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

Marnie Thompson
RPM

Dylan Wiliam
Institute for Education, London

Introduction

Teaching and learning aren't working very well in the United States. A lot of effort and resource, 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 as one measure of the effectiveness of aggregate current practice is sobering. Nationally, graduation rates hover below 70% (Barton, 2005), certainly not the hallmark of an educated society. Worse, for the students who are most likely to land in low performing schools—poor kids and kids of color—graduation rates are even more appalling. The Schott Foundation (Holzman, 2006) reports 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 go 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 these kinds of 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 the daily, minute-to-minute interactions that take place 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 and Wiliam, 1998; Elmore, 2004; 2002; Fullan, Hill and Crevola, 2006). And we have to figure out how to do this at scale, if we are at all serious about improving the educational outcomes of all students, especially students now stuck in chronically low performing schools.

Scaling up a classroom-based intervention isn't 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 U.S.) is a different kind of challenge. Not only is the sheer number of classrooms daunting, the complexity of the systems in which classrooms exist, the separateness of these classrooms, and the private nature of the activity of teaching means that each and every teacher has to "get it" and "do it" right, all on their own. No one else can do it for them, just as no one else can do students' learning for them. No matter how good the intervention's theory of action, no matter how well designed its components, the design and implementation effort will be wasted if it doesn't actually improve teachers' practices—in all the diverse contexts in which they work, and with a high level of quality. This is the challenge of scaling up.

This paper is the opening paper in a symposium dedicated to discussing one promising intervention into the "black box"—a minute-to-minute and day-by-day approach to formative assessment that deliberately blurs the boundaries between assessment and instruction, called *Keeping Learning on Track*—and our attempts to build this intervention in a way that tackles the scalability issue head on. While *Keeping Learning on Track* is in many ways quite highly developed, we are in midstream in our understanding and development of 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 *Keeping Learning on Track* intervention, this paper also offers a theoretical framework that we call "Tight but Loose," as a tool that can 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 within five "place-based" stories that are presented in the next papers in the symposium. By comparing context-based differences in program implementation and examining the outcomes achieved, it is possible to discern "rules" for implementing *Keeping Learning on Track* and more general lessons about scaling up classroom-based interventions. These ideas are taken up in a concluding paper in the symposium, 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 *Keeping Learning on Track*, 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, drawing in 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 *Keeping Learning on Track* across diverse contexts.

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

Keeping Learning on Track is fundamentally a sustained teacher professional development program, and as such, it has deep roots in the notion of capacity building described by Elmore (2004; 2002). We were led to teacher professional development as the fundamental lever for improving student learning by a growing research base 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 and Pianta, 2005; Hanushek, Kain, O'Brien and Rivken, 2005; Wright, Horn and Sanders, 1997). Through this logic, we join Elmore and others—notably Fullan (2001) and Fullan, Hill, *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.

Keeping Learning on Track 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;

2. A process component (*how* we support teachers to learn about and adopt assessment for learning 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 assessment for learning 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 and MacGilchrist, 2001; Wilson and 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 *Keeping Learning on Track* 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 "assessment for learning" (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 *Keeping Learning on Track* 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 Assessment for Learning

Knowing that teachers make a difference is not the same as knowing how teachers make a difference. From the research summarized briefly 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 and 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 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 turns out to be relatively undemanding. In the US, it costs around $25,000 to produce one standard deviation increase in one student's achievement. This estimate is based on the fact that one year's growth on tests used in international comparisons, such as TIMSS and 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 and 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, like investing in teacher professional development, can show a significant advantage over recurrent costs such as class-size reduction.

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 assessment for learning, 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 two 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, 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 (1998) 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 and Wiliam, 2007 for more on this point).

A subsequent intervention study (Black, Harrison, Lee, Marshall and Wiliam, 2003) involved 24 math and science teachers who were provided professional development designed to get them to utilize 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 and Black, 2004). Other small-scale replications (Clymer and 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 symposium, the most important are summarized briefly below.

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 and 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 (forthcoming in 2007) and Wiliam and Thompson (2006).

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 and Wiliam, 2006, p. 91)

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 U.S., reform efforts around formative assessment face a severe problem, due to the use of the term "formative assessment" (and, more recently, "assessment for learning") 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 assessment for learning or formative assessment. In this process, our original view 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 assessment for learning, 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 which 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 *Keeping Learning on Track*, 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 her or himself, it is possible to "unpack" the "big idea" of formative assessment into five key strategies, as shown in Table 1.

	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	

Figure 1: Deriving the five key strategies of assessment for learning

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

The five strategies certainly bring the ideas of assessment for learning closer to being 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) versus "know why" (knowledge of universal truths). For a fuller discussion of this contrast, see Wiliam (2003). We argue in this paper that the scalability of a complex intervention requires both, because helping teachers "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 and 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.

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 assessment for learning 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 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, 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 non-threatening 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 x 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 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 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 the 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 "="; if the assignment is better than the last one, she gives it a "+"; and if the assignment is not as good as the last one, she gives it a "−". 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 "Pre-Flight Checklist" or Rubric. Students trade papers and check each other's work against a "pre-flight 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 pre-flight checklist is a list of the required, basic components for an assignment, such as "title page, introduction, 5-paragraph explanation, conclusion." The pre-flight 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. We 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 pre-designated 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 ten techniques represent fewer than ten percent of the techniques now catalogued in the *Keeping Learning on Track* program. See Leahy, Lyon, Thompson, and Wiliam (2005) for descriptions of additional techniques used by U.S. teachers in enacting assessment for learning. Wiliam (2007) lists several more and goes into a great deal of detail on the empirical basis behind many specific techniques.
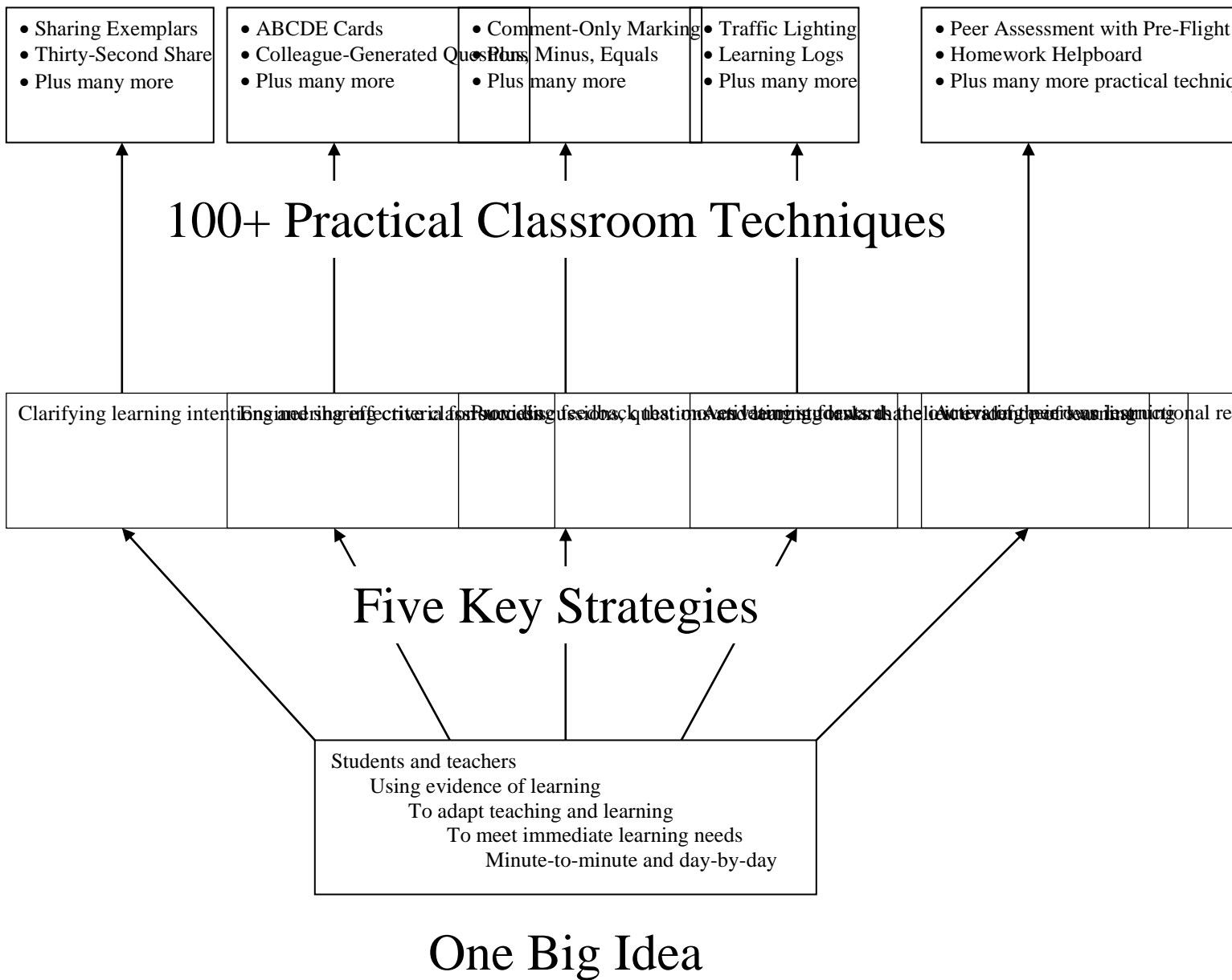
We note that all the techniques are decidedly low-tech, low cost, and usually within the capabilities of individual teachers to implement. In this way, they differ dramatically from large-scale interventions like class-size reduction or curriculum overhauls, which can be quite expensive and difficult to implement because they require school or system level changes. We also note that 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, Harrison et al., 2003; Leahy, Lyon 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 grades math teachers may not work well at all in a 2nd grade writing lesson. It is even true that what works for 7th grade pre-algebra in one classroom may not work for 7th grade pre-algebra in another classroom, even if it's right down the hall—because of student or teacher differences. Given this variation, it's important to provide teachers 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 one-hundred-plus practical techniques, we have devised the graphic in Figure 2, which appears on the next page.

Figure 2. How the Big Idea, five key strategies, and practical techniques of *Keeping Learning on Track* relate

• Sharing Exemplars • Thirty-Second Share • Plus many more	• ABCDE Cards • Colleague-Generated Questions • Plus many more	• Comment-Only Marking • Plus, Minus, Equals • Plus many more	• Traffic Lighting • Learning Logs • Plus many more	• Peer Assessment with Pre-Flight • Homework Helpboard • Plus many more practical technic

100+ Practical Classroom Techniques

Clarifying learning intentions and sharing criteria for success	Engineering effective classroom discussions, questions...	Providing feedback that moves learners...	Activating students as the owners of their own learning	Activating students as instructional re

Five Key Strategies

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

One Big Idea

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 assessment for learning 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 assessment for learning 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 within 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 assessment for learning. 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 de-professionalization that have occurred in many school faculties.

Without effective professional development systems to teach teachers to "do" assessment for learning, 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 and Yoon, 2001). The challenge is to develop models of professional development and scalable systems of delivery that faithfully disseminate the content of assessment for learning, 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, *Keeping Learning on Track* supports two distinct phases of professional learning: 1) initial exposure and motivation, and 2) ongoing guided learning, practice, reflection, and adjustment. These phases and the research and development process that led to their current structure within *Keeping Learning on Track* are explained in detail in an earlier paper by Thompson and Goe (2006), from which much of the following is drawn.

In phase 1, teachers and school leaders are exposed through an interactive, two-day workshop or seminar to an overview and basic information about assessment for learning, 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 Big Idea that unifies and drives the five strategies of *Keeping Learning on Track*;

- The five assessment for learning strategies of *Keeping Learning on Track;*

- 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 assessment for learning come alive in their classrooms. In the course of the introductory workshop, participants get direct experience with dozens of the *Keeping Learning on Track* 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 *Keeping Learning on Track* teacher learning communities.

Woven throughout the workshop are presentations on the research base for *Keeping Learning on Track*: how we know assessment for learning and sustained teacher learning communities work to change teacher practices and improve student learning. To motivate interest in assessment for learning 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 U.S. (Clymer and Wiliam, 2006/2007; Wylie, Thompson, Lyon and 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 assessment for learning 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 daily life in schools do not naturally allow teachers and school leaders to develop expertise in complex interventions like assessment for learning. Thus, a primary process by which *Keeping Learning on Track* attempts to effect these changes in teachers' practice is via school-embedded teacher learning communities. These have the potential to provide teachers with the information and support they need to develop their practice in deep and lasting ways, and are designed to build school capacity to support individual and institutional change over time.

Developing teacher expertise through teacher learning communities

Teacher learning communities embody critical process elements needed 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 and Dean, 2003), takes place over a period of time rather than being in the form of one-day workshops (Cohen and Hill, 1998), and involves teachers in active, collective participation (Garet, Porter et al., 2001; Ingvarson, Meiers and Beavis, 2005).

There are, of course, many professional development structures that would be consistent with this research base, but we believe that teacher learning communities, 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 assessment for learning.

There is a growing evidence base on how to build and sustain teacher learning communities (Borko, 1997; Borko, 2004; Borko, Mayfield, Marion, Flexer and Cumbo, 1997; Elmore, 2002; Garmston and Wellman, 1999; Kazemi and Franke, 2003; McLaughlin and Talbert, 1993; 2006; Putnam and Borko, 2000; Sandoval, Deneroff and 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 school and classroom, 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 of its treatment of learning as 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 assessment for learning is no different. An expert in assessment for learning 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 non-essential details. She is then able to swiftly compare that situation with her intended goals for the lesson, her knowledge of the content being taught, her developmental knowledge of students in general and these students in particular, and other relevant schema. Guided by the results of these comparisons, she then selects her 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 non-experts 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-non-expert 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 non-experts 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 non-experts. There were subtle but essential details of real classroom life that were either absent from the staged depictions or out-of-sync. The non-experts 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 assessment for learning, as it is for most complex teaching behaviors. Learning to "do" assessment for learning 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 assessment for learning 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). Teacher learning communities 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 teacher learning communities is usually warmly greeted by teachers—who generally wish for more collegiality in their professional lives—we did not elect to employ them 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 needed 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 chess players playing literally thousands of "practice games." As Ross (2006) points out, experts practice differently from non-experts, 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 teacher learning communities for *Keeping Learning on Track* because their grassroots character lends itself to scaling up the intervention. As Black *et al*. (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 assessment for learning? 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 teacher learning communities to scale *with fidelity* to the original assessment for learning 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 teacher learning communities seem to be particularly functional vehicles to support teacher learning about assessment for learning. First, the practice of assessment for learning 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 teacher learning communities 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, teacher learning communities are 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 assessment for learning 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, teacher learning communities redress a fundamental limitation of assessment for learning, which is its (perhaps paradoxical) generality and specificity. The five assessment for learning 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. Teacher learning communities provide a forum for supporting teachers in converting the broad assessment for learning strategies into "lived" practices within their specific subjects and classrooms.

This "bonus" feature of teacher learning communities focused on assessment for learning—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 and Franke, 1992; Gitomer, Latham and 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 in observations of 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 teacher learning communities focused on assessment for learning 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, teacher learning communities 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 proofs" 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 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 support teachers need to take such risks.

Because "learning by doing" is integral to the development of expertise in the complex realm of assessment for learning, 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 assessment for learning 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 the

facilitator and advocate for the teacher learning community. 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 like they know more than they do. They want the "breathing room" to ask for help in the places 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 assessment for learning, unless the coach has previously developed expertise in assessment for learning through their use and refinement in their own classroom. Since we know that such expertise is rare in classroom teachers, we have reason to believe that few coaches and principals developed this expertise before they left the classroom. This is not to say that coaches and principals can play no useful role in supporting teacher learning communities focused on assessment for learning, 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 assessment for learning 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 peripherality 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. They suggest that "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 teacher learning communities, 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.

Teacher learning communities 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 teacher learning communities consistently and effectively is not as simple as changing federal and state regulations and funding frameworks for professional development.

Significant structural barriers to teacher learning communities 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 teacher learning communities, it is important to say that we are not endorsing teacher learning communities 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 teacher learning communities 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.

Strong guidance for ongoing learning about for assessment for learning, through modules

The most popular notions of professional learning communities assert the need for teachers to meet and plan collaboratively, and generally insist that "data" occupy a central place in 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 *Keeping Learning on Track* teacher learning communities than is typically the case in the professional learning communities 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, comprised of directions and materials for 90-120 minutes worth of group study, represent an attempt to scaffold in enough content that there is a decent shot at maintaining fidelity, one of the challenges of using teacher learning communities to scale up any intervention. Recognizing that there are no formative assessment experts in most teacher learning communities, each module provides explicit guidance for the conduct of a monthly teacher learning community meeting. Each module contains an agenda, detailed leader notes with guidance for timing and discussion points, plus 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). *Keeping Learning on Track* offers enough modules to cover two 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 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 how 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 their most recent experiences trying out assessment for learning techniques. Knowing that they will be asked to report on their most recent efforts was shown in Black, Harrison, et al. (2003) be a helpful, even necessary, spur to action. This is balanced by an activity near the end of the module: a segment devoted to Personal Action Planning. This is a time for teachers to describe on paper their next steps in trying out and refining assessment for learning 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 are to practice assessment for learning 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 how to improve student learning through assessment for learning.

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 teacher learning community 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 assessment for learning 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, 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 *Keeping Learning on Track*, 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 more as "bait" 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 leader 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 isn't "coverage of the material" (an extension of the pressure they are under in their own classrooms).

Extended support and guidance for teacher 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 teacher learning community modules were our first step in this direction. These proved to be necessary but insufficient—as it turns out, sustaining teacher learning communities has its own complexity and context-driven peculiarities, much like assessment for learning. The development, pilot-testing, and refinement of the assessment for learning modules goes a long way toward ensuring that *when* teacher learning communities meet, they maintain a strong and faithful focus on helping teachers adopt assessment for learning 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 teacher learning communities *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 teacher learning community 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 assessment for learning, where the content focus was the institutionalization of teacher learning communities. 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 teacher learning communities is the lack of time within regular school hours when teachers can meet to discuss teaching and learning or 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 teacher learning communities, for 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 declines. In one district's implementation (described in Wylie, Thompson et al., 2007), there is a required, monthly, day-long workshop for all the teacher learning community leaders in the district. In other districts, the support meetings for teacher learning community leaders are less frequent or not as long, but the gist is the same. In these meetings, the teacher learning community leaders function in two roles successively: first, as teachers learning about applying assessment for learning in their own classrooms, and second, as advocates and facilitators for teacher learning communities.

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 assessment for learning. 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's up and running; a protocol for reviewing the next module in a sequence; exercises to develop coaching skills for facilitators; 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 teacher learning community leaders is spent in a kind of "whole-group" How's It Going session, in which the leaders report on their own experiences using assessment for learning. A *Keeping Learning on Track* 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 assessment for learning. 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 a teacher 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 assessment for learning 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 assessment for learning.) So the *Keeping Learning on Track* 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 assessment for learning technique called "hot seat questioning."

Once the exact nature of what is being done is made clear, the leader can then move onto 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 assessment for learning 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 her 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), they come to the next meeting having already sorted through these preliminaries, able to present a much more concrete and grounded narrative of their assessment for learning 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 they were 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 assessment for learning, 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 *Keeping Learning on Track* 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 teacher learning community.

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 *Keeping Learning On Track* 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, *Keeping Learning on Track* 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 *Keeping Learning on Track*, the three-step process looks like this:

(A) Teachers learn extensively and deeply about minute-to-minute and day-by-day assessment for learning via an initial workshop and sustained engagement in teacher learning communities →

(B) Teachers make minute-to-minute and day-by-day assessment for learning a central part of their everyday teaching practice, implementing the Big Idea and five strategies of assessment for learning through judiciously chosen practical techniques →

(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.

We will explicate each of these three "super-steps" in turn, expanding on the theory of action of each step.

Super-step (A): Teachers learn extensively and deeply about minute-to-minute and day-by-day assessment for learning via an initial workshop and sustained engagement in teacher learning communities

There are three important underlying aspects to this super-step: 1) Teachers are learners; 2) There has to be a complete and correct transmission of the knowledge base for minute-to-minute and day-by-day assessment for learning; 3) The nature of the learning required to become proficient at assessment for learning 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 taken care of 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 assessment for learning. 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 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 assessment for learning. This second aspect may seem self-evident, but the development, refinement, extension, and documentation of the knowledge base of *Keeping Learning on Track* 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 eight 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 kinds 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 assessment for learning 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 assessment for learning 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 assessment for learning 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 like they are unable to handle anything more complex than a script. It takes time and practice for teachers to internalize the assessment for learning 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 super-step.

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 assessment for learning. First, let's look more closely at the How's It Going segment of a *Keeping Learning on Track* 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 U.S., 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 and Goe, 2006; Thompson and Goe, 2006), teachers knowing they will be expected to report out on their most recent efforts is often what motivates them to follow through on trying out some aspect of assessment for learning. 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 assessment for learning 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 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 teacher learning community 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 *Keeping Learning on Track* 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 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:

- 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 *Keeping Learning on Track* knowledge base, including the problem of competition from other initiatives, must be detected and minimized.

- It takes time to develop expertise in assessment for learning, 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 two-year commitment to this process is doomed to fail, so they shouldn't 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 assessment for learning.

- Meetings need to last long enough that every teacher gets focused, individual attention for the specific problems they are facing in their classroom.

- Attendance at these meetings is necessary but insufficient—merely showing up doesn't cut it. Rather, active and full participation is a requirement. Everyone must report on his or her latest assessment for learning efforts. Reports and feedback must be detailed and bend toward a rigorous application of the theory of action of assessment for learning. And everyone must commit—in writing—to their next steps in instituting assessment for learning in their classroom.

- There is an unavoidable upper limit to the number of teachers in a teacher learning community, because if there are too many members, there simply won't be enough time for each teacher to get individual attention for their unique problems in instituting assessment for learning.

- Since the essential parts of a teacher learning community meeting are the reflective self-reports 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 *Keeping Learning on Track*-sanctioned learning community that follows a *Keeping Learning on Track* 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 assessment for learning a central part of their everyday teaching practice, implementing the Big Idea and five strategies of assessment for learning 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 assessment for learning central in their teaching practice. We would have to concede that the *Keeping Learning on Track* 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. "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" (Fullan, 1991, 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: 1) The paired notions that assessment for learning becomes central to practice (centrality) and is used minute-to-minute and day-by-day (frequency), and 2) 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 assessment for learning 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 assessment for learning *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 assessment for learning is second nature and interwoven into everything they do. And this can only be achieved through a sustained program of support, such as outlined in the first step in the theory of action.

While 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 (Wiliam and Thompson, 2006) just doesn't provide enough information on enough days to have much of a chance to make a difference in students' learning. It's reasonably good at helping teachers make larger decisions about things like curricular emphases, but it's 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 isn't one with a the big payoffs that are so often quoted from Black and Wiliam's 1998 meta-analysis on the effects of formative assessment.

Minute-to-minute and day-by-day assessment for learning, 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 goes up as well. Furthermore, when implemented in a "dense" way, such that it really does become central to practice, assessment for learning is an intervention that changes the entire classroom culture to one focused on learning (Black, Harrison 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 assessment for learning 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 assessment for learning expressed in *Keeping Learning on Track* 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: 1) we "require" teachers to take professional responsibility for their classrooms in regard to each of the five strategies; and 2) 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:

- 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 doesn't 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 assessment for learning 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: 1) the particular ways in which teaching is made more responsive to the immediate learning needs of students, and 2) 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 assessment for learning techniques represented within *Keeping Learning on Track* though Wiliam (forthcoming in 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 their answers to questions asked by the teacher. This requires every student in the class to show their 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 doesn't happen except when there is thinking going on (Bransford, Brown and 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. Her next decisions will be taken in the dark, with the result that she may move on before students "get it" or dwell too long on a topic that everyone already understands. Even if she doesn't move on at the wrong time, she will be in the dark about the exact nature of any misconceptions students harbor if she doesn't 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 she receives 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 won't 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 isn't 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 re-distribute 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 assessment for learning 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 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, like "think-pair-share," students serve as sounding boards for one another, whereby thinking can be developed and deepened. In techniques like "best composite paper" or

the various "jig-saw" techniques, student thinking is pooled to make a joint work product that is better than 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 she can concentrate her 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 their 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 assessment for learning 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 they 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 "surface" 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 their 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, Brown 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:

- The professional learning portion of the intervention needs to support teachers effectively to have a reasonably good idea of how to implement the individual techniques and help them understand well why they work, so that the mechanism of each technique can be brought to bear.

- Teachers' persistent practice of selected assessment for learning 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 like 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 assessment for learning on 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 *Keeping Learning on Track,* 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 programs' 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 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 kids 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 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) do a good job of explaining the reasons that a strong theory of action (what they call "explication of the innovation mechanism" (p. 5)) must be coupled with clarity on the intervention itself (the "descriptive knowledge base" (p. 4)). They point to the problem of an innovation's surface features obscuring its true underlying mechanism. For example, they cite the way that "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).

They then apply this idea to the distortions they have seen in the way lesson study has been translated into practice in the U.S., 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 U.S. 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 *Keeping Learning on Track* 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 *Keeping Learning on Track*—a testament to the pervasiveness of this problem (more about these in a later section of this paper).

Factor 2: A comprehensive notion of what it means to scale up an intervention across diverse contexts

As Coburn (2003) points 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, she goes 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 like assessment for learning. Coburn proposes a needed expansion of the notion of scaling up that is rooted in the notion of "consequential change," by which she means change that makes a difference for teaching and learning. She discusses four inter-related dimensions to scale: 1) depth, 2) sustainability, 3) spread, and 4) 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 *Keeping Learning on Track.* For example, she discusses *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 *Keeping Learning on Track*.

She then turns to the dimension of *sustainability*, focusing on the idea of consequential change sustained over time. She points 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 teacher learning communities, 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 takes up the idea of *spread*, which she defines 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 *Keeping Learning on Track*, spread within might show itself in a classroom when a teacher who has been applying assessment for learning to one subject begins applying it across all the subjects 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 takes 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 cautions against an under-conceptualization of this dimension as being only about "buy-in" or acceptance of the reform. Rather, she emphasizes a real shift in knowledge about and authority/capacity for extending the reform. The *Keeping Learning on Track* 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 et al., 2007 and other papers in this session).

The expanded notion of scale that Coburn proposes 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're 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 assessment for learning—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) doesn't 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).

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 thinks 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 like they are sensitive to systemic particularities (if only because they have tend 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 assessment for learning as a central part of their teaching practice. Where the system works to support that, then leave it alone. Where the system is in the way, change it. We agree wholeheartedly with Elmore (2004; 2002) 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 like "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, negotiating 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 come in handy.

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 *Keeping Learning on Track*. 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 its 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 points to the reform's origins in late 70s and early 80s empirical studies that 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 goes 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 Essential 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 examines this history while outlining five competing, seldom explicit, criteria that are used for judging a reform's success or failure. He states 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 says:

> [T]here 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, Perry *et al.* 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 five and twelve year olds who got into the program by lottery, compared to students who were not admitted in the lottery.

Cossentino's commentary on the study and Montessori is quite enthusiastic. She begins by highlighting the deep empirical and theoretical work that stands behind the Montessori method:

> [C]ontemporary 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 U.S.—in its commitment to the theory of action of a single approach. Cossentino recognizes this in her commentary and then speaks 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 systemwide understanding of the core technologies of teaching and learning. In Montessori schools, this means deep knowledge of what Montessori is (and isn't). 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 in place in a small number of (mostly private) schools in the U.S. 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). Heck, as proponents of a different complex intervention with a deep theory of action and knowledge base, we're in exactly the same boat. We're 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 *Keeping Learning on Track*: 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's 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 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 *Keeping Learning on Track*. 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 symposium. 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 *Keeping Learning on Track*. 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 *Keeping Learning on Track*? That is the text, or at least the sub-text, of the next papers in this symposium, which will relate a set of place-based stories of *Keeping Learning on Track* in implementation 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 white boards to boost student engagement and to get information on what student thinking looks like, then you have to regularly expect every student to hold theirs 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 white boards 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 *Keeping Learning on Track.* 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 assessment for learning. 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 assessment for learning. That is why we build in the explicit expectation that teachers participate in learning communities focused on assessment for learning. 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 assessment for learning.

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 assessment for learning 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 assessment for learning in their own classroom, get feedback, and come up with a plan 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 twenty 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 their 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 *Keeping Learning on Track* with other reforms, and they looked for techniques that mapped nicely onto these. In essence, they wanted to use the fact that *Keeping Learning on Track* 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 assessment for learning needs to learn how it applies in all five of the strategies, not just the one or two that hold immediate appeal for them. They don't have to use every technique, not at once, and not ever. But they do have to, over time (the span of one to two 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 *Keeping Learning on Track* 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 *Keeping Learning on Track* is to relate to certain system policies and practices, like parent communications, report cards, and the like. Because the "loose list" list is likely to include a lot things outside the classroom, it's easy to think of these as "systemic" issues, and then to jump to the idea that we are "loose" about systemic issues. But that doesn't 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 teacher learning community 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 *Keeping Learning on Track* "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, in this paper, we have discussed in detail one particular intervention—a professional development program entitled *Keeping Learning on Track*. 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 ten 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 teacher learning communities 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.

References

Balfanz, R. and N. Legters (2004). Locating the dropout crisis: Which high schools produce the nation's dropouts? Where are they located? Who attends them?, Johns Hopkins University.

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. <u>Creating powerful thinking in teachers and students: diverse perspectives</u>. J. N. Mangieri and C. C. Block. Fort Worth, TX, Harcourt Brace College**:** 161-186.

Black, H. (1986). Assessment for learning. <u>Assessing educational achievement</u>. D. L. Nutall. London, Falmer press**:** 7-18.

Black, P., C. Harrison, C. Lee, B. Marshall and W. Dylan (2002). Working inside the black box: Assessment for learning in the classroom. London, King's College, Department of Education and Professional Studies.

Black, P., C. Harrison, C. Lee, B. Marshall and D. Wiliam (2003). <u>Assessment for learning: Putting it into practice</u>. Maidenhead, UK, Open University Press.

Black, P. and D. Wiliam (1998). "Assessment and classroom learning." <u>Assessment in Education: Principles, Policy and Practice</u> **5**(1): 7-73.

Black, P. and D. Wiliam (1998). "Inside the black box: Raising standards through classroom assessment." <u>Phi Delta Kappan</u> **Phi Delta Kappan International (online)**(Oct 1998).

Black, P. and D. Wiliam (2006). Developing a theory of formative assessment. <u>Assessment and learning</u>. Thousand Oaks, Sage**:** 81-100.

Black, P. and D. Wiliam (2007). "Large-scale assessment systems: design principles drawn from international comparisons." <u>Measurement: Interdisciplinary Research and Perspectives</u> **5**(1).

Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. <u>Educational evaluation: new roles, new means: the 68th yearbook of the National Society for the Study of Education (part II)</u>. R. W. Taylor. Chicago, University of Chicago Press. **68:** 26-50.

Borko, H. (1997). "New forms of classroom assessment: Implications for staff development." <u>Theory into Practice</u> **36**(4 (Autumn 1997)): 231-238.

Borko, H. (2004). "Professional development and teacher learning: Mapping the terrain." <u>Educational Researcher</u> **33**(8): 3-15.

Borko, H., C. Mayfield, S. F. Marion, R. Flexer and K. Cumbo (1997). Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development, CSE Technical Report 423. Los Angeles, CRESST.

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

Brousseau, G. (1984). The crucial role of the didactcal contract in the analysis and construction of situations in teachng and learning mathematics. <u>Theoruy of mathematics education: ICME 5</u>

topic area and miniconference. I. H. G. Steiner. Bielefeld, Germany, Institut fur Didaktik der Mathematik der Universitat Bielefeld. **54:** 110-119.

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.

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. and D. Wiliam (2006/2007). "Improving the way we grade science." Educational Leadership **64**(4): 36-42.

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

Coburn, C. (2003). "Rethinking scale: moving beyond numbers to deep and lasting change." Educational Researcher **32**(Number 6): 3-12.

Cohen, D. K. and H. C. Hill (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).

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, Center for the Study of Teaching and Policy University of Washington.

Darling-Hammond, L., D. J. Holtzman, S. J. Gatlin and J. V. Heilig (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: 6-11.

Dweck, C. (2000). <u>Self-theories: Their role in motivation, personality and development</u>. Philadelphia, Psychology Press.

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). <u>School reform from the inside out: Policy, practice, and performance</u>. Cambridge, MA.

Elmore, R. F. (2002). Bridging the gap between standards and achievement: The imperative for professional development in education, Albert Shanker Institute.

Fennema, E. and M. L. Franke (1992). Teachers' knowledge and its impact. <u>Handbook of research on mathematics teaching and learning</u>. D. A. Grouws. New York, Macmillan Publishing Co.**:** 147-164.

Fullan, M. (1991). <u>The new meaning of educational change</u>. London, Cassell.

Fullan, M. (2001). <u>Leading in a culture of change</u>. San Francisco, Jossey-Bass.

Fullan, M., P. T. Hill and C. Crevola (2006). <u>Breakthrough</u>. Thousand Oaks, Corwin Press.

Garet, M. S., A. Porter, L. Desimone, B. Birman and K. S. Yoon (2001). "What makes professional development effective? Results from a national sample of teachers." <u>AERJ</u> **38**(4): 914-945.

Garmston, R. and B. Wellman (1999). <u>The adaptive school: a sourcebook for developing collaborative groups</u>. Norwood, MA, Christopher-Gordon Publishers.

Gitomer, D. H., A. S. Latham and R. Ziomek (1999). The academic quality of prospective teachers: The impact of admissions and licensure testing. Princeton, New Jersey, Educational Testing Service.

Hamre, B. K. and R. C. Pianta (2005). "Academic and social advantrages for at-risk students placed in high quality first grade classrooms." <u>Child Development</u> **76**(5): 949-967.

Hanushek, E., J. F. Kain, D. M. O'Brien and S. G. Rivken (2005). The market for teacher quality, NBER working paper 11154. Washington, DC, National Bureau of Economic Research.

Hanushek, E. A. (2004). Some simple analytics of school quality. Washington, DC, National Bureau of Economic Research.

Hayes, V. P. (2003). Using pupil self-evaluation within the formative assessment paradigm as a pedagogical tool, Unpublished EdD Dissertation, University of London.

Hendry, C. (1996). "Understanding and creating whole organizational change through learning theory." Human Relations **49**(5): 621.

Hill, H. C., B. Rowan and D. L. Ball (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., M. Meiers and A. Beavis (2005). "Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes & efficacy." Education Policy Analysis Archives **13**(10).

Jepsen, C. and S. Rivkin (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. and M. L. Franke (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. German-American Science and Math Education Research Conference, Kiel, Germany.

Kluger, A. N. and A. DeNisi (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.

Lave, J. and E. Wenger (1991). Situated learning: Legitimate peripheral participation. New York, Cambridge University Press.

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

Lewis, C., R. Perry and A. Murata (2006). "How should research contribute to instructional improvement? The case of lesson study." Educational Researcher **35**(3): 3-14.

Librera, W. L. (2004). Ensuring quality teaching and learning for New Jersey's students and educators.

Lillard, A. and N. Else-Quest (2006). "Evaluating Montessori education." Science **313**(5795): 1893-1894.

Lyon, C., C. Wylie and L. Goe (2006). <u>Changing teachers, changing schools</u>. Annual Meeting of the American Educational Research Associaton, San Francisco.

Ma, L. (1999). <u>Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States</u>. Mahwah, NJ, Erlbaum.

McLaughlin, M. and J. Talbert (1993). Contexts that matter for teaching and learning: Strategic opportunities for meeting the nation's educational goals. Palo Alto, CA, Stanfoird University Center for Research on the Context of Secondary School Teaching.

McLaughlin, M. and J. E. Talbert (2006). <u>Building school-based teacher learning communities: Professional strategies to improve student achievement</u>. 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, 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, National Commission on Mathematics and Science Teaching for the 21st Century.

Natriello, G. (1987). "The impact of evaluation processes on students." <u>Educational Psychologist</u> **22**(2): 155-175.

Nonaka, I. and H. Takeuchi (1995). <u>The knowledge-creating company: how Japanese companies create the dynamics of innovation.</u> New York, New York, NY: 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, NCDPI.

NSDC, N. S. D. C. (2001). Standards for staff development, NSDC, National Staff Development Council.

Perrenoud, P. (1998). "From formative evaluation to a controlled regulation of learning: Towards a wider conceptual field." <u>Assessment in Education: Principles, Policy and Practice</u> **5**(1): 85-102.

Putnam, R. T. and H. Borko (2000). "What do new views of knowledge and thinking have to say about reserach on teacher learning?" <u>Educational Researcher</u> **29**(1): 4-15.

Ramaprasad, A. (1983). "On the definition of feedback." <u>Behavioral Science</u> **28**(1): 4-13.

Reeves, J., J. McCall and B. MacGilchrist (2001). Change leadership: Planning, conceptualization, and and perception. <u>Improving school effectiveness</u>. J. Macbeath and P. Mortimer. Buckingham, UK, Open University Press**: 122-137.**

Rodgers, C. (2002). "Defining reflection: Another look at John Dewey and reflective thinking." <u>Teachers College Record</u> **104**(no 4): 842-866.

Rodriguez, M. C. (2004). "The role of classroom assessment in student performance on TIMSS." <u>Applied measurement in education</u> **17**(1): 1-24.

Ross, P. E. (2006). "The expert mind." <u>Scientific American</u> **295**(no 2): 64-71.

Sandoval, W., V. Deneroff and M. L. Franke (2002). <u>Teaching, as learning, as inquiry: moving beyonfd activityu in the analysis of teaching practice</u>. American Educational reserach Association, New Orleans.

Schein, E. H. (1996). "Culture: The missing concept in organization studies." <u>Administrative Science Quarterly</u> **41**(2).

Scriven, M. (1967). The methodology of evaluation. <u>Perspectives of curriculum evaluation</u>. R. W. Tyler, R. M. Gagne and M. Scriven. Chicago, Rand McNally. **1:** 39-83.

Slavin, R. E. (1995). <u>Cooperative learning: Theory, research and practice</u>. Boston, Allyn & Bacon.

Thompson, M. and L. Goe (2006). <u>Models for Effective and Scalable Teacher Professional Development</u>. Annual Meeting of the American Educational Research Association, San Francisco.

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.

Wiliam, D. (2003). The impact of educational research on mathematics education. <u>Second International Handbook of Mathematics Education</u>. A. Bishop, M. A. Clements, C. Keitel, J. Kilpatrick and F. K. S. Leung. Dordrecht, Netherlands, Kluwer Academic Publishers**: 469-488.**

Wiliam, D. (forthcoming in 2007). Keeping learning on track: Classroom assessment and the regulation of learning. <u>Second Handbook of Research on Mathematics Teaching and Learning, a project of the National Council of Teachers of Mathematics</u>. F. K. Lester. Greenwich, CT, Information Age Publishing.

Wiliam, D., C. Lee, C. Harrison and P. Black (2004). "Teachers developing assessment for learning: Impact on student achievement." Assessment in Education:Principles, Policy, and Practice.

Wiliam, D. and M. Thompson (2006). Integrating assessment with learning: What will it take to make it work? The future of assessment: Shaping teaching and learning. C. A. Dwyer. Mahwah, NJ, Lawrence Erlbaum Associates.

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

Wright, S. P., S. P. Horn and W. L. Sanders (1997). "Teacher and classroom context effects on student achievement: Implications for teacher evaluation." Journal of Personnel Evaluation in Education **11**: 57-67.

Wylie, C., M. Thompson, C. Lyon and D. Snodgrass (2007). Keeping learning on track in an urban district's low performing schools. American Educational Research Association. Chicago, IL.