire organization—be that school, district or state—seems to be a necessity that is unlikely to change site to site. The program implementers have no desire to wait for the perfect districts to come along ready to adopt the KLT program but have become more appreciative of the importance of clear and ongoing communication at and among various levels within a district. In addition, questions can be posed to interested districts to help them examine their own contexts in order to best select where to begin with a program such as KLT. For example, knowing that the elementary schools have just embarked on a whole-school reading initiative might mean that the initial implementation of KLT is focused at the middle schools. The Tight but Loose framework itself now provides a way to help districts think about their initial plan against the theory of action and provides a framework for evaluating progress after the implementation has begun.

A State-Sponsored Pilot Project in Selected Schools in Vermont

Cynthia Tocci (ETS, Princeton, NJ) and Gayle Taylor (Vermont Department of Education)

Introduction

“How do you know how your students are doing?” is the guiding question asked by the commissioner of education in Vermont. A singular question guides what local systems do throughout the state with regard to assessment, instruction, and professional development. Of course, local systems have different roadmaps identifying the paths they take in response to the question, but the State wanted to provide guidance, and a common understanding in the area of classroom assessment. In order to support teachers and principals, the Department of Education (DOE) initiated the Vermont Formative Assessment Pilot Project (FAPP). The goal of the initiative is to help teachers know how their students are doing on a real-time basis in order to change their instruction to effect teaching and learning.

In this paper, the structure and implementation of an assessment for learning (AfL) professional development program called *Keeping Learning on Track*[®] (KLT) in a state-sponsored project will be described. State-wide scalability of an initiative coming from the DOE has similar issues as in a district-wide implementation, yet, also some unique issues. Districts and states have similar challenges of establishing and maintaining communication to, from, and among all the users in multiple locations, supporting and monitoring the implementation of the program, and getting buy-in from both early adopters and future users. When trying to implement an intervention state-wide, these challenges are accentuated due to the greater geographical distribution of the participating sites, the politics of dealing with multiple superintendents and curriculums, and trying to create teams at each site as well as making those teams feel as if they are part of a much larger team. The context of this state-wide project required modifications to the training template and the process of implementing KLT. This paper will discuss how KLT was adapted, delivered, and put into practice in this state’s reform configuration.

For this intervention, Vermont identified nine schools, the majority of which were either in school improvement or had not made Adequate Yearly Progress (AYP) for at least 1 year. Additionally, in order to leverage a coaching model that was successful in other projects, the State hired nine independent consultants to be FAPP coaches and assigned one to each school. In this particular implementation, coaches are the liaison between the teacher learning community (TLC) and the principal as well as the school and DOE. The coaches formed a professional learning community (PLC), and this paper will also detail their role and participation.

Since the distribution of the schools represented a fairly large geographical area, a different type of support system was required for implementing KLT. Therefore, the details of the support system and the mid-year outcomes of using a coaching model in a state-wide implementation will be presented.

Implementation of the AfL Initiative

Conceptualization of a Pilot

In order to support the commissioner’s quest to improve student learning, the DOE decided to focus on formative assessment. They selected professional development focused on AfL, because it would directly impact classroom teacher knowledge of what students know and do not know in order to inform instruction. The DOE extended the commissioner’s question,

“How do you know how your students are doing?” to “How do you know on a real-time basis how your students are doing in order to change your instruction to affect their learning?” The DOE designed assessments that are instructionally sensitive, of necessity, because of their responsibilities for state assessment. However, they knew that would not be enough to impact change at the classroom level. As the group developed its model of a comprehensive assessment system, the director used a ring stacking toy as a visual, where the bottom ring, the largest, was classroom assessment. In addition to the group creating the state assessment, another group was working on the Comprehensive Local Assessment System (CLAS) and its guiding principles. The classroom component was a strong focus of their discussions. The leader had the group read a book about AfL (Black et al., 2003) which informed their guiding principles.

As the guiding principles were being developed, the CLAS group, started by the standards and assessment team, invited representatives from other DOE teams to their discussions. The larger group decided that if they were going to impact assessment at the classroom level, they would need to provide professional development for both teachers and principals that would lead to systemic change. Using the framework articulated by Black and Wiliam, the groups conceptualized a pilot project and issued a request for proposals. ETS submitted a response proposing to partner with the DOE to implement an adaptation of a professional development program called KLT. This intervention focused on a minute-to-minute and day-by-day approach to formative assessment that deliberately blurred the boundaries between assessment and instruction. This multi-year, professional development program used a hybrid model consisting of workshops, school-based TLCs, and supporting materials.

Funding for FAPP was secured by combining monies from a collection of resources: Title I school improvement funds, Gear Up partnership fund, other local funds, and staff resources from across the department. The FAPP, while much longer in the planning stages, was essentially to be implemented between May 2006 and June 2007. The goals of FAPP were to support the implementation of AfL strategies in the classroom that reveal and contribute to the learning of each student, and to create sustainable TLCs in participating schools. The main steps of the pilot included a series of planning sessions with the DOE and ETS, the selection and participation of nine schools, the use of external coaches as TLC leaders, delivery of workshops for coaches and school teams, participation in monthly school-based TLC meetings, a series of whole group implementation meetings, and an evaluation of pilot implementation. Each of the steps will be elaborated below.

Planning Sessions

The DOE established a FAPP planning team to partner with ETS to design and implement the pilot. Since assessment was at the core of the DOE’s work and they had some examples of teams involved in initiatives that did not seem well aligned in the past, the DOE decided to intentionally begin a process to ensure alignment. Thus, the DOE recognized that the size and composition of the planning team is critical to its success and to the implementation of FAPP, and thus recruited staff from across the department. The team consists of eight personnel with a variety of responsibilities and expertise, spanning areas such as standards and assessment, curriculum, and new initiatives.

ETS was contracted to provide professional development and technical assistance for the project. Three ETS staff members, including Dylan Wiliam, attended 2 all-day planning team meetings in March and May 2006, and had many conversations and e-mail exchanges between meetings and prior to the workshops. The purpose of the meetings was to discuss and create

adaptations to the KLT program for the FAPP. The inclusion of external coaches (see more details below) and a different configuration of the workshops required extensive conversations about what components of the KLT program must remain as designed and how others might be adapted to meet the needs of this state-wide pilot yet maintain the integrity of the program. The DOE content specialists were asked to join the planning team and ETS staff at the May meeting. The content specialists brought new perspectives and teacher needs to the front. Most districts were implementing other initiatives, many requiring content-specific changes, and the specialists wanted a role in the FAPP so they could help teachers unite the various efforts. It was decided that they would attend the workshops and be available to coaches, teachers, and principals during the workshops and throughout the pilot year.

The responsibilities of the planning team did not end after the planning meetings. All members attended every workshop as engaged participants, not just as observers. After the summer workshops, ETS staff attended one all-day planning team meeting to plan the first of two all-coach and all-school team implementation workshops. The goal of this meeting was to design the workshops, as these were not part of the regular KLT program. ETS staff had developed a number of activities and tried them with other pilot districts, but complete workshop days did not exist. Outlines for the workshops were completed and the details of the actual activities were developed in subsequent weeks. Again, there were conference calls and e-mail exchanges prior to the delivery of the workshops. One other difference in this part of the pilot was that Dylan Wiliam was no longer participating; he had accepted a position in England. The remaining two ETS staff members were joined by two state-based consultants who had participated in the summer workshops and were hired by ETS to assist with development and implementation of the pilot. Since the consultants were familiar with both KLT and local issues, they were able to provide valuable assistance to ETS staff.

In addition to the joint planning meetings, the DOE planning team met throughout the pilot to monitor its implementation, discuss the blending of other initiatives, examine any issues, and plan efforts to scale the program during the next school year. These meetings helped all involved to stay focused on the goals of the commissioner, the DOE, CLAS, and KLT, and have both a common language and a common understanding about AfL. The discussions with each other and with ETS helped to define what parts of the program must remain within a tight framework and what could be modified.

Selection and Participation of Schools

The DOE learned from past experiences that a pilot should include enough schools to have a reasonable sample, but not so many that the necessary amount of start-up support cannot be provided. They established selection criteria for the schools and sent letters of invitation to more schools than would be selected. The schools had to volunteer and complete an application with questions about the school, teachers, and the school's average knowledge and use of formative assessment. Nine schools were selected to participate, the majority of which were either in school improvement or had not made AYP for at least 1 year.

There were five elementary schools, one K-8 school, one middle school, and two high schools. The schools ranged in size and student demographics. They represented the varying geographical locations, economic status, and population density found in the state. Each of the schools had more than one initiative being implemented that school year. A common issue for schools and districts was competing initiatives. The struggles faced by the Cleveland School

District to address this critical issue when implementing KLT are reported elsewhere (Wylie, Thompson, Lyon, & Snodgrass, 2007).

Each school committed to sending the principal and a team of teachers to the all-school team workshops totaling five days (see below for more information). In addition, this group would form a school-based TLC that had to meet at least once per month for 2 hours with the TLC leader. The purpose of these monthly meetings was for the teachers to share their experiences implementing their individualized action plans and engage in activities designed to deepen their understanding of A/L strategies and techniques (for more information, see Thompson & Wiliam, 2007). The membership of the TLCs varied from school to school in size and composition. In some cases, they were grade-level or subject-area teams, while in other cases, there were cross-grade or subject teams, and in one case, the community was the entire school. The number of participants in a TLC ranged from four to seven. In most cases the teachers volunteered for the pilot; however, in some cases they were volunteered by their principal.

Use of External Coaches

KLT calls for one of the teachers in the school-based TLC to assume the role of a TLC leader. The leader uses modules created by ETS developers to facilitate the monthly TLC meetings. The program provides training for the TLC leaders to deepen their understanding of A/L concepts and the KLT program, and to provide them with tools for leading a group of their colleagues as they try to make changes to their practice. The purpose of having a classroom teacher as the leader is, as with other participants in a TLC, to have his/her own laboratory to try A/L techniques and bring those experiences to the meeting. Although one member, and sometimes two members, of a TLC is/are designated as a leader, the intention is that the leader is a learner along with the other TLC members.

The DOE decided to hire coaches for the TLC leader role instead of relying on on-site teachers, recognizing that, in order for schools to have the best chance for successful implementation during the pilot, the DOE needed to provide them with as much support as they could. An external facilitator, at least initially, was something they could provide for the schools. The intent was that the coach would decrease the burden on the school and ensure that schools would stay with the FAPP and implement it as intended. It also gave the DOE entry into the school's experience. The DOE's plan was to transfer the role of TLC leader to one of the regular TLC teachers for the implementation of year-two modules, and to use the coaches the following year to support the state's effort to scale the program.

The FAPP coach model was a deviation from the model used by other KLT pilot sites. The DOE and ETS staff had extensive conversations about the pros and cons of the two models and eventually came to an agreement of what the FAPP model would entail, including selection criteria, responsibilities, training, and support for the coaches.

The DOE recruited, interviewed, and selected the coaches. Each coach had extensive experience in education, had taught in a K-12 setting, and had been extremely competent in other coaching/mentoring situations. Each school was assigned a coach. Responsibilities of the coaches included providing classroom-level support for the teachers as they implemented new A/L ideas in their classrooms, facilitating the monthly TLC meetings, consulting with the principal as he or she supported the TLCs, championing the implementation of A/L in the school, and finally participating in a PLC for the coaches. The coaches were required to keep an online journal for the purpose of documenting the work of FAPP in the schools and participate in an

online community designed to support the coaches to examine and refine coaching practice and share effective practices with other coaches. The experiences and challenges of this model are addressed in the sections below.

Workshops for Coaches and School Teams

The KLT model provides a 2-day introductory workshop for all teachers, principals, and other school- and district-based personnel who will be responsible, in any manner, for implementing the program. The workshop covers KLT's One Big Idea, Five Key Strategies, the research supporting the program, and examples of the many AfL techniques to be used in the classroom, and provides time for the development of individual teacher action plans. The Introductory Workshop is followed by a 2-day TLC leader workshop. Participants include all teachers identified as TLC leader candidates and other school- and district-based personnel. The workshop focuses on TLC research, group facilitation and supportive accountability skills, materials used during the monthly TLC meetings, and school implementation plans for KLT.

The FAPP training model scheduled workshops for the coaches on 2 consecutive days in June and 1 day in July. Then the workshop for all school teams (the TLC teachers and principal from each school and the assigned coach for that school) took place during 3 consecutive days in July. The DOE planning team and ETS supplemented the original KLT training and ran the sessions, with Dylan Wiliam leading most of the learning activities. The coaches experienced the TLC leader training with the additional day of time devoted to learning the how to navigate the online community software, to discuss the DOE's expectations of the coaches, and to develop coaching skills, while the school teams experienced the Introductory workshop with the additional day devoted to working with the coach, including having the first TLC meeting using KLT's Module 1 activities.

The DOE commitment to this pilot was exemplified, yet again, by having all participants stay overnight at the training site, thus providing an opportunity for all teachers, principals, coaches, DOE planning team and content specialists, and ETS staff to network and share ideas. As another indicator of support, the DOE invited the superintendents of the participating schools and superintendents near the training area to a special evening presentation. The commissioner of education and Dylan Wiliam presented an overview of the FAPP initiative. They were videotaped so clips could be used to disseminate information about FAPP throughout the state.

The way the planning team and ETS grappled with the issues raised by the use of external coaches and the length and sequencing of the workshops was to have ETS deliver a modified version of the introductory and TLC leader workshops in the 3 days before the coaches were joined by their school teams. This way the coaches would have a little greater knowledge about AfL, and could establish themselves as resources with their teams early in the process. Another modification was to have the coaches lead their teams through the first module about building a teacher learning community and refining action plans. During the same time, the DOE and ETS staff met with the principals to work on each school's FAPP implementation plan and discuss how the principals could support the pilot this year and sustain and expand the work in future years.

Monthly School-Based TLC Meetings

KLT is designed as a 2-year program, with TLCs meeting at minimum 2 hours a month to share their classroom experiences with the AfL techniques, to engage in new activities to expand and deepen their understanding of AfL, and to refine their action plans. The FAPP

coaches scheduled these meetings with their school teams and the school principals. Depending on the school, the TLCs met either during the day or after school.

The coaches facilitated the meetings, using the KLT modules as a guide. The modules included suggested methods of leading the discussions and a script and handouts for leading the new learning portion of the meeting. Since the first module was completed during the summer training, the TLCs did not meet to complete the second module until late September. This variation provided enough time for the teachers to try an AfL technique with their classes.

The size of the TLCs varied from four to seven teachers. The coaches and team members of the larger groups reported that 2 hours was extremely rushed to cover all on the agenda, because it took longer for each person to report on his or her experiences with using an AfL technique. The TLC leader was responsible for setting the stage in providing supportive accountability for all TLC members to be responsible to try AfL techniques with their class and report to the group. They were also responsible for giving a gentle push to forward the participants' thinking and reflection about formative assessment and how it could be used to inform instruction. In a few cases, the teacher learning community did not get to the new learning. As shared with the coaches, the concentration on teachers sharing experiences and providing supportive accountability was critical to changing classroom AfL practice, and if anything in the meeting needed to be sacrificed or modified, it was the new learning segment.

Implementation Meetings

The implementation of any new program that required change in teacher practice required ongoing support for all participants. The DOE decided that one form of support would involve bringing together coaches for two coaches-only meetings as well as the coaches and school teams together on two different occasions during the school year. The goals of the four all-day meetings were to get a sense of how participants were progressing, to provide opportunities for learning new concepts, and to provide opportunities to network with others.

All of the new learning activities focused on developing a vision of a KLT classroom. The activities for the coaches' meetings focused on how the classroom behaviors of teachers and students might change over time by creating continuums of practice for selected AfL techniques. The purpose of the activities was to provide coaches with a tool to assist teachers to advance along the continuums. Many of the coaches were asked by the teachers to visit and observe them implementing the techniques. The continuum of practice activity was designed to help the coaches to focus on the AfL aspects witnessed during an observation. The activities for the school team meetings focused on how both the classroom behaviors and environment would change, and what might need to be done to help the students understand these changes. The consensus during the initial planning meetings was that principals would also need to have a common understanding of the changes that might occur in the classrooms and be able to look for them during regular classroom walkthroughs and observations. Both the coaches' and school teams' meetings provided opportunities for the DOE and ETS staff to check in with FAPP participants to hear how implementation was going and to clarify and present KLT information.

What's Loose and What Are the Challenges of This Implementation?

One way to think about the specific, site-based enactment of KLT in the Vermont schools is to consider the framework of Tight but Loose (Thompson & Wiliam, 2007). It is instructive to review this specific story to identify ways in which the developers of the program, along with district staff responsible for its support, had a clear idea of what it was they were trying to enact

and why. This is codified in the program's theory of action, which is what we consider to be the tight part. However, holding tightly to certain principles is different from implementing an entire project with strict rigidity, since there are ways in which the program must be able to take account of, build on, and support specific local contexts. It is important, however, that any looseness be played out in a very conscious way. Conscious loosening involves a thoughtful discussion of the local context, proposed changes to the standard implementation of a program, and the effect that those proposed changes may have on the theory of action. The remainder of this section is structured by identifying the loose aspects of the implementation. Many times it is the challenge of implementing a program within a particular context that prompts considerations of alternatives or of what to loosen in a program. When one considers that any single school will encounter challenges, competing initiatives, and other things that tend to cause a less than easy and smooth implementation of a new program, it can be assumed that these problems will be plentiful when the program is being introduced to a large district or across a state. This section will discuss some of the challenges.

The greatest diversion from the program's guidelines for implementation involved the decision to use external coaches, instead of teachers, as TLC leaders. These coaches may have more time to prepare for the TLC meeting, but at first that did not seem to be sufficient to make up for the challenges of using outside coaches. To begin with, the coaches were not members of the schools. On one hand, a coach coming into a school can work outside of personality conflicts that might exist between members of the school, but this individual is perhaps disadvantaged by not being familiar with local issues and initiatives and not knowing the TLC teachers, principal, and other staff. Furthermore, not being in the school on a daily basis could also be a disadvantage. However, being outside of the school system may work in the coach's favor in terms of acting as a champion of the TLC within the school.

Another issue that frequently emerged with any TLC leader was the newness of the role. The leaders, especially the coaches in Vermont, were considered experts in other situations, yet with AfL, they were novices. Even though they did have time to learn the contents of the KLT material, they did not have a laboratory or classroom to try out AfL techniques, and thus did not get to experience change in their own practice. Not experiencing the real efforts needed to make change a part of practice may have affected the leader's interactions and credibility with the TLC teachers.

Considering the content and process of the program, the DOE anticipated a few other conditions that might be a challenge and would require a discussion among the project partners. The additional challenges for this pilot included the following:

- Geographical dispersion of the coaches and schools would require a process for sharing information and experiences. The questions that the partners grappled with included how would communication keep flowing among the DOE, ETS, coaches, and school teams, how often, and for what purpose.
- Supportive accountability is a key aspect of KLT, but how might it be perceived in the field when the DOE is funding the initiative? How would the funding agency monitor the implementation of an initiative without supervising it and the main players? The agency needed honest and accurate information about the implementation, yet, would the school teams be able to give it? Would school teams be less than forthcoming because they were worried about how the DOE would perceive their actions and results? What would the DOE do if the information was negative?

- Some principals may have wanted to participate in TLCs. The original design of KLT reserved membership in a TLC to teachers who had at least one class in which they could try out the AfL techniques on a regular basis. If principals were added to the TLCs or invited to observe, how would it affect the dynamics of the group? Would teachers feel they were being evaluated on what they tried and their successes or lack of successes?

The anticipation and discussion of these challenges resulted in a set of solutions that maintained the integrity of the program, yet met the needs of the context.

Using Coaches

There was some apprehension about using external coaches to lead the TLCs. The DOE addressed the issue with careful planning of how the concept and coaches would be introduced to the school teams during the summer training sessions. The trainings were split so that only the coaches could meet and be given an overview of the AfL concepts in order to have greater participation in the all-school teams workshop. During the first training session, the coaches participated in an activity where they identified pros and cons of being an external coach and generated ideas about how to leverage the positive aspects in order to manage the less-than-positive aspects. During the all-school teams workshop, the coaches led two activities to convey the roles and responsibilities of the coach. They participated in a skit presenting illustrations of what a coach is and is not. In addition, the coaches sat with their assigned schools at the all-school teams training. The workshop schedule included a 2-hour segment on the last afternoon so that each coach could complete the first module with his or her teacher learning community.

Mid-year feedback on the monthly TLC meetings led by the coaches was positive. Teachers used their TLCs for sharing and as a means of reflection and obtaining collegial support for the implementation of formative assessment. A dominant theme reported was the sharing that happens among colleagues: sharing of ideas, questions, and successes; learning from each others' experiences and problem-solving challenges; having the time to be focused on their goals for formative assessment; and the feeling of a safe and non-threatening environment are the most commonly mentioned benefits of the TLCs. Principals also observed the benefits, claiming that teachers were collaborating more. One principal, attending TLC meetings, reported that they had the potential to be a strong source of embedded professional development.

With regard to using coaches, teachers and principals have provided primarily positive feedback about having a coach assigned to their TLC. Positive attributes mentioned about the coaches were their knowledge and skills, their supportive manner, and their ability to keep the group focused and on track with the FAPP goals. The negative attributes mentioned were their lack of investment in the local context and the little depth they brought to those particular discussions. Opinion about the value of an internal versus external coach had varied. The benefits of having an outside skilled person to facilitate as opposed to a person knowledgeable of the local system and present were the main attributes of the debate. In most cases, the pilot appeared to have worked, with the positive experiences outweighing the negatives.

Keeping It Tight With Ongoing Support

The DOE wanted to establish methods for communicating with the coaches and school teams during the pilot. To this end, the DOE formed a PLC for all coaches and itself. The department established an online community for the PLC on Teacher Workplace (TWP) in order to provide electronic resources and exchange information. The coaches were required to submit information about their TLC meetings, participate in discussions, and share with each other and

the DOE. In particular, they were asked to keep an on-line journal after each school visit, in part based on writing prompts, documenting their work in FAPP schools. The coaches participated in FAPP threaded discussions, approximately two discussions a month for 8 months. Each coach would be expected to post a minimum of one response to the discussion prompt and two replies to colleagues' posts in each of the discussion folders per month. They also contributed to and used content-related resources in TWP folders, where coaches submitted their own questions and/or questions from teachers in FAPP schools for the DOE content specialists. Each of the DOE planning team members was assigned to a school in order to monitor the entries.

Coaches used the online community to support each other in facilitating their teacher learning communities and to report their preparation and reflection on each module and on the TLC meetings. The DOE reported that it was an easy way to provide information to all coaches and to better understand what was happening in the field. The DOE brought forth issues raised by the PLC to the planning team to incorporate the issues into the implementation meetings.

In addition, the DOE decided to bring just the coaches together with DOE and ETS staff in November and February. The agenda for those meetings followed the agenda that a district would use if it brought together its TLC leaders. It included a How's It Going? piece for coaches to report at least one thing that was going well and one thing that was not going as well as they had hoped. That piece was followed by activities that deepened understanding of AfL techniques, supported TLC facilitation skills, and addressed some common issues. The DOE also planned for two all-school team meetings in December and April. The purpose of these meetings was to have participants share their experiences in implementing formative assessment in their classrooms and deepen their understanding of the underlying concepts.

The feedback from the coaches' meetings was very positive. The coaches, the DOE, and ETS reported that a deeper understanding of AfL was gained, and new facilitation techniques were learned. Similar findings were reported for the all-school team implementation meetings. Teachers had a positive response to meeting with other FAPP teams, the DOE, and ETS. Again, the predominant theme was the sharing aspect of the December meeting. The school team members noted that they appreciated learning about other teams' experiences, and in many cases, their own experiences were validated by the others' experiences. Modeling of FAPP by the facilitators was mentioned as a benefit. One thing evident was that teachers were seeking additional resources to broaden their understanding and sustain their work with formative assessment. Another thing that came up in all the implementation meetings was recognition from the field regarding the unified support they received from the DOE. Their openness and their willingness to listen to experiences and work at making necessary adaptations were recognized by most participants.

Regarding student understanding and use of formative assessment, teachers reported a pattern of student reactions, including increased student engagement, most notably of those students who have historically not participated; increased student ownership or self-accountability of learning; increased respect for and listening to peers; and student adjustment to less grading and more comment-only marking. A few teachers commented on having mixed success, with students being more reflective of their learning and responding more to each other. Teachers reported some challenges to students as they implemented formative assessment, including students breaking old habits and their discomfort and frustration in using new techniques. Principals observed that students engaged in AfL techniques and noted some changes in student behavior in terms of responsibility and accountability.

Conclusions

Planning a state-wide implementation is never easy, even when using an already developed program. The Vermont DOE approached the implementation with very purposeful choices about who should be involved, what levels of support would be needed, and how the entire initiative would be communicated and implemented. They began by creating an internal planning team with membership representing various interests of the DOE. They decided to begin small by selecting only nine schools from a list of volunteers. They used external coaches, a deviation from ETS's KLT model, who would transition their TLC leader role to a TLC teacher at end of the 1st year. The DOE also provided ongoing support for all involved and planned to scale the program after making modifications based on formative feedback.

The tight aspect of this professional development program remains its content and process of conducting workshops for establishing common understanding of AfL concepts and the specifics of the KLT program, TLCs meeting on a regular basis with a curriculum to follow and a leader, and ongoing support and accountability that focused on changing behaviors around the central concept of KLT. The lessons learned revolved around what could be the loose aspects of the program. They found that using coaches to lead TLCs (which was not part of the KLT model) was mostly successful. Modifying the training to get all participants ready for the pilot seemed to work. Providing ongoing communication and support to keep participants focused and help them with implementation of KLT was useful. And letting principals visit the TLC meetings was a possibility, if they understood the participation guidelines. To date, nothing has emerged as not working. Participants implemented KLT, and positive changes in teacher practice and student learning were noted.

The preliminary results of the FAPP were positive, and the FAPP teachers and principals were seeking information about specific next steps and long-range plans for expansion and sustainability of FAPP. There was some interest from non-TLC teachers within FAPP schools about formative assessment and the FAPP project. Teachers and principals began to intentionally share the FAPP work with their faculty, school boards, and parents. As expansion of the project was being considered, teachers and principals provided feedback on potential indicators for effective project growth. The result was that the Vermont DOE plans to support an expansion of the program next year. They will train the coaches and selected DOE staff to lead the workshops for all new participants, and support Year 2 of KLT modules for the current TLCs.

Epilogue

Since this paper was first presented at AERA in the spring of 2007, the pilot schools have continued with their school-based TLCs into the 2nd year. In order to begin the process of scaling up beyond the original participating schools, the DOE chose to use a training of trainers model, whereby ETS trainers conducted a training session with the original nine coaches and a dozen DOE staff members, thus developing local capacity within the state. Coaches were paired with someone from the state department, and these training pairs conducted five training sessions in the summer with approximately 140 teachers. In most of the pilot schools, additional TLCs were established due to teacher interest. In addition, new TLCs in 16 additional schools were established. The goal for the summer of 2008 is to further increase the number of training sessions.

Responsibility for the organization of these sessions has been handed over to the regional educational services agencies, while the state DOE maintains responsibility for providing the trainers and an online community for TLC leaders and the trainers. The shift of responsibility for

delivery of the training sessions is part of the strategy to address scaling and sustainability. The shift in reform ownership is needed for any program to be successful beyond the initial implementation.

The DOE are champions for classroom formative assessment as critical part of their efforts to improve student achievement throughout the state. As part of implementing the KLT, the Vermont DOE in part dealt with the issue of competing reforms by taking time in introductory workshops to explain bigger assessment picture in the state and where KLT fits in. Furthermore, they made sure that participants heard this message from a DOE staff member. Finally, the state monitored all of the training sessions and is collecting feedback and information about KLT's implementation in both the pilot and the new schools. This data will be reviewed with the Tight but Loose principles in mind.

Scaling Up Across Diverse Contexts: Lessons Learned From Five Implementations of the *Keeping Learning on Track*[®] Program

Dawn Marie Leusner (ETS, Princeton, NJ), Judith Ellsworth (University of Wyoming at Laramie), and Laura Goe (ETS, Princeton, NJ)

Introduction

This paper is the closing one in a report dedicated to thinking through the difficult issue of scaling up educational reforms in ways that preserve their effectiveness, even as contexts vary widely. The report's opening paper (Thompson & Wiliam, 2008) presented a detailed overview of the components and theory of action for a teacher professional development program that serves as the exemplar program for this discussion. The *Keeping Learning on Track*[®] (KLT) program is designed to help teachers adopt minute-to-minute and day-by-day formative assessment into their everyday teaching practice, via an initial workshop followed by sustained engagement in school-based teacher learning communities (TLCs). The opening paper also outlined a framework, called tight but loose for thinking about how to scale up reforms like the KLT program across diverse contexts while maintaining reasonable levels of fidelity to the core principles of the intervention. As Thompson and Wiliam said, "The tight but loose formulation combines an obsessive adherence to central design principles (the tight part) with accommodations to the needs, resources, constraints, and particularities that occur in any school or district (the loose part), but only where these do not conflict with the theory of action of the intervention" (p. 5).

That paper is followed by five stories of implementation of the KLT program across diverse settings (Goe & Mardy, 2008; Lyon, Cleland, & Gannon, 2008; Maher & Wiliam, 2008; Tocci & Taylor, 2008; Wylie, Thompson, Lyon, & Snodgrass, 2008). The KLT program and its predecessor versions have been implemented in a wide variety of locations and contexts within the last 3 years. From single schools, to entire districts, to a state-sponsored program in target schools, five case studies in five states (Maryland, New Jersey, Ohio, Pennsylvania, and Vermont) are presented in this report. Collectively, these show numerous context-sensitive variations in delivery method, fidelity, and impact, and provide grounding for a discussion of the tight but loose framework.

In this paper, we attempt to survey the preceding papers in a way that leads to a better understanding of the tight but loose framework, both as it is applied to the KLT program and also for what it has to tell us about scaling up classroom-focused school reforms more generally. Our review of the five place-based papers shows that the five implementations exemplify varying degrees of success, with regard to staying tight where they needed to and in dealing with the need to be loose in places where flexibility and accommodation to local realities are essential.

An important test of any framework is its ability to explain or illuminate real-life events and situations, so our discussion of the tight but loose framework will focus primarily on how it can help us understand what took place within the five place-based stories presented in the report. Consequently, we have searched these papers for convergent and divergent themes that are relevant to the ideas bound up in the tight but loose framework. Similarities and differences across the sites will be discussed to help illustrate how implementations of the KLT program might look in other settings as well as how tight but loose was applied across the implementations—in terms of what was kept tight, what was handled more flexibly, and how these two sets of implementation decisions interact with each other.

Keeping Learning on Track's Theory of Action

According to Hawley and Valli (2001), professional development ideally should be focused on the needs of both teachers and students, be a coherent part of a school-wide plan for change, accommodate teachers' voices in the planning and development of professional development activities, and encourage collaboration within the school. Moreover, Hawley and Valli asserted that professional development should undergo evaluation to determine its impact on student learning and on teaching practice. The authors state that these components of professional development represent a consensus view in the field about what professional development should be. The KLT program reflects this type of approach.

As discussed by Thompson and Wiliam (2008), the KLT program's theory of action mirrors the three-step model common to all professional development programs. This common three-step model can be summarized as follows:

1. Teachers learn about a better way to teach through professional development
2. Teachers adopt the better approach to teaching
3. Student learning is improved because of these improvements in teaching

The KLT program addresses each of these steps:

1. Teachers learn extensively about minute-to-minute and day-by-day assessment for learning (A/fL) via workshops and sustained engagement in TLCs
2. Teachers make minute-to-minute and day-by-day A/fL part of their everyday teaching practice, implementing the One Big Idea and Five Key Strategies through carefully chosen classroom techniques
3. Student learning improves as a result of the particular ways in which their teaching is made more responsive to student learning

This is an extremely simplified version of the KLT theory of action. See the opening paper (Thompson & Wiliam, 2008), where the theory of action for each of these three steps is fully explicated.

The Standard Implementation Model for the Keeping Learning on Track Program

To better illustrate how the implementations described in the previous papers adhere to or differ from the standard implementation model for the KLT program, we will outline what has come to be seen (through several iterations) as the standard model. The first stage is to introduce teachers to the One Big Idea and Five Key Strategies of minute-to-minute and day-by-day A/fL via a 2-day workshop. This introductory workshop explains the empirical research base and shares a variety of practical techniques associated with each strategy that teachers can implement in their classrooms. To see this in practice, the strategies and a sample of techniques are modeled during the workshop to help teachers internalize how these can be applied within their own schools. Teachers also engage in an assortment of hands-on learning activities to help them deepen their understanding of A/fL. Additionally, teachers are introduced to the idea and value of TLCs as a vehicle to reflect upon their experiences in applying A/fL within their everyday practice and learn how sharing with their colleagues helps to build supportive accountability, which in turn builds their personal and collective capacity.

The second stage is to identify a subset of teachers to participate in the TLC Leader Workshop. They will serve as the TLC leaders. These teacher-leaders participate in an additional 2-day workshop where they are introduced to the research supporting the significance of school-based TLCs as well as their structure and content in relation to the KLT program. They also engage in several activities designed to help them learn to provide supportive accountability to their peers. These activities help to deepen the dialogue that occurs within TLC meetings, while creating a safe, open environment where teachers can question and challenge each other as they advance their own understanding of A/JL. Additionally, leaders meet two or more times a year to reflect on their own experiences of implementing the KLT program, discuss the facilitation of their TLCs, and plan for upcoming meetings.

The Tight but Loose Framework

The Tight but Loose framework is a way of thinking about how to scale up reforms, interventions, and programs so that they maintain their effectiveness even as they are adapted to different contexts. There are some components of a program or reform that must be implemented exactly as intended (the tight part—obsessive adherence to the central design principles), while there are others that can be adapted to fit unique local contexts (the loose part—accommodations to the needs, resources, constraints and particularities that occur in any school or district). The key, as mentioned earlier, is that the accommodations must not conflict with the theory of action of the intervention.

From this description of the framework, it is clear that a key requirement of the Tight but Loose framework is having a comprehensive and well-articulated theory of action. Without a strong and explicit theory, it will be impossible to determine where looseness is acceptable and where it must be resisted. Thompson and Wiliam (2008) further stated, “not only do teachers need to understand [the theory of action] to make it work, they (and the systems that surround them) have to understand what is *not* part of the theory of action, so they can make good decisions about which pieces of the intervention they must hold onto in the face of contextual challenges, and which pieces they can be flexible about” (p. 36).

Each school’s culture, environment, goals, strengths, and challenges are different, but these do not necessarily rule out the successful implementation of a program or reform. Successful implementation includes not only realizing that every school setting is unique, with its own issues, priorities, and constraints, but also requires preparing for differences in advance, to the extent possible, since they will profoundly affect the implementation approach.

Thompson and Wiliam (2008) suggested that the KLT program (and indeed, many other reforms intended to improve teaching and learning) can be implemented successfully in a wide variety of settings, provided that the tight and loose components are conveyed explicitly and are clearly understood by the adopting school, district, or state. Thus, the variations in implementation fidelity (and the ensuing variations in impact) depicted in the stories here help us gain a sense of just how well the components and theory of action were explained, understood, and adopted in real schools and districts.

In the following sections, we will contrast a number of aspects of the implementation of the KLT program across the five implementation sites, in order to document the ways in which schools, districts, and states adapted this program to their local contexts, as well as to describe the challenges that they faced. In many cases, these challenges are not unique—they are likely to hamper implementations of other types of reforms as well. Thus, this discussion should be useful for thinking about program implementation generally.

Description of the Five Sites

As part of the larger report, authors have described each implementation in considerable detail. A brief summary will be provided of implementation factors in each site, followed by a table comparing the implementations.

It is important to note that each implementation occurred during different phases of ETS's development of the KLT program. The implementation descriptions are presented in chronological order and include the program elements that were available at the time of implementation. Through this ongoing process of implementation, ETS researchers and developers learned not only what additional components and structures were necessary as they worked in different sites over several years, but also what supports were required to ensure a successful implementation.

Ho-Ho-Kus

One of the earliest U.S. implementations of the model of AfL pioneered in Britain was in Ho-Ho-Kus, a small, K-8, upper-middle class New Jersey school district. The district has only one school—a K-8 school with about 600 students. Ho-Ho-Kus is a school characterized by high achievement, strong parental involvement, highly qualified (according to NCLB criteria) mostly experienced teachers, and strong central support for professional development.

The Ho-Ho-Kus implementation was teacher driven, where three teachers who participated in bi-monthly half-day sessions with Dylan Wiliam were key factors in the school-wide spread of the program. These three teachers kept learning logs in the 1st year in which they recorded their attempts to implement AfL and reflected more deeply upon their experiences through more formalized action plans in the 2nd year. While all three teachers talked with other Ho-Ho-Kus teachers about AfL, one of the teachers became a particularly enthusiastic advocate for AfL and stirred interest among other teachers in the school/district. Support for this implementation was high within the school, as evidenced by the school/district provision of classroom coverage for teachers who volunteered to attend the workshops, as well as a culture that encouraged teacher collaboration and an administration and community that were open to and supported innovation. The program also was perceived by the teachers as complementary to another initiative (a learner active technology-based instruction initiative) that was already in place in the district.

However, as with any intervention, this implementation also faced a number of challenges. For example, teachers had difficulty scheduling time to meet formally during the school day to talk about AfL on their own, and there was no formal guidance provided for TLCs (ETS staff was still in the process of figuring out what TLCs should look like and what they needed to be successful). Still, teachers occasionally managed to find time to share strategies and techniques for AfL outside the school day (in after-school sessions led by Diane Mardy), and some teachers reported that they regularly discussed AfL in teacher-led, grade-level team meetings. While the program was vocally supported by the administration and teachers took ownership of the program, there was little tangible support for sustainability—no additional funding was offered to continue working with the new KLT program that formalized AfL strategies and techniques. Thus, teachers are now on their own to sustain and internalize what they learned. Nevertheless, during the 2 years that AfL was formally presented, a great deal of interest and enthusiasm was engendered among teachers in Ho-Ho Kus.

So how did the Tight but Loose framework apply to Ho-Ho-Kus? The strategies and techniques were implemented with considerable fidelity, (evidenced by yielding powerful results, according to several teachers), because the transmission of the basic principles of AfL was strong and clear (coming directly from one of its key developers) and the transmission was well-received by a group of committed, capable teachers. The fact that the program was almost entirely teacher driven meant that the strategies and techniques implemented were left entirely to the judgment of individual teachers, supporting the theory of action.

However, other aspects of the implementation did not match the program's current recommendations, largely because this implementation was one of the first efforts at a development of the teacher change portion of the theory of action. The basic philosophy of the AfL portion of the program and the continued sharing of practice among teachers were basic guiding principles, but the structure and content for TLCs and hence structures for keeping teachers on track were not yet formalized.

It is not clear that all the strategies and techniques being utilized in the context of the learning community in Ho-Ho-Kus were aligned with those of the developers. For example, some teachers adapted some techniques (such as Traffic Lights and Popsicle Sticks) as classroom management and behavioral tools, and these were enthusiastically adopted by other teachers. The teachers' justification, however, suggests a deep understanding of AfL principles—they focused on student self-assessment strategies such as traffic lights to engage students in monitoring and reporting their own behavior such as using the “stop, slow, go” light colors to indicate how they thought their behavior in the lunch room was that day. Additionally, these teachers used a group of all-student response techniques from the questioning strategy to assign classroom duties. Again, teachers justified such a use of this strategy because learning appropriate classroom behavior, including routines, is also important, not just learning academic subjects. The Ho-Ho-Kus teachers seemed to be as concerned about students' citizenship as they were about their learning, as reflected in their discussions of how and why they adapted AfL techniques. While teachers were certainly encouraged to be innovative and adapt techniques to meet the daily routines of classroom life, the focus on behavior suggests that the teachers moved beyond the fundamental goal of AfL—to collect and use evidence of student learning to adapt instruction to meet students' learning needs in real time.

Our sense is that without the guiding materials and notion of supportive accountability that came to be embodied in later versions of the program, it might be difficult for teachers who lost the true focus of the program to return to the central idea of the intervention. Furthermore, because we had not yet begun to stress the importance of sustained collegial support in the form of TLCs, the school and district, while vocally supportive, did not provide opportunities for teachers to meet during regular hours. Thus, teachers simply did not get the level of interaction around the topic of AfL that we believe was needed to keep a new approach to teaching at the forefront of their thinking and practice. On the plus side, however, several teachers reported discussing AfL strategies and techniques in their grade level groups. However, this less structured approach may not yield reliably sustainable results, a central issue in scaling up (Coburn, 2003).

Despite these mixed results in terms of maintaining fidelity (the tight part), the Ho-Ho-Kus effort did result in some notable improvements in teaching, judging from Goe and Mardy (2007). As we mentioned above, this is partly due to the clear transmission of the message of AfL. We would also have to attribute some of the improvements in Ho-Ho-Kus to the way in which the program's leaders fulfilled the loose portions of the tight but loose framework, as

applied to the KLT program. That is, in areas of program implementation that are explicitly or implicitly outside the theory of action, Ho-Ho-Kus was able to capitalize on local strengths. Specifically, the district's pre-existing culture of professional collaboration and commitment to improvement served this implementation well. Within this culture, the choice of which strategies and techniques to take up was left wholly to individual teachers, something that has not been easily accomplished in many other implementations. Also, the teachers in this district were accustomed to being left in charge of their own professional development, and took it seriously enough that they arranged to meet and share best practices in A/fL on their own time. In those meetings, they did their best to keep a good strong focus on A/fL, because they were serious about their professional commitment to growth and improvement.

An A/fL TLC simply might not have existed at all without a teacher leader who coordinated and led the meetings and the contribution of the underlying, pre-existing culture of ongoing professional development and growth, in spite of having no time to meet during the day. This is an example of how the loose portion of Tight but Loose works; the timing, location, compensation structure, and other logistical features of the TLCs are left up to the implementers to decide for themselves, and this was a case where the implementers managed to draw on local capacities to make it happen. On the other hand, the limitations of this particular way of fulfilling this aspect of program implementation (albeit one the theory of action is loose about) have already been made clear—without tangible, sustained support from the administration, the teachers' interest alone may not be sufficient to build a sustainable program. It should be reiterated that the district was not directly asked to provide support such as in-school meeting opportunities, since this was before the KLT program, with its emphasis on TLCs, was developed.

This illustrates an important aspect of the loose side of the Tight but Loose formulation: just because the theory of action is not explicit about how something should be done does not mean that all approaches are equally viable. The loose side of the equation is important because it allows the program to flex to meet the conditions and needs of a particular context, but that flexibility comes with the responsibility of finding an approach that actually works, given the conditions in a particular context. In Ho-Ho-Kus, we would say that this approach, which lacked the important component of TLCs, only worked to a point.

St. Mary's County

In the state of Maryland, school districts are organized on a county basis. St. Mary's County is a peninsula with a population of roughly 100,000 and a population density approximately half the Maryland average. St. Mary's County enrolls close to 17,000 students across its 24 rural schools (16 elementary, four middle, and four high schools).

St. Mary's County Public Schools introduced a mentoring model, where the KLT program was implemented as part of the teacher induction program to investigate the relevance and feasibility of A/fL for beginning teachers and explore different kinds of models for district-wide implementation. Another early implementation, the program complemented St. Mary's newly developed mission statement and countywide action plan for "using assessment to guide instruction" (SMCPC, 2004).

Teachers volunteered in pairs, and the program cast both the novice teacher and the more experienced mentor teacher in learner roles. The director of professional and organizational development supported teachers with monthly meetings where they shared their experiences, engaged in new learning, and devised action plans similar to the more formalized procedures

now included in the KLT program training materials (Thompson & Wiliam, 2007). Teachers were also encouraged to keep reflective logs, which they did in varying degrees of detail. The teacher pairs met more or less regularly outside the school day between meetings.

The mentoring in this program was designed to support the instructional action plans developed by the teachers. Each mentoring team was given guidance in the coaching process, and formal training was offered. Mentors were given direction to provide feedback relative to the strategies in the actions plans and not on extraneous observations of what may occur in the classroom.

The county systemically provided funding for release time to offer teachers the opportunity to observe and coach each other. However, teachers found it difficult to make the needed arrangements for peer observations, so these did not occur on a monthly basis as originally intended. Still, each pair did manage to observe one another once or twice during the year.

The program continued into the next school year and was extended to new teachers across the county. In the spread of the program, some aspects began to be integrated into more general school improvement processes. For example, team action plans were developed and implemented on a quarterly basis for each department or grade-level team that identified specific goals, assessments of those goals, professional development needs, and extension/remediation approaches as a result of the assessments (Maher & Wiliam, 2007).

How did the Tight but Loose framework apply to St. Mary's? First, the knowledge transfer of the basic principles of AfL was clear, coming directly from one of its key developers. Thus, the strategies and techniques were implemented with some degree of fidelity. Second, the choice of which strategies and techniques to implement was left up to the discretion of each individual teacher, another tight tenet of the KLT program. Third, there was a time and place—the monthly meetings—for teachers to share and reflect upon their experiences and develop action plans outlining what they would attempt next, including any support they might need.

What we have learned in our research to date is that teachers themselves can provide effective leadership for their peers because they are going through the same learning and change process and they have essential insights into the pace of change, the kinds of dilemmas faced, and the types of support that make sense, all within the context of the classroom (Thompson & Wiliam, 2007). Therefore, another aspect of the program that it is important to be tight about is to have a teacher who is participating in the KLT program lead the discussions of AfL. St. Mary's is moving in this direction by enlisting teachers who have participated in the 1st and 2nd year to facilitate the monthly meetings as a teacher-driven professional development activity.

Huntington

Huntington is a K-12 district in a suburb of Philadelphia, Pennsylvania, and consists of five elementary schools, a middle school, and a high school. The district serves over 5,000 students in a predominantly upper-middle class community.

This was a two-tiered implementation, in that ongoing evaluations and conversations regarding the effectiveness of the first tier informed the planning and implementation of the second tier. That is, the first implementation failed, but lessons learned ensured a more successful implementation the second time around. A comparison of both implementations is discussed below. This implementation is also characterized as a knowledge-sharing model where a third party provider (The Math & Science Partnership of Greater Philadelphia or MSPGP) collaborated with ETS and the district to set up the KLT program district wide. Huntington was

viewed as a good choice for district-wide scaling, since it was performing well academically and had strong central support for the program.

ETS's purpose for this implementation was to move responsibility for the program further away from the original developers by relying on MSPGP staff to be the primary contact and method of support. The MSPGP took on the tasks of planning the implementations, communicating with the district, and providing ongoing guidance and support, while ETS provided the content and expertise in A₁L. The initial pilot spanned 1 school year and the second implementation, which was much more successful, spanned 2 school years.

In addition to supporting the Tight but Loose framework, this case study also demonstrated the necessity for sharing a program's theory of action. The structures and support mechanisms required for this particular implementation were not adequately described during the initial implementation. Thus, the MSPGP did not have access to the knowledge that ETS developers had about the structures that may be necessary to support a district, its teachers, and teacher leaders in this initiative. The standard implementation model now explicitly provides the theory of action through its workshops, TLCs, and ongoing follow-up support for teacher leaders.

ETS had to examine how knowledge is shared between and among organizations to provide the necessary knowledge to third-party providers. According to the knowledge transmission cycle of Nonaka and Takeuchi (1995), sharing of knowledge requires that internalized knowledge of one person or group be made explicit so that it can be shared with another person or group. In addition, not only does new knowledge need to be made explicit, but in order to fully understand the importance of the information, the other person or group must make it operational. Once operational, the new knowledge can be combined with existing knowledge structures, internalized, and made accessible. In this case, it was important to consider not only how knowledge was shared between ETS and the MSPGP, but also how knowledge was shared between the MSPGP and the district.

So how did Tight but Loose play out in Huntington? An analysis of the strengths and weaknesses of the two implementations provides support for keeping several aspects of the KLT program tight (see Lyon et al., 2007, for additional details).

First, providing ongoing support with the sole focus of A₁L is essential. In the first implementation, follow-up occurred during monthly meetings with curriculum supervisors. Supervisors had difficulty finding time to include A₁L discussions with an already packed agenda; thus these conversations did not occur with any regularity. In the second implementation, TLCs were utilized, where time was solely devoted to A₁L. This resulted in a stronger implementation and more regular follow-up.

Second, teachers need to lead the TLCs. In the first implementation, A₁L discussions were facilitated by curriculum supervisors who did not have sufficient training and materials to lead these meetings and who in many instances did not have their own classrooms in which to implement the techniques. One of the mechanisms by which teachers develop expertise is by sharing their experiences with the group, discussing their successes and challenges, and refining a plan for the next month. Having recognized this problem, adjustments were made in the second implementation, including training a core team of teacher leaders to run the monthly sessions. These teacher leaders were provided with release time to deepen their own understanding of A₁L before being made responsible for leading the meetings with participating teachers the following year.

Third, TLCs should contain four to eight participants. During the second implementation, due to substitute shortages and scheduling constraints, the four leader groups were combined to create two larger communities with approximately 15 participants in each. As a result, the How's It Going? discussions were not as productive. The groups became too large, and participants did not all know one another. This made it difficult for every teacher to share and for meaningful discussions to develop. It is important for the groups to remain small to allow for the exchange of knowledge to occur and to allow each group to develop an identity and establish a culture of safety and support.

Finally, a core principle of the program that must remain tight is that teachers must choose the specific techniques that will work for them, given their teaching style, students, and curriculum. In the first implementation, participants were collectively required to choose one strategy to focus on. This limited the teachers' ability to use their professional judgment and limited the group's exposure to the variety of techniques that teachers may have adopted. In the second implementation, teachers decided for themselves what strategies and techniques they would implement in their classrooms, which resulted in better teacher buy-in of the program.

Although analysis of the two implementations provided support for tight aspects of the KLT program, it also highlighted several areas where the program can be flexible, or loose, to respond to local needs. For example, the standard implementation provides a 2-day workshop for teacher leaders. The teacher leaders then learn the new content along with their peers. In the second Huntington implementation, however, teacher leaders were provided training over the course of 4 months. This allowed leaders to not only learn the new content, but also to deepen their own understanding of AfL, internalize the information, and prepare to present the materials before being responsible for delivering them to the entire group.

A second example is that the standard implementation calls for 2-hour monthly teacher learning community meetings. In the second implementation, Huntington utilized district-wide professional development days for these meetings. This use of time is appropriate as long as the meetings center solely on AfL and scheduling is done to ensure that there is enough time between meetings for teachers to try new techniques, but not so much time between meetings that the focus on AfL is lost.

This case study provides a description of one model of implementation involving a third party provider. It indicates that for a third party provider model to be successful, they must have a strong relationship with the district, the ability to secure support from the district and teachers, and a certain tenacity to see the program through to the end (Lyon et al., 2007). They must also be true advocates for the program, which the MSPGP certainly are, having recognized the benefits of AfL in the district's math and science teaching.

Cleveland

Cleveland, Ohio, is a large urban school district with a number of hard-to-staff schools, low student achievement, and difficulty finding sufficient numbers of highly-qualified, experienced teachers. The district has been in improvement status for several years, and as such there have been a number of different strategies and reforms implemented that focused on boosting student achievement. The KLT program was a district-sponsored implementation championed by a central office administrator across 10 targeted schools identified by the district as in need of school improvement. This was the first non-local implementation of the full standard program where ETS worked across the miles and did not provide any form of regular hand-holding. Strengths of this implementation included an extremely committed program

advocate, district funding, teachers sustained in collegial relationships, and ongoing teacher leader support throughout the school year. However, this implementation faced a number of challenges inherent in many school districts.

Principals were not able to attend the introductory workshop due to competing priorities and so were not cognizant of the level of teacher support that was needed. Therefore, there was no clear and consistent communication between the schools. Also, while advocates were assigned to each school, their roles were limited. Furthermore, they did not directly report to the central office administrator. Changes in leadership and district policies, and difficulty locating substitute teachers also produced barriers to effective implementation.

However, due to the implementation of the KLT program, these 10 schools saw improved state assessment scores in math and reading compared to other schools in the district not participating in the program (see Wylie, Thompson & Wiliam, 2007). The program expanded to five additional schools the following year, and seven of the 10 schools maintained TLCs that continue to meet in Year 2 due to these results.

It is instructive to consider this specific story for the ways in which the developers of the program, along with district staff responsible for its support, had a clear idea of what it was they were trying to enact and why. This is codified in the program's theory of action, which as we mention, is considered to be the tight part. It is important, however, that any looseness be played out in a very conscious way. Conscious loosening involves a thoughtful discussion of the local context, proposed changes to the standard implementation, and the effect that those proposed changes may have on the theory of action. On the other hand, unconscious loosening may cause the program to lose integrity. Unconscious loosening is often a result of pressures or constraints in districts that are outside the control of reform leaders. The process of periodic reviews, looking for evidence of both conscious and unconscious looseness, can support a retightening or revision process (see Wylie, Thompson, Lyon et al., 2007).

So how did Tight but Loose play out in Cleveland? The district was consciously tight in not dictating which strategies and techniques teachers should use. Initially, there was a misunderstanding that a particular technique was being mandated, and there was significant consternation among the group, since the introductory workshop stressed that a basic tenet of the program is not telling teachers what to do (Thompson & Wiliam, 2007). District staff who saw the power of the technique incorporated it into other professional development work. The intent had never been to suggest that this technique was mandatory, but this was how it was perceived by the teachers in that workshop. The central office administrator reiterated to her staff the importance of allowing teachers to choose which techniques they would use.

ETS was also consciously tight in one area—not providing the full glossary of techniques to the teachers until their teacher learning community meeting discussions were formative. At the onset, the group discussions and classroom observations were described by trained observers as indicating superficial use of the techniques (i.e., using whiteboards as scrap paper for example, rather than using them to ascertain understanding among the entire class). It was not that ETS did not want to share the more than 100 techniques with these teachers, but that it felt that they needed to deepen their understanding of the few techniques modeled in the initial workshops before they were provided with the full list.

On the other hand, the KLT program was able to be flexible to accommodate complementary district efforts that enhanced the goals and methods of the program. For example, the central office administrator had done some prior work that resulted in seven principles that defined the professional development work that she and her staff did in the

district. The research and work that they drew upon was not exclusively related to formative assessment, but it did not conflict with the central tenets of the intervention, and therefore it created a sense of ownership of the program among district staff. This is an example of being loose in order to build on local strengths. Another way in which ETS accommodated the needs of the district was through the development, along with district staff, of the structure for the learning community leader meetings.

An example of unconscious loosening is how schools were selected for the intervention. As mentioned earlier, this implementation began with district staff selecting the schools that would participate. The KLT program began with the 10 lowest performing schools. Funding was available which was not available to other schools in the district because of the status of these schools, and thus these schools were selected. Given a choice, ETS might not have selected these schools, but rather might have started with a set of schools on the basis of their ability to demonstrate school-level administrative support and commitment to the project. This is because a tight component of the program is that teachers have time to meet in their TLCs, with assurance that the time will be protected for all teachers to fully participate.

Another unconscious loosening was the size of the TLCs. ETS recommends establishing groups of four to eight. However, due to substantial last minute changes that had district teachers moving between schools and grade levels, the meetings provided an opportunity for teachers who did not know each other to come together in a collegial, focused and productive environment. While ETS reiterated the rationale that in larger groups not everyone would get an opportunity to share what they had been doing, nor would there be sufficient time for them to get meaningful feedback and support, the groups pushed back and ensured that everyone would have a voice. However, based on evidence from teacher leader meetings, not everyone in each group was taking time for professional sharing and getting feedback from the group about how they were implementing A/L in their classrooms, even though these are critical aspects of the theory of action. While it was clear from the leaders that they themselves understood the importance of everyone sharing and that it was an important aspect of accountability, the size of the groups and the time allotted to the meetings were prohibitive of these kinds of discussions.

An additional concern about larger groups was that an individual's sense of personal responsibility and commitment to the group would be diminished. A teacher may think that it does not matter whether they show up to some meetings if there are 15 others in the group. While ETS was not able to get a consistent record of attendance, between teacher leader reports and teacher learning community observations, there appeared to have been a downward trend in attendance as the year progressed.

Another way in which TLCs deviated from the standard implementation was in regard to the length and frequency of the meetings. Due to a bargaining agreement with custodians, teachers had to vacate the buildings within 90 minutes of the end of the school day. Even though the district offered funds (left over from the substitute kitty) to pay custodians' overtime rates, no school took the district up on the offer. Therefore, groups could only meet for an hour or so after school. Some groups met twice a month for an hour at a time and some met for 45 to 60 minutes every week. ETS did not object to the ways that groups adjusted but did suggest that groups try for two 90-minute meetings, since they would lose time at the start and end of the meeting on two occasions rather than at a single monthly meeting.

In spite of all of the struggles, the KLT program expanded to five additional schools in the following year. In addition, some teachers joined a teacher learning community even though they had not attended the introductory workshop. Either teachers heard something of interest

from those who attended, or perhaps they were primarily interested in the stipend, but regardless, no one wanted to discourage them from exploring AfL. For more information, see Ellsworth et al. (2007); Wylie, Lyon, et al. (2007); Wylie, Thompson, Lyon, et al. (2007); and Wylie, Thompson, & Wiliam (2007).

Vermont

The state of Vermont sought assistance in providing guidance and a common understanding in the area of classroom assessment. In order to support teachers and principals, the Department of Education (DOE) formed the Vermont Formative Assessment Pilot Project and contracted with ETS to carry out this work (see Tocci & Taylor, 2007). A number of schools across the state were invited to apply, the majority of which were in school improvement status or had not made adequate yearly progress for at least 1 year. Nine schools that volunteered were selected to participate in this project.

Vermont was a state-sponsored implementation in the nine targeted schools (one Pre-K to 8, five elementary, one middle, and two high schools) that used a coaching model in its implementation of the KLT program. In addition to meticulous planning, strategic preparation, and long-term planning, there was also unified backing within the DOE in support of this work. Nine coaches, who were external consultants hired just for this project, were each assigned to a participating school. These coaches served as a liaison between the teacher learning community and principal as well as their respective school and the DOE.

Coaches scheduled teacher learning community meetings with their school team and principal, and teachers often asked their coaches to observe their implementation in class. A formative evaluation was also conducted to provide mid-year and year-end reports to the DOE to help them make any necessary modifications to their implementation.

However, the geographical distance between schools posed some initial problems with communication. There was concern among DOE staff that teachers might not fully participate in the program and that the program would not be sustained beyond the pilot year with the state as the funding source. For instance, how would the DOE perceive teacher actions and results? What would the DOE do if information shared within the confines of teacher learning community meetings was viewed as negative? Some principals wanted to participate in the teacher learning community meetings—would their presence affect the dynamics of the group? Would teachers refrain from sharing or asking for help in fear they were being evaluated? How would teachers react to outside coaches leading the TLCs?

While there were mixed reviews of using coaches as teacher learning community leaders, feedback was primarily positive. The DOE established effective methods of communication in the form of a professional learning community for coaches, and they provided an online community for electronic resources and information exchange. Each DOE planning team member was assigned to a school to monitor the entries into this system and discuss next steps.

So how did Tight but Loose play out in Vermont? First, the tight components of this implementation were the content and process of conducting the workshops in establishing a common understanding of AfL concepts and the specifics of the KLT program. Second, the TLCs met on a regular basis and followed a curriculum with a leader. These communities provided ongoing support and accountability that focused on changing behaviors around the central concept of the program.

The greatest divergence from the program's guidelines for implementation involved the decision to use external coaches instead of teachers as teacher learning community leaders. As

noted earlier, learning alongside one's peers in implementing AfL in one's own classroom helps to build collective expertise among the group and allows for supportive accountability. Not experiencing the real efforts needed to make change a part of practice may have affected the coaches' interactions and credibility with the teachers. However, coaches communicated with principals about emerging leadership within the TLCs and groomed those teachers to take on the leader role in Year 2.

Another divergence was letting principals sit in on the TLC meetings. This is problematic because those who are not attempting to make changes in their own teaching practice can never be full participants in the community. They can be of substantial help to the community--brokering ideas, acting as advocates, and facilitating the community's learning--but they should refrain from holding themselves up as experts, unless they "walk the walk." Principals who attended the teacher learning community meetings were happy to see teachers collaborating more, and they have reported that these meetings have the potential to be a strong source of embedded professional development. The preliminary results of the Formative Assessment Pilot Project were positive, and the Vermont DOE plans to support an expansion of the program next year.

Site Summary

Each of the five sites came into contact with the KLT program in different ways, but each of them had similar motivations for becoming involved with the program—they all believed that formative assessment could help them improve student learning. As each site initiated contact with ETS after seeing or reading the research on formative assessment, ETS partnered with them and arranged to deliver professional development if they allowed us to include them in ongoing research activities, because of the ETS research agenda on the implementation, institutionalization, and impact of the KLT program. This reciprocal appeal to work together resulted in numerous productive working relationships, all at various stages of the development of the program.

Looking across the implementations, ETS provided ongoing support at varying levels throughout each site's implementation, presented the workshops in a variety of ways (in terms of delivery, not content), and in most sites, provided teacher learning community content. Across the five sites, there have been many forms of initial delivery and follow-through of the KLT program. ETS provided input that included meeting regularly with small groups of teachers, meeting regularly with pairs of teachers (mentors/mentees), working with third-party providers, speaking to hundreds of teachers at a time, and speaking/working with administrators and teacher trainers.

Each of these delivery models met with varying levels of success and each had their own set of challenges. In most cases, teachers volunteered to join their site's intervention, while in others participation was encouraged as part of a school, district, or statewide effort. Thus, in some cases the school, district, or state committed staff to the effort, while in other cases, teachers volunteered themselves.

As ETS researchers observed and took field notes, we determined from studying and comparing these implementations and what followed that successful implementation is largely dependent on the level of ongoing program support, that is, support that extends well beyond the initial immersion workshop. As a result, across the three most recent implementations (Huntington, Cleveland, and Vermont), there was more support for leaders, not just teachers. This was largely a function of ETS scaling up and letting go of our role as the primary deliverer of content. As we handed the content over, we had to support those leaders. This refers back to the idea that Thompson and Wiliam (2007) discuss earlier as scaling up and sustainability. Table 1 summarizes each implementation.

Table 1
Summary of Site Implementations

Where	What	Strengths	Challenges	Successes
Ho-Ho-Kus, NJ	A district-wide implementation in a K-8 school district	<ul style="list-style-type: none"> • committed advocate • school culture that encouraged teacher collaboration • admin and community that supported innovation • classroom coverage for teachers to attend workshops • program complemented another district initiative 	<ul style="list-style-type: none"> • no formal TLC structure and informal efforts hampered by scheduling issues • no ongoing admin support; teachers on their own to sustain the program • teachers adapting techniques in ways not intended 	<ul style="list-style-type: none"> • enthusiasm and effort stayed high through word-of-mouth • teacher-led, grade-level team meetings to talk about classroom experiences implementing A/L • teachers took ownership of strategies and techniques
Saint Mary's County, MD	A mentoring model in 24 public schools	<ul style="list-style-type: none"> • program complemented mission statement and countywide action plan • dedicated leader of monthly meetings • release time for teacher pairs to observe each other • teacher pairs met between monthly meetings 	teacher pairs encountered difficulty in arranging classroom observations (funding in the form of release time was available)	<ul style="list-style-type: none"> • cast both the novice teacher and mentor in learner roles • action planning integrated into general school processes • program expansion to additional schools throughout the county
Huntington, PA	A capacity-building model with a third party provider in a district-wide implementation (7 schools)	<ul style="list-style-type: none"> • strong central support for professional development • external funding via a third party provider • release time for teachers to meet, stipends or flex hours to participate 	<ul style="list-style-type: none"> • knowledge transfer from ETS to third party to district • change in admin • lack of buy-in and follow-up • TLCs led by non-teaching staff 	<ul style="list-style-type: none"> • district evaluated initial shortcomings when implementing a second year • local capacity-building from ETS to third party provider • successful implementation using a third party provider
Cleveland, OH	A district-sponsored implementation in 10 selected low-performing schools	<ul style="list-style-type: none"> • a dedicated central office administrator • support in the form of program funding, stipends, supplies, etc. • collegial teacher relationships • ongoing teacher leader support throughout the school year • complemented another district initiative 	<ul style="list-style-type: none"> • across schools, lack of principal buy-in, support and communication • assigned advocate support in each school did not materialize • changes in leadership • competing priorities • shortage of substitutes • meeting constraints (teachers vacated schools to honor custodian contract) 	<ul style="list-style-type: none"> • improved state test results in math and reading • 7 out of the 10 schools continued in 2nd year • program expanded to 5 additional schools • teachers not exposed to workshops participating in TLCs
Vermont	A coaching model in a state-wide implementation across 9 schools in need of improvement	<ul style="list-style-type: none"> • meticulous planning, implementation and follow-through • each coach assigned to a school • ongoing DOE support and accountability in improving teacher practice 	<ul style="list-style-type: none"> • communication across participating schools (geographical distance) • scalability while maintaining quality • communicating the program to non-participating schools 	<ul style="list-style-type: none"> • capacity-building: coaches transition their TLC leader role to teachers and become trainers to sustain/expand program • classroom observations initiated by teachers • program continued into 2nd year

What the Five Sites Had in Common

While sites varied in context, each of the five stories had common themes. First, a single champion stood out as the advocate who took charge of promoting the KLT program in a district or a school (i.e., a teacher leader in Ho-Ho-Kus, the director of professional development in St. Mary's, the MSPGP in Huntington, the central office administrator in Cleveland, and the DOE in Vermont). These were people in leadership positions pushing and pulling to ensure conditions supported teachers' participation and their efforts to implement the KLT program in their classrooms. Clearly, leadership plays an important role in program success. In these case studies, having a respected advocate in the program resulted in the rapid spread of KLT strategies and techniques across all grade levels and subjects.

Other literature has noted that an intervention or program can fail or succeed based, in large part, on the qualities of a particular individual leading it. For example, in a study of the implementation of reading interventions, researchers found that five facilitators had very different perceptions of their roles in the schools and programs they worked with. Their particular stance both impacted the program and was impacted by the program (Fevre & Richardson, 2000). The application of this study to the KLT program suggests that the person facilitating the implementation has a critical role to play, which impacts the success or failure of the program and, in turn, can impact facilitator perception of the role over time.

A second common thread across sites was the fact that teacher participation was voluntary. While some schools were conscripts (Cleveland and Huntington), the teachers in those schools were not mandated to participate, though full participation was encouraged. It is notable that in these volunteer sites, the participating teachers' enthusiasm became the catalyst for others becoming interested in the KLT program, either in the schools themselves or through program expansions to new schools and districts. Teachers sharing their successes using KLT strategies and techniques in the classroom provided a powerful impetus for other teachers to learn more about the practices and try them in their own classrooms. The success of the KLT program in volunteer sites should not be interpreted as a suggestion that the program will not work well under conditions where it is mandated. Rather, schools, districts, or states that were well prepared to introduce and implement KLT found it more manageable and garnered more enthusiasm from teachers. There is, however, as in any professional development program, the potential of mandated practice to impact teachers' interest in and attitudes about KLT. Such reactions would depend on how the teachers were supported and the resources that were provided.

How the Five Sites Differed

Besides the commonalities across sites, there were also definable differences based on necessary adaptations. These differences can be discussed within two main categories, delivery and support at the local level. The ways in which program delivery was adapted varied from site to site as a result of widely differing levels of teacher buy-in, administrative support, available resources, and district backing. There were persistent challenges encountered throughout the implementation. Thus, the ways in which collaboration with local educators enabled ETS developers to meet site-specific needs without compromising the program's theory of action was crucial to program sustainability.

The chief difference in the delivery of the KLT program centered around what happened after the introductory workshops were presented, or the kick-off of the program. In Ho-Ho-Kus, ETS staff personally led a hybrid form of a teacher learning community session, where, during

several visits over 2 years, teachers were given a half day of professional development time to voluntarily attend meetings. In Cleveland, ETS and district staff worked collaboratively in the KLT program implementation, providing professional development opportunities throughout the year. At the other sites, ETS staff were involved mainly in a research and advisory capacity and not as much in delivering content. In Huntington, ETS staff led several large workshops throughout the year, but the main work of implementing the program fell to the MSPGP. In Vermont, the DOE consulted with ETS staff throughout the year, and ETS staff frequently went onsite to assist in strategizing sessions. These various mechanisms are examples of a way in which the KLT program was designed to be loose; there is not one right way of delivering the KLT program at the local level. While the content is the same, differences in context may lead to a range of workable solutions that will ensure successful implementation.

The funding sources as well as the dollars invested in support of the KLT program varied widely, from district funding in Cleveland, to state funding in Vermont, to third party providers such as the MSPGP in Huntington carrying much of the load. In some sites, teachers were given release time during the day to conduct TLCs, while at other sites, teachers had to meet on their own time, with some sites offering a stipend. Meeting after school was attempted in many of the sites, but some teachers had so many after school commitments that it was virtually impossible to pin down a day and time during which teachers could meet. In addition, leadership at the various sites ranged from a volunteer teacher providing leadership on his or her own without additional compensation or release time to fully funded positions for leading TLCs and providing active support for teachers as they implemented KLT strategies and techniques. At the state and district locations, dedicated staff were given the time to develop and implement activities and communication outlets as well as to manage interference from external forces (usually, competing interests that would take professional development time away from the KLT program.) Additionally, in two of the sites, ETS collaborated with external groups that were liaisons between the developers and schools. These external groups brought their own resources to the table. In the case of Huntington's second implementation, for example, the TLCs were facilitated by the MSPGP, who championed the KLT program and worked to ensure that teachers had time and resources to implement the KLT program in their classrooms.

Application of the Tight but Loose Framework Across Sites

The KLT program content, in the form of the workshop materials, remained mostly static across all the sites. The continued sharing of practice, although varied in structure and frequency by site, was another tight component across each implementation site.

In the initial Huntington implementation, the district had each TLC choose a particular strategy. However, all of the sites heeded to the tightness of allowing teacher choice in what techniques they would try in their classrooms; leaving this to the teacher's judgment allows them to base their decisions on their knowledge of their classes and their students.

Although the program's content in the workshops stayed tight, the manner and mechanism in which the workshops were delivered varied across the sites and represents a loose approach to meeting local needs. For example, Vermont's implementation included a modified 3-day introductory workshop, which differs from the standard 2-day model. Additionally, this was intended for coaches and not for teachers as originally designed.

Additionally, in most cases, teachers grouped themselves within their own TLCs and met on their own terms. Whether teams were grouped by grade level, subject area, and/or school location, groups were structured in a way that made the best sense for teachers in their schools.

Also, groups met at times that were most conducive to teacher schedules and/or other school commitments. Some TLCs met during the school day and others after school. Meetings were held in 2-hour blocks on a monthly basis, 1-hour blocks bi-weekly, and/or in district-wide professional development days that were scheduled intermittently. This flexibility is an important aspect of respecting teachers' time. The resources available to teachers, including release time and substitute support, varied dramatically among the sites, and implementers took these circumstances into account in order to capitalize on local strengths.

ETS has determined which aspects of the program need to remain tight and what pieces can be adapted to meet local needs working in different ways with each site, as described below:

- The content and process of the program must be focused on the core content. That is, the workshops, the materials, and the activities are non-negotiable. The materials are research based and when implemented as such have been shown to be successful in classrooms across all grade levels and disciplines in a wide range of teaching and learning contexts (ETS, 2007). However, the way in which the workshop is delivered (not its content) can be adapted to meet local situations.
- In order to foster supportive accountability, practitioners need to lead TLCs. This does not mean that the teacher leader has to be an expert in *A_fL*, but rather that the leaders need to understand the trials and tribulations of classroom teaching so that they can grow along with the rest of the group. In the case of Vermont, outside coaches acted in the role of interim leaders as they moved teachers toward these leadership roles. Teacher leadership can also take different forms, such as individual leaders or leadership teams with teacher leaders working together to plan and lead meetings or, in some cases, taking turns as meeting leaders.
- The maximum number of teachers participating in a TLC should be eight. Four to eight teachers participating in a TLC has been found to be a reasonable number for teachers to effectively deepen their understanding of *A_fL* principles. As found in Huntington and Cleveland, it is important for the groups to remain small to allow for the exchange of knowledge to occur and to allow the groups to develop an identity and establish a culture of safety and support. There is, however, no prescribed configuration for TLC membership. Groups that make up TLCs can be composed of various combinations of grade levels, subject areas, and even cross-school groupings. As described earlier, participation has been most commonly voluntary, although there are no restrictions to mandated participation based on local context.
- TLCs must meet regularly. In an article about getting teachers to collaborate, Pappano (2007) noted that teachers often meet for the purpose of collaboration but end up chatting instead, which is unlikely to further academic goals. She suggested that leadership, structure, and goals are the keys to ensuring that meetings do not get sidetracked. On this score, the KLT program is tight, insisting that TLCs have a leader, a clearly delineated structure (through the modules) and goals that are in keeping with the program as reflected in the modules and in teachers' particular interests and priorities as they share classroom applications. Furthermore, the KLT program insists that these meetings be held sacred, i.e., that they are not to be co-opted by administrators wanting to make announcements or discuss particular issues, and they are not to be sidetracked by conversations that have no relationship to improving teaching and learning. Modules,

however, do not have to be used in a specified order but can be used at different times of the year, depending on the priorities and interests of the group. Meeting duration can also be flexible depending on local contexts (e.g., a regular monthly 2-hour session or two bi-monthly 90- or 60-minute sessions). Ideally, module requirements are most effectively addressed in the longer meeting timeframes.

- Teacher choice of what techniques are tried in the classroom is a critical aspect of the KLT program. Every classroom is different, and teachers need the flexibility to implement what makes the most sense for them.

Time is the most critical resource needed for teachers to engage in the KLT program, including regular and honored times for TLC meetings, time for teachers to reflect on the ideas of A/L as they relate to their own context, time to share successes and challenges and to receive peer feedback, and time and support to plan and prepare for changes in their classrooms in order for this work to take hold. The successful implementation of the KLT program requires multiple levels of informed support at the district level, the building administrative level (principals), and from teachers in the form of supportive accountability.

Conclusion

The challenges faced by every district, school, and classroom are different. In much the same way, an effective professional development program cannot be based on a one size fits all model. The ways in which institutions approach program implementation can vary dramatically, depending on their own local circumstances. As can be seen in the five stories discussed in this paper, local contexts, support, resources, policies, and other factors influenced the implementations of KLT in very different ways.

As the KLT program was being developed, issues of scalability remained at the forefront of discussions and research. Many different methods of introducing and delivering the KLT program were tried, both in response to varied local requests and in a deliberate effort to improve our understanding of what supports were required at the local level to ensure the success of the implementation. The KLT program was deliberately designed to ensure that components did not depend on the developers themselves, but could take advantage of site-specific internal and external supports and capabilities. Each of the sites we studied had a strong commitment at the administrative level and provided support for the KLT program, but the type, intensity, and duration of support varied considerably.

The Tight but Loose theory as it is applied to the five place-based stories shows, in real settings, where some aspects of the program need to remain tight and where flexibility—the looseness—is able to meet the local needs of each individual setting. For example, the KLT program's theory of action insists that professional development activities must be sustained over time. Little (1993) pointed out that traditional forms of teacher professional development (one-off workshops with no follow-up, for instance) lack key elements that could impact classroom practices. She noted that focus, intensity, and continuity were needed to make lasting change, and few professional development delivery models incorporated these aspects. The KLT program has made concerted efforts to be tight on focus, intensity, and continuity. For example, the content of the professional development offering is not open to change, though the delivery mechanism can be adjusted. In addition, continuity is crucial to the KLT program's theory of action, which is why TLCs are an integral and non-negotiable component of successful implementation of the program (Thompson & Wiliam, 2007). However, the KLT program is loose about some aspects of TLCs, such as how the groups are formed (by grade, by subject,

etc.), how often they meet (monthly, bi-weekly, etc.) and the duration of the meetings. While KLT provides a standard implementation model that offers suggestions on how to structure TLCs, ultimately, the model can be adapted to meet the local needs of each individual implementation site.

This was an important element of KLT that was deliberately built into the program—the flexibility to make certain accommodations to address and meet needs within a unique local context. For instance, there is substantial flexibility in creating teacher leader community group structures and scheduling convenient meeting times. In fact, we have seen TLCs based on grade level, subject area, building location, and/or a combination of all three. Also, TLC members have been creative in finding times to meet. While the recommended framework suggests meeting once a month for 2 hours, we have seen TLCs meet twice a month for 1 hour because 2-hour blocks of time were simply impossible to find. As a core principle of the theory of action, teachers are given free reign to choose what strategies and techniques they want to implement in their classrooms. A resulting advantage to this is that, depending on what each teacher selects, members learn together about the different strategies and techniques that comprise AfL. Several accommodations can also be made in the scheduling of the professional development workshops to better suit local needs. For example, while having the workshops run successively (2 days of introductory workshops followed by 2 days of leader workshops) is considered ideal, we have seen success in delivering the initial introductory workshops months before giving the leader workshops. There are pros and cons to this plan. While not optimal in terms of retention of the content, it could be argued that it recharges teachers' enthusiasm and eagerness when KLT is re-introduced as the school year is about to commence.

In studying these five sites, we confirmed that there is not just one model for achieving a successful implementation—there are probably as many models as there are unique sites. The key is conveying to the schools, districts, or states during the introduction and implementation which components of the program they must strenuously adhere to (tight) and which things may adapt to their local contexts (loose). High-level breakdowns in implementing the KLT program, such as a failure (either by ETS and/or the school/district) to apply or clearly communicate the theory of action when reaching implementation decisions, provided learning opportunities for ETS developers, allowing them to evaluate the breakdowns and then articulate how such failures could be avoided in the future. Such breakdowns were also learning opportunities for the sites themselves. Therefore, the role of evaluation represents another loose aspect of the program. Evaluation information was collected in Huntington, Cleveland, and Vermont. Here, the implementations were assessed periodically, and where needed, necessary corrections were made mid-stream—but only when the modifications did not hamper the theory of action. Thus, these changes were made to strengthen the local variation of the program. It is important, however, to monitor any modifications closely and to be prepared to alter these later if the changes become barriers to implementation.

This collection of studies on five very diverse sites demonstrates that context is a crucial consideration when implementing professional development, and that the design of a program should be sufficiently malleable to accommodate variations. These papers also demonstrate that flexibility in approaches to delivering professional development can be achieved without compromising the fidelity of the program. In other words, flexibility can be achieved without sacrificing the core principles of the theory of action. Essentially, program design must be rigorous, but not rigid.

In summary, while the scale of implementation (the magnitude of the rollout) differed drastically across sites, it was not the sheer numbers of teachers involved that posed barriers to implementation. Rather, it was how each site introduced the program, the level of commitment and backing provided by the site's leadership over time, and the structures available to support teachers' sustained implementation that played critical roles in the success or failure of a delivery model in actually changing teachers' practices. In other words, those with strong commitment at the district or state level survived the inevitable bumps along the way while others struggled.

As can be seen in this discussion, certain aspects of the KLT program are non-negotiable. However, other aspects of the program are adaptable to meet local needs. We learned together what aspects of the KLT program must be tight in order to maintain the program's fidelity and what aspects could be loose in order to meet individual circumstances. Where changes do not impact authenticity of the content and fidelity of implementation, they are generally acceptable. However, we continue to hold true to our beliefs about the things that cannot be changed.

Tight but Loose: Through the Looking Glass

*Margaret Heritage (CRESST, University of California, Los Angeles) and
W. James Popham (Professor Emeritus, University of California, Los Angeles)*

Paper 1: Superb Stage-Setting

Paper 1, in just about any sort of multi-paper writing, is supposed to set the stage for the papers to follow. The first paper of this report by Thompson and Wiliam (2008) accomplishes such a mission in a remarkable manner. Not only are we given a lucid, compelling rationale for the selection of a tight but loose strategy for scaling up any sort of school reform, but we also receive a candid recounting of why ETS chose to employ that particular scaling-up approach when implementing KLT. We also learn about the innards of KLT itself, with descriptions well stocked with illuminating examples. This opening paper, masterfully written, is a tour de force in every sense of that phrase. It represents an exceptional analysis by two thoughtful writers who have blended their own research and that of others with personal experience-honed insights about the realities of educational practice.

The KLT reform is one of consequential changes in classrooms, and Thompson and Wiliam (2008) clearly lay out those changes in terms of the classroom contract: teachers focus on learning rather than teaching; students become active participants in learning rather than passive recipients; and the teacher/student relationships is characterized by collaboration. The proposed changes in the contract are grounded in a strong empirical research base. Something that will be especially welcome to practitioners is that the KLT program actually takes big research ideas and translates them into the specifics of classroom action. As the authors rightly point out, this is not always the case. Practitioners are routinely told by researchers in general terms what they should do, but are often left in the dark about the specific ways in which they should or could do it in their classrooms. The KLT model is a powerful example of research-based practices that can be implemented in the classroom, combined with a process by which teachers can become experts in the implementation, as opposed to merely engaging in “rote application of declarative or procedural knowledge” (p. 13).

In a kind of reverse reform effort, instead of beginning with system reform, the starting point of the KLT implementation was fully centered on the classroom, albeit the classroom within the context of the system. The focused classroom intervention served to shed light on the system problems that needed to be changed for the intervention to work, and equally important, showed those parts of the system that worked well to support the implementation. This perspective is another example of how the KLT program was made practical and accessible to practitioners. Teachers and administrators were able to understand the system reforms needed in order to make changes in teaching and learning, which intuitively makes more sense than starting with the system changes they think will be necessary to provide the appropriate context for change.

As the authors point out near the end of their lengthy analysis, they have attempted to describe their “preliminary ideas for a framework for thinking about school reform at scale” (p. 42). But beyond their extensive consideration of the particular implementation strategy advocated here, namely, the tight but loose model, we also learn a great deal about the specifics innards of the KLT reform and, in particular, why it was chosen as a vehicle for encouraging the use of formative assessment in more classrooms. Thus, a reader of Thompson and Wiliam (2008) can, in a very real sense, adopt a sort of bifocal analysis, that is, a more distant view of the virtues of the scaling-up strategy being advocated, as well as a close-up scrutiny of the particular

formative assessment implementation model being recommended. We both left the paper with the clear impression that the authors had done a solid job of supporting a tight but loose scalability strategy and a KLT incarnation of classroom formative assessment.

Because the analysis is so extensive, it would be possible to deal at some depth with many of the topics addressed by Thompson and Wiliam (2008). It is also tempting to do so. Yet, were we to succumb to this temptation, the length of our commentary on Paper 1 might well exceed that paper's nontrivial length. Accordingly, we will select only a handful of the numerous points made by the authors, recognizing all too well that many other points were more than worthy of our attention.

Turning first to the tight but loose scaling-up strategy, Thompson and Wiliam (2008) provided a persuasive case for the need to rely on this sort of approach when trying to implement reforms in what they accurately characterize as a "bewildering diversity of schools and school systems" (p. 42). Clearly, given the substantial differences among school contexts, and in particular the dramatic differences among the human beings who staff our schools, any attempt to scale up a reform that failed to address these diversities would be destined for disaster. Yet, although there is clearly a need for a certain amount of looseness when installing a reform, a danger is always present that the looseness needed to adapt to extant diversities might eviscerate those key elements of the reform that make it worth installing in the first place. And thus, of course, we see why there must be a tight element of a reform intervention, an element that cannot, under any circumstances, be loosened, or the worth of the reform initiative wafts away. Thompson and Wiliam build a powerful case for the virtues of the tight but loose scaling-up strategy, and they do so without even involving the KLT approach that they and their colleagues devised.

We found their candor to be particularly appealing when they get around to the application of the tight but loose strategy to KLT itself. It is so easy for advocates of any intervention, particularly those who have been intimately involved with the generation of that intervention, to come across as though there not only are clear-cut answers to every problem that can be raised, but that those answers are readily discernible to anyone who is not a fool. Yet Thompson and Wiliam forthrightly concede that, at this stage of development, they are "far less sure of the things we believe we must be tight about" (p. 40) than they are with respect to what must be involved in formative assessment itself. It is refreshing to find the architects of an intervention revealing sensible uncertainties about how to best implement it.

Turning to KLT, we see a series of carefully spelled out rationales for the incorporation of certain elements of that intervention. To illustrate, the explanation of why it was that TLCs were chosen to be an integral part of the KLT intervention was especially persuasive. Similarly, the description of the components constituting the KLT modules is powerfully given. One leaves Thompson and Wiliam's (2008) description of each module's key parts, the bookend activities *How's it Going?* and *Personal Action Plan* and the teacher learning activities, with a head-nodding approval of the need for those components.

Because the contribution of these KLT modules to promoting the fidelity of the KLT intervention, we hope that there has been adequate tryout and revision of the module materials. It seems that there are now at least 18 such modules in existence, enough for 2 school years' worth of monthly meetings. Neither in this paper nor in the subsequent papers do we receive a particularly detailed description of the status of these materials, for instance, their development status, the degree to which they have been field tested, and any sort of evaluative evidence regarding their effectiveness. Hopefully, future reports regarding KLT will feature additional

information about these significant materials. TLCs can, of course, survive sans printed materials of this sort. However, one suspects that the tight part of KLT will be far tighter if the TLCs can rely on sets of well developed and properly evaluated modules.

Paper 2: No Hocus-Pocus in Ho-Ho-Kus

In Goe and Mardy (2008), we learn about the attempt of educators in the Ho-Ho-Kus School to incorporate assessment for learning (A_fL). Goe and Mardy deliberately try to provide readers with differing perceptions regarding what transpired during a 2-year effort to implement A_fL in this K-8 New Jersey school. What emerges in the paper is an honest, unvarnished description of why, in this instance, A_fL seems to have been embraced by many teachers in the Ho-Ho-Kus school.

Both authors recognize that the educational setting in which this intervention was introduced is, in many ways, ideally suited to implement the tight but loose approach to implementing classroom formative assessment devised by Wiliam and his ETS colleagues. For openers, the in-person presentations of Dylan Wiliam assured the accurate introduction of the tight elements of A_fL (i.e., those research-ratified features of A_fL that must be adopted unswervingly by a school's teachers). In addition, the collegial climate of this school, the release time for its teachers to participate, and the leadership role taken on by one of the school's most veteran teachers appeared to be positive factors in the way A_fL was received by the school's faculty. A_fL seems to be working in The Ho-Ho-Kus School, and factors such as these surely played a part in making it successful.

Goe and Mardy (2008) leave us wondering, though, about the long-term sustainability of the effort. We learn about what happened over the course of the 2-year focus on A_fL, but the true test of efficacy of the program is how A_fL strategies continue to evolve over time, and the degree to which they weather changes in leadership and staffing to remain embedded in the everyday practices of the school. We also learn about Mardy's hopes for increases in standardized test scores as a result of instantiating formative assessment practices, but do not find out if her hopes were fulfilled. Certainly, 2-years of implementation may not be sufficient time to see changes in student test scores. However, the goal of increases in student achievement on all measures should not be lost from sight.

Despite the success of the Ho-Ho-Kus implementation, there are potential cautions associated with this tale that, at the very least, ought to be recognized. When one adopts a form of narrative inquiry such as the one reported in Thompson and Wiliam (2008), a form of inquiry that, as the writers point out early on, is an amalgam of diverse perspectives, approaches, and methods, what sometimes gets left out are the potential negatives, which are only lurking in this instance, but might scurry to the surface in other settings.

The installation of A_fL in Ho-Ho-Kus followed an eminently sensible roadmap which, although there may have been minor setbacks along the way, appears to have worked pretty much the way it was supposed to work. But what if, in an attempt to implement A_fL in another school, the collegiality level of the school was much lower, or the seasoned teacher who assumed a leadership roll had been less well accepted. What if the person who first lays out the nature of the A_fL tight aspects does a less than lustrous job? Our point is a simple one: when any attempt is made to scale up interventions such as A_fL, it is imperative to recognize that—even if educators assiduously adhere to a well-conceived installation template—the particular people involved will almost always make the difference in whether the intervention succeeds or fails.

In this instance, the AfL installation template seems to have been well conceived, and the resultant story is one of successful implementation. But realism requires us to recognize that, in other settings, people may trump plans. All that any organization such as ETS can do is establish a set of procedures, some tight and some loose, that will increase the likelihood that students will be properly taught. In Ho-Ho-Kus, it appears that AfL is leading teachers to improved instruction.

Paper 3: Implementing AfL at the District Level: A Qualitative Look

Maher and Wiliam (2008) described a small-scale qualitative study, conducted over 3 years with pairs of teachers (novice and veteran) in Maryland's St. Mary's County Public Schools (SMCPS), focused on the feasibility of AfL for beginning teachers and for district-wide implementation. The basis for the study was twofold: the authors' view that veteran and novice teachers find AfL equally difficult to develop, and their presumption that if teachers benefit from the tight but loose framework of KLT professional development, then the same should be true for school districts.

Maher and Wiliam (2008) provided a straightforward appraisal of the results of the study. We learn about both the successes and challenges of implementing AfL, through which a chicken and egg problem surfaces. Several teachers report the benefits of AfL techniques on their students' attitudes, motivation, autonomy, and levels of engagement in class. By contrast, a number of teachers comment on the difficulty of implementing AfL but attribute such difficulties to problems with students' behavior. On the one hand, then, AfL seems to improve students' behavior, yet on the other hand, students' behavior is an impediment to AfL implementation. Either way, this paper underscores the importance of the context for AfL. Establishing classroom norms of independence, self-reflection, peer assessment, and responding to feedback are all critical to the success of AfL.

To grade or not to grade is an issue that often features prominently in teachers' concerns about AfL. Maher and Wiliam (2008) sheds some interesting light on the topic, ranging from one teacher who realized how much she had changed her views on grading when explaining to a colleague that student progress did not need to be linked to a grade, to students' becoming more reflective about their learning when they knew they were not going to receive a grade. Positive reactions were registered by both students and their parents to the use of comments rather than grades. This should provide some reassurance to those who worry about the impact of AfL on student motivation or about parent reactions as they move to the use of descriptive feedback about student learning.

Although the study in Maher and Wiliam (2008) was not designed to explore the impact of AfL on student achievement, the authors discuss changes in student scores on state-mandated tests. Some impressive gains were recorded. Because AfL was only one of a number of initiatives being implemented by the district, the authors properly recognize that the achievement scores cannot be solely laid at the door of AfL. Nonetheless, their conclusion that AfL can play a part in systematic school improvement seems reasonable. Given the looseness of the investigative procedures, the most useful yield from this study is its collection of perceptive comments from teachers and students rather than any evidence of students' test-score increases.

In terms of a district-wide model, we do not learn very much about what it actually takes to implement AfL from the district perspective. More on this would be useful. In the end, though, AfL is about improving teacher practices in the classroom. When Juliet, in her final reflection,

comments, “I am a better instructor now that I have refined my thinking about assessment,” (Maher & Wiliam, 2008, p. 65) she really gets to the heart of the matter.

Paper 4: If at First You Don’t Succeed

Because a prominent concern of the writers of this set of papers is the scalability of a particular AfL model, the focus by Lyon, Cleland, and Gannon (2008) on implementing AfL in a school district is particularly illuminating. We learn from Lyon et al. how a suburban, high-performing Philadelphia school district (Huntington Township) attempted to install the KLT program over a 2-year period. More specifically, two successive attempts to implement this version of AfL are recounted, the first of which was decisively less successful than the second. What’s most informative about this paper is its candid depiction of what went right and what went wrong in both of those implementation efforts.

Given that the ETS team’s implementation strategy was predicated on a tight but loose approach, it is particularly interesting to see how the educators in the Huntington Township District responded to the tight and the loose features of the KLT approach. One is often struck with the necessity for implementers of a tight but loose approach to AfL to make important judgments about just how loose a tight feature of the KLT program can be.

In this pilot study by Lyon et al. (2008), however, another complexity adds to our difficulty of interpreting the events associated with the implementation. A third party, rather than the ETS developers of the KLT program, assumed the major responsibility for guiding the program’s implementation. That is, the Math Science Partnership of Greater Philadelphia, an organization funded by the National Science Foundation which serves 46 school districts in or near Philadelphia, served as the third party implementation agency. Reliance on a third-party implementer, of course, also illuminates the challenges of using an installation approach in which the original architects of an intervention are not the purveyors of that intervention.

As is often the case in real-world investigations of this sort, we are left with a complex picture to interpret. Although we have an opportunity to see how a group other than the originators of the KLT program might be able to help a school district install AfL, the presence of another key group (in addition to the KLT developers and the school district educators) adds potential complexity to our ability to make sense out of what transpired. Nonetheless, if any version of AfL is going to be installed widely, it will most surely be necessary for third-party installers to play a part in such expansion activities.

In their reporting of what took place in both of the implementation efforts, the authors give us a forthright accounting of the strengths and weaknesses of the initial and second Huntington installations. There seems to have been no wart-removal efforts on the part of the writers as they recount what went on in the Huntington Township District, for we learn of some procedural mistakes that, retrospect reveals, were decisively wrong-headed. The second implementation of AfL in Huntington shows a significant effort to correct many of the foibles of the first implementation.

There are three striking features that arise from the description of the AfL implementation efforts. The first is the need for teachers, who are actually working in classrooms and attempting to implement the strategies and techniques themselves to be leaders of learning communities. Teachers were more successful than curriculum developers in these leadership roles. Second, although teachers learned more effectively from their peers, the amount of training and support needed for the teacher leaders was significant, and would be a major consideration in scale-up efforts. A third feature, and one that would certainly impact scalability, is that despite what we

know from the literature about the value of sustained professional development, and despite the commitment of the district to the implementation over a period of time, still there were problems in finding the time needed for the TLCs and for the teacher leaders to meet. One wonders if the moment will ever arrive when schools and districts can fully commit to an effort without the interference of competing priorities, scheduling, and substitute shortages.

Equally revealing in Lyon et al.'s (2008) report of a district-level attempt to infuse AfL into teacher's classroom instruction is the importance of such prosaic procedural concerns as inflexible scheduling that preclude teachers and administrators from meeting, or how much time is available for the How's it Going? interactions of learning communities once the size of those communities has been increased. Although analysts of AfL implementations can become enthralled with seductive theoretical concerns about what needs to be tight and what can be loose, the success or failure of real-world applications of AfL often depends on tiny but terrifically important procedural factors. Lyon et al. remind us that in any serious effort to scale up AfL in school districts, apart from a general installation strategy such as that embodied in the Tight but Loose approach of ETS architects, the proof of the pudding is apt to depend on the specific human beings involved and the particular procedural choices they make.

Paper 5: A Comedy of Errors

The central theme of Wylie, Thompson, Lyon, and Snodgrass (2008) is interplay between the tight part and the loose part of the tight but loose theory of action of KTL during an implementation of the program in 10 of Cleveland's lowest performing schools. Wylie et al. have described the tension between maintaining the fidelity of KLT (the tight part) and the need to accommodate to the local context (the loose part) to have some semblance of a successful implementation.

Although Cleveland Municipal School District (CMSD) is hundreds of miles from ETS, which meant that a critical component of the implementation would be to build district capacity, ETS decided to undertake the initiative because of a highly competent advocate at the district office, who was prepared to do whatever was needed to make the work a district priority. Optimism was further fueled when a seven-page memorandum of understanding, outlining the time, resources and support required for successful implementation was agreed upon by ETS and the district.

Notwithstanding the advocacy of this informed and committed person and the clarity of the extensive memorandum, the authors detail a dizzying array of problems that the implementation (and the advocate) faced, most of which could have been avoided. Given the importance of getting buy-in from the school level leadership for the initiative, perhaps the most remarkable event in the whole catalogue of errors was that the district scheduled a district-wide student head-count, which required all principals to stay at their schools, on the day a critical KLT workshop for school and district administrators was scheduled. None of the principals attended the workshop, a fact that was to hamper the implementation throughout.

In a comedy of errors spirit, the implementation resulted in some happy resolutions. Sometimes ETS accommodated to some of the local constraints by loosening the approach, either to build on district strengths or to work around problems on the ground, such as the lack of substitutes. At other times, ETS personnel did not veer from the tight part of the approach, as when they resisted teacher demands to provide a glossary of techniques, which in their view would lead to superficial implementation and restrict opportunities for rich conversations among teachers.

Valuable lessons were learned from maintaining tightness or permitting looseness and were ultimately incorporated into the KLT model. However, no reader of Wylie et al. (2008) can fail to be dismayed by the extent of the system-inherent challenges put in the way of the ETS implementation, in spite of what, at first, seemed like positive conditions for the initiative in the district. What ETS experienced in CMSD has echoes in other reports in this set and presents us with an important reality about system-wide change: systems can be obdurate, and approaches to changing them require both adherence and adaptation.

Paper 6: State-Level Scaling Up

In Tocci and Taylor (2008), we learn about a substantial effort to install AfL in a sample of pilot schools representing an entire state, Vermont. Tocci and Taylor have described an attempt to implement a variant of the KLT program in nine schools as part of the Vermont Formative Assessment Pilot Project. Although attempting to implement this AfL model at the state level, as could be anticipated, gives rise to challenges of increased scale not encountered when trying to establish AfL at the school or district level, we found it interesting that the same sorts of tensions between the tight and loose elements of the tight but loose implementation strategy are seen at all three levels, namely, the school, the district, and—in this instance—the state.

As always, the success of implementing AfL, as is true when implementing any substantive educational reform effort, is heavily dependent on the caliber and commitment of the individuals involved in the initiative and the financial and other resources devoted to the implementation. In this case, the resources allocated by the Vermont Department of Education (DOE) as well as the numbers and seniority levels of the DOE staff assigned to the project appear to be consonant with what seems warranted for state-level initiatives of this magnitude. It is also clear that the state's chief state school officer, the Vermont commissioner of education, was strongly supportive of the formative assessment initiative and, indeed, took an active part in a number of activities intended to launch the program properly.

Nine schools were selected from a number of schools whose staffs had volunteered to take part in the pilot. Most of the schools had been experiencing difficulty in satisfying the accountability requirements of the No Child Left Behind Act (NCLB). A variety of grade ranges were represented in the participating elementary, middle, and high schools serving as pilot schools.

Although the KLT program calls for the creation of school-based TLCs led by one of the teachers the school, Vermont's DOE preferred to use external coaches as TLC leaders. This departure from the usual KLT installation approach was adopted in the belief that DOE needed to supply the nine schools with "as much support as they could" so that the perceived burden on the schools would be minimized. At the pilot project's conclusion, a positive appraisal was reached about using external coaches as TLC leaders rather than having one of the school's teachers play that role. It appears, however, that during the pilot project itself there was considerable disagreement about this particular loosening of the KLT model.

We were impressed with the considerable up-front preparation by DOE staff who, working with ETS personnel, devoted considerable energy to coming up with a sensible conceptualization of how the KLT approach would work best for their particular state. Moreover, during the pilot tryout in the nine schools, a serious attempt was made to formatively evaluate the project itself. Two research coaches were appointed to provide mid-year and year-end

formatively focused appraisals of the pilot project's activities. This pilot project, clearly, was regarded as a high-priority undertaking by leaders of the Vermont DOE.

The tight but loose nature of the KLT model anticipates the likelihood that when any large-scale intervention is going to be installed in the real world with real educators, such real-world educators are apt to bring their own experience-honed preferences to the installation. In this instance of state-level AfL implementation, we see that key leaders in Vermont's DOE possessed certain preferences that, in some cases, ran counter to the typical way in which ETS staff members preferred to install their KLT program. Because the KLT program was installed with reasonable success in the nine Vermont pilot schools, this suggests that the Tight but Loose implementation strategy is sufficiently robust to accommodate such educator preferences—even at the state level.

Paper 7: To Tighten or To Loosen?

Through a detailed review of the five sites where *KLT* was implemented, Leusner, Ellsworth and Goe (2008) have identified a number of convergent and divergent themes that contribute to our understanding of the Tight but Loose framework. In their analysis of contrasting aspects of the implementation, they point to the times when the program must be tight, for example, the content and the processes of KLT, and when the reins can be loosened, for instance, in the order of the modules or meeting duration. They conclude that context is a crucial consideration for professional development, and that flexibility can be achieved without sacrificing core principles of the KLT program.

Leusner et al. (2008) have noted that scalability was at the forefront of the KLT research and discussion. In this regard, we think there are several issues referred to in the paper that might warrant further attention. The first is the issue of culture. Years of research document the role of system culture, (shared beliefs, customs, and behaviors) as either an inhibitor or a facilitator of change. In essence, the KLT program is about changing teacher behavior and, by extension, student behavior in the classroom. In the site where there was a pre-existing culture of professional collaboration and a commitment to improvement, the KLT implementation benefited.

These two cultural norms appear to be essential to the success of the KLT program. In the absence of such extant norms, then the job of the program is to change culture in order to impact behavior, and the research on culture tells us that cultural change is a slow process. The implementations reviewed in Leusner et al. (2008) are of a relatively short duration. It seems to us that bringing these classroom reforms to scale could often require considerable efforts to change culture before we see real and sustained changes in behavior. One wonders what other tightening or loosening of the framework might be required to make more fundamental cultural changes.

The second issue is that of voluntary teacher participation, which was common to each of the five sites. However, as far as scalability goes, we are left with the question of whether participation can always be voluntary. Granted, volunteers had the effect of becoming catalysts for other teachers' interest in AfL, in some instances, and resulted in increased teacher participation. But if participation is always voluntary, will reforms ever be achieved at scale? And if participation isn't voluntary, the same question lingers. Perhaps this is where culture comes in: if culture changes, then will all want to participate?

Third, another common element of implementation, intended to create teacher buy-in, was that teachers were able to determine what strategies and techniques they wanted to use in

their classrooms. While this seems sensible in terms of the intended purpose and also to ensure that teachers achieve early successes by starting with what is comfortable and doable for them (one assumes that they will not first take on something that is harder to do), the authors are silent on the specifics of how teachers moved beyond these initial steps into areas that were more challenging for them. Where new practices become more difficult to implement is essentially where rubber meets road. To make AfL a full and complete component of teaching and learning, teachers will need to take on challenges, and we would like to know, especially in relation to scaling up, what needed to be tightened, loosened, or indeed changed altogether once teachers were faced with strategies and techniques they perceived to be more difficult.

Finally, there is the issue of knowledge transmission, especially, but not exclusively, with third party-providers. We learn that ETS had to figure out how to share and make explicit internalized knowledge with a third party and, in turn, how the third party had to internalize the knowledge and make it operational. Then, the third party must repeat the same process, sharing newly acquired knowledge with the program participants. When we consider the potential for key understandings to be lost in translation, the knowledge sharing aspect of implementation could have a substantial impact on scale-up efforts.

Conclusion

We think these authors are to be congratulated on what they have accomplished in this set of papers. Thompson and Wiliam (2008) had provided a sound theoretical, empirical, and analytic framework for the following papers, spelling out, in compelling detail, the rationale for the scale-up strategy of the KLT program. The next six papers give us real stories about real contexts in which the KLT program was implemented, providing honest appraisals of the program's successes and challenges, and outlining the genuine obstacles to implementation inherent in the education system.

Taken as a whole, these papers represent an important contribution to our understanding of AfL, of effective professional development, and of the complexity of issues associated with scaling up. We are left in no doubt that the task of profoundly changing teacher behavior is a daunting one. But daunting as the task is, because the authors' illuminate the scale-up problems and possible solutions so clearly, these papers leave us with a feeling of optimism and a sense that it can be done.

Looking beyond this set of papers, we hope that in the future we will learn about the sustainability of the participating teachers' AfL practices after the initial 2-year implementation; the extent to which AfL becomes embedded in the daily practices of the schools, including the classrooms where teachers did not volunteer for the program; and the degree to which system changes needed to support the TLCs are instantiated. Finally, as the authors make clear throughout, the goal of AfL practices, and of scaling up the KLT program, is to impact student learning and to raise student achievement. It is, of course, still early days in terms of seeing effects on achievement, but we look forward to further research results about the effects of from KLT implementation on achievement. In the meantime, these authors have done a great service to the field, and we say "Bravo" to all of them.

References

- Balfanz, R., & Legters, N. (2004). *Locating the dropout crisis: Which high schools produce the nation's dropouts? Where are they located? Who attends them?* Baltimore, MD: Center for Research on the Education of Students Placed at Risk. (ERIC Document Reproduction Service No. ED484525).
- Barton, P. E. (2005). *One-third of a nation: Rising dropout rates and declining opportunities*. Princeton, NJ: ETS.
- 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, H. (1986). Assessment for learning. In D. L. Nutall (Ed.), *Assessing educational achievement* (pp. 7-18). London: Falmer Press.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). *Working inside the black box: Assessment for learning in the classroom*. London: King's College, Department of Education and Professional Studies.
- 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. J., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7-73.
- Black, P. J., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-148.
- Black, P., & Wiliam, D. (2006). Developing a theory of formative assessment. In J. R. Gardner (Ed.) *Assessment and learning* (pp. 81-100). Thousand Oaks, CA: Sage.
- Black, P., & Wiliam, D. (2007). Large-scale assessment systems: Design principles drawn from international comparisons. *Measurement: Interdisciplinary Research and Perspectives*, 5, 1-53.
- Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. In R. W. Taylor (Ed.), *The 68th yearbook of the National Society for the Study of Educational Evaluation: Part 2. New roles, new means* (pp. 26-50). Chicago: University of Chicago Press.
- Borko, H. (1997). New forms of classroom assessment: Implications for staff development. *Theory Into Practice*, 36(4), 231-238.
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, 33(8), 3-15.
- Borko, H., Mayfield, C., Marion, S. F., Flexer, R., & Cumbo, K. (1997). *Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development* (CSE Technical Report No. 423.). Los Angeles, CA: CRESST.
- Bransford, J., Brown, A., & Cocking, R. (1999). *How people learn: Brain, mind, experience, school*. Washington, DC: National Academy of Sciences.
- Brookhart, S. M. (2005, April). Research on formative classroom assessment. In *Formative classroom assessment: Research, theory, and practice*. Symposium conducted at the annual meeting of the American Educational Research Association, Montreal, Canada.
- Brousseau, G. (1984). The crucial role of the didactical contract in the analysis and construction of situations in teaching and learning mathematics. In H. G. Steiner (Ed.), *Theory of mathematics education* (Occasional paper no. 54., pp. 110-119). Bielefeld, Germany: Institut fur Didaktik der Mathematik der Universitat Bielefeld.

- Brown, M. L. (1983). *Graded tests in mathematics: The implications of various models for the mathematics curriculum*. British Educational Research Association. London: King's College London Centre for Educational Studies.
- Chase, S. E. (2005). Narrative inquiry: Multiple lenses, approaches, voices. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 651-679). Thousand Oaks, CA: Sage.
- Churchman, C. W. (1979). *The systems approach and its enemies*. New York: Basic Books.
- Churchman, C. W. (1982). *Thought and wisdom*. Seaside, CA: Intersystems Publications.
- Clymer, J. B., & Wiliam, D. (2006/2007). Improving the way we grade science. *Educational Leadership*, 64(4), 36-42.
- 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.
- Coburn, C. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, 32(6), 3-12.
- Cohen, D. K., & Hill, H. C. (1998). *State policy and classroom performance*. Philadelphia: University of Pennsylvania Consortium for Policy Research in Education.
- Cossentino, J. (2007). Evaluating Montessori: Why the results matter more than you think. *Education Week*, 26(21), 31-32.
- Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438-481.
- Cuban, L. (1998). How schools change reforms: Redefining success and failure. *Teachers College Record*, 99(3), 453-477.
- Darling-Hammond, L. (1999). *Teacher quality and student achievement: A review of state policy evidence*. Seattle WA: Center for the Study of Teaching and Policy, University of Washington.
- Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education Policy Analysis Archives*, 13, 42.
- Division of Abbott Implementation. (2005). *Excerpt from the most recent filing of Abbott Regulations regarding secondary education*. Trenton, NJ: New Jersey Department of Education.
- DuFour, R. (2004). What is a "professional learning community"? *Educational Leadership*, 61(8), 6-11.
- Dweck, C. (2000). *Self-theories: Their role in motivation, personality and development*. Philadelphia: Psychology Press.
- ETS. (2007). *Keeping Learning on Track: TLC leader kits*. Princeton, NJ: ETS.
- Ellsworth, J., Martinez, C., Lyon, C., & Wylie, E. C. (2007, April). *Leading while learning: Supporting the leaders of Keeping Learning on Track TLCs*. Paper presented at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago.
- Elmore, R. (2002). *Bridging the gap between standards and achievement: Report on the imperative for professional development in education*. Washington, DC: Albert Shanker Institute.
- Elmore, R. (2004). *School reform from the inside out: Policy, practice, and performance*. Cambridge, MA: Harvard Educational Review.

- Fennema, E., & Franke, M. L. (1992). Teachers' knowledge and its impact. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 147-164). New York: Macmillan.
- Fevre, D. L., & Richardson, V. (2000). *Staff development in early reading intervention programs: The facilitator* (CIERA Rep. No. 3-011). Ann Arbor: Center for the Improvement of Early Reading Achievement.
- Fullan, M. (1991). *The new meaning of educational change*. London: Cassell.
- Fullan, M. (2001). *Leading in a culture of change*. San Francisco: Jossey-Bass.
- Fullan, M., Hill, P. T., & Crevola, C (2006). *Breakthrough*. Thousand Oaks, CA: Corwin Press.
- 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 Educational Research Journal*, 38(4), 914-945.
- Garmston, R. & Wellman, B. (1999). *The adaptive school: A sourcebook for developing collaborative groups*. Norwood, MA: Christopher-Gordon Publishers.
- Gitomer, D. H., Latham, A. S., & Ziomek, R. (1999). *The academic quality of prospective teachers: The impact of admissions and licensure testing*. Princeton, NJ: ETS.
- Goe, L., & Mardy, D. (2007, April). A teacher-driven implementation of assessment for learning in New Jersey. In J. Ciofalo (Chair), *Tight but loose: Scaling up teacher professional development in diverse contexts*. Symposium conducted at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago.
- Goe, L., & Mardy, D. (2008). A teacher-driven implementation of Assessment for Learning in New Jersey In E. C. Wylie (Ed.), *Tight but loose: Scaling up teacher professional development in diverse contexts* (ETS Research Rep. No. RR-08-29, pp.45-54). Princeton, NJ: ETS.
- Hamre, B. K., & Pianta, R. C. (2005). Academic and social advantages for at-risk students placed in high quality first grade classrooms. *Child Development*, 76(5), 949-967.
- Hanushek, E. A. (2004). *Some simple analytics of school quality*. Washington, DC: National Bureau of Economic Research.
- Hanushek, E., Kain, J. F., O'Brien, D. M., & Rivken, S. G. (2005). *The market for teacher quality*. NBER working paper No. 11154. Washington, DC: National Bureau of Economic Research.
- Hawley, W. D., & Valli, L. (2001). The essentials of effective professional development: A new consensus In D. Boesel (Ed.), *Continuing professional development* (pp. 1-17). Washington, DC: U.S. Department of Education, National Library of Education.
- Hayes, V. P. (2003). *Using pupil self-evaluation within the formative assessment paradigm as a pedagogical tool*. Unpublished doctoral dissertation, University of London.
- Hendry, C. (1996). Understanding and creating whole organizational change through learning theory. *Human Relations*, 49(5), 621-641.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2), 371-406.
- Holzman, M. (2006). *Public education and Black male students: The 2006 state report card*. Schott Educational Inequity Index. Cambridge, MA: The Schott Foundation for Public Education.

- Ingvarson, L., Meiers, M., & Beavis, A. (2005). Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes & efficacy. *Education Policy Analysis Archives, 13*(10). Retrieved May 5, 2008 from <http://epaa.asu.edu/epaa/v13n10/>.
- Jepsen, C., & Rivkin, S. (2002). *What is the tradeoff between smaller classes and teacher quality?* (NBER Working Paper No. 9205). Washington, DC: National Bureau for Economic Research.
- Kazemi, E., & Franke, M. L. (2003). *Using student work to support professional development in elementary mathematics: A CTP working paper*. Seattle, WA: Center for the Study of Teaching and Policy.
- Kilpatrick, J. (2003). *Teachers' knowledge of mathematics and its role in teacher preparation and professional development programs*. Paper presented at the German-American science and math education research conference, Kiel, Germany.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254-284.
- Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–196). Cambridge, UK: Cambridge University Press.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.
- Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute-by-minute and day-by-day. *Educational Leadership, 63*(3), 18-24.
- Lewis, C., Perry, R., & Murata, A. (2006). How should research contribute to instructional improvement? The case of lesson study. *Educational Researcher, 35*(3), 3-14.
- Librera, W. L. (2004, December). *Ensuring quality teaching and learning for New Jersey's students and educators*. Paper presented to the New Jersey State Board of Education.
- Lillard, A., & Else-Quest, N. (2006). Evaluating Montessori education. *Science, 313*(5795), 1893-1894.
- Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educational Evaluation and Policy Analysis, 15*(2), 129-152.
- Leusner, D. M., Ellsworth, J., & Goe, L. (2008). Scaling up across diverse contexts: Lessons learned from five implementations of the *Keeping Learning on Track* Program. In E. C. Wylie (Ed.), *Tight but loose: Scaling up teacher professional development in diverse contexts* (ETS Research Rep. No. RR-08-29, pp. 105-124). Princeton, NJ: ETS.
- Lyon, C., Cleland, D., & Gannon, M. (2007, April). Letting go of the reins: Learning about scalability in the context of one district-wide implementation. In J. Ciofalo (Chair), *Tight but loose: Scaling up teacher professional development in diverse contexts*. Symposium conducted at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago.
- Lyon, C., Cleland, D., & Gannon, M. (2008). Letting go of the reins: Learning about scalability in the context of one district-wide implementation. In E. C. Wylie (Ed.), *Tight but loose: Scaling up teacher professional development in diverse contexts* (ETS Research Rep. No. RR-08-29, pp. 67-78). Princeton, NJ: ETS.

- Lyon, C., Wylie, E. C., & Goe, L. (2006, April). *Changing teachers, changing schools*. Paper presented at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, San Francisco, CA.
- Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Erlbaum.
- Maher, J., & Wiliam, D. (2007, April). *Keep Learning on Track in new teacher induction*. In J. Ciofalo (Chair), *Tight but loose: Scaling up teacher professional development in diverse contexts*. Symposium conducted at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago.
- Maher, J., & Wiliam, D. (2007, April). *Keeping Learning on Track in new teacher induction*. In E. C. Wylie (Ed.), *Tight but loose: Scaling up teacher professional development in diverse contexts* (ETS Research Rep. No. RR-08-29, pp. 55-66). Princeton, NJ: ETS.
- Marzano, R. J. (2003). *What works in schools: Translating research into action*. Baltimore, MD: ASCD.
- McLaughlin, M., & Talbert, J. (1993). *Contexts that matter for teaching and learning: Strategic opportunities for meeting the nation's educational goals*. Palo Alto, CA: Stanford University Center for Research on the Context of Secondary School Teaching.
- McLaughlin, M., & Talbert, J. E. (2006). *Building school-based TLCs: Professional strategies to improve student achievement*. New York: Teachers College, Columbia University.
- National Center for Education Statistics. (2005). *Past and projected elementary and secondary public school enrollments: Public elementary and secondary school enrollment in prekindergarten through grade 12, by grade level and region, with projections: Various years, fall 1965–2015*. Retrieved March 12, 2007, from <http://nces.ed.gov/programs/coe/2006/section1/table.asp?tableID=432>.
- National Commission on Mathematics and Science Teaching for the 21st Century. (2000). *Before it's too late: A report to the nation from The National Commission on Mathematics and Science Teaching for the 21st Century*. Washington, DC: Author.
- National Staff Development Council. (2001). *Standards for staff development*. Oxford, OH: Author.
- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, 22(2), 155-175.
- Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. New York: Oxford University Press.
- North Carolina Department of Public Instruction. (2005). *Report and recommendations from the State Board of Education Teacher Retention Task Force*. Raleigh, NC: Author.
- 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.
- Pappano, L. (2007). More than 'making nice': Getting teachers to (truly) collaborate. *Harvard Education Letter*, 23(2), 1-3.
- Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning: Towards a wider conceptual field. *Assessment in Education: Principles, Policy and Practice*, 5(1), 85-102.
- Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher*, 29(1), 4-15.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28(1), 4-13.

- Reeves, J., McCall, J., & MacGilchrist, B. (2001). Change leadership: Planning, conceptualization, and perception. In J. Macbeath & P. Mortimer (Eds.), *Improving school effectiveness* (pp. 122-137). Buckingham, UK: Open University Press.
- Rodgers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record*, 104(4), 842-866.
- Rodriguez, M. C. (2004). The role of classroom assessment in student performance on TIMSS. *Applied Measurement in Education*, 17(1), 1-24.
- Ross, P. E. (2006). The expert mind. *Scientific American*, 295(2), 64-71.
- Sandoval, W., Deneroff, V., & Franke, M. L. (2002). *Teaching, as learning, as inquiry: Moving beyond activity in the analysis of teaching practice*. New Orleans, LA: American Educational Research Association.
- Schein, E. H. (1996). Culture: The missing concept in organization studies. *Administrative Science Quarterly*, 41(2), 229-240.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39-83). Chicago: Rand McNally.
- Slavin, R. E. (1995). *Cooperative learning: Theory, research and practice*. Boston: Allyn & Bacon.
- St. Mary's County Public Schools. (2004). *Fulfilling the promise in every child*. Leonardtown, MD: Author.
- Stiggins, R. J. (2001). *Student-involved classroom assessment* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.
- Stiggins, R. J., Arter, J. A., Chappuis, J., & Chappuis, S. (2004). *Classroom assessment for student learning: Doing it right—using it well*. Portland, OR: Assessment Training Institute.
- Thompson, M., & Goe, L. (2006, April). *Models for effective and scalable teacher professional development*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Thompson, M., & Wiliam, D. (2007, April). *Tight but loose: A conceptual framework for scaling up school reforms*. In J. Ciofalo (Chair), *Tight but loose: Scaling up teacher professional development in diverse contexts*. Symposium conducted at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago.
- Thompson, M., & Wiliam, D. (2008). Tight but loose: A conceptual framework for scaling up school reforms. In E. C. Wylie (Ed.), *Tight but loose: Scaling up teacher professional development in diverse contexts* (ETS Research Rep. No. RR-08-29, pp. 1-44). Princeton, NJ: ETS.
- Tocci, C., & Taylor, G. (2007, April). *A state-sponsored pilot project in selected schools in Vermont*. In J. Ciofalo (Chair), *Tight but loose: Scaling up teacher professional development in diverse contexts*. Symposium conducted at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago.
- Tocci, C., & Taylor, G. (2007, April). A state-sponsored pilot project in selected schools in Vermont. In E. C. Wylie (Ed.), *Tight but loose: Scaling up teacher professional development in diverse contexts* (ETS Research Rep. No. RR-08-29, pp. 125-134). Princeton, NJ: ETS.

- U.S. Department of Education. (2005). *Highly qualified teachers: Improving teacher quality state grants: ESEA title II, part A, non-regulatory guidance*. Washington, DC: U.S. Department of Education.
- Ulrich, W. (2002). *An appreciation of C. West Churchman*. Retrieved February 19, 2007, from http://www.geocities.com/csh_home/cwc_appreciation.html
- Wiggins, G., & McTighe, J. (2000). *Understanding by design*. New York: Prentice Hall.
- Wiliam, D. (2003). The impact of educational research on mathematics education. In A. Bishop, M. A. Clements, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Second international handbook of mathematics education* (pp. 469-488). Dordrecht, Netherlands: Kluwer Academic Publishers.
- Wiliam, D. (2007). *Keeping Learning on Track: Classroom assessment and the regulation of learning*. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1051-1098). Greenwich, CT: Information Age Publishing.
- Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy, and Practice*, 11(1), 49-65.
- Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53-82). Mahwah, NJ: Lawrence Erlbaum Associates.
- 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.
- Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11, 57-67.
- Wylie, E. C., Lyon, C., Martinez, C., & Ellsworth, J. (2007, April). *How formative assessment ideas translate from talk to practice*. Presented at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago.
- Wylie, C., Thompson, M., Lyon, C., & Snodgrass, D. (2007, April). *Keeping Learning on Track in an urban district's low performing schools*. In J. Ciofalo (Chair), *Tight but loose: Scaling up teacher professional development in diverse contexts*. Symposium conducted at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago.
- Wylie, C., Thompson, M., Lyon, C., & Snodgrass, D. (2008). *Keeping Learning on Track in an urban district's low performing schools*. In E. C. Wylie (Ed.), *Tight but loose: Scaling up teacher professional development in diverse contexts* (ETS Research Rep. No. RR-08-29, pp. 79-92). Princeton, NJ: ETS.
- Wylie, E. C., Thompson, M., & Wiliam, D. (2007, April). *Keep Learning on Track: Impact on student learning*. In J. Ciofalo (Chair), *Tight but loose: Scaling up teacher professional development in diverse contexts*. Symposium conducted at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago.