24 Expertise and Expert Performance in Teaching

James W. Stigler
Department of Psychology, University of California at Los Angeles, California

Kevin F. Miller
School of Education, University of Michigan, Michigan

Introduction

There are approximately 3.5 million full-time elementary and secondary teachers working in US public and private school classrooms (US Department of Education, 2016). Given their potential to impact the lives of young people, it would seem to be of great importance to understand what makes one teacher more effective than another, what expertise in teaching looks like, and how it develops. Despite the fact that teaching is one of the oldest and largest of human professions, however, we still lack a clear conception of what it means to be an expert teacher. This is not because researchers have not tried to pin this down. They have, and we will try in this chapter to make a contribution to this effort. But it is worth discussing, at the outset, why these questions have proven so difficult to answer.

Perhaps the biggest challenge to understanding expertise in teaching is that teaching is not an individual endeavor in which the teacher him- or herself is the only actor. Teaching is a complex system of interacting elements, and effective teaching requires that all of these elements work together to produce the desired outcomes. Later we will consider the nature of this system. But for now, consider simply that teachers cannot achieve their goals without the cooperation of students. Chess experts don't require the cooperation of the chessboard. The chessboard, the musical instrument, and so on are invariants in many domains of expertise. But in education, teachers are in a very real sense at the mercy of students, and policy-makers, and curriculum designers, and so on. This is something we will need to address as we consider what it means to be an expert teacher.

Another obstacle to understanding expertise in teaching is the "pseudo-expertise" one develops as a student. Before anyone starts their formal training as a teacher, they already have experienced well over 10,000 hours as students in classrooms, making teaching the profession with the most intensive and lengthy apprenticeship of any. One consequence of this experience is that everyone in our society, including teachers, thinks they already know what an expert teacher is, without any serious consideration of the research. This leads to bias in research on teaching, with a distinct lack of research designed to investigate theories that "everyone knows aren't true." It also may reduce variation in teaching methods within our culture, which makes it more difficult to explore alternatives.

Finally, our work is hampered by a lack of a consensus on the aims of education. We have a wide array of desired educational outcomes, yet teaching practices that prove effective for one outcome might be ineffective, or worse, for others. Similarly, practices that yield impressive short-term results (e.g. high scores on year-end standardized tests) may have negative effects on long-term outcomes such as career advancement and satisfaction decades later. Work by Jackson and colleagues (Jackson, 2012; Jackson, Rockoff, and Staiger, 2014), for example, has shown that teachers who produce the strongest gains on achievement tests are not the ones who succeed at reducing absences and suspensions, variables shown to predict future educational attainment and adult earnings. This complicates the straightforward strategy of studying teaching expertise by studying acknowledged experts.

Although it is tempting to rally round a single measure of student outcomes (e.g. the currently popular "value added" models of student achievement), we must be careful not to get too focused on the metric instead of on the underlying process the metric is intended to reflect. Campbell (1979) warned of this in what came to be known as "Campbell's law":

> The more any quantitative social indicator is used for social decision making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor. (Campbell, 1979, p. 85)

These concerns don't invalidate the effort to understand and measure student educational outcomes or to understand how teachers develop the skill to promote these outcomes, but they do help explain why understanding teaching expertise is not a simple matter of identifying and studying acknowledged experts.

In this chapter we try to take a broader approach to understanding the nature and development of expertise and expert performance in teaching. Because the literature on expertise in teaching is limited, we instead try to integrate a number of ideas and findings from literatures as diverse as cross-cultural comparisons of teaching, cognitive psychology, and systems improvement, among others. At a minimum, we hope to start a new conversation about what expertise looks like in the highly complex domain of classroom teaching.

The Nature of Teaching

A Definition of Teaching

Coach John Wooden famously said, "Everyone's a teacher, to someone" (Gallimore & Tharp, 2004, p. 119). The ubiquity of teaching means that we need to restrict the domain we seek to describe. In this chapter we will narrow the focus somewhat to include only classroom teaching. Despite this narrow focus, classroom teachers make up the largest segment of public sector employment, with 3.5 million teachers in the United States (US Department of Education, 2016) and more than 29 million in the world (UNESCO Institute for Statistics, 2015).

We base our definition of teaching on one offered by Lampert (2003; see also Ball & Forzani, 2007). Lampert sees teaching as "working in relationships," specifically, the relationships among the three core elements of a classroom lesson: the teacher, the students, and the content that is being taught. Teachers must manage the content and the students, but most importantly they must manage the relationship between students and content, over time. These relationships define the "problem spaces" in which teachers work. So, for example, teachers must relate to students, and must collaborate with students to get work accomplished, which generally means getting students engaged with studying the content. Problems arise in the management of each of these relationships, and teachers, over time, develop routines for handling the problems that recur.

Although we find this model to be a useful starting place for a full definition of teaching, we will broaden it in three ways. First, teaching always implies some goal for students (e.g. a learning goal) and some sequence of events designed to achieve that goal. Although some educators may prefer a more democratic view of relationships within the classroom, we believe it is important to see the teacher as an actor with a specific agenda for what she wants students to learn. This goal-directed nature of teaching was cited explicitly in some early definitions of teaching. For example, Thorndike (1906) defined teaching as the methods used to help students achieve the learning goals valued by society. The importance of making this agenda explicit has grown over the past 10 or 20 years as states and districts have focused more on creating and implementing clear academic content and performance standards for what students should know and be able to do at each grade level. Despite the recent influence of national and local standards, however, the teacher must still decide what learning goals the students in her class can and should achieve – at the end of the lesson, at the end of the unit, or at the end of the whole school year.

A second modification we make to Lampert's model of teaching is to broaden "teaching" to include the planning and reflection that go on before and after the lesson. It is common in the United States to overemphasize the importance of what happens during the lesson – the classroom performance of the teacher – and de-emphasize the intellectual work outside the classroom – what happens during planning and reflection. In fact, the effectiveness of a classroom lesson can be determined as much or more by the plan as by the on-the-fly decisions made by the teacher during the lesson. Planning is a teacher's single most powerful leverage point for improving the quality of what happens during a classroom lesson. Likewise, the teacher's later reflection on, and analysis of, problems that occur as the lesson unfolds can lead to revisions in the plan for next time, revisions that yield improvements in instruction. Reflection, or analysis, as we will call it, can be thought of as planning for the future informed by evidence gathered during the lesson.

A final element in our definition of teaching is time. Classroom teaching is highly constrained by time. Teaching is often implemented as a series of lessons, and these lessons are separated in time. Each lesson is often limited to a set amount of time. And there are only a finite number of lessons in the school year to cover the prescribed curriculum. Thus, in some sense, planning of individual lessons and sequences of lessons is a zero-sum game: as new activities are added, others must be deleted.

The centrality of time in teaching becomes even clearer when we consider students. Berliner (1990) noted that the amount of time students spend actively engaged in learning is often a small fraction of the time they spend in school. The amount of such engaged time can vary greatly across classrooms, and this can be a strong predictor of student learning (Fisher et al., 1980). Teachers' skillful management of time is a key factor in determining student learning. That management, and the planning and decisions that undergird it, are an important aspect of teaching.

Thus, we offer this definition of teaching: *Teaching consists of the interactions of teachers, students, and content, in classrooms, that are intended to achieve some goal or goals for students, within a specified period of time (e.g. a classroom lesson), together with the planning that takes place before, and the analysis that takes place after, the lesson.*

Teaching Is a System

More than many domains of expertise, teaching is a complex system with many moving parts. An expert pianist, for example, can have near total control over the product of her expertise, reliably performing beautiful music time after time. She can depend on a great deal of constancy in the environment. Unlike students, the keys on the piano on which she performs can be counted on to stay in place and respond consistently; the Mozart concerto will be the same each time she performs it; and the audience will apply a common set of standards as they appreciate and evaluate her work.

Teaching, in contrast, is constrained by a number of variables, many beyond the teacher's control. First there is the matter of students: a teacher can succeed only if students cooperate and engage in the tasks and activities the teacher assigns. According to Cohen, a leading education researcher, this is one of the key predicaments that define the teaching profession (Cohen, 2011). But it is not just the students that co-determine the results of teachers' work: curriculum, textbooks, assessments, policies, parents, peers, and so on, all interact and contribute to the nature and outcomes of the system we call teaching.

The fact that teaching is a complex activity that must be considered from a systems theory perspective presents challenges to those who would describe and study expertise in teaching. A teacher may appear to be an expert in one environment, yet seem more like a journeyman when observed in a different environment. The success of a lesson depends in part on a teacher's preparation and planning and her ability to assess students' understanding and respond to their needs. But it also depends on factors that are largely out of her hands, such as the previous experiences of those students as well as institutional culture, routines, and support for learning.

The question of whether expert performance "belongs" to the expert or to the organizational context is a matter of continuing controversy. Mueller and Dyerson (1999) have argued that expertise in complex institutions requires organizations that can develop and take advantage of individual skills; absent that, human experts may fail to make a contribution commensurate with their abilities. Others, such as Hoffman (1998), have argued that the question of whether knowledge resides in institutions or individuals is largely an artificial one that relates to how expertise is used rather than whether or not individuals can be experts. In this sense, an expert teacher in a dysfunctional school system would not be some kind of oxymoron, but rather a waste of human resources. Because larger cultural factors are central to both the development and practice of expert teachers, we would argue that expert teaching and its development require institutional supports and only make sense in a larger cultural context.

Teaching Is a Cultural Activity

Not only is teaching a complex system, but it is also a cultural system, what some have called a *cultural activity* (Gallimore, 1996; Stigler & Hiebert, 1999). Our own thinking on this emerged from the Third International Mathematics and Science Study (TIMSS) video studies, conducted during the 1990s (Stigler & Hiebert, 1999, 2004). In these studies, videotapes of national samples of eighth-grade mathematics and science lessons were collected in seven countries: Japan, Hong Kong, the Netherlands, Switzerland, Czech Republic, Australia, and the United States. The study had two goals: to document what "average" teaching looks like at a national level, and to compare teaching in the United States with teaching in countries that are high achieving when compared to the United States based on assessments of mathematics and science achievement.

Findings from these studies indicated a striking homogeneity of teaching practices within countries, but marked differences in practices across countries. Even within a country as diverse as the United States – racially, ethnically, linguistically, and socio-economically – a national sample of eighth-grade mathematics teachers appeared to be following a common script, despite the fact that teachers are given high levels of autonomy and control over the methods they use. The US lesson script appears designed to produce what Skemp (1987) called an "instrumental" understanding of mathematics. Teachers walk through example problems, then supervise students as they practice solving similar problems, the goal being for students to remember the steps used by the teacher, and then to be able to execute the steps without errors.

In Japan, by contrast, the lessons follow a different cultural pattern. Japanese teachers typically begin by presenting students with a difficult problem to work on, one they have not seen before. The teacher does not instruct students how to solve the problem, but simply lets the students struggle to find a way to solve it on their own. Because students have not been told how to solve the problem they usually come up with a variety of different solution methods – some correct, some incorrect – which they then discuss in class. Through these discussions, teachers focus on what Skemp would call a "relational" understanding of mathematics, working to deepen connections with core underlying mathematical ideas (see also Schwartz, Chase, Oppezzo, & Chin, 2011).

Cultural routines such as these are not created deliberately. Instead, they evolve slowly over time as cultures adapt to an ever-changing environment. Modern schooling is itself a cultural invention, perhaps one of the most successful cultural inventions of our modern era. As the world economy shifted from agriculture to industrial production, schools developed to prepare workers who could fit into the jobs and living arrangements that were emerging. Schools, in a sense, went viral, and now they are everywhere. And significantly, the cultural routines of teaching that developed in the United States more than a century ago appear to be mostly unchanged (Cuban, 1990; Hoetker & Ahlbrand, 1969).

Cultural activities are learned implicitly, which makes teaching quite different from most other domains of expertise. As a cultural activity, teaching is more like dinner-time conversation than it is like flying an airplane. The routines of dinner-time conversation are learned from growing up in a family and observing how others behave at meal times. People do not take a course or read a manual to learn this. They learn to participate in cultural activities by observing and imitating others. There is evidence that people learn to teach in just this way – by observing their teachers during their 13 years of schooling before entering college and by imitating what they remember (Lortie, 1975).

Cultural activities can be hard to see simply because they are so widely shared within a cultural group. The teaching methods that are prevalent in US schools today appear natural to those of us who grew up here, but may appear strange to those who experienced a different tradition of schooling. Breaking free of our cultural lenses is one of the main benefits of cross-cultural comparison. When Japanese students are asked to solve problems they have never seen before, they struggle and often appear confused, yet teachers do not intervene to simplify the problems or resolve the confusion. When we observe this in Japan it suddenly draws attention to how uncomfortable US students and teachers are with the experience of confusion. These different cultural routines are supported by wider cultural beliefs about the role of confusion in learning.

Finally, cultural activities are hard to change because they are multiply determined. Many factors conspire to keep things as they are. We know from cognitive science research that confusion is actually a critical part of deeper learning (D'Mello, Lehman, Pekrun, & Graesser, 2014). Yet if a US teacher tried to induce confusion in their students (let's call it "productive confusion") using teaching routines similar to those used in Japan, she would no doubt get a lot of pushback. Students would complain ("We haven't had that!"), parents would complain ("It's not fair to expect students to work on problems you haven't taught them how to solve"), textbooks would not introduce material in the right order, and so on. The cultural nature of teaching raises problems for educational change, but it also has implications for the nature and development of teaching expertise.

Expertise and Cultural Activities

Cultural activities are implicitly learned and often operate outside of our awareness. This raises some interesting issues in terms of expertise. Most participants in cultural activities are neither experts nor novices; they are simply operating within the normal limits of variation for their cultural system. In Ericsson's terms (Ericsson, 2008), they have reached a (perhaps premature) level of automaticity, a level at which they are "good enough."

An important consequence of the cultural nature of teaching is that one can be an expert in implementing teaching routines that are not themselves optimal. Hatano and Inagaki (1986) distinguished between two courses of expertise. Routine experts become efficient at implementing relatively set routines (an example might be someone working in a fast food restaurant) without necessarily understanding why they work or being able to reproduce them in a different setting. Adaptive experts (such as a sushi chef) deal with constantly changing problems and need to develop both an understanding of why things work as they do and an ability to alter their approach as circumstances change.

It seems obvious that teaching involves adaptive expertise, because students present constantly evolving challenges to teachers. Yet to the extent teaching is nested within a set of cultural beliefs, the range of exploration of possible strategies will be limited. Those cultural beliefs vary across cultures but also change over time. Resnick and Resnick (1977) showed that societal expectations on literacy have changed dramatically over time. Strategies for teaching that work when only a low level of literacy is expected for most students will not be as successful when we expect much more from all students. This in turn has implications for what constitutes expertise in teaching literacy.

Whether we should expect expert teachers to be able to transcend the limits of their cultural teaching routines is an open and challenging question. We have now discussed some of the obstacles to a straightforward analysis of expertise in teaching and described a model of what teaching entails. In the remainder of this chapter we will discuss what this means for studying and developing teaching expertise.

Expertise in Teaching

The first step in studies of expertise typically involves identifying a set of experts and analyzing how they differ from novices. In the case of teaching, three main approaches have been used, each with important limitations. These are: (1) comparing beginning and experienced teachers, (2) studying teachers who have been identified as experts through a process of nomination or certification, and (3) looking at student outcomes to identify expert teachers.

Experience as a Proxy for Expertise

A large body of literature supports the idea that expertise requires practice over long periods of time, but experience alone does not guarantee the development of expertise. Studies of the effect of teacher experience on student achievement generally find positive but small relationships between a teacher's years of experience and their students' learning (e.g. Nye, Konstantopoulos, & Hedges, 2004; Rockoff, 2004); some studies show benefits during the first several years, but none after that (e.g. Hanushek, 2003; Rivkin, Hanushek, & Kain, 2005). The problem with conflating experience and expertise has long been recognized (e.g. Berliner, 1986). Despite this, experience still has been the main variable used to indicate expertise in teaching, at least until recently. Given that most practitioners in a domain will plateau before reaching the highest levels of expertise, these modest relationships probably underestimate the effects of expert teachers on their students' learning.

Studying What Recognized Experts Do

The second approach to understanding expertise involves studying people who have been recognized as experts. One obstacle to this approach is that, at least in the United States, teaching is an activity that is often observed only by students. Berliner (1986) noted that judging of skill and outcomes in athletics, livestock, dogs, crops, and other fields takes years of systematic training and practice, but we lack anything similar in the evaluation of teaching.

Efforts to systematically observe elementary and secondary school teaching have a long pedigree, going back at least to a program in the 1880s and 1890s that John Dewey participated in at the University of Michigan (Williams, 1998). More recently, the National Board for Professional Teaching Standards (NBPTS) has developed a process for certifying teachers. To date more than 112,000 teachers have completed their certification process, which involves submitting a multimedia teaching portfolio and taking a three-hour assessment examination.

An early evaluation of the program (Bond, Smith, Baker, & Hattie, 2000) looked at a set of 65 experienced teachers divided into groups based on their performance on the assessment, with approximately equal numbers passing and failing the NBPTS certification assessment. Interviews and classroom observations showed consistent differences favoring the designated expert teachers in, among other areas, the depth and challenge of problems set for students, the ability of teachers to anticipate and plan for classroom problems, and the depth of their representation of classroom situations. Student writing samples from the NBPTS certified teachers demonstrated higher understanding than did those from the comparison group. A later study by Hogan and Rabinowitz (2009) compared NBPTS certified and novice teachers and found similar differences in the depth of their representation of classroom problems.

Efforts to show that NBPTS certified teachers have better student outcomes than do their non-designated peers have shown more mixed results. Cavalluzzo (2004) compared student achievement in mathematics in a large urban school district in classes taught by Board-certified and other teachers, finding consistent effects favoring the certified teachers. Goldhaber and Anthony (2007) and Clotfelter, Ladd, and Vigdor (2007) found that students of NBPTS certified elementary school teachers in North Carolina did better than those of uncertified teachers, with effects on the order of 0.05 standard deviations. Goldhaber and Anthony reported a larger effect, of 0.11 standard deviation for students eligible for free or reduced cost lunch.

A much smaller effect was reported by Harris and Sass (2009), who looked at data from Florida. They also reported larger effects for teachers who had received NBPTS certification when it first became available. They argued that this suggests that the certification process initially identified a set of committed and effective teachers, but that as it became more widespread the certification process was less successful in identifying more effective teachers.

The NBPTS has provided a systematic way of designating some teachers as at least relatively expert. There is some evidence that the teachers so designated differ from their peers in ways consistent with current ideas about effective teaching, but less consistent and convincing evidence that their students do better than do those of other teachers. This suggests that we might do better by simply identifying expert teachers based on how their students perform.

Identifying Experts Based on Student Achievement

An inductive approach to identifying experts involves looking at results and designating as experts those teachers whose students develop faster than expected. As discussed at the start of this chapter, this approach is at the heart of the "value added" approach to teacher assessment. Finding ways to connect teacher expertise to student outcomes will be critical to improving education, but efforts to do this to date have shown how complicated this seemingly straightforward approach is.

One recent attempt to identify practices associated with student learning was the colossal

Measures of Effective Teaching (MET) project sponsored by the Bill and Melinda Gates Foundation. The study was the most ambitious study of teaching ever undertaken: more than 20,000 videotaped lessons were collected from 3,000 elementary and middle-school teachers' classrooms in seven urban school districts across the United States. Students were surveyed, and students' test scores – both on state standardized tests and on supplemental tests of higher-order thinking in both mathematics and literacy – were collected and matched to the video data. Even more impressive, in the second year of the study a subsample of 800 teachers were randomly assigned, within schools, to different classrooms of students, and again, students' learning at the end of the year was measured.

This was the first large-scale study to identify teacher effects on students' learning using random assignment. The measures of teacher effectiveness based on students' learning in year one (non-random assignment) were highly predictive of the randomly assigned students' learning at the end of year two (Kane, McCaffrey, Miller, & Staiger, 2013), and importantly, the size of the teacher effect in year one was the same as that in year two. Clearly, some teachers are more proficient than others at producing student gains on standardized state tests. But disappointingly, the observational measures applied to the videos of classroom teaching yielded very little of note, predicting almost none of the variance in student learning at the end of the year (Kane & Staiger, 2012).

These findings are in some ways reminiscent of the early work on chess (de Groot, 1965, 1966). Despite numerous attempts to find differences in the way chess masters and weaker players play chess (e.g. number of moves considered, search heuristics, depth of search, etc.), the only reliable difference turned out to be in the quality of the move: chess masters make better moves in any given situation, and thus win more games, than the weaker players.

In the case of teaching, we can imagine several additional reasons such correlations would be low. One might be the alignment between teachers' goals and the measures used to indicate student outcomes. If standardized tests aren't measuring the student learning outcomes of highest priority to teachers, then you would not expect a correlation between teachers' behaviors in the classroom and students' end of year test scores. But perhaps even more important is the contextual nature of teaching. A move that might be best in one situation might be precisely the wrong move to make in a different situation. For example, a critical remark to one student may be just what he needs to engage him in digging deeper on a problem. But the same remark may come as a crushing blow to a different student, who needs more encouragement.

In this sense, teaching is more like driving to work than like shooting a rocket ship to the moon. Most of the work involved in shooting a rocket ship can be done in advance. The trajectory can be calculated with near-perfect accuracy, such that when the button is pushed, everything unfolds as predicted. Driving to work, on the other hand, requires constant adjustment to the expected and unexpected variations that occur. Taking a left at a certain intersection may usually be best, but not always. Unexpected obstacles or weather conditions require a response from the driver. The expert commuter knows when to adjust, when to go around. Teaching, thus, is more like driving to work. With goals in mind, teachers must constantly read the situation, monitor progress, and make necessary adjustments.

This analysis leads us to reject the idea that expertise in teaching can be defined in terms of decontextualized "best practices." Our view is that correlations between teacher actions and student learning are low not because we haven't yet identified the right set of best practices, but because teaching itself is contextual, meaning that such correlations will always be low.

Further support for this view comes from the TIMSS video studies. If expertise in teaching is defined by a set of best practices, then one would expect the practices used by teachers to be similar across the highest achieving countries. For example, because Japan is a high-achieving country in mathematics, we might expect that Japanese teaching routines (as described earlier) would be similar to those used in other high-achieving countries.

In fact, however, this proved not to be the case. Although students in the Netherlands, Czech Republic, Switzerland, and Hong Kong – as well as Japan – all score relatively high on international mathematics tests, teaching methods across these countries vary markedly from one another. Features of teaching seen as desirable by education reformers in the United States – for example, the use of manipulatives, real-world problem scenarios, and group work – were found in some, but not all, of the high-achieving countries. Almost everything coded in the TIMSS video studies varied among the high-achieving countries (Stigler & Hiebert, 2004).

We interpret this finding, again, as further evidence for the cultural and contextual nature of teaching. In Japan, a teacher can pose a difficult problem at the start of a lesson, and students will immediately set to work on it, even if they find it frustrating and uncomfortable. If an American teacher adopted this same practice the results might be quite different. In a project aimed at helping American teachers to increase student discussion in mathematics lessons (Wang, Miller, & Cortina, 2013), we found that helping teachers develop skills at leading mathematical discussions and giving them daily feedback was not sufficient to promote change. We had to include some "professional development" for students as well, because they did not know how to listen to and engage with each other's mathematical ideas. Teaching expertise exists within a cultural matrix, which means that expert teachers in different cultures may act very differently.

The Construct of Learning Opportunities

If teacher expertise cannot be equated with a set of best practices, what does explain the considerable teacher-level variance in student learning outcomes? We believe a clue may be found by digging deeper in the TIMSS videos of classrooms in the high-achieving countries. In fact, we do see commonalities, not at the level of what teachers *do* but in the kinds of *learning opportunities* they manage to create and sustain for students. Teaching routines differ among the high-achieving countries. But although they use different routines, and the actions of teachers differ, all appear to have found ways to create a common set of learning opportunities for students.

Based on results from the TIMSS video studies, and on our reading of research on teaching and learning more broadly, we propose three distinct types of learning opportunities that are necessary to produce high levels of learning in mathematics and, we believe, in other subjects as well:

- *Productive struggle* – This can be simplified to the aphorism *no pain no gain*: deep learning requires some element of struggle. Despite the romantic view of learning as ideally fun and enjoyable, students learn more when they are engaged in hard intellectual work which, though rewarding in the end, is not necessarily enjoyable in the moment (see, for example, Bjork & Bjork, 2011).
- *Explicit connections* – Evidence from the discovery learning literature suggests that just struggling with core content will not necessarily lead to learning (Kirschner, Sweller, & Clark, 2006). It is also important to help students make explicit connections between the problems they are working on and the concepts they need to understand. The source of these connections (individual discovery, listening to peers, or teacher summaries) may matter less than whether or not they are explicitly made.
- *Deliberate practice* – Finally, struggle and connections need to be sustained over time,

through practice; and not by repetitive practice but by deliberate practice, as demonstrated in the literature on expertise, with the opportunity to vary strategies and get informative feedback.

Whereas the most common paradigm for researching the effects of teaching on learning looks for the simple effect of teacher actions/behaviors on student outcomes, we propose that a slightly more complex model will be required: teacher actions and behaviors create learning opportunities for students, which, in turn, produce student outcomes. Under this model, expert teachers are not defined as those who employ a set of best practices, but instead those who (1) have the ability to assess students' current knowledge state both prior to and during instruction, (2) formulate clear learning goals, (3) consider a large number of strategies and routines in their repertoire, (4) make good judgments about which strategies are most appropriate in any given situation, and (5) are able to implement the strategies effectively to create learning opportunities for students. This model is represented in Figure 24.1.

Are these additional layers really necessary? We believe they are. Besides the fact that simple correlations of teaching actions with student learning have proven less than fruitful, there is another reason: learning, especially the kind of deeper learning called for in current education standards, takes time. Unlike chess, where the final outcome can be known in the space of an hour or so, student learning of core content standards often takes place over months and years. Experts need to orient their actions simultaneously toward long-term learning outcomes and to more immediate indicators of learning. These more immediate indicators, we believe, are best found in a theoretically motivated construct such as learning opportunities – similar to what Lipsey (1993) has called "small theories."

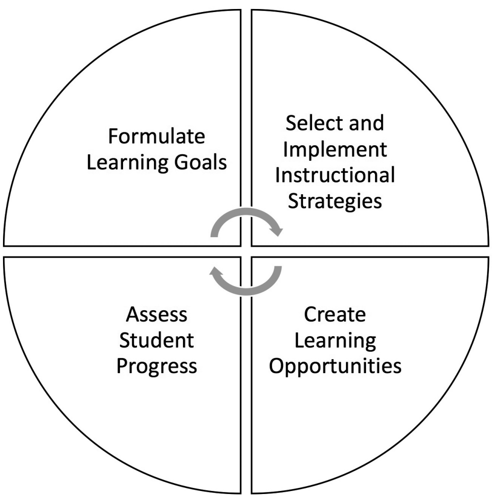

Figure 24.1 A model of expertise in teaching.

Expertise in the Classroom

What does it take for teachers to create these learning opportunities in the classroom, especially given the complex nature of teaching? First, teachers must have clearly defined goals for what they want students to learn, and the ability to assess the gap between students' current knowledge and where they are trying to help students go next. Expert teaching is not just performing the acts of teaching. It is a highly contextualized endeavor in which teachers must create the precise learning opportunities that will move students to the next level of learning and development. Doing this requires great skill at formative assessment, which is one of the core elements of expertise in teaching (Black & Wiliam, 1998).

Based on these assessments, which must be continually updated during instruction, teachers then must create learning opportunities targeted toward students' current needs. Doing this will require knowledge, skill, and judgment.

Knowledge. Much research has focused on what expert teachers need to know in order to teach effectively. Clearly they need to know the subject that they are teaching. But this begs the question of what we mean by "know." A great teacher of high school physics does not need to know physics as well as a professional physicist might know it. In fact, the professional physicist might be a below-average teacher of physics to high school students, or even to university students (see Feldon, 2007). Knowledge for teaching is not the same as knowledge for some other purpose.

What do teachers need to know? Shulman (1987) coined the term *pedagogical content knowledge* to refer to the special knowledge teachers need in order to teach effectively. However, pedagogical content knowledge is just the tip of the iceberg. Shulman proposed a taxonomy of teacher knowledge that still proves useful today (Shulman, 1987, p. 8). It includes:

- content knowledge;
- general pedagogical knowledge, with special reference to those broad principles and strategies of classroom management and organization that appear to transcend subject matter;
- curriculum knowledge, with particular grasp of the materials and programs that serve as "tools of the trade" for teachers;
- pedagogical content knowledge, that special amalgam of content and pedagogy that is uniquely the province of teachers, their own special form of professional understanding;
- knowledge of learners and their characteristics;
- knowledge of educational contexts, ranging from the workings of the group or classroom, the governance and financing of school districts, to the character of communities and cultures; and
- knowledge of educational ends, purposes, and values, and their philosophical and historical grounds.

Much of the research that has been inspired by Shulman's taxonomy has focused on pedagogical content knowledge. Hill, Ball, and colleagues (Ball et al., 2006; Ball, Thames, & Phelps, 2008; Hill & Ball, 2004) have invested considerable effort in further subdividing, and measuring, pedagogical content knowledge, developing measures referred to as the Mathematics Knowledge for Teaching (MKT) scales.

Many researchers have studied the relationship between pedagogical content knowledge and student learning, but as with research on teacher practice, the correlations have been disappointingly low (Hill, Rowan, & Ball, 2005). One reason for these low correlations must certainly be the complex systemic nature of teaching; with so many variables in the mix, the correlation of any single one would be expected to be low. Another reason may be that the most common measures of pedagogical content knowledge consist of paper and pencil multiple-choice items, which may be measuring inert knowledge (Bransford, Goldman, & Vye, 1991; Whitehead, 1929). As with the case of teaching practices, the key point may not be whether or not you know something, but whether you are able to access and apply the knowledge when you need it to improve students' learning opportunities.

Kersting and colleagues have developed an innovative measure of teacher knowledge that they view as a measure of "usable" knowledge for teaching. This measure, called the Classroom Video Analysis (or CVA) assessment, requires teachers to view video clips of authentic classroom episodes and then comment on what they see in the video in terms of interactions among the teacher, students, and content being taught. Coding of teachers' open responses yields a measure of teacher knowledge that correlates with Hill and Ball's MKT measures (Kersting, Sherin, & Stigler, 2014). More important, however: the CVA measures have been shown to be far better predictors of student learning than have the paper and pencil MKT measures (Kersting, Givvin, Thompson, Santagata, & Stigler, 2012; Kersting et al., 2016).

Skill. Despite the importance of teacher knowledge, there is a performance component as well to teaching. It's not enough to just know what to do, but you also need to be able to do it well in a variety of situations. Most teacher education and professional development programs for teachers have focused on making teachers more knowledgeable; few actually give teachers opportunity to practice the skills of teaching. But some recent work has recognized the importance of skilled implementation for creating learning opportunities in classrooms.

One such project is the one led by Lampert, Franke, and Kazemi (Lampert et al., 2013). These researchers have identified a set of instructional routines that are within the capabilities of preservice teachers, yet still complex enough to represent the work of teaching. Such routines (which they refer to as activities) include Choral Counting, Launching and Using Word Problems, and so on. Participants in the program engage in "rehearsal" of these instructional routines through a process that very much resembles deliberate practice.

Participants in the program begin by planning mini-lessons that incorporate one of the instructional routines, and then implement the lesson with their peers. After getting feedback from peers, they implement the lesson in a real classroom and record the result on video. They later take these videos back to the university and analyze their implementation, getting feedback from experts and peers. By using such routines as a site for deliberate practice, teachers in training are able to integrate the skills of teaching with the concepts and knowledge they are being taught in their teacher education program. The goal is to develop adaptive knowledge of the routines – the ability to implement them effectively to achieve instructional goals in a range of contexts.

Instructional routines are not the only aspects of teaching that require skilled implementation. The skills of managing emotional connections with students are perhaps equally critical, especially if we expect students to engage in the hard work of productive struggle. Csikszentmihalyi, Rathunde, and Whalen (1997), in their influential study of talented teenagers, identified the relationship that students had with their early teachers or coaches as an important theme. Activities such as playing a violin or running long distances are not initially intrinsically rewarding, but having a teacher one does not want to let down can keep students engaged until they develop enough skill for the activity itself to be rewarding.

This is an aspect of teaching that is often downplayed in US discussions. We interviewed elementary school teachers in China and the United States about their ideas of what contributed to their students' success in learning mathematics (Correa, Perry, Sims, Miller, & Fang, 2008). Chinese teachers were much more likely than their US counterparts to emphasize the importance of their relationship with their students (as one teacher put it, "First of all . . . I think I'll let the students love me first. The students need to love their teacher before they love the subject, so I should develop a good relationship with them").

Judgment. Finally, just knowing *that* and knowing *how* are not enough. Because teaching is highly contextual and complex, teachers must also have the ability to decide which of many possible strategies they should pursue in any given time and place. They need to be able to size up a situation, decide which strategy to employ, and then adapt it to achieve their specific instructional goal. In other words, teachers need judgment. They can be highly knowledgeable in all the ways Shulman (1987) describes, and highly skilled at implementing a wide variety of instructional strategies. But unless they make good decisions about when and how to employ their knowledge and skill, their knowledge may not serve to support students' learning.

Expert teachers' judgments must be based, in part, on an expert reading of the situation, something that has been studied across a wide range of domains of expertise. Experts in fields from chess to radiology to electronic circuit design all appear highly skilled at extracting structure from the world as it relates to their domains of expertise (Hoffman, 1998; Kellman & Garrigan, 2009). Sometimes referred to as "situation awareness"

(Endsley, 1995), experts are able to quickly recognize the important features of complex situations and determine their consequences, whereas those who are less expert are more likely to miss key features, or spot them more slowly.

These kinds of effects have also been found for teachers. Sabers, Cushing, and Berliner (1991) found that experienced teachers were better than novices at shifting their attention among multiple views of a classroom and identifying important events. In our lab, we have used mobile eye tracking methods to compare experienced and novice teachers teaching the same subjects to the same students (Miller, 2011). Experienced teachers are more focused in their attention to the important aspects of the classroom environment, such as students and curriculum materials, but some of the most interesting findings involve some of the trade-offs that beginning teachers need to make to tame the complexity of having to attend to a classroom of students at the same time they are trying to present a coherent lesson. A particular trade-off (Cortina, Miller, McKenzie, & Epstein, 2015) involves the balance of attention to individual students versus to the class as a whole. Beginning teachers who are rated as being highly responsive to individual students do not do a good job of dividing their attention evenly among the students as a whole, as shown by eye tracking data. Experienced teachers do not show this trade-off, with some teachers managing to be both responsive to individual students and attentive to the class as a whole. Situation awareness in teaching requires developing the kind of quick categorization of what's important in a classroom situation, as measured by the Sabers et al. (1991) task and the measures developed by Kersting and colleagues (described above). To be an expert teacher, one must be able to perceive structure in educational contexts, then link that structure to underlying concepts and principles of teaching and learning and to the repertoire of strategies and routines that might be used to achieve the immediate goal.

How Teachers Become Experts

If there is a lesson to be learned from the huge literature on teacher professional development it is this: professional development, for the most part, doesn't work. But then, most of what counts as professional development consists of haphazard, voluntary, and brief workshops that are disconnected from the daily work of teaching (Birman et al., 2007). Such activities may succeed in making teachers more knowledgeable, but because teaching is a complex cultural system, training just the teacher, in a time and place divorced from the ecology and culture in which they operate, is highly unlikely to improve the performance of the system as a whole.

Indeed, a small number of carefully designed experimental studies show that when professional development is intensive, ongoing, and job-embedded, focused on students' learning of the curriculum being taught, and aligned with the school's improvement goals and priorities, such programs can produce significant improvements in student learning (Cohen & Hill, 2000; Darling-Hammond, Wei, Andree, Richardson, & Orphanos, 2009; Saunders, Goldenberg, & Gallimore, 2009; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). In other words, when other critical elements of the system of teaching are taken into account, and when teachers are given time to work on improving implementation within the context of the system in which they are working, teaching will improve. What does not work is training teachers apart from the setting in which they are expected to practice.

There is a large literature on how to improve complex systems that include human actors, in fields ranging from automobile manufacturing to healthcare (e.g. Langley et al., 2009; Rother, 2009). We also know quite a bit about what it takes to change cultural routines (e.g. Gallimore, 1996; Feldman & Pentland, 2003), as well as what it takes to develop expertise across a wide variety of domains (Ericsson, Krampe, & Tesch-Römer, 1993; Hatano & Inagaki, 1986). Rarely, until recently, however, have these strategies and methods been applied to improving teaching (Bryk, Gomez, Grunow, and LeMahieu, 2015). In the next part of the chapter we hope to make some of these connections.

Teachers versus Teaching as the Focus of Improvement

Feldman and Pentland (2003), in their work on organizational routines, made a useful distinction. Routines, they write, consist of two related parts: "One part embodies the abstract idea of the routine (structure), while the other part consists of the actual performances of the routine by specific people, at specific times, in specific places (agency)" (p. 95).

Teaching can be thought of in a similar way. The structure consists of the cultural routines, and the system in which the routines have evolved and function. Agency consists of the way the routines are implemented in different classrooms by different teachers. Improving teaching can thus be accomplished in two distinct ways: improving the system itself, including the routines, and improving the expertise of the teacher as she implements the cultural routines that she has inherited.

As it turns out, the processes of improving the performance of complex systems, changing cultural routines, and developing expertise have a lot in common. In all cases, awareness is a critical part of the process – seeing clearly what the current system looks like. Improvement scientists (cf. Langley et al., 2009; Rother, 2009) talk about learning to "see the system," which is never easy, but it is even more difficult when the system includes cultural routines. Building awareness of current routines (or skills in the case of expertise) is often better accomplished in collaboration with others – peers, team members, stakeholders, or even just a coach – because it requires that the routines be described explicitly – put into words.

With a clear understanding of the system as it currently works (what Rother, 2009, calls the "current condition"), the next step is to establish a "target condition," that is, what you want to see working differently (or more expertly) than it is now. Then, obstacles are identified – what is keeping you from reaching the target condition – and a next step is identified: what will you work on next to move the system incrementally closer to the next target condition? In systems improvement approaches such as the one described by Rother, this sets off a series of PDSA (Plan/Do/Study/Act) cycles, a process in which a change idea is planned and implemented, then analyzed to see what is learned.

Importantly, developing the expertise of individual actors is not usually a goal of systems improvement. The goal, in fact, is usually just the opposite: to create a system that performs reliably at a high level without relying on high levels of expertise. If at all possible, the system should be designed so that all human actors, within a normal range of skill, can function effectively. According to Gawande (2007), this is the reason the field of obstetrics eventually gave up forceps in favor of the Caesarean section as a way to deliver babies in trouble: almost any physician could successfully execute a Caesarean section, whereas only those with high levels of expertise could use forceps effectively. When used expertly, forceps produce better results than a Caesarean. But when used by less skilled doctors, the results of a forceps delivery can be disastrous, and all obstetricians must begin as novices.

The advantage of improving the system of teaching (as opposed to only the expertise of teachers) is obvious: changes in the system, even if small and incremental, can lead to long-term, large-scale improvements in student outcomes. Teachers come and go; most actually stay in the profession for only a few years (Simon & Johnson, 2015; Ingersoll & May, 2012). But system improvements that can be captured and accumulated over time have a far longer lifespan. Thus, it makes sense to shift at least part of our focus from teach*ers* to teach*ing*. But how can we accumulate system improvements so that they can be shared with each new generation of teachers?

Morris and Hiebert (2011) put forth an interesting proposal, that instructional objects (e.g. curriculum materials, lesson plans, etc.) be seen not only as important components of an instructional system, but as a mechanism for storing professional knowledge. In their vision, educators will work together to continuously improve the system of teaching, capturing what they learn, as much as possible, in improvements to the actual instructional materials themselves. The shared instructional objects thus become a sort of knowledge base to guide teaching.

Yet, despite the importance of improving the system of teaching – the curriculum, resources, methods, routines, and so on – there will always be a need, at the same time, to improve the expertise of teachers. As pointed out earlier, teaching is more like driving to work than it is like shooting a rocket to the moon: successful teaching will always include a large element of *implementation*. Because teaching is so highly contextual, and because so much judgment is required, it will always be necessary to develop individual teachers' expertise to effectively implement routines for all students in all situations.

Experience by itself will not lead to expertise. Instead, opportunities for deliberate practice will be required, to which we now turn.

Creating the Conditions for Deliberate Practice of Teaching

A large body of work has documented the critical role of deliberate practice in the development of adaptive expertise. Deliberate practice is not to be confused with repetitive practice. Anyone with lots of experience has also gained a lot of repetitive practice. But deliberate practice is something else. Deliberate practice is usually a designed experience – often by a teacher or coach – not one that happens naturally (Ericsson, 2006). It usually requires a set of training tasks that are outside the current comfort range of the learner, but not so far that they can't be mastered in a matter of hours.

According to Ericsson (2006), deliberate practice always involves conscious concentration on the skill, and informative feedback, both from the performance itself and from a coach or peers. This requirement for concentration "sets deliberate practice apart from both mindless, routine performance and playful engagement, as the latter two types of activities would, if anything, merely strengthen the current mediating cognitive mechanisms rather than modify them to allow increases in the level of performance" (Ericsson, 2006, p. 694). Hatano and Inagaki (1986) argue that the connection of concepts and understanding to the practice of routines is what makes the routines adaptive and able to be applied effectively in novel situations.

There are a number of characteristics of teaching that work against successful implementation of deliberate practice. First among these might be the cultural assumption that teaching is not something generally subject to improvement. This belief might arise from a number of sources. Unlike other domains of expertise, the outcomes of teaching (e.g. student learning) are affected by numerous factors, the teacher being only one. Especially in environments where achievement is low, it may be difficult to convince teachers that the work they do can have a significant and direct impact on outcomes. If teachers don't see their own skill as something that needs to be improved, it will be hard to engage them in deliberate practice.

No matter how motivated teachers are to improve, there are other things that make it hard for them to engage in deliberate practice of teaching. Chief among these is the fact that, unlike many domains of expertise, teaching is one in which most of the teacher's time is spent in *performance*, not in practice. Fadde and Klein (2010) point out that in music and sports, for example, performance happens rarely, leaving lots of time for practice in between performances. But many professions, for example, management, primarily consist of performance on a daily basis. In these kinds of professions – and teaching is clearly one – Fadde and Klein (2010) propose that *deliberate performance*, not deliberate practice, is the most feasible route to expertise.

Fadde and Klein (2010) define deliberate performance as "the effort to increase domain expertise while engaged in routine work activity" (p. 6). Starting with features of deliberate practice (e.g. repetition, timely feedback, task variety, and progressive difficulty), they propose four types of exercises that professionals can engage in while on the job. These exercises include estimation (e.g. predicting how many students will use each of two different strategies for approaching a math problem), experimentation, extrapolation, and explanation. Fadde and Klein also propose that coaching is a much-needed resource for anyone embarking on a path of deliberate performance. Coaching would help to bring hidden routines to awareness, to make explanations explicit, and to provide feedback in domains where feedback is often difficult to come by.

Thus, it is difficult but not impossible to create the conditions for deliberate practice of teaching. Our challenge is to create settings in which teachers can practice the implementation of a teaching routine or strategy; see the concrete results of their actions in evidence of students' learning; reflect on the cause–effect relationships between their teaching, the learning opportunities they create for students, and the evidence of what students learn; and get feedback from knowledgeable others. We turn now to two examples of where this has happened.

Lesson Study as a Lab for Deliberate Practice

Much has been written about the Japanese practice of lesson study (Lewis, Perry, & Hurd, 2004; Lewis, Perry, & Murata, 2006; Stigler & Hiebert, 1999). In lesson study, teachers meet regularly in groups of four to six to work on planning, implementing, and improving specific lessons from the curriculum. The goal of lesson study is not just to improve a particular lesson. Instead, the lesson becomes a vehicle for working on an improvement goal that goes well beyond a single lesson. So, for example, a lesson study group might be working on deepening students' understanding of a particularly challenging concept, or on improving their ability to elicit students' thinking through questioning.

Similar practices are at work in other Asian countries, and the practice is also growing in popularity in the United States. Lesson study is relevant here for at least two reasons. First, it appears to have been inspired by the early work of Deming, a pioneer of improvement science who, coincidentally, spent much of his career working in Japan (Gabor, 1990; Kenney, 2008). As practiced in Japan, lesson study is first and foremost a research and development process, much like the PDSA cycles of improvement science (Langley et al., 2009). Second, lesson study appears to be aligned almost perfectly with the requirements for deliberate practice. In this sense it has the potential to greatly accelerate the development of expertise in teaching, as well as the system of teaching itself.

The lesson functions similarly to the designed practice tasks described by Ericsson (2006). In lesson study, the practice of teaching is slowed down, often for weeks, as teachers discuss, first, what they want students to take from the lesson, and then what kinds of evidence they can collect to indicate students' thinking and learning during the lesson. The group then plans, in great detail, a lesson that they hypothesize will achieve the goals they have set, and develops explicit hypotheses and predictions for both the processes and outcomes they will observe. Once they have created a detailed plan – a process that often takes place over several weeks – one of the teachers in the group will teach the lesson while the others observe, clipboards in hand, collecting evidence that then will be used to revise and improve subsequent versions of the lesson.

Debriefing sessions follow the implementation of the lesson. In these sessions, teachers share the evidence they have collected, and analyze its significance for both understanding and improving the learning process that unfolds during the lesson. They focus special attention on places where their predictions were wrong, and seek to explain these deviations with cause–effect hypotheses. For example, they may observe some subset of students whose work indicates a misconception of the concept being taught. They then may think carefully about the task that was assigned, and even the specific questions the teacher asked as the work progressed, seeking to explain how the misconception could have emerged, and to find an incremental improvement that they can test in the next version of the lesson, which generally will be taught by a different member of the group. Through all of this the focus is kept on the lesson that the group has designed, not on the individual teacher who taught it, a strategy that helps to mitigate any defensiveness teachers may feel that could slow down their own learning.

Ericsson has noted that premature automaticity may well be one of the biggest enemies to the development of expertise. Teaching, because of its heavy reliance on cultural routines, may be even more subject to this barrier to improvement than are other domains of expertise. Lesson study provides a means of disrupting the normal "good enough" routines of teaching. By spending weeks or months on a single lesson, each and every part of the lesson – both the parts that are planned ahead of time and the parts that are adapted at the time of implementation – is subject to deep analysis, revision, and practice, all in a context in which feedback is readily available. Lesson study is thus a context or lab in which deliberate practice can take place.

Analysis as the Key to Developing Expertise in Teaching

Even if lesson study can provide a setting for deliberate practice of teaching, it does not appear to provide enough opportunities for such practice to produce high levels of expertise over time. But what if the main things being practiced during lesson study are not the skills of teaching generally, but the specific skills teachers would need to learn from their own experience? In other words, perhaps lesson study is providing teachers with skills that can transform their daily work in the classroom into further opportunities for deliberate practice – something similar to Fadde and Klein's (2010) theory of how deliberate performance might yield growth in expertise.

A primary candidate for what these skills might be are analytic skills: teachers who participate in lesson study are learning to observe and analyze practice in ways that could directly improve their ability to read a classroom situation, select an appropriate strategy, and implement the strategy effectively, all while monitoring evidence of students' thinking and learning. Specifically, teachers in lesson study are practicing:

- Careful observation and analysis of students' thinking and learning during a classroom lesson.
- Generation of cause–effect theories (what Lipsey, 1993, calls "small theories") that link a teacher's actions to students' thinking and learning during the lesson, and that may explain failures in learning.
- Generation of alternative teaching strategies that may, if one's theory is correct, lead to improved outcomes for students.
- Testing of alternative strategies, and using what is learned to revise one's own theories of classroom teaching and learning.

If these specific skills can be applied by teachers to their own teaching, then daily experiences in the classroom can become a site for deliberate practice (or deliberate performance) and thus a possible mechanism for the continuous improvement of teaching over time. Teachers become more aware of their own teaching routines, and then develop and test changes in routines that might better address the needs of their students. Through iterative application of these skills, teachers will become better at reading situations, deciding which strategy to use, and adapting the strategy to meet their specific instructional goals.

Evidence supporting this theory comes from work, reviewed above, by Kersting and colleagues in which teachers' analyses of classroom video clips are used as indicators of pedagogical content knowledge (Kersting, Givvin, Sotelo, & Stigler, 2010). In this simple paradigm, teachers are asked to watch a series of short video clips depicting authentic classroom events, and are then asked to comment on the interactions among the teacher, students, and mathematics. Teachers' analytic skills, measured in this way, have been shown to significantly predict students' learning from pre- to post-test in a variety of mathematical domains, and the effect has been shown to be mediated by an observational measure of instructional quality (Kersting et al., 2012).

Although the relationship is correlational – i.e. nothing was manipulated in the study – other research lends further support to the idea that teachers' analytic skills directly lead to improvements in instructional quality and student learning. Roth et al. (2011), for example, showed that teachers who participated in a year-long professional development program focused on the analysis of lesson videos (project STeLLa: Science Teachers Learning from Lesson Analysis) produced high quality instruction and higher levels of student learning than did a group of comparison teachers. And analysis of practice has become a key component of many teacher learning programs.

Conclusion

Teaching is a complex socio-cultural system, both hard to see and hard to change. Improving teaching requires that both the routines of teaching and the expertise of the teacher be improved. Although the aim of improving teaching routines is to enable an average teacher to produce the desired outcomes, the nature of teaching will always necessitate a certain level of adaptive expertise on the part of the teacher. Teachers must have knowledge (knowing that); they must have skill (knowing how); and they must have judgment – the ability to size up a situation, see its structure the way an expert physicist sees the structure of a physics problem, and then bring the right knowledge to bear so as to achieve the instructional goals.

The models of teaching and of the development of expertise in teaching proposed here share the key feature that both require cycles involving analysis of a problem, planned activity, and assessment of the results that then informs future practice. This approach has proven successful in improving performance in a number of important domains, and we believe that it, rather than the promulgation of a decontextualized set of best practices, holds the most promise for improving the quality of teaching and learning.

Perhaps because of the complexity and culturally nested nature of teaching, the fields of teacher education and teacher professional development have lagged in bringing principles of systems improvement and development of expertise to the task of improving teaching. Teaching presents a challenge for models of expertise – it is difficult to identify experts and the realities of teaching make it difficult to engage in deliberate practice. The concept of deliberate performance and work on improving performance in other complex fields suggest that there are promising ideas from other fields that can be usefully applied to understanding and improving the development of teachers.

References

Ball, D. L., Bass, H., Hill, H., Sleep, L., Phelps, G., & Thames, M. (2006). Knowing and using mathematics in teaching. Paper presented at the Learning Network Conference, Washington, DC, January.

Ball, D. L., & Forzani, F. M. (2007). 2007 Wallace Foundation Distinguished Lecture – What makes education research "educational"? *Educational Researcher*, 36, 529–540.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59, 398–407.

Berliner, D. C. (1986). In pursuit of the expert pedagogue. *Educational Researcher*, 15, 5–13.

Berliner, D. C. (1990). What's all the fuss about instructional time? In M. Ben-Peretz & R. Bromme (eds.), *The nature of time in schools: Theoretical concepts, practitioner perceptions* (pp. 3–35). New York: Teachers College.

Birman, B. F., Le Floch, K. C., Klekotka, A., Ludwig, M., Taylor, J., Walters, K., . . . & O'Day, J. (2007). *State and local implementation of the No Child Left Behind Act. Volume I, Title I: Teacher quality under NCLB: Interim report*. US Department of Education (ERIC Document Reproduction Service No. ED497970).

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). New York: Worth.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80, 139–144.

Bond, L., Smith, T., Baker, W., & Hattie, J. A. (2000). *The certification system of the National Board for Professional Teaching Standards: A construct and consequential validity study*. Greensboro, NC: Center for Educational Research and Evaluation, The University of North Carolina at Greensboro.

Bransford, J. D., Goldman, S. R., & Vye, N. J. (1991). Making a difference in people's ability to think: Reflections on a decade of work and some hopes for the future. In R. J. Sternberg & L. Okagaki (eds.), *Influences on children* (pp. 147–80). Hillsdale, NJ: Erlbaum.

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Cambridge, MA: Harvard Education Press.

Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning*, 2, 67–90.

Cavalluzzo, L. C. (2004). *Is National Board certification an effective signal of teacher quality?* Alexandria, VA: CNA Corporation Working Paper.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, 26, 673–682.

Cohen, D. (2011). *Teaching and its predicaments*. Cambridge, MA: Harvard University Press.

Cohen, D., & Hill, H. (2000). Instructional policy and classroom performance: The mathematics reform in California. *The Teachers College Record*, 102, 294–343.

Correa, C. A., Perry, M., Sims, L., Miller, K. F., & Fang, G. (2008). Connected and culturally embedded beliefs: Chinese and U.S. teachers talk about how their students best learn mathematics. *Teaching and Teacher Education*, 24, 140–153.

Cortina, K. S., Miller, K. F., McKenzie, R., & Epstein, A. (2015). Where low and high inference data converge: Validation of CLASS assessment of mathematics instruction using mobile eye tracking with teachers. *International Journal of Science and Mathematics Education*, 13, 389–403.

Csikszentmihalyi, M., Rathunde, K., & Whalen, S. (1997). *Talented teenagers: The roots of success and failure*. Cambridge University Press.

Cuban, L. (1990). Reforming again, again, and again. *Educational Researcher*, 19, 3–13.

Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession*. Washington, DC: National Staff Development Council.

de Groot, A. D. (1965). *Thought and choice in chess*. The Hague: Mouton.

de Groot, A. D. (1966). Perception and memory versus thought: Some old ideas and recent findings. In B. Kleinmuntz (ed.), *Problem solving*. New York: John Wiley.

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, 29, 153–170.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37, 32–64.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, R. R. Hoffman, & P. J. Feltovich (eds.), *The Cambridge handbook of expertise and expert performance* (pp. 685–706). Cambridge University Press.

Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: A general overview. *Academic Emergency Medicine: Official Journal of the Society for Academic Emergency Medicine*, 15, 988–994.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100, 363–406.

Fadde, P. J., & Klein, G. A. (2010). Deliberate performance: Accelerating expertise in natural settings. *Performance Improvement*, 49, 5–14.

Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly*, 48, 94–118.

Feldon, D. F. (2007). The implications of research on expertise for curriculum and pedagogy. *Educational Psychology Review*, 19, 91–110.

Fisher, C. W., Berliner, D. C., Fully, N. N., Marliave, R. S., Cahen, L. S., & Dishaw, M. M. (1980). Teaching behaviors, academic learning time and student achievement: An overview. In C. Denham & A. Lieberman (eds.), *Time to learn* (pp. 7–32). Washington, DC: US Department. of Education, National Institute of Education, Program on Teaching and Learning.

Gabor, A. (1990). *The man who discovered quality: How W. Edwards Deming brought the quality revolution to America*. New York: Penguin Books.

Gallimore, R. (1996). Classrooms are just another cultural activity. In D. Speece & B. Keogh (eds.), *Research on classroom ecologies: Implications for children with learning disability* (pp. 229–250). Mahwah, NJ: Erlbaum.

Gallimore, R., & Tharp, R. (2004). What a coach can teach a teacher, 1975–2004: Reflections and reanalysis of John Wooden's teaching practices. *The Sport Psychologist*, 18, 119–137.

Gawande, A. (2007). *Better: A surgeon's notes on performance*. New York: Metropolitan Books.

Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National board certification as a signal of effective teaching. *Review of Economics and Statistics*, 89, 134–150.

Hanushek, E. A. (2003). The failure of input-based schooling policies. *Economic Journal*, 113, F64–F98.

Harris, D. N., & Sass, T. R. (2009). The effects of NBPTS-certified teachers on student achievement. *Journal of Policy Analysis and Management*, 28, 55–80.

Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, & K. Hakuta (eds.), *Child development and education in Japan* (pp. 262–272). New York: W. H. Freeman.

Hill, H. C., & Ball, D. L. (2004). Learning mathematics for teaching: Results from California's mathematics professional development institutes. *Journal for Research in Mathematics Education*, 35, 330–351.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42, 371–406.

Hoetker, J., & Ahlbrand Jr, W. P. (1969). The persistence of the recitation. *American Educational Research Journal*, 6, 145–167.

Hoffman, R. R. (1998). How can expertise be defined? Implications of research from cognitive psychology. In R. Williams, W. Faulkner, & J. Fleck (eds.), *Exploring expertise* (pp. 81–100). New York: Macmillan.

Hogan, T., & Rabinowitz, M. (2009). Teacher expertise and the development of a problem representation. *Educational Psychology*, 29, 153–169.

Ingersoll, R. M., & May, H. (2012). The magnitude, destinations, and determinants of mathematics and science teacher turnover. *Educational Evaluation and Policy Analysis*, 34, 435–464.

Jackson, C. K. (2012). Recruiting, retaining, and creating quality teachers. *Nordic Economic Policy Review*, 1, 61–105.

Jackson, C. K., Rockoff, J. E., & Staiger, D. O. (2014). Teacher effects and teacher-related policies. *Annual Review of Economics*, 6, 801–825.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment*. Seattle, WA: MET Project of the Bill and Melinda Gates Foundation.

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: MET Project of the Bill and Melinda Gates Foundation.

Kellman, P. J., & Garrigan, P. (2009). Perceptual learning and human expertise. *Physics of Life Reviews*, 6, 53–84.

Kenney, C. (2008). *The best practice: How the new quality movement is transforming medicine*. New York: Public Affairs.

Kersting, N. B., Givvin, K. B., Sotelo, F. L., & Stigler, J. W. (2010). Teachers' analyses of classroom video predict student learning of mathematics: Further explorations of a novel measure of teacher knowledge. *Journal of Teacher Education*, 61, 172–181.

Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, 49, 568–589.

Kersting, N. B., Sherin, B. L., & Stigler, J. W. (2014). Automated scoring of teachers' open-ended responses to video prompts bringing the classroom-video-analysis assessment to scale. *Educational and Psychological Measurement*, 74, 950–974.

Kersting, N. B., Sutton, T., Kalinec-Craig, C., Stoehr, K. J., Heshmati, S., Lozano, G., & Stigler, J. W. (2016). Further exploration of the classroom video analysis (CVA) instrument as a measure of usable knowledge for teaching mathematics: Taking a knowledge system perspective. *ZDM*, 48, 97–109.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41, 75–86.

Lampert, M. (2003). *Teaching problems and the problems of teaching*. New Haven, CT: Yale University Press.

Lampert, M., Franke, M. L., Kazemi, E., Ghousseini, H., Turrou, A. C., Beasley, H., . . . & Crowe, K. (2013). Keeping it complex using rehearsals to support novice teacher learning of ambitious teaching. *Journal of Teacher Education*, 64, 226–243.

Langley, G. J., Moen, R., Nolan, K. M., Nolan, T. W., Norman, C. L., & Provost, L. P. (2009). *The improvement guide: A practical approach to enhancing organizational performance* (2nd edn.). San Francisco: Jossey-Bass.

Lewis, C., Perry, R., & Hurd, J. (2004). A deeper look at lesson study. *Educational Leadership*, 61, 19–22.

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

Lewis, C., & Tsuchida, I. (1998). A lesson is like a swiftly flowing river. *American Educator*, 22, 12–17.

Lipsey, M. W. (1993). Theory as method: Small theories of treatments. *New Directions for Evaluation*, 57, 5–38.

Lortie, D. C. (1975). *Schoolteacher: A sociological study*. University of Chicago Press.

Miller, K. F. (2011). Learning from the experience of others: What education can learn from video-based research in other fields. In M. Sherin, V. Jacobs, & R. Phillip (eds.), *Mathematics teacher noticing: Seeing through teachers' eyes* (pp. 51–65). New York: Routledge.

Morris, A. K., & Hiebert, J. (2011). Creating shared instructional products: An alternative approach to improving teaching. *Educational Researcher*, 40, 5–14.

Mueller, F., & Dyerson, R. (1999). Expert humans or expert organizations? *Organization Studies*, 20, 225–256.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26, 237–257.

Resnick, D., & Resnick, L. (1977). The nature of literacy: An historical exploration. *Harvard Educational Review*, 47, 370–385.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73, 417–458.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94, 247–252.

Roth, K. J., Garnier, H. E., Chen, C., Lemmens, M., Schwille, K., & Wickler, N. I. (2011). Videobased lesson analysis: Effective science PD for teacher and student learning. *Journal of Research in Science Teaching*, 48, 117–148.

Rother, M. (2009). *Toyota Kata: Managing people for improvement, adaptiveness and superior results*. Boston: McGraw-Hill.

Sabers, D. S., Cushing, K. S., & Berliner, D. C. (1991). Differences among teachers in a task characterized by simultaneity, multidimensional, and immediacy. *American Educational Research Journal*, 28, 63–88.

Saunders, W. M., Goldenberg, C. N., & Gallimore, R. (2009). Increasing achievement by focusing grade-level teams on improving classroom learning: A prospective, quasi-experimental study of Title I schools. *American Educational Research Journal*, 46, 1006–1033.

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, 103, 759–775.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57, 1–23.

Simon, N. S., & Johnson, S. M. (2015). Teacher turnover in high-poverty schools: What we know and can do. *Teachers College Record* 117, 1–36.

Skemp, R. R. (1987). *The psychology of learning mathematics*. London: Routledge.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: Simon & Schuster.

Stigler, J. W., & Hiebert, J. (2004). Improving mathematics teaching. *Educational Leadership*, 61, 12–17.

Thorndike, E. L. (1906). *Principles of learning*. New York: A. G. Seiler.

UNESCO Institute for Statistics (2015). *A Teacher for Every Child: Projecting Global Teacher Needs from 2015 to 2030*. October, Number 27, page 1. Montreal: UNESCO. Institute for Statistics. http://uis.unesco.org/sites/default/files/documents/fs27-a-teacher-for-every-child-projecting-global-teacher-needs-from-2015-to-2030-en.pdf.

US Department of Education (2016). *Digest of Education Statistics*, 2014 (NCES 2016–006).

Wang, Z., Miller, K. F., & Cortina, K. S. (2013). Using the LENA in teacher training: Promoting student involvement through automated feedback. *Unterrichtswissenschaft*, 4, 290–305.

Whitehead, A. N. (1929). *The aims of education and other essays*. New York: Macmillan.

Williams, B. A. (1998). *Thought and action: John Dewey at the University of Michigan*. Ann Arbor: University of Michigan Bentley Historical Library Bulletin #44.

Yoon, K. S., Duncan, T., Lee, S. W. Y., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement*. Issues and Answers. REL 2007-No. 033. Austin, TX: Regional Educational Laboratory Southwest.