

Learning Sciences
Dylan **Wiliam** Center

Sustaining Formative Assessment with Teacher Learning Communities

Dylan Wiliam and Siobhán Leahy

Introduction

In Dylan Wiliam's [*Embedded Formative Assessment*](#) (Solution Tree, 2011) and in our forthcoming *Handbook for Embedded Formative Assessment* (LSI, 2015), we discuss a number of techniques that teachers can use to increase student engagement, to collect evidence about student achievement, and to make their teaching more responsive to the needs of their students. Most of these ideas have been around for many years, and are effective in raising achievement. Yet they are not in widespread use. This suggests that actually implementing these techniques is more difficult than it might appear, and therefore we need to think carefully about how to support teachers in developing their use of formative classroom assessment. That is the focus of this paper.

The 5 Strategies of Embedded Classroom Formative Assessment

1. Clarifying, Sharing, and Understanding Learning Intentions and Success Criteria
2. Eliciting Evidence of Learners' Achievement
3. Providing Feedback that Moves Learning Forward
4. Activating Students as instructional Resources for One Another
5. Activating Students as Owners of their Own Learning Sciences International

From Embedded Formative Assessment (2011)

Content, then process

In "Keeping Learning on Track: Classroom Assessment and the Regulation of Learning" (2007), Dylan Wiliam pointed out that while investment in teacher professional development has been a feature of the educational landscape for many years, evidence that it has made any difference to student achievement has been depressingly thin. Michael Fullan put it like this: "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).

In *Embedded Formative Assessment*, we saw that part of the reason is that much teacher professional development has been focused on what is easy to deliver, rather than what makes a difference to student outcomes. Policymakers have behaved rather like desperate hitters who come up to the plate and try to hit a home run off every pitch. The result is, of course, a lot of strike-outs. What we need instead is "small ball:" get a walk to first, steal second, get bunted over to third, and score on a sacrifice fly.

But in professional development, like in small ball, the details matter. Many people have advocated professional learning communities (PLCs) as the answer, but this is to put process before content. PLCs are good ways to achieve some goals, but much less effective for others. For example, if we wanted to increase teacher subject knowledge, then professional learning communities would be a rather inefficient way to do that. Some

form of direct instruction, by experts, would be more effective. That is why process should come *after* content. First we should decide what kinds of changes in teaching will make the largest impact on student outcomes, and then—and only then—we should work out the best way to secure these changes.

“We should determine the content of teacher learning first, and the process should then be chosen to meet that end.”

In other words, we should determine the content of teacher learning first, and the process should then be chosen to meet that end. If we start with process, we are, in

effect, saying something like, “Professional learning communities (or coaching, or whatever else is the flavor of the month) are the answer. Now what was the question again?” Or, as Abraham Maslow put it: “He who is good with a hammer tends to think that everything is a nail.”

I call this fundamental principle of teacher professional development “Content, then process.” To be successful, teacher professional development needs to concentrate on both content *and* process (Reeves, McCall & MacGilchrist, 2001; Wilson & Berne, 1999), but the content must come first. In other words, we need to focus on *what* we want teachers to change, or change about what they do, and then we have to understand *how* to support teachers in making those changes

The content element was the focus of *Embedded Formative Assessment*, and it has two components. First, *evidence*: the research evidence suggests that classroom formative assessment can have significant impact on how much students learn. Indeed, the evidence suggests that attention to classroom formative assessment can produce greater gains in achievement than any other change in what teachers do. Second, *ideas*: I have shared a number of practical techniques that teachers can use to begin to develop their practice of formative assessment.

However, knowing what will help teachers most is only part of the solution, and this is where the quotation from Michael Fullan above is so important. The professional development that he described was not misguided in its aim. Much of the content of that professional development was entirely appropriate. What was not given enough thought was how to support teachers in making changes to their practice when they returned to their classrooms—the *process* of teacher change.

As a result of extensive work with teachers and administrators over the last ten years, trying to promote and support the development of classroom formative assessment, I have concluded that there are five key process components: choice, flexibility, small steps, accountability and support (Wiliam, 2006). Each of these is discussed in turn below.

Choice

In the late 1960s, Meredith Belbin, a researcher at the University of Cambridge’s Industrial Training Research Unit, began studying the way that people interacted while playing business games at the Administrative Staff College at Henley-on-Thames in England (which later became Henley Management College). He observed that when teams were effective, members of the team were able to cover eight team roles:

- Company worker
- Chairman
- Shaper
- Plant
- Resource investigator
- Monitor-evaluator
- Team worker
- Completer-finisher

Belbin observed that individuals had strong preferences for the roles they played in groups. While each individual might have two or three roles they were happy to play, when the team needed individuals to play other roles, individuals were rarely able to sustain focus in these “non-preferred” roles for long. The most effective teams were therefore those where the preferences of the members covered the eight team roles (later on Belbin added a ninth role: specialist).

The other important point that Belbin made was that each role had key strengths *and* allowable weaknesses. In other words, when an individual had strengths, they would also have weaknesses that were, in effect, the other side of the same coin. Someone who has lots of imaginative ideas may not be very good at thinking through the practical implications of the ideas, and someone who is very detail-oriented may have a tendency to worry about small, and possibly irrelevant details. A summary of strengths and weaknesses of the eight roles is shown in table 8.1.

	Principal strengths	Allowable weaknesses
Company worker	Disciplined, hard-working	Lack of flexibility
Chairman	Valuing contributions	Not particularly creative
Shaper	Drive	Impatience
Plant	Thinking “outside the box”	Impractical
Resource investigator	Openness to new ideas	Short attention-span
Monitor-evaluator	Hard-headed	Poor motivator
Team worker	Responsive to others	Not good in crises
Completer finisher	Detail-oriented	Obsessive

Table 8.1: Key strengths and weakness of Belbin’s eight team roles

Although the significance of Belbin’s work was not realized for many years, it has recently generated a great deal of interest in what is called “strengths-based capability development.” The main point here is that in the past, much professional development

has focused on people's weaknesses. Indeed, it is very common to find reports from performance evaluations divided into two lists, headed "strengths" and "areas for development." Belbin's work highlights the fact that, for some individuals, their weaknesses might be so egregious as to require immediate attention, but for most, the organization would benefit more from helping those individuals become truly outstanding at the things they are already good at.

In a similar way, it is often assumed that to improve, teachers should work to develop the weakest aspects of their practice, and for some teachers, these aspects may indeed be so weak that they should be the priority for professional development. But for most teachers, the greatest benefits to students are likely to come from teachers becoming even more expert in their strengths. In early work on formative assessment with teachers in England (Black, Harrison, Lee, Marshall & Wiliam, 2003), one of the teachers, Derek (this, like the names of all teachers, schools, and districts mentioned in this paper, is a pseudonym) was already quite skilled at conducting whole-class discussion sessions, but he was interested in improving this practice further. A colleague of his at the same school, Philip, had a much more "low-key" presence in the classroom, and was more interested in helping students develop skills of self-assessment and peer-assessment. Both Derek and Philip are now extraordinarily skilled practitioners—among the best I have ever seen—but to make Philip work on questioning, or to make Derek work on peer-assessment and self-assessment would be unlikely to benefit their students as much as supporting each teacher to become excellent in his own way.

Furthermore, when teachers themselves make the decision about what it is that they wish to prioritize for their own professional development, they are more likely to "make it work". In traditional "top-down" models of teacher professional development, teachers are given ideas to try out in their own classrooms, but often respond by blaming the professional developer for the failure of new methods in the classroom (e.g., "I tried what you told me to do, and it didn't work"). However, when the choice about the aspects of practice to develop is made by the teacher, then the responsibility for ensuring effective implementation is shared.

Viewed from this perspective, choice is not a luxury but a necessity. If teachers are to develop their practice in the way that will make most difference to their students, they need choice, because for the vast majority of teachers, only they know what aspects of their practice will be most productive to develop. Of course, this choice must be exercised within a framework that provides some assurance that the changes will be beneficial, which is why the five strategies of formative assessment discussed in *Embedded Formative Assessment* are so important. However, within this framework, the choice of what aspects of their practice to prioritize for development must be left to the practitioners, who know more about their own classrooms and their own students than anyone else.

Flexibility

As well as choice of what to prioritize in their development, teachers will also need to modify ideas developed by other teachers to make them work in their own classrooms. Part of the reason for this is differences between teachers in their style of teaching, but it is also important to recognize that there are differences from school to school and class to class. What works in one context may not work in another because schools differ in their openness to experimentation and their appetite for risk. The expectations of the students are also important. Where students are insecure they may be uncomfortable with changes that will be beneficial but that make new and unfamiliar demands of them (Spradbery, 1976). As a result, teachers may need to modify the way that techniques are introduced. The expectations of parents can also constrain what is possible, at least in the short term.

Teachers therefore need the flexibility to be able to “morph” the classroom formative assessment techniques with which they are presented to fit their own classroom context (Ginsburg, 2001). The danger in this is that a teacher may so modify an idea so much that it is no longer effective—what Ed Haertel has described as a “lethal mutation” (Brown & Campione, 1996).

For example, collaborative learning can produce significant increases in student achievement, but only where it is implemented in a way that emphasizes group goals *and* individual accountability. Many teachers are reluctant to ensure that students are individually accountable for contributing to the work of the group, and they do not implement collaborative learning in a way that increases student achievement. What is needed, therefore, is a way of allowing teachers flexibility, while at the same time constraining the use of that flexibility so that modifications to the original ideas do not unduly weaken their effectiveness.

It is important to make the distinction between the *strategies* of classroom formative assessment on the one hand, and the *techniques* that can be used to enact these strategies in classrooms on the other. The five key strategies of formative assessment are always smart things for a teacher to be doing. However, the actual techniques teachers use to enact these in their practice requires careful thought. Some techniques work better in some school subjects than others, and some work better with some students than others. It is also important to note that the teacher’s belief in the value of a technique is important.

For example, many teachers are comfortable using popsicle sticks with fifth-graders but think they are a little too childish for eighth-graders. Having said that, teachers of Advanced Placement courses in high schools have found popsicle sticks a useful strategy for random questioning, because when 18-year-olds are asked to answer a question, their first reaction is “why me?.” With the popsicle sticks, the answer is, “It’s your unlucky day. Deal with it. Now what’s the answer to the question?”

Teachers should be free to adopt whichever of these techniques they wish, thus providing choice—the first of the five process requirements. By anchoring the techniques to (at least) one of the five key strategies, we provide a means by which teachers can modify the techniques, but still provide a reasonable assurance of fidelity to the original research, and therefore a reasonable guarantee that the techniques will be effective in increasing student achievement.

This represents a radically different model of the “dissemination” of educational research. In the standard model, researchers research, find out things and then tell teachers what to do. In other words, the researcher gains insights into classrooms and learning, and attempts to communicate those insights to teachers as knowledge. For example, giving task-involving feedback is better than giving ego-involving feedback. According to the traditional model of dissemination (see, for example, English, Jones, Lesh, Tirosh, & Bussi, 2002, p. 805), all we need to do then is to tell teachers to give more task-involving feedback and less ego-involving feedback. Of course, it’s not as simple as that, because whether feedback is ego-involving or task involving depends on the context in which it is given, the attitudes of the recipient, and the relationship between the donor and the recipient of the feedback.

“One of the main reasons that educational research has had so little impact on educational practice is because the very hardest task of all—working out how to implement research findings in real contexts—has been left almost entirely to teachers, and this is both unfair and foolish.”

One of the main reasons that educational research has had so little impact on educational practice is because the very hardest task of all—working out how to implement research findings in real contexts—has been left almost entirely to teachers, and this is

both unfair and foolish. But I believe that by starting with the research literature, it is possible to build up a theory of formative assessment that can then be manifested in a set of classroom techniques. Because they are derived from research evidence, these techniques form a set of “validated practices” that may not work in every setting, but *are* likely to be effective in *most* contexts, because of the framing provided by the five strategies and the research evidence on which they are based. Teachers then take these techniques, and through regular use, incorporate them into their own theories of classroom practice, and thus change their own ideas about practice.

Years ago, the late Millard Fuller, founder of “Habitat for Humanity” pointed out that in matters of environmental awareness, it is generally easier to get people to *act* their way into a new way of *thinking* than it is to get them to *think* their way into a new way of *acting*. The two approaches to research and practice described above can be thought of in the same way. Traditional approaches to the dissemination of educational research are, in effect, trying to get teachers to think their way into a new way of acting. In contrast, encouraging teachers to adopt new practices that they then incorporate into their routine teaching is a way of getting them to act their way into a new way of thinking.

This is an important insight because changing what teachers think, if it does not change what they do, will not benefit students. Students benefit only when teachers change what they do in classrooms. It is as if we have been training quarterbacks by getting them to read books and watch film. Such activities may have some value, but ultimately, they have to be able to put these ideas into practice, and that is the hardest part. Indeed, I do not think it is too much of an overstatement to say that the reason for the relative ineffectiveness of most professional development over the last quarter century is because it has been based on an incorrect diagnosis. We have assumed that the “problem” is that teachers don’t know enough, and so we have herded them into rooms to hear experts supply the missing knowledge, and then have assumed that, with their deficits in knowledge rectified, teachers will be more effective.

It hasn’t worked because the diagnosis is wrong. Teachers don’t lack knowledge. What they lack is support in working out how to integrate these ideas into their daily practice, and this takes time, which is why we have to allow teachers to take small steps.

Small steps

It is extremely important for schools to improve—and quickly. For that reason, it is hardly surprising that policymakers, politicians, and administrators want to get teachers developing their classroom formative assessment practices as quickly as possible. However, the research evidence shows that teachers are slow to change their classroom practices. Indeed, many people have gone as far as to claim that teachers are resistant to change—that teachers cling to a set of professional habits, which to a very real extent, represent a core part of each teacher’s professional identity, which is why they are unwilling to change. If we want to support teachers in developing their practice, it is important for us to understand why changes in practice are so slow. Is it just resistance to change, or something deeper?

It is important to understand the reasons because what we do to support change will depend on the reasons for the slowness of previous attempts. If teachers are resistant to change because they are happy with the way things are, and don’t want to change, then we could, for example, look for ways of overcoming resistance. This is often the rationale for incentive schemes—the idea is that teachers will adopt new ideas if they are paid to do so. However, the evidence from careful evaluations of incentive schemes show that they are not particularly successful (see, for example Springer, Ballou, Hamilton, Le, Lockwood, McCaffrey, Pepper & Stecher, 2010).

A far more likely reason for the slowness of teacher change is that it is genuinely difficult, because high-level performance in a domain as complex as teaching requires automatizing a large proportion of the things that teachers do. For learner drivers, shifting gears, using the turn indicator, checking the rear-view mirror, and steering all at the same time seems impossibly complicated—and undertaken consciously, they are. Experienced drivers have practiced these activities so many times that they become automated and thus take up

little of the available resources for cognitive processing. However, as anyone who tries to change the way they drive—for example to reduce the extent to which they “ride the clutch”—has discovered, these automated procedures are extremely hard to change.

In fact, as David Berliner has shown, expert performance in teaching shares many of the properties of expertise in other areas (Berliner, 1994). After reviewing research on the nature of expertise in teaching, he offered eight propositions on the nature of expertise:

Experts excel mainly in their own domain: A teacher might be very skilled at teaching seventh grade math, but much less effective at teaching math to seventh-graders with special educational needs. Expertise tends to be very specific.

Experts often develop automaticity for the repetitive operations that are needed to accomplish their goals: In my teaching on pre-service teacher education programs, I found that novice teachers take around four hours to plan an hour’s instruction, while expert teachers plan lessons of higher quality in five minutes or less. In other words, planning a lesson is something an expert does as much as 50 times faster than a novice. The more often we do things, the better and faster we get at them.

Experts are more sensitive to the task demands and social situation when solving problems: When asked to plan a sequence of instruction, expert teachers tend to want more details about the physical situation in which they will be teaching, and want details of the experience, ability, and backgrounds of the students they will be teaching. Also, expert teachers will often use humour at the beginning of an instructional sequence, but then become more serious, while novices are likely to project the same emotional tone throughout.

Experts are more opportunistic and flexible in their teaching than novices: Novices tend to try to use “all purpose” solutions, whereas experts are more likely to use a solution that is relevant only to the specific problem at hand. Experts are able to “steer” responses made by students towards the lesson objectives, whereas novices either find it difficult to deviate from their planned course, or, when they do, tend to get side-tracked.

Experts represent problems in qualitatively different ways than novices: Many studies have found that the difference between experts and novices is not so much the amount of knowledge (although this is important) but how that knowledge is organized. Novices tend to organize their knowledge according to surface features, whereas experts are more likely to use deep structural features. For example, when asked to describe a student, an expert teacher is more likely to mention aspects related to learning needs, whereas a novice teacher is more likely to mention aspects of personality that are not directly relevant to instructional planning.

Experts have fast and accurate pattern recognition capabilities. Novices cannot always make sense of what they experience. For example, an experienced teacher can walk

into a classroom and see that a particular arrangement of tables will not work for what they have planned, even though they may not be able to say why.

Experts perceive meaningful patterns in the domain in which they are experienced.

When shown a still photograph of a classroom, and asked what is going on, novice teachers are likely to respond with descriptions (e.g., “Four students are sitting round a table”) whereas experts are more likely to make inferences about what is going on (e.g., “It looks like some kind of group task in a science lab”).

Experts begin to solve problems slower but bring richer and more personal sources of information to bear on the problem that they are trying to solve. When given details of a student, and asked to plan how to help the student progress, novices are likely to begin planning quite rapidly, whereas experts tend to take much longer (in one study cited by Berliner, novices began planning after three minutes, while experts took 10 minutes to begin). Part of the reason for experts’ longer planning appears to be because they search through their previous experiences to find similar situations they have met before (which of course the novices don’t have), but are also careful to explore whether the current situation is sufficiently similar to the previous case for the previous case to be a good guide to what to do.

It is only by observing novice teachers that one realizes how expert the best teachers are. Expert teachers, with a few commands and a wave of the arms, get students into a classroom, seated, and at work in a matter of seconds, whereas with novice teachers, the same process can have the appearance of herding cats.

These polished routines are the result of hundreds, thousands, and sometimes even hundreds of thousands of repetitions. They are what get teachers through the day—without them, their jobs would be impossible. But the “automaticity for repetitive operations that are needed to accomplish their goals” that Berliner mentions also means that once established, these routines are hard to change.

A few years ago, I was working with teachers in HoHoKus, a K-8 school district in northern New Jersey. One sixth-grade teacher was trying to use “No hands up except to ask a question,” but she found this very difficult because every time she asked a question, she would begin the question by asking, “Does anyone...?” or “Has anyone...?” She asked me, “Why am I finding this so hard?” We sat down and worked out that in the 22 years she had been teaching, she had probably asked half a million questions in her classroom. When you’ve done something one way half-a-million times, doing it any other way is going to be very difficult.

Teaching is even more extreme than driving a car in this respect, because teachers come to the profession with a series of “scripts” of how classrooms should operate “hard-wired” into their minds from their time as students. These scripts, such as requiring students to raise their hands if they have an answer to a teacher’s question, seem natural, but of course they are learned, and they get in the way of learning (Wiliam, 2005).

Moreover, many of the changes in practice associated with implementing formative assessment are not just difficult to change because they are habituated—they also contradict widely distributed and strongly held beliefs about, for example, the value of grades for motivating students. Even when teachers are convinced of the value of approaches such as “comment-only grading” they are often dissuaded from trying them out by more senior colleagues who dismiss innovations as fads advocated by ivory tower academics who don’t know what real teaching is.

The consequence of all this is that in implementing any professional development model, we have to accept that teacher learning is slow. In particular, for changes in practice—as opposed to knowledge—to be lasting, they must be integrated into a teacher’s existing routines, and this takes time. Many people involved in professional development are familiar with the experience of encouraging teachers to try out new ideas, and seeing them being enacted when they visit teachers’ classrooms, only to learn shortly afterwards that the teachers have reverted to their former practices.

We have to allow anyone who is changing something at which they are already expert to take “small steps.” The collection of routines that teachers establish to get through the day are their greatest asset, but at the same time a liability, because getting better involves getting a little bit worse, at least for a while. When Tiger Woods wanted to change his golf swing after winning the 1997 Masters golf tournament, he withdrew from competition, and practised his new swing until he was happy with it. But no teacher is going to get three months to practise new techniques in front of a mirror before trying them out in the classroom, which is why changing one’s teaching is such a daunting prospect. It involves disrupting the things that get you through the day. No wonder one teacher described changing her classroom routines as “engine repair, in flight.”

A second reason that teachers need to take small steps is that expertise cannot be reduced to words, so that teachers need to “feel” their way when trying out new ideas. The fact that expertise is more than just following a set of rules was neatly illustrated in a study carried out by Klein and Klein (1981). The researchers prepared six video clips of someone performing cardio-pulmonary resuscitation (CPR). Five of the video clips were of students who were just learning CPR and one was of an experienced paramedic. The six clips were shown to a number of experienced paramedics, students learning CPR, and CPR instructors, and each participant was asked, “Who would you want doing CPR on you if you needed it?” The experts chose the expert 90% of the time. The students chose the expert 50% of the time. The instructors—the people who teach other people how to do CPR—chose the expert only 30% of the time. The researchers concluded that the instructors were looking for people who enacted the “rules” of CPR that they themselves taught, and often, the expert paramedics did not do this. However, it is important to note that experts could still identify experts. It was not as if each expert had a completely idiosyncratic way of doing CPR. Rather it appears that expertise in CPR was not adequately captured by the “rules” of CPR taught by instructors.

What this suggests is that expertise cannot always be reduced to words. In many situations, we know more than we can tell (Polanyi, 1966 p. 4). And yet, for as long as professional development has been recognized as essential for teachers, the endeavour of helping teachers improve has been treated as one of the provisions of knowledge. Most professional development has been based on the assumption that teachers lack some important knowledge, and experts are brought in to remediate this apparent weakness, presumably on the assumption that once these gaps in knowledge have been put right, the teachers will be able to return to their classrooms and teach better. Much professional development has been ineffective because the diagnosis is faulty. Teachers don't tend to lack knowledge so much as the ability to translate their intentions into action, and here, we have much to learn from WeightWatchers®.

WeightWatchers ought to be the least successful organization on the planet, because everyone who wants to lose weight knows what to do: eat less and exercise more. That's it. There is no secret third rule of weight loss that is revealed upon payment of a subscription to WeightWatchers. For example, perhaps if I stirred my breakfast cereal anti-clockwise, it might turn the dextrose in the sugar into left-handed sugar molecules that would not be absorbed by my body? Unfortunately not. It's just eat less, and exercise more. But WeightWatchers have realized that they are not in the knowledge-giving business; they are in the habit-changing business. People who want to lose weight know what they need to do. What they need are structures and supports that help them do what they want to do.

In the same way, if we are serious about helping teachers improve their practice, we have to help teachers change their classroom habits. And just in the same way that weight-loss programs need to be focused on weight loss that is sustained, we also need to support teachers to make changes in their practice that lasts.

I have lost count of the number of times that I have described a teaching technique to a group of teachers only for one of them to say to me, "I used to do that. It was good." Teachers are bombarded with new ideas, which they are encouraged to try out, but before they have time to consolidate these new ideas into their practice, some newer idea comes along. This is why the idea of "sharing good practice" can be so dangerous as a model for teacher professional development. Of course teachers need new ideas, but most teachers already have more good ideas than they will ever be able to incorporate into their teaching. What teachers need is help creating new habits, and this is a challenge in all areas of human endeavor—even medicine.

In the 19th century, many women died shortly after childbirth—the disease was named "puerperal fever" which was more a label than a diagnosis, since literally it just means a fever of a woman in childbirth. Ignaz Semmelweis, a doctor at the Vienna General Hospital, noticed that the prevalence of this disease had increased sharply after the introduction of "pathological anatomy" in which doctors carried out post-mortem examinations to investigate the cause of death, while waiting for women in labor to give birth. However, no such increase had been seen where the mothers were attended by

midwives (who were of course not involved in post-mortems). His suggestion that doctors wash their hands in a weak solution of calcium hypochlorite (common household bleach) was widely reviled at the time, even though the introduction of this procedure cut the mortality rate from puerperal fever at Vienna General Hospital from an average of around 15% to less than 5% (Semmelweis 1861/1983). A century and a half on, it is now accepted that good hand hygiene is essential to avoiding cross-infection in hospitals. Except, a century and a half later, we don't seem to have made much progress.

The surgeon Atul Gawande (2007) has pointed out we have made far greater progress in advanced techniques of surgery than in apparently “easier” areas such as basic hygiene. He quotes the example of washing hands after contact with patients, where the compliance rate in many hospitals is below 50% even though it is widely accepted that much higher rates are required to control the spread of resistant forms of staphylococcus aureus such as vancomycin-resistant Staphylococcus aureus (VRSA) and methicillin-resistant Staphylococcus aureus (MRSA) (p. 15).

In fact, a review of hand-hygiene studies by Pittet (2001) reviewed studies of hand hygiene and found that compliance rates of below 50% was the norm. The problem is not ignorance, nor willful disobedience. In most hospitals, a strict procedure is specified for hand washing, but, Gawande points out,

Almost no one adheres to this procedure. It seems impossible. On morning rounds, our residents check in on twenty patients in an hour. The nurses in our intensive care units typically have a similar number of contacts with patients requiring hand washing in between. Even if you get the whole cleansing process down to a minute per patient, that's still a third of staff time spent just washing hands. Such frequent hand washing can also irritate the skin, which can produce dermatitis, which itself increases bacterial counts. (Gawande, 2007 p. 18)

If we are going to help teachers change their classroom habits, we need to recognize that this is going to be hugely challenging, and is going to require both support and accountability, which are the subjects of the next two sections.

Accountability

All teachers need to improve their practice; not because they are not good enough, but because they can be better. For that reason, I think it is entirely appropriate for teachers to be held accountable for making improvements in their practice. In developing their practice, teachers should develop those aspects of their practice that is likely to be of most benefit to their students; in other words, they should be accountable to the evidence about what is likely to benefit students. This is not meant to imply a slavish following of the latest research findings, but that teachers should be literally accountable—they should accept that they should expect, and be able, to render an

account of why they have chosen to develop one aspect of their practice rather than another.

In our work with teachers, we have found it helpful to engage them in detailed planning of what changes they plan to make in their teaching. This process could be called “action planning” but it is important to note that our experience is that this is best done with a highly structured approach—very different from the tokenistic “action planning” that occurs at the end of many teacher professional development events.

Of course there are many different protocols that might be adopted for action planning, but our experience of working with teachers developing their practice of formative assessment suggests that the following features are particularly important:

1. *The action plan should identify a small number of changes that the teacher will make in their teaching:* When teachers try to change more than two or three things in their practice at the same time, the result is often that their classroom routines deteriorate significantly, and they then fall back on those routines with which they feel comfortable or “safe.” Like the story of the tortoise and the hare, teachers who try to change too many things at the same time end up making less progress than those who make small, gradual, manageable changes.
2. *The plan should be written down:* Writing the plan down makes it more likely that the teacher thinks the plan through while writing it down, makes the ideas more concrete, and also creates a record that means that teachers are less likely to forget what they planned to do.
3. *The plan should focus on the five key strategies of formative assessment:* As noted above, teachers should prioritize changes that are likely to benefit students, and although there are other changes that might benefit students, the robustness and the coherence of the research on the effects of formative assessment suggests that this should be the starting point for all teachers.
4. *The plan should identify what the teacher plans to reduce, or give up doing to make time for the changes:* Most teachers are working as hard as they can, so if these changes are treated as an addition to their load, they are unlikely ever to be implemented. To make time for these changes, the action plan must identify something that the teachers are currently doing that they will stop doing, or do less of, to make time available for the changes. Asked to make such clear priorities, people often hope that they can make the necessary changes by being more efficient in their use of time, but this is usually hopelessly optimistic. The only way to make time available for new things is to reduce, or stop doing entirely, things that they are currently doing, in order to create time for innovation.

Support

The last process element—support—is closely related to accountability. Indeed, some authors have described them as a single feature of effective learning environments for

teachers: supportive accountability (Ciofalo & Leahy, 2006). The central idea is the creation of structures that, while making teachers accountable for developing their practice, also provide the support for them to do this. Support and accountability can therefore be thought of as two sides of the same coin.

The closeness of the relationship between support and accountability can be seen when we consider the process of “action planning” which can be thought of both as providing a measure of accountability, and as providing support for the teacher to change their practice. Some teachers find it demeaning, or even insulting, to be asked to put down in writing what they are going to try out in their classroom, but in looking back at the process, teachers have said time and again how useful the idea of committing their plans to paper was, both in terms of making their plans concrete, and in terms of making a commitment to their peers. Tim, a math teacher at Spruce Central High School, was reflecting, after participating in teacher learning communities for a year, on the experience of making a written commitment to the rest of group towards the end of each meeting:

I think specifically what was helpful was the ridiculous NCR [No Carbon Required] forms. I thought that was the dumbest thing, but I’m sitting with my friends and on the NCR form I write down what I am going to do next month. Well, it turns out to be a sort of “I’m telling my friends I’m going to do this” and I really actually did it and it was because of that. It was because I wrote it down. I was surprised at how strong an incentive that was to do actually do something different [...] that idea of writing down what you are going to do and then because when they come by the next month you better take out that piece of paper and say “Did I do that?” [...] Just the idea of sitting in a group, working out something, and making a commitment [...] I was impressed about how that actually made me do stuff. (Lyon, Wylie & Goe, 2006 p. 20)

Clearly, creating this ‘supportive accountability’ could be done in a number of ways. One way would be to assign each teacher a coach, but this would be expensive, and it is by no means clear that an adequate supply of appropriately skilled coaches would be available. For that reason, between 2003 and 2006 Siobhán Leahy and I working with colleagues at the Educational Testing Service, developed and piloted a number of models for supporting teachers (for extended accounts of these early developments, see Thompson & Goe, 2008; Wylie, Lyon & Goe, 2009; Wylie, Lyon, & Mavronikolas, 2008). What we learned is summarized in the following section.

Supporting formative assessment with teacher learning communities

One of our earliest models involved a facilitator meeting every two or three weeks with groups of four to six high school teachers to discuss the changes they were attempting to make in their practice. As a result of this work, it became clear that a two- or three-week cycle did not allow enough time for the teachers involved to plan and implement changes in their practice in time for reporting back at the next meeting. Those who have never

been teachers cannot understand this. Surely, they say, a teacher could try out a new idea in two or three weeks? But because they have never been teachers, they do not understand the complexity and fragility of a teacher's working life. Moreover, many of the techniques require careful planning, and don't work with all subject matter. If a teacher wants to try out getting the students to use "traffic lights" to undertake a self-assessment, then this needs to be done in a lesson where the learning intention and the success criteria can be expressed clearly in language that is accessible to students. If a teacher wants to use the ABCD cards for a hinge-point question, she needs to decide which lesson to try this out in, where in the lesson to check on the students' understanding, and may need to arrange time to meet with another teacher to discuss the design of the question. Four weeks appears to be a minimum period of time for teachers to plan and carry out a new idea in their classrooms.

On the other hand, implementations that involved meetings that occurred at intervals of six weeks or more appeared to lose momentum. This led to the adoption of a monthly cycle of meetings, and in trying out this model in literally hundreds of schools in dozens of districts over the last eight years, we have not come across any evidence that suggests that intervals between meetings of approximately four weeks is not an optimum, at least in respect of changes in practice related to formative assessment.

Originally, we had assumed that schools would be able to find two hours for each of the monthly meetings, and while this was clearly possible in some districts, in others it was not, so we looked carefully at ways of reducing the length of the monthly meeting. After experimentation with different lengths of meetings (including meetings as short as 60 minutes), we concluded that 75 minutes should be an absolute minimum. In some schools, because of the lack of a single 75-minute slot, teachers tried having two 40-minute slots, but this appeared to be much less successful than a single 75-minute slot.

Our experiences with meetings for a small number of participants had also led us to conclude that the minimum number of participants needed for a meeting to be effective was around eight. Meetings with fewer than eight participants often required significant input from the group's leader or facilitator, particularly when occasional absences due to illness and other factors reduced the size of the group further. While such intensive support from the facilitator might provide an effective learning environment for those attending, such a model would be unlikely to be scalable.

On the other hand, where the group was much larger than 12 (as was often the case in our early work in the Cleveland Municipal School District), there was not enough time to hear back from each member of the group. In interviews, many participants in teacher learning communities have told us that it was the fact that they knew that they would be required to give their colleagues an account of what they had been doing that made them prioritize working on changing their classroom practice over all the pressing concerns of a teacher's life (Ciofalo & Leahy, 2006). Given that on any day, one or two teachers are likely to be absent for any number of reasons, to ensure that a group has at least eight participants, a group size of 10 to 12 is recommended.

As well as design guidelines for the size of group and frequency of meetings, we also explored the extent to which it was necessary for teachers to share particular assignments (e.g., early grades or subject-specialisms in secondary schools). It has been our experience that teachers greatly value meeting in mixed-subject groups, in order to get ideas from teachers of different subjects or different ages. However, we had also observed many instances of a teacher rejecting suggestions from other members of the group with a claim that the suggested idea would not work for her or his own subject-specialism. In order to balance these tensions, we have explored models where the teachers do not all come from the same subject specialism, but, in order to provide some disciplinary support, we ensure that for each teacher in the group, there was at least one other with the same age- or subject-specialism. To date, we do not have any data that suggests that any particular method of constituting a group is better than another, although we are aware that the scope for deep conversations about subject matter are likely to be limited where the group is made up of individuals with different subject specialisms (Grossman, Wineburg & Woolworth, 2000). One model that some schools have adopted, which appears to work well, is to begin with mixed or hybrid models, and then, after a year or two, to move towards more specialist groupings, making it possible to “deepen the conversation.”

One final design feature of the monthly meetings of the teacher learning communities was related to their structure. We were aware that in most approaches to teacher professional development, novelty was often regarded as paramount, in order to keep things “fresh.” The disadvantage of such an approach is that participants arrive at the meeting not knowing the roles they are expected to play, so that the organization of the teachers’ learning, rather than the learning itself, was foregrounded.

On the other hand, we were aware from the work of Lee Shulman on professional learning in law, medicine and other professions that they adopt standard ways of organizing the learning of professionals, and these “signature pedagogies” are enduring and widespread.

Another interesting question about these pedagogies is: Why do they persist? Why do they last so long? I think the answer is that even though I can point out the flaws in every one of these signature pedagogies—and each is flawed—by and large, they work. They achieve the ends for which they were “designed”, or in a kind of Darwinian sense, they survive the competition with alternative pedagogical means. (Shulman, 2004 p. 15)

For this reason, we realized that there could be considerable benefits of adopting a standard structure for these monthly meetings. The fact that each meeting follows the same structure means that participants come to the meeting knowing the roles they are to play, both in terms of reporting back on their own experiences, and providing support to others. We explored a number of possible different models, but the model presented below has worked well in all the different settings in which it has been tried.

Introduction (5 minutes)

Agendas for the meeting are circulated and the learning intentions for the meeting are presented.

Starter activity (5 minutes)

Participants engage in an activity to help them focus on their own learning.

Feedback (25 minutes)

Each teacher gives a brief report on what they committed to try out during the “personal action planning” section at the previous meeting, while the rest of the group listen appreciatively and then offer support to the individual in taking their plan forward.

New learning about formative assessment (25 minutes)

In order to provide an element of novelty into each meeting of the TLC, and to provide a steady stream of new ideas, each meeting includes an activity that introduces some new ideas about formative assessment. This might be a task, a video to watch and discuss, or a ‘book study’ in which teachers will discuss a book chapter relevant to formative assessment that they have read over the past month.

Personal action planning (15 minutes)

The penultimate activity of each session involves each of the participants planning in detail what they hope to accomplish before the next meeting. This may include trying out new ideas, or it may simply be to consolidate techniques with which they have already experimented. This is also a good time for participants to plan any peer observations that they plan to undertake. It is our experience that if the participants leave the meeting without a definite date and time to observe one another, the peer observation is much less likely to take place (Maher & Wiliam, 2007).

Summary of learning (5 minutes)

In the last five minutes of the meeting, the group discusses whether they have achieved the learning intentions they set themselves at the beginning of the meeting. If they have not, there is time for the group to decide what to do about it.

These meetings provide both support and accountability. Many teachers have spoken about the usefulness of these meetings for providing advice about how they might move forward when they are “stuck,” but they also create a strong measure of accountability for teachers to actually implement their plans.

So much for the easy bit...

The research on formative assessment and the design principles for the teacher learning summarized above create a very clear vision of what should be happening in classrooms, and what kinds of teacher professional development might help move teachers towards

such practice. However, this clarity only takes one so far in the design of products that might be distributed at scale.

Perhaps the most interesting finding from this early phase of development was that many schools appropriated elements of the program to support their existing plans. Despite having attended presentations where the research basis for formative assessment was discussed at some length, and where it was shown that there was little or no research to support other innovations in learning that were attracting interest at the time (e.g., “brain gym” or learning styles), many schools reported that they liked the idea of teacher learning communities, but had decided to use them to support teachers in whatever was the school’s current priority for professional development (e.g., differentiated instruction, personalization, learning styles). Of course, this appropriation of resources is hardly surprising, impossible to prevent, and may be very positive in its outcomes, but what was surprising was that most of those who had transformed the innovation beyond recognition appeared to believe that they were implementing the materials *as intended*.

A case study in one district

Cannington is a local authority (school district) in Greater London, covering an area of approximately 10 square miles, and serving a diverse population of approximately 200,000 residents, with three times as many of its residents from minority ethnic communities as the country as a whole.

In July 2007, a philanthropic organization made available some funds for the establishment of a research project designed to raise student achievement in mathematics, science, and modern foreign languages—subjects that supervisory staff in Cannington had identified as priorities for development—although how this might be done was not identified at the time the philanthropic donation was made.

In November 2007, a presentation was made to a meeting of the principals of the secondary schools in Cannington, proposing the establishment of three teacher learning communities in each secondary school, one focusing on mathematics, one focusing on science and the third, on modern foreign languages, to provide support for teachers in their development of classroom formative assessment practices. Members of the project team attended meetings of the Cannington principals over the subsequent months to provide updates on progress, and in July 2008, a series of three one-day training events was held—one for each school subject—for teachers in the participating schools. The number of teachers from each school attending each of the events is shown in Table 1.

School	Number of teachers attending training event for:		
	Mathematics	Science	Modern Languages
Ashtree School	1	1	0
Cedar Lodge School	5	1	3
Hawthorne School	4	10	5
Hazeltree School	7	12	2
Larchtree School	1	0	0
Mallow School	6	7	3
Poplar School	11	3	1
Spruce School	7	8	5
Willowtree School	2	5	2
Totals	44	47	21

Table 1: Numbers of teachers attending each training event

The training day consisted of a brief summary of the research on formative assessment, an overview of the five “key strategies” of formative assessment, and an introduction to approximately 30 of the techniques that teachers might use to implement formative assessment in their classrooms. The day concluded with details on the creation of the three subject-specific school-based teacher learning communities that would be established to provide ongoing support in each school. The training session also provided guidance on the role of the leader of each of the three teacher learning communities to be established in each school.

The reactions of the teachers to the training was extremely positive, and at the end of the training day, the participants from six of the nine schools appeared to have a firm plan for implementation. One school (Ashtree) had decided to delay participation in the project for a year, and Hazeltree School had earlier that month decided to create mixed-subject, rather than subject-specific, teacher learning communities, as they felt that this was more in keeping with the professional development work that had already taken place at the school. However, since the funding had been provided specifically for supporting the teacher of mathematics, science, and modern foreign languages, it was agreed that this would in effect mean that Hazeltree would be withdrawing from the project, although they continued to receive all materials necessary for supporting teacher learning communities. Larchtree School had only sent a single teacher (to the mathematics session), but the teacher concerned appeared confident that she would be able to “cascade” the training to other teachers in the mathematics department, and possibly also to the other subjects.

While it was not possible for each teacher of mathematics, science, and modern foreign languages in each school to attend the one-day workshop, all teachers of these subjects in the secondary schools in Cannington were provided with a short (30-page) booklet outlining the major principles of formative assessment, together with specific details on how they could be applied in their subject (Hodgen & Wiliam, 2006; Black & Harrison, 2002; Jones & Wiliam, 2007) and a complete set of handouts for participants for each of the nine monthly meetings scheduled to take place over the coming year.

In order to provide a simple channel of communication between the teachers in the school and the project, six expert facilitators (two for each subject-specialism) were appointed. Their major role was not to “drive” the implementation of the program in the schools, but rather to respond to requests from TLC leaders and administrators within the school on the use of the materials. Each facilitator kept a log of their contacts with teachers at the school, which provided the main source of evidence on the extent to which the TLCs were functioning as intended.

Given the involvement of the principals of the school at each step of the process up to this point, and their investment in releasing significant numbers of teachers to attend the initial workshops, we expected that teacher learning communities would be established quickly; but for a variety of reasons, adoption was extremely patchy.

At the end of seven months, logs provided by the facilitators were coded by two different raters, with each rater asked to rate the progress made by each teacher learning community on a scale from 1 (little or no progress) to 4 (good progress). When the ratings generated independently were compared, in no case did the ratings differ by more than one point, and agreed ratings of two raters are shown in table 8.3.

School	Progress		
	Mathematics	Science	Modern Languages
Ashtree School*	-	-	-
Cedar Lodge School	1	1	2
Hawthorne School	2	2	4
Hazeltree School*	-	-	-
Larchtree School	4	1	1
Mallow School	3	1	2
Poplar School	1	3	3
Spruce School	4	3	3
Willowtree School	1	1	4
Average	2.3	1.8	2.8

*Did not participate in project (Ashtree deferred for a year, Hazeltree implemented a different model).

Table 8.3: Extent of progress of teacher learning communities in each school

Given the lack of correlation between the investment made by each school in sending teachers to the training and the progress made by the TLCs, we decided to investigate the reasons for the lack of progress.

Two features of this table are particularly worth noting. First the greatest progress appears to have been made in Modern Foreign Languages, which was the subject with the least-well attended initial session. Second, with the exception of Spruce School, there does not appear to be any tendency for the best-progressing TLCs to be in the same school.

In the schools that were participating in the project, only nine of the 21 planned TLCs were making reasonable progress (defined as a rating of 3 or 4 in Table 2). A more careful

analysis of the facilitator logs indicates that the major cause of the poor progress made by the 12 TLCs making slow or no progress (defined as a rating of 2 or 1 in Table 2) was that the teachers had not been given time to meet (this was also the case for some of the more successful TLCs, who had decided to commit their own personal time to these meetings because of their interest in, and commitment to, the project). Where time within their contracted hours had been made available for members of TLCs to meet, the TLCs were going well, with considerable enthusiasm for the meetings, and in particular, the teachers appeared to value the opportunity to talk about practice in a structured way.

The difficulty that TLCs had in finding time to meet is, at first sight, rather surprising. While none of the schools in Cannington is on the list of approximately 600 schools in England designated by the government as “National Challenge” schools for their failure to achieve the key benchmark of 30% of their students achieving proficiency (grade C or above) in five subjects in the national school-leaving examination for 16-year-olds, there is considerable pressure to improve results. Public transport links in Cannington are good, so students can easily travel from one side of the municipality to the other; thus, parents have a great deal of choice in secondary schools, and this choice is at least informed, if not driven, by examination results at age 16. All the principals say that improving academic outcomes, as measured by success on national examinations taken at ages 16 and 18 is one of their top priorities, and yet despite the research evidence suggesting that formative assessment could make more difference to these academic outcomes than anything else, it appears as if it was difficult for the principals and other school administrators to prioritize the development of formative assessment.

It is even more surprising when one considers that the principals had made a commitment to the program six months before its commencement, had been kept fully informed of the resource requirements necessary for the monthly meetings, and had made a considerable investment in the project by committing an average of 12 teacher days so that teachers could attend the introductory workshop.

When the principals of the participating schools were asked whether they were still committed to the project, they all said they were, and that in fact, the establishment of teacher learning communities remained a priority. But of course, this was not the only priority. Like every other school, the schools in Cannington have a number of very important initiatives in place. All of the initiatives are good ideas, but when everything is a priority, nothing is, and this illustrates perhaps the most important point about our efforts to improve schools—the fact that we have to be selective.

In my conversations with school leaders and other administrators, I have been struck by how often they see their task as ensuring that everything that goes on in their schools adds value to student learning. They search for things that are not adding value, or even detracting from student learning, so that they can eliminate them. It sounds like a recipe for school improvement, but it's not.

Almost without exception, people working in education want students to learn, and they spend their time on things that they think will help their students learn better. If we spend

our time looking for unproductive practices—things that do not help students to learn—we will produce very little change to our schools, because most of what everyone does in schools is valuable. The only way we can improve schools, therefore, is to stop people doing good things—in order to give them time to do even better things. This is the uncomfortable reality behind the trite phrase “work smarter not harder,” and it is why change in schools is so difficult.

Teachers are professionally invested in what they do. To a very real extent, teachers are what they do, and so asking them to change what they do involves changing who they are. When I suggest to teachers that spending less time on one activity (e.g., grading) to spend more time on another that is likely to have a greater impact on student learning (e.g., planning good questions to use in class), the response is often, “Are you saying what I’m doing is no good?” I’m not. I am suggesting that there may be other ways of spending time that may have a greater impact on student learning.

If we are to realize the potential of classroom formative assessment to improve student achievement, we must create an environment in every school in which each teacher expects to improve their practice continuously, and in which the improvements are focused on those things that make the greatest difference to student achievement. This will mean discontinuing practices which, while effective, are not as effective as other uses that could be made of the same resources—stopping people doing good things, to give them time to do even better things.

To learn more about implementing and sustaining teacher learning communities in your school, see Dylan Wiliam and Siobhán Leahy’s *Embedding Formative Assessment Professional Development Pack: a two-year professional development pack for schools and colleges: teacher learning communities in action*. [Details here](#).

To speak to our consultants about *Embedded Classroom Formative Assessment professional development in 2015*, or for further information about setting up teacher learning communities, call Learning Sciences Dylan Wiliam Center: 1.855.226.5595.