education policy analysis archives

A peer-reviewed, independent, open access, multilingual journal

Arizona State University

Volume 18 Number 19 August 20th, 2010 ISSN 1068–2341

Using Assessments for Instructional Improvement: A Literature Review[1]

Viki M. Young
SRI International

Debbie H. Kim
Northwestern University

Citation: Young, V. M., & Kim, D. H. (2010). Using assessments for instructional improvement: a literature review. *Education Policy Analysis Archives, 18* (19). Retrieved [date], from http://epaa.asu.edu/ojs/article/view/809

Abstract: The current educational reform policy discourse takes for granted the central role of using data to improve instruction. Yet whether and how data inform instruction depends on teachers' assessment practices, the data that are relevant and useful to them, the data they typically have access to, and their content and pedagogical knowledge. Moreover, when one considers teachers' organizational contexts, it is clear that school leadership and support for using data, capacity-building strategies, and the norms of adult learning and collaboration circumscribe opportunities to examine relevant data and to improve instructional practice in response. This literature review examines teacher as well as organizational practices and characteristics as they pertain to formative uses of assessment. We identify

[1] Accepted under the editorship of Sherman Dorn. This article was supported in part by the Center on Continuous Instructional Improvement at the Consortium for Policy Research in Education (CPRE) and funded by the William and Flora Hewlett Foundation. All views expressed are solely those of the authors and do not necessarily represent the views of CPRE or the Hewlett Foundation.

opportunities for important research to illuminate how and under what conditions teachers and schools as organizations can use data to inform instruction.
Keywords: educational reform; formative assessment; organizational theory.

El uso de evaluaciones para la mejora de la enseñanza: Una revisión bibliográfica
Resumen: Los discursos recientes sobre las reformas de las políticas educativas dan por sentado el papel central que tendría la utilización de "datos" para mejorar la instrucción. Sin embargo, cómo los datos podrían ayudar a mejorar la enseñanza depende también de otros factores, tales como las prácticas de evaluación de los profesores, que datos son relevantes y útiles para ellos, que los datos sean usualmente accesibles, y el conocimiento sobre contenidos y pedagogía de los profesores. Además, al considerar los contextos de trabajo de los docentes, es evidente que el liderazgo en las escuelas, el apoyo para la utilización de los datos, las estrategias institucionales de capacitación y los estándares de aprendizaje de los adultos y de colaboración, limitan las oportunidades de examinar los datos pertinentes y mejorar las prácticas docentes. Esta revisión de la literatura investiga las prácticas y características docentes y organizacionales que se relacionan con usos formativos de la evaluación. Identificamos oportunidades para investigaciones relevantes que pueden aclarar cómo y bajo qué condiciones los maestros y las escuelas pueden utilizar datos de las evaluaciones para mejorar la enseñanza.
Palabras-clave: reforma educativa; evaluación formativa; teoría organizativa.

A utilização das avaliações para a melhoria do ensino: revisão da literatura
Resumo: Os discursos recentes sobre a reforma da políticas educativas pressupõem um papel central dos "dados" na melhoria do ensino. No entanto, a forma como os "dados" podem ajudar a melhorar o ensino também depende de outros fatores, como as práticas de avaliação dos professores, quais dados são relevantes e úteis para eles, que os dados sejam geralmente acessíveis, e conhecimento, conteúdo e pedagogia dos professores. Além disso, considerando os contextos de trabalho dos professores é claro que a liderança nas escolas, o apoio à utilização de dados institucionais, as estratégias de formação e as normas de aprendizagem e colaboração para os adultos limitam as oportunidades de analisar dados relevantes e melhorar as práticas docentes. Esta revisão da literatura investiga características e práticas educacionais e organizacionais relacionadas com usos formativos da avaliação. Identificamos oportunidades para pesquisas relevantes que podem esclarecer como e em que condições os professores e as escolas podem utilizar os dados da avaliação para melhorar o ensino.
Palavras-chave: reforma avaliação educativa, formativa, teoria organizacional.

Introduction

The No Child Left Behind Act (NCLB) enshrined the logic of data-driven decision-making in education. Data-driven decision-making—a manufacturing principle ushered in by the total quality management (TQM) movement beginning in the 1980s (e.g., Deming, 1982)—later influenced service functions and industries and eventually the public sector in the 1990s.[2] Paying homage to data use, NCLB uses phrases such as "evidence-based decisions" and "scientifically based

[2] See, for example, a handbook series for which Joseph Juran is editor-in-chief. It applies quality management to areas as diverse as customer service (Fuchs, 1999), decision-making (Redman, 1999), and government (Gore, 1999). See also Juran (1992) on TQM in goods and services, and Ingram, Louis, and Schroeder (2004), Leonard (1996), and Schmoker and Wilson (1993) on applying Deming's principles to education.

research" 111 times, according to one count (Mann & Shakeshaft, 2003). The importance of using "data" is now taken-for-granted as an essential strategy for educational improvement. School district superintendents surveyed in summer 2005 consistently reported data use as the most important strategy for guiding decisions to improve student achievement (*Education Week*, 2005, cited in Coburn & Talbert, 2006).

At least during the accountability policy era stemming from standards-based reform, schools' and districts' initial forays into data-driven practices have relied on standardized test scores; those scores constitute the most prevalent data and the data that predominate, given state and federal accountability measures. However, educators are quick to note that annual standardized test scores have only limited usefulness in the classroom. The results are not timely, often not available until after students have moved on to another teacher; the test items may not be aligned with the curriculum; and because annual results are too infrequent, teachers' own experiences with the students quickly supersede the information provided by those test scores.

To respond to these criticisms, district administrators and school leaders have begun to implement timelier and more curriculum-aligned assessment programs. They intend for results from the assessments to inform teachers' and schools' instructional decisions throughout the school year. *Formative assessment* are the watch-words in the current policy environment. Black and Wiliam's (1998) literature review underscores the potential value of using assessments for formative purposes. They report a meta-analysis that obtained a mean effect size of 0.92 for studies in which teachers pursued explicit procedures for reviewing data and determining next steps based on the analysis, compared with a mean effect size of 0.42 for studies in which teachers used data at their discretion (Fuchs & Fuchs, 1986, cited in Black & Wiliam, 1998). In their meta-analysis of the effects of instructional cues, student participation, and corrective feedback, Lysakowski and Walberg (1982) also report an average effect size of almost a standard deviation (0.97) for the 94 studies included in their analysis. Among those studies, the 20 studies focused on corrective feedback resulted in a mean effect size of 0.94.[3]

Despite such promise, knowledge of how teachers use various forms of assessments for instructional improvement and of the organizational conditions that support their use remains limited. This literature review takes stock of what the field knows and to offer suggestions for where the field needs to go. First, we ask what we know about how formative assessment data and other ways of gathering evidence about students' progress influence teachers' practice. Second, we ask what is known about the policy, school, and classroom conditions that increase (or decrease) the influence of formative assessment data and other data about student progress on teachers' instructional decisions. Finally, we ask what we know about the use of formative assessment data at the school and district levels.

Methods

We used several electronic databases, including ERIC and JSTOR, to search for relevant literature. Initial search term sets included [*teacher*, *formative*, *assessment*]; [*formative*, *assessment*, *practice*]; [*classroom*, *assessment*]; [*formative*, *assessment*, *instruction*]; and various author and project names known to the authors or cited in other articles. After culling the returned articles for relevance, we identified additional articles by pursuing references cited in the articles we reviewed. The full bibliography can be found at the end of the text.

[3] One of the 20 studies had a mean effect size between -1.49 and -1, while the other 19 studies had positive effect sizes ranging from intervals of 0–0.49 to 3–3.49.

The articles we selected for this literature review are mainly based on empirical research. Many of the articles we did not pursue from the returned list of articles were not empirical studies. We reviewed some additional papers that explicate the rationale behind data-driven decision-making for contextual background. Most of the research-based articles describe case studies that used a variety of data sources and that were often conducted as part of action research (studies by participants on their own reform efforts) or research on particular interventions. Some of the research-based articles also report on surveys of teachers' assessment practices. For each research-based article, we prepared a cover sheet that summarizes the article in terms of its overall design, main findings, and conclusions. The cover sheets also capture our evaluation of research quality in terms of validity, reliability, and generalizability.

The rest of this review consists of five major sections. We discuss definitional issues next, followed by a general discussion of teachers' assessment practices. The third section focuses on a series of studies designed to improve teachers' assessment techniques and the influence of such professional development on instruction, as well as school and district uses of data. In the fourth section, we consider the organizational conditions that are described in the numerous articles on teachers' assessment practices. The closing section considers the potential areas for future research emerging from the review.

Formative Assessments or Formative Uses?

Formative Assessments and Other Terms

The majority of articles included in this review were published between 1980 and 2008. These articles describe formative uses of a broad range of assessments and use different but related terms. Each term is situated in a particular policy era. In articles from (roughly) the 1980s, *classroom assessment* generally refers to teachers' assessment practices as distinct from any testing mandated by the district or state. *Performance assessment* gained visibility as systemic reform—more commonly referred to as standards-based reform—ascended in the late 1980s through the mid-1990s. At the time, assessment reformers argued that performance assessments, rather than traditional paper-based tests with closed-ended items, would better reflect what students need to do to demonstrate that they had met content and performance standards. *Data* generally refers to annual large-scale standardized test scores. *Data* and *formative assessment* are the most recent additions to the data-driven decision-making dictionary; they first became prevalent in education policy language in the late 1990s.

Corresponding to the policy eras, the articles reviewed also use different units of analyses. Literature on classroom assessment generally focuses on individual teachers. Performance-based assessment articles tend to discuss systems (e.g., how to implement large-scale performance-based assessments). More recent literature uses an organizational perspective to investigate how schools and teachers use data, generally in the form of standardized test scores, as well as other types of assessments and information. Only with the latter set of articles do organizational conditions as they pertain to *data-driven* or *data-informed* decision-making come to the forefront in the research.

In conducting this literature review, we quickly recognized that *formative assessment* is imprecise and analytically inadequate. Varying characteristics are implicitly or explicitly ascribed to formative assessments; for example, being curriculum-embedded, occurring during the school year rather than at the end, being less standardized, or entailing a performance. Wininger and Norman (2005) studied 20 of the most commonly used educational psychology textbooks and found that the definition, uses, and the stated importance of formative assessment, along with associated terminology, varied considerably among the texts. The functions of formative assessment that the

textbooks described ranged from guiding instruction to providing feedback to students. Many of the texts used the term *formative evaluation* to describe formative assessment, whereas others used *formative assessment* or *informal assessment* (Wininger & Norman, 2005). Perhaps not surprisingly, another author reporting on teachers' confusion between summative and formative assessments pointed out that "to avoid conflict and to clarify misconceptions, teachers would value clear guidance about what constitutes formative assessment" (Neesom, 2000, p. 7). Brookhart (2007), in her review of the literature, took an expansive definition to include assessments serving to further student learning, that "formative and summative assessment need not be mutually exclusive" (p. 45).

Using the term *formative assessment* is fraught with confusion. Are assessments formative because policymakers intend them to be formative or because they share technical properties that improve their applicability to teachers' work? If we pursue a definition of formative assessment that is rooted in the format of the assessment or in the intentions of policymakers—but not necessarily the intended users, teachers and principals—how would we analyze situations in which teachers do not use formative assessments as intended? More importantly, what would we call assessments that teachers use to inform instruction but that policymakers do not intend to be formative (e.g., year-end standardized tests). And because assessments range in their degree of formality, where should boundaries be drawn to include or exclude certain types of informal assessments? Shifting the perspective to that of a practitioner (i.e., the types of data that are formative for teachers' purposes) provides more analytic purchase.

Formative Uses of Assessment

Assessment practices in the classroom serve many functions. They aid in planning instruction, shaping instruction as it unfolds, gauging student achievement, and evaluating curriculum (Herman & Dorr-Bremme, 1983; Shavelson & Stern, 1981; Stiggins, 1991). When teachers assess their students, they often attempt to gauge their knowledge and skills acquisition, as well as different social factors, including student participation, interaction, and attendance (Cizek, Fitzgerald, & Rachor, 1995/1996; Herman & Dorr-Bremme, 1983). These data contribute to teachers' decisions in planning lessons, grouping students for instruction, diagnosing the strengths and weaknesses of individual students, and reassigning students to different instructional groups throughout the year (Herman & Dorr-Bremme, 1983; Stiggins & Bridgeford, 1985). The purposes that assessment data are supposed to serve and the occasions for using the data are not necessarily defined *a priori*; teachers choose how and when to use the information gathered through assessments. They certainly have structured evaluative events, e.g., end-of-unit tests; equally certainly, they make use of a range of data and impressions extemporaneously as a student struggles with a particular lesson.

Because assessments may differ substantially but still be considered formative, Wiliam and Black (1996) and more explicitly Wiliam and Leahy (2006) argue that the term *formative* should describe practitioners' uses of assessments rather than the assessments themselves. This definition recognizes that different types of assessments—and indeed other forms of data—may be formative if teachers appropriately apply them to inform instructional choices, regardless of administrators' or policymakers' intentions for those tests. Focusing on formative activities also takes the practitioner's perspective into consideration; both the helpful aspects of assessments and their limitations emerge as a result of her goals for improvement.

This practitioner approach, however, begs the question of the assessment practices that teachers use and the kinds of decisions that are driven by assessment results. Torrance and Pryor (2001) identify two approaches to formative uses of assessment that teachers might take, convergent and divergent. The purpose of convergent assessment "is to find out *if* the learner knows,

understands, or can do a predetermined thing. It is characterized by detailed planning, and is generally accomplished by closed or pseudo-open questioning and tasks" (p. 616). Divergent assessment "emphasizes the learner's understanding rather than the agenda of the assessor. … to discover *what* the learner knows, understands and can do. It is characterized by less detailed planning, where open questioning and tasks are of more relevance" (p. 617; emphasis in original). Hattie and Timperley (2007) further offer three questions—"Where am I going? How am I going? and Where to next?" (p. 88)—that help define the purpose for formative assessment.

Shavelson (2003) also offers a definition of formative assessment and phases of instruction. He describes three types of formative assessment: "on-the-fly," "planned-for-interaction," and formal and embedded in curriculum. "On-the-fly" formative assessments occur when "teachable moments" take place in the classroom. "Identification of these moments is initially intuitive and then later based on cumulative wisdom of practice" (Shavelson, 2003, p. 4). Formative assessments that fall under "planned-for-interaction" are deliberate and involve questioning designed to discern and improve students' knowledge acquisition. Many of the studies we review below follow this latter model of assessment. Formal, curriculum-embedded assessments are intended to create "teachable moments." These assessments can be built into content units and are meant to illuminate students' progress toward subgoals that cumulatively lead to achieving the overall learning goals for a given unit (Shavelson, 2003).

Teachers, however, may not view these formative uses of assessments as integral to their instruction; or, if they do, the general lack of training associated with assessments is likely to result in a struggle to do it well for all but a few individuals who might have a natural orientation towards reflection and evaluation. Wininger and Norman (2005) noted practitioners' inconsistent understanding of the role of assessments in instructional decision-making. Neesom (2000) found that teachers consider formative assessment as beyond their normal instructional obligations, and Daws and Singh (1996) indicated that few teachers explicitly use assessments formatively as part of their instructional practice, despite their general awareness of the assessments' potential advantages. In their school survey, Daws and Singh (1996) found that teachers perceive assessment as primarily summative and fail to leverage assessment activities for formative purposes. For example, teachers' reasons for marking student work typically entail assigning grades rather than using that work to identify appropriate activities for subsequent instruction. Similarly, teachers keep pupil folders not so much for assessment purposes but "as a bureaucratic exercise to satisfy what is perceived by teachers to be an external accountability requirement" (Daws & Singh, 1996, p. 97).

If teachers traditionally have not used assessment results to inform instruction-related decisions, what are teachers' assessment practices in general? Have teachers received training in using assessments formatively? If so, what conditions facilitate such uses?

What Are Teachers' Assessment Practices?

More than 50 years ago, teachers reported that their training in testing and measurement was insufficient (Noll, 1955), and that sentiment persists today. In 1985, nearly three-quarters of surveyed teachers expressed concerns about their self-created tests, with their most common concern being the need for improving the tests (Stiggins & Bridgeford, 1985). In 1993, approximately one-third of surveyed teachers "indicated [that] they were very interested in becoming more proficient in interpreting test scores and student assessment in general" (Impara, Plake, & Fager, 1993, p. 115). In both the 1999–2000 and 2003–04 nationally representative Schools and Staffing Survey (SASS), approximately one-third of teachers with fewer than 5 years of experience reported that, in their first year of teaching, they were either "not at all" or only "somewhat"

prepared to assess students.[4] And in 2007, more than 80% of surveyed educators agreed or strongly agreed that "improving my ability to use data will help me become a better educational professional" (Wayman, Cho, & Johnston, 2007, p. 21). In short, across multiple studies conducted over decades, a significant proportion of teachers report uncertainty or a desire for improvement in their assessment practices. With such seemingly shaky foundations, what are teachers' assessment practices in the classroom?

Nature of Teachers' Assessment Practices

Teachers incorporate multiple types of assessment into their instruction, and they do not rely on a single source of information. Assessment types range from formal testing techniques (e.g., teacher-made tests, standardized tests, and homework) to more informal, "on-the-spot" assessments (e.g., student behavior, perceived student effort, teacher expectation, informal observation, and interaction cues) (Cizek et al., 1995/1996; Fleming & Chambers, 1983; Herman & Dorr-Bremme, 1983; McMillan, 2002; Stiggins & Bridgeford, 1985). Teachers commonly use formal assessments to measure content knowledge, typically in terms of factual recall and other rote learning achievements (Cizek et al., 1995/1996; Fleming & Chambers, 1983); however, teachers also employ a combination of formal assessments and informal, observational assessments (Cizek et al., 1995/1996; McMillan, 2002). Teachers typically assess to assign grades, which constrains the types of assessments they use (Brookhart, 2007). Moreover, teachers mediate test results with more impressionistic information (Cizek et al., 1995/1996; Shavelson & Stern, 1981). For example, when asked about sources of information they considered in assigning final grades, a large majority of teachers reported "formal achievement measures (e.g., tests, assignments, etc.)" and/or "other informal measures (e.g., impressions of effort, conduct, teamwork, etc.)" (Cizek et al., 1995/1996, p. 167). Teachers place great value on the informal information about student progress that they glean from their everyday classroom interactions. Even in the context of district support and expectations for using interim assessments to make instructional decisions, they interpret the formal assessment results in light of what they know about the students and are seldom surprised by the scores (Goertz, Oláh, & Riggan, 2009).

A relatively recent study found that virtually all teachers who were surveyed as part of a K-12 comprehensive school reform program valued internally developed (school-based) assessments (Supovitz & Klein, 2003). Between 94% and 97% of staff rated student portfolios, Running Records (oral reading assessment), and open-ended assessments as useful for instruction, with two-thirds to three-quarters of teachers rating them as "highly useful." Over three-quarters also reported district and state standardized tests as useful; however, slightly more than half of all teachers surveyed rated them "somewhat useful" rather than "highly useful" (Supovitz & Klein, 2003, p. 13).

Another study examining teachers' assessment practices found that the vast majority of respondents "comfortably" used spontaneous performance assessment.[5] Perhaps because of the nature of performance assessment, fewer than half of the assessments had written criteria (Stiggins & Bridgeford, 1985). Teachers tended to rely heavily on their own mental record-keeping to store

---

[4] 31.3% in 1999–2000 and 34.3% in 2003–04. SASS is administered by the National Center for Education Statistics. http://nces.ed.gov/surveys/sass/

[5] Stiggins and Bridgeford (1985) define performance assessment as "the observation and rating of student behavior and products in contexts where students actually demonstrate proficiency" (p. 273). They define spontaneous performance assessment as assessment that "arises spontaneously from the naturally occurring classroom environment and leads the teacher to a judgment about an individual student's level of development" (p. 273). Spontaneous performance assessment corresponds to Shavelson's (2003) "on-the-fly" assessment.

and retrieve information while assessing their students (Stiggins & Bridgeford, 1985). Also, fewer than half of the respondents considered multiple performance observations before making a judgment on any given performance assessment, and even fewer rated performances without knowing the student's identity or checking their own ratings against other test scores (Stiggins & Bridgeford, 1985). In part, these results may obtain because teachers are seeking information that objective tests do not necessarily provide (e.g., insights into students' procedural knowledge). Or, for instance, to proceed to the next part of a lesson, teachers need to know whether students have grasped enough of the concept. Thus for the vast majority of teachers in this study who use spontaneous performance assessments, those assessments do not appear to be systematic, criterion-driven, and cumulative. However, teachers' interactions with students over time contribute to impressions of individual students. Herman and Dorr-Bremme (1983) argue that teachers "accord the highest importance to their own observations of students' work and to their own clinical judgment" (p. 12).

A majority of teachers report that they develop their own tests, quizzes, and examinations (Cizek et al., 1995/1996; Impara et al., 1993; McMillan, 2002). They create their own tests in a majority of instances; commercial publishers provide the remainder (Cizek et al., 1995/1996). The great majority of teachers cite the tests they develop as crucial or important to their instruction- and grading-related decisions (Herman & Dorr-Bremme, 1983; Impara et al., 1993). Notwithstanding their routine design of classroom tests, teachers express interest in becoming more proficient in assessment (Impara et al., 1993).

As is the case with objective tests more generally, teacher-made tests tend to assess students' low-level recall of declarative knowledge rather than critical thinking or ability skills (McMunn, McColskey, & Butler, 2003–04). Teachers tend to use short-answer tests over essay questions, matching items over multiple-choice or true-false questions, and "more test questions to sample knowledge of facts than any of the other behavioral categories studied" (Fleming & Chambers, 1983, p. 32). (Those tendencies, however, may have changed in the decades since the study was conducted.) Even when teachers' written questions test students' ability to recall inferences, they do not demand comparative or evaluative responses from students (Stiggins, Griswold, & Wikelund, 1989). Overall, teacher-made tests appear to measure content over process and leave little room "to test behaviors that can be classified as ability to make applications" (Fleming & Chambers, 1983, p. 32).

To supplement their formal assessments, teachers use published tests, including tests provided by curriculum publishers and standardized tests. Over the course of elementary, middle, and high school, the use of published tests drops as the grade level increases (Cizek et al., 1995/1996; Herman & Dorr-Bremme, 1983; Stiggins & Bridgeford, 1985). For the major tests that teachers use for assigning grades, they emphasize self-developed and publisher-provided tests fairly evenly. For high school teachers, however, approximately three-quarters of the major tests they use are self-developed, and approximately one-quarter are publisher-provided (Cizek et al., 1996/1996). Stiggins and Bridgeford (1985) suggest that the increase in teacher-made tests at higher grade levels may be due to teachers' perceived need to tailor tests to unique classroom conditions at those higher levels. Moreover, a desire for greater "quality control" may lead teachers to use assessment measures that they believe are more accurate in grading and judging their students—that is, tests they create and modify themselves (Stiggins & Bridgeford, 1985). A substantial percentage of teachers also believe that the quality of available tests is not satisfactory (Herman & Dorr-Bremme, 1983), and an even larger proportion of teachers do not feel that standardized tests can be used to enhance instruction (Impara et al., 1993). This dissatisfaction with published tests, including those provided in textbooks, calls into question teachers' perceptions of the value of commercially available interim

and benchmarks tests, as districts increasingly adopt more frequent testing under the logic of data-driven decision-making.

Teachers tend to test frequently on their own. According to one survey, approximately three-quarters of teachers test their students at least once a week through minor assignments or major tests that count toward a grade (Cizek et al., 1995/1996). These numbers describe the occurrence of assessments and do not include the occurrences of spontaneous observational assessments that are integral to overall assessment practices. To the extent that informal assessments are spontaneous, ongoing, and integrated into questioning techniques and everyday observations, teachers' frequency in using them may be harder to gauge. The relatively sparse information available about teachers' informal assessment practices, juxtaposed against the value that teachers place on them, signals a gap in research that may be worth pursuing. In particular, with the emphasis of NCLB on improving large-scale assessment performance, the gap between what teachers find useful and what the policy environment prizes may be ever widening.

Expertise for Formative Assessment Practices

Using assessments formatively in the classroom is not a beginner's skill. It takes a range of foundational content knowledge, pedagogical understanding, instructional skill, and classroom management to effectively use or implement formative assessment practices.

Teachers' knowledge of student learning and subject matter. Teachers' assessment practices tend to reflect their understanding of students' learning processes and the content they teach. Learning and thus instruction are not linear processes. Although content knowledge may conceptually build in a sequential fashion, not all students grasp content in the same way. Learners' insights towards achieving a deeper understanding of a formal body of knowledge can be sporadic and disjointed. For assessments to be formative—that is, for assessment to be instructionally relevant and the basis for instructional change—teachers need to be able to identify appropriate assessment data (e.g., classroom discourse, observations, tests), use those data to gauge students' emerging conceptions and individual learning trajectories, and then adjust instruction accordingly. Determining students' emerging ideas aids the teacher in knowing what parts of previous instruction need additional emphasis, and how to scaffold and tailor subsequent instructional activities. This approach also allows the teacher to gauge the strength of students' developing content knowledge.

Studies have found that teachers who have strong content knowledge can flexibly adapt to a student's place in his or her knowledge acquisition trajectory (Aschbacher & Alonzo, 2004; Duschl & Gitomer, 1997; Fennema, Franke, Carpenter, & Carey, 1993). Teachers with a strong grasp of the content they are teaching are also more adept at considering their students' learning in direct relation to the content rather than in general development terms (Johnston, Afflerbach, & Weiss, 1993). When a teacher's knowledge of subject matter is both deep and flexible, she can break down concepts, find different entry points for different students, and repackage topics to match students' apparent understanding and misconceptions as evidenced in their work, oral responses, or other assessments. In the Cognitively Guided Instruction project, researchers sought to understand the impact of teachers' knowledge of children's thinking on student learning. They also explored how teachers used their knowledge about children's mathematical thinking while making instructional decisions (Fennema et al., 1993). The study found that exemplary teachers used their knowledge of problem types to broaden the curriculum and to tailor instruction to students. Those teachers did not base their decisions on a formal hierarchy of mathematical concepts; rather, they were able to disassemble and reassemble their content knowledge on the basis of student needs (Fennema et al., 1993). In science, teachers' improved understanding of the concepts in particular lessons helped them use science notebooks to assess students' understanding and provide feedback to them

(Aschbacher & Alonzo, 2004). The researchers concluded that the value of the science notebooks as an assessment tool depended on the strength of the teachers' science content knowledge (Aschbacher & Alonzo, 2004).

Another study found that teachers who focus on students' conceptual understanding in their assessment practices was related to teachers' diagnostic and analytic abilities (Goertz et al., 2009). Those teachers with an orientation towards students' conceptual understanding also tended to respond to assessment results with instructional rather than organizational changes. That is, they might have provided additional ways of representing mathematical concepts or tried to tap into students' prior knowledge, as opposed to using the assessments results to determine which subjects to reteach, how to group students, or identify specific students for additional supports (Goertz et al., 2009). As Stiggins (1991) argues, teachers must "possess (a) a clear and highly differentiated vision of understanding of the achievement target to be attained by students and (b) a thorough understanding of the full range of assessment alternatives available to assess the target of interest" to engage in sound assessment practices (p. 8). Similarly, the notion of learning progressions—where teachers understand the building blocks and the sequencing that are necessary for students to master specific concepts or learning objectives—can undergird the points at which teachers assess their students' progress, as well as the content or skills appropriate along that progression (Heritage, 2008). Others have emphasized that teachers' capacity to use data is inextricably linked to their instructional knowledge (Datnow, Park, & Wohlstetter, 2007). Data analysis can indicate the areas within which teachers should focus more effort but cannot tell them what to do. For that, teachers must have or be supported in developing the content knowledge and pedagogical tools to respond to the data analysis. Together, these studies indicate that understanding the content goal of a lesson, having a deep base of subject matter knowledge to draw on, and having a framework about common student misconceptions help teachers craft more flexible instructional approaches to accommodate students' emerging grasp of subject matter (Fennema et al., 1993; van Zee & Minstrell, 1997). Such versatility in instruction allows teachers to take advantage of assessment information.

As discussed earlier, teachers favor spontaneous performance assessments. These *ad hoc* opportunities depend on a certain amount of expertise to identify and capitalize on them in the moment. Facilitating a student's leap between content understanding and meaning-making and reasoning requires the ability to manage the flow of information and student ideas. Teachers can use dialogue with students to probe and redirect their ideas toward the learning goal (Duschl & Gitomer, 1997). Doing so requires an appropriate amount of guidance—guidance that allows the students to do their own thinking while still channeling emerging student conceptions toward the learning goal (Aschbacher & Alonzo, 2004; Fennema et al., 1993). Achieving the balance between supporting the student and allowing appropriate struggle requires experience.

Teachers' expertise in managing classroom interactions. Teachers tend to develop assessment practices only after they enter the profession. Although assessment is an integral teaching function, teaching candidates receive weak assessment training during their pre-service programs (Cizek et al., 1995/1996; Impara et al., 1993). Thus assessment falls under on-the-job training. And because teachers' work traditionally takes place in isolation from other teachers (Little, 1990; Lortie, 1975), teachers may struggle on their own to improve the tests they use, and to determine what they want to assess and what students need to do. When schools or external partners introduce specific assessment practices as a reform *per se*, they may place additional burdens, including time constraints, on teachers (Hall, Webber, Varley, Young, & Dorman, 1997; Stiggins & Bridgeford, 1985; Wayman et al., 2007). Teachers' ability to incorporate these new assessment practices depends

in part on their everyday classroom management skills to regulate the flow of activity and interactions within the classroom community.

In a project designed to help teachers create and implement performance-based classroom assessments, for example, teachers reported difficulty in record-keeping and time management (Borko, Mayfield, Marion, Flexer, & Cumbo, 1997). The support provided by researchers facilitated the teachers' understanding of what to observe and what information to record. In schools with less classroom-based support, participants' discussions of record-keeping were disconnected from classroom events and what they might mean for instruction (Borko et al., 1997). Another study followed teachers' attempts to implement formative practices consistent with the Foundation Approaches in Science Teaching (FAST) curriculum (Herman, Osmundson, Ayala, Schneider, & Timms, 2006). Teachers who had existing strategies and routines that engaged students and held them accountable succeeded in presenting the curriculum at a pace appropriate for the students (Herman et al., 2006). For example, exemplary teachers had effective routines in place to assure that students participated in discussion, which laid a foundation for engaging students with new questioning techniques. Similarly, teachers who managed instructional time effectively and had routines that helped students stay on task were better able to accommodate assessment-related activities such as recording observations or holding conferences with individuals or small groups (Black, Harrison, Lee, Marshall, & Wiliam, 2004). As Hattie and Timperley (2007) put it, "to make the feedback effective, teachers need to make appropriate judgments about when, how, and at what level to provide appropriate feedback and to which of the three questions [Where am I going? How am I going? Where to next?] it should be addressed." (p. 100). The teachers' success in implementing the new formative assessment practices was thus dependent on the strength of their existing classroom management and structures.

Teachers' preparation in assessment. Despite the interconnectedness between teachers' assessment practices and their content knowledge and classroom management skills, teachers' pre-service and in-service assessment training has been weak (Impara et al., 1993; Stiggins, 1991). Only one fifth of teachers reported receiving any in-service training related to constructing adequate tests or using test results to improve instruction (Herman & Dorr-Bremme, 1983). Those teachers who reported receiving a class or in-service training in testing and measurement were more comfortable interpreting standardized test information (Impara et al., 1993). Nonetheless, the majority of states do not require teachers to demonstrate competence in assessment to earn a teaching license. Only 14 states had assessment competency requirements in 2002, making assessment courses a lower priority in teacher preparation (*Education Week*, 2002, cited in McMunn et al., 2003–04).

Even where assessment courses are offered, they are not always relevant to teachers' instructional needs (Cizek et al., 1995/1996; Impara et al., 1993). Stiggins (1991) examined the "areas of mismatch between the assessment problems [that] teachers face and the type of assessment training they receive" (p. 7). He provided three categories of assessment purpose: as a means of informing instructional decisions, as a teaching tool, and as a behavioral monitoring mechanism to keep students focused. Few teacher preparation courses addressed assessment as a means of informing decisions, and no courses addressed the teaching tool and behavioral monitoring aspects of assessment (Stiggins, 1991). Gullickson (1986) also reported that a survey of teachers and professors teaching pre-service educational measurement courses revealed that teachers placed greater priority on "nontest evaluation activities (i.e., rating scales, observation, sociograms, anecdotal reports, and class discussion), [and] formative and summative evaluation" (p. 350) than did professors. Given the degree to which teachers report relying on nontest information of student progress (discussed above), this gap in emphasis potentially represents a significant need for teacher

preparation reform. Practitioners complain that "college courses in tests and measurement were not relevant to their needs in the classroom" (Impara et al., 1993, p. 116).

To develop teaching candidates' competency in science performance tasks, Morrison and McDuffie (2003) introduced science performance assessment activities in their elementary science methods course. With the supports provided by the project (e.g., mentor relationships and the placement of pre-service teachers in classrooms), the pre-service teachers were able to learn and implement many aspects of performance assessment. The careful and detailed instruction provided to the participants, along with placement in mentors' classrooms, allowed study participants to become more adept in assessment than the typical pre-service teacher. Nevertheless, they did not develop much skill in analyzing children's thinking or developing inquiry-based instruction within the time constraints of the course. Improvements shown by the study participants were considerable but limited in scope. "Significant and deliberate efforts" were provided by the intervention team but led only to the "beginnings of a foundation being constructed" (Morrison & McDuffie, 2003, p. 26), suggesting that even with intentional supports, pre-service training might fall short of providing teachers with adequate training. In addition, the course conditions that this intervention provides are rarely if ever found in the typical pre-service training, indicating limitations in scaling up such a program.

In-service training in assessment was similarly rare (Cizek et al., 1995/1996; Herman & Dorr-Bremme, 1983; Impara et al., 1993). One study found that a majority of teachers surveyed preferred in-service training to the other options presented, suggesting that teachers would like formal, in-person training to hone their assessment skills (Impara et al., 1993). For those teachers who reported having some pre- or in-service training in assessment, the most recent training was typically more than 6 years before the study (Impara et al., 1993). With the institutionalization of data-driven decision-making, related training may now be more frequent. Schools are offering in-service training in analyzing results from state and benchmark assessments, as reported by 43% of sampled teachers in a 2007 national survey (Means, Padilla, DeBarger, & Bakia, 2009). Such training notwithstanding, more than half of the teachers reported on the same survey that they would like additional professional development on "how to develop diagnostic assessments for [their] class[es]" (58%) and "how to adjust the content and approach used in [their] class[es] in light of student data" (55%) (Means et al., 2009, p. 30). Case studies of teachers in districts emphasizing data-informed decision-making further revealed that they were able to read common graphs and tables generated by typical data systems but faltered in manipulating the data or making comparisons, suggesting some gaps in teachers' preparation in analyzing assessment results (Means et al., 2009). Thus overall, teachers receive little formal training in maintaining and enhancing the assessment skills that they learn—by default in many cases—on their own. Overall, pre- and in-service assessment training for teachers is weak and not consistently applicable to classroom practice.

The gap between the emphasis placed on assessment in teacher training and the central function of assessment in instruction is disconcerting. Even though pre-service education has traditionally had a weak socializing influence on teachers' practice (Lortie, 1975), the relatively scant attention paid to assessment in pre-service and in-service training promotes highly individualized practice that depends on a teacher's own conceptions of teaching. As McMillan (2002) argues, "While measurement instruction usually includes consideration of student outcomes or objectives, much more is needed on helping teachers conceptualize deep understanding and reasoning, and the kinds of evidence that are needed to document them" (p. 42).

TEACHERS' BELIEFS ABOUT AND CONCEPTIONS OF TEACHING

Teachers' beliefs about and conceptions of teaching influence all aspects of their teaching—assessment included. Teachers' beliefs about subject matter (Young, 2006), valid assessment techniques (McMillan & Nash, 2000), and their role in instruction (Torrance & Pryor, 2001) filter out assessment practices and results that are inconsistent with those beliefs and attune with ones that are consonant.

Assessment data come from multiple and varied sources, including teacher observations, interactions with students, teacher-made tests, and publisher tests. How much weight a teacher assigns to the different types of data depends on the teacher's beliefs about their educational significance. For example, in weighing the value of data from a district oral fluency test, teachers who observed correlations between their students' comprehension and fluency scores tended to give credence to the results, whereas those who identified many exceptions to the rule tended to downplay the fluency scores and focus on comprehension strategies (Young, 2008). Similarly, teachers' educational philosophies influence how they react to instructional reform efforts. In one effort that sought to help teachers design and implement classroom-based performance assessments, researchers observed that teachers tended to ignore new ideas and practices that were incompatible with their own philosophies (Borko et al., 1997). The teachers' beliefs acted as the filter through which new ideas were perceived, interpreted, and executed. Recognizing the instrumental role of teachers' beliefs, the staff development team stated that if they were to "embark on another staff development effort, we would build in explicit attention to beliefs as well as practices" (Borko et al., 1997, p. 27).

At the school level, Ingram, Louis, and Schroeder (2004) found persistent cultural assumptions that undermined continuous improvement efforts. Teachers use their own "personal metric" for evaluating their instructional effectiveness, and these metrics often differ from those used in external accountability systems. They "base their decisions on experience, intuition and anecdotal information (professional judgment)" instead of systematically collected information (p. 1281). The researchers pointed out that these norms pose a formidable barrier to efforts to orient teachers toward systematic data analysis and data-based continuous improvement.

Reforming Assessment Practices to Improve Instruction

Against this backdrop of teachers' routine assessment work, reformers have designed and implemented a number of professional development efforts aimed at deepening teachers' understanding and use of embedded assessments to make instructional decisions. Fewer studies have examined school- and district-level initiatives to use data as a reform.

Teacher-Focused Intervention Studies

The British Assessment Reform Group (1999) advocates principles of assessment for learning that are ambitious and that it acknowledges are at odds with common practice today. For example, the group advocates assessment embedded in teaching and learning, teachers sharing learning goals with students, the use of assessments to help students know the standards they are striving for, and pupils assessing their own performance and reflecting on data with their teachers, among other characteristics. Many articles describe the design and effects of interventions that seek to improve teachers' formative uses of assessments and that incorporate various tenets that the Assessment Reform Group proposes. These studies tend to take the form of action research, in which the researchers design, implement, and study the training; indeed, the researchers often act as

the key facilitators and trainers for teachers. The studies take a broad view of what might be useful assessment information, and cite evidence of student thinking and dialogue with students as informal assessment data. On the whole, these studies point to the importance of teachers' ability to clarify learning goals, recognize student ideas, conceptions, and misconceptions as instructional cues, and to shape classroom environments conducive to students' engagement in ongoing assessment.

Clear learning goals. Clear learning goals improve teachers' ability to use assessments for formative purposes. At a basic level, learning goals help teachers identify what they need to assess and provide direction in planning instructional activities as follow-up to the assessment results. From this perspective, using assessments formatively entails gathering data about students' emerging conceptions and knowing how to effectively use their ideas in pointing them toward the learning goal.

Focusing on the learning goal is seemingly a fundamental principle of lesson planning and classroom teaching, but it is nonetheless difficult on a moment-to-moment basis. Participants in the Gillingham Partnership Formative Assessment Project attempted to develop skills in giving students oral feedback using the learning objectives of the lesson as a frame of reference. Roughly 40% of the teachers "found it difficult to focus their oral feedback on the point of the lesson, often being diverted by an urge to 'comment on everything' or distracted by classroom interruptions and events, or finding it impossible to ignore features like 'bad handwriting'" (Clarke & McCallum, n.d.-a, p. 12). Teachers participating in the same study had similar insights when asked to provide written feedback on student work. They found themselves providing feedback on parts of the students' work products that were not necessarily germane to the students' achievement of the learning goals (Clarke & McCallum, n.d.-a).

Two other examples illustrate oral responses as assessment information, and the importance of holding steadfast to the learning goals. Van Zee & Minstrell (1997) provided a case study of the *reflective toss*,[6] where an expert teacher allowed a student to freely express her thinking—correct ideas and misconceptions alike—and used questioning to direct her reasoning toward the learning goal. For example, the student proposed an alternate method of finding the mean without believing it would produce the same results as calculating the average. The teacher followed her unexpected reasoning by asking her to clarify what she meant by "average", and then asked, "Now would that come out the same as this if you did this [referencing average vs. mean]?" (van Zee & Minstrell, 1993, p. 242). Through questioning, the teacher was able to help the student recognize her misconceptions. Duschl and Gitomer (1997) describe a similar model entitled the *assessment conversation*.[7] In the assessment conversation, the teacher elicits various students' ideas, acknowledges the ideas in relation to the unit or lesson goal, and then uses the diverse student ideas to discuss which ones better satisfy standards of substantiated reasoning. Rather than appealing to authority such as the teacher or text, the teacher poses questions that allow the students to evaluate the relative quality of the ideas that they and their peers presented.

The reflective toss and the assessment conversation are founded on the same instructional principle—they allow students to arrive at the learning goal through their own reasoning. The teacher encourages the students to verbalize their ideas, facilitates the connection between their reasoning and the learning goal, and through classroom discourse promotes students' grasp of the intended learning goal. Although these examples are based on oral responses as assessment, the broader principle applies to teachers' analysis of student work as assessment information as well.

---

[6] Van Zee and Minstrell (1997) define the *reflective toss* as an exchange between teacher and student that typically consists "of a student statement, teacher question, and additional student statements" (p. 228).
[7] Duschl and Gitomer (1997) define the *assessment conversation* as "a specially formatted instructional dialogue that embeds assessment into the activity structure of the classroom" (p. 39).

That is, as Herman et al. (2006) put it, having a clear learning purpose enables teachers to "use the [students'] developmental trajectories to focus their instruction, their thinking about student progress, and their informal responses to it" (p. 26).

Studies of developing teachers' use of student work, oral responses, or classroom discourse as assessment information also find that instructional approaches based on dialogue and exchange are further improved by clear criteria that define success in meeting the learning goals. Open, transparent discussion of success criteria in relation to the learning goal enables teachers and students to develop a shared understanding of what is necessary to "move learning forward, with a realization that learning has to be done *by* the student and cannot be done *for* the student" (Harrison, 2005, p. 259; emphasis in original). This explicit learning-goal orientation in turn enables students to participate in their own learning process (Borko, Flory, & Cumbo, 1993; Clarke & McCallum, n.d.; Harrison, 2005; Herman et al., 2006) and enhances the responsibility they feel toward each other and themselves in monitoring their learning (Fennema et al., 1993). Including students in the development and discussion of the assessment criteria, coupled with clear learning goals, allows students to "more easily identify and understand the reasons behind their successes and improvements" (Clarke & McCallum, n.d., p. 62).

Focus on student needs. Studies of interventions aimed at changing teachers' assessment practices report teachers' increased focus on student needs as a result of engaging with new forms of assessment (Driscoll, 1999; Harrison, 2005; Torrance & Pryor, 2001). Through assessment data, teachers understand their students' needs better, and alter and plan their instruction accordingly (Hall & Hewitt-Gervais, 2000; Harrison, 2005). Data gathered from various sources including portfolios (Hall & Hewitt-Gervais, 2000) and teacher-student questioning (Harrison, 2005) provide evidence of student needs and the basis for how teachers differentiate instruction for their students.

An increased emphasis on student needs implies a change in the nature of teacher-student interactions. With such an emphasis, teachers relinquish the sole right to determine the pace and content of instruction: students' needs and progress become teachers' primary consideration in deciding whether to move on in the curriculum, how to review or reteach, and to whom to direct instruction. To attend to students' needs rather than strive for curriculum coverage requires more flexibility in teachers' instructional approaches. Instructional plans are vulnerable to change in response to emerging student needs, as well as to the amount of time teachers allocate to particular activities. Duschl and Gitomer (1997) note, however, that teachers' curriculum coverage routines are difficult to break, even when diverse student responses indicate uneven levels of understanding.

Several projects attempted to alter teachers' conceptions of teaching to a more student-centered approach. By extension, the projects required teachers to broaden their definition of assessments and to use those assessments formatively; that is, to identify and address student needs. Teachers participating in one such project realized at the beginning that their view of formative assessment was narrow— mainly as a formal requirement. Through self-evaluation and discussion with colleagues, participants incorporated student observations as the foundation for their assessment practices. From observation, they determined what their students knew and understood, and they began to realize the importance of student conceptions in their teaching (Torrance & Pryor, 2001).

Another study designed to help teachers develop an appreciation for what is worthwhile in mathematics included "better understanding of students' needs and more appropriate instructional goals and curriculum design" in its definition of assessment (Driscoll, 1999, p. 82). Participating teachers concluded that a "worthwhile task" includes consideration of student context and background as clues about student needs (Driscoll, 1999). Another study found that teachers' willingness to include their students' reasoning and needs in their instructional practice led to richer

student involvement and higher quality teacher-student interactions (Harrison, 2005). Moreover, when the students became accustomed to verbalizing their thinking and defending their reasoning, they began to assume responsibility for their own learning, strengthening their understanding of learning processes and products (Fennema et al., 1993). Students can be partners in teachers' attempts to use data for instructional purposes, shifting the teachers' role to one of facilitator instead of provider or judge (Tunstall & Gipps, 1996, p. 399).

The classroom learning environment. The social environment of the classroom emerges as an important factor and outcome in how teachers implement different kinds of assessment practices. The social environment particularly applies to interventions promoting oral responses or student work as evidence of student reasoning and understanding, both of which can inform instruction. In this regard, the studies reported that students must feel comfortable with making their thinking public and risking an incorrect answer (Black & Harrison, 2001; Fennema et al., 1993; Harrison, 2005; van Zee & Minstrell, 1997). In the Formative Assessment Project at King's College in the U.K., for example, students participated in critical discussions in which their ideas were alternately built on and challenged (Black & Harrison, 2001), and they examined the strengths and weaknesses of their own and their peers' work. Doing so helped them internalize the meaning of high performance within that classroom context (Harrison, 2005). At least one study, however, argued that a comfortable learning environment that facilitates student sharing is insufficient to improve assessments, instruction, and learning. The study concluded that students also need feedback on the quality of their answers (Duschl & Gitomer, 1997).

These studies of particular interventions have sought seamlessness between instruction and assessment that does not reflect the descriptions of teachers' habitual assessment practices reviewed earlier. The two sets of studies thus reveal a gap between the traditional teacher-made paper-and-pencil tests (complemented by frequent spontaneous performance assessments that lack systematicity) on the one hand, and integrated, constructivist instruction (which in theory elicits continuous evidence of student learning, misconceptions, and needs) on the other. This gap raises many questions about the resources and support needed to spread the lessons of these intervention studies to a significant proportion of the teaching force.

School and District Uses of Assessment

Much of the assessment literature focuses on within-classroom, individual teacher practice. Little has been written about school and district uses of data for formative purposes *per se,* except for more recent work on school-based inquiry. However, because those projects surfaced with the prevalence of large-scale standardized test beginning in the late 1990s, many of those studies focus on analyzing annual achievement results for school improvement planning. The central thrust of that work indicates that disaggregated analyses identify achievement gaps between student subpopulations—revelatory in themselves (Lachat & Smith, 2005; Stokes, 2001)—and help teachers, leaders, and administrators concentrate resources. In one sense, identifying which students need the most help can be formative, resulting in teachers' or grade-level teams' organizing their lessons differently or using targeted strategies for different student needs (e.g., to group students and provide appropriate interventions) (King & Amon, 2008; Marsh, Pane, & Hamilton, 2006; Supovitz & Klein, 2003; Young, 2006). Due to the annual nature of the assessments, however, schools tend to use those types of data to inform decisions at the beginning of the school year, after which the data can become obsolete.

Studies that discuss school or district efforts to use data formatively had been relatively scarce, although use of data is fast becoming a central tenet of many school- or district-level theories of change. McLaughlin and Mitra (2003) identify stages of development for schools attempting to

build "cycle of inquiry" practices. Partnered with the Bay Area School Reform Collaborative (BASRC), initially an Annenberg Challenge initiative, the schools received external funding to support a school-based facilitator or inquiry coach, external professional development to learn about the cycle of inquiry processes, and annual "critical friends" feedback. The BASRC cycle of inquiry mirrored typical continuous improvement cycles depicted in business literature.[8] McLaughlin and Mitra (2003) elaborate developmental social and organizational conditions associated with schools that are novice, intermediate, or advanced in engaging in the cycle of inquiry. Among the distinguishing dimensions, advanced schools become learning communities that initiate their own professional development, exhibit distributed leadership structures, and focus their inquiry on an issue of central importance. This focused effort becomes the spine for bringing coherence across multiple reform efforts. Moreover, faculties in schools advanced in inquiry practices accept inquiry as iterative, schoolwide, and linked to classroom instruction (McLaughlin & Mitra, 2003). Advanced schools also continuously "seek better forms of data" (McLaughlin & Mitra, 2003, p. 17). The BASRC schools included in this study predominantly used annual student outcomes as data but sought additional data to inform instructional practice. They recognized that questions about instructional practice occur in shorter cycles than that described by annual data.

Supovitz and Klein (2003) categorize the purposes for which schools use a range of assessment information. They note that school leaders used student performance results to design professional development and to set goals and targets. As with other uses of data, the challenge these school leaders faced lay in gathering timely information for the current school year and acting on that information during the same year. Current student performance data that indicate student needs for which teachers might want more support do not necessarily represent the needs of prior or subsequent student cohorts (Supovitz & Klein, 2003). Similarly, leaders establish annual goals, often accountability measures set by the district superintendent in addition to legislated requirements, and leaders look for midcourse information about whether teachers and students are progressing toward those year-end goals (Supovitz & Klein, 2003). However, it is setting interim targets that may constitute formative action, by reorienting teachers to short-interval achievements such as specific reading levels for each quarter (Supovitz & Klein, 2003). Teachers can then use information about which students are meeting interim goals to identify both target students and timely interventions.

Kerr, Marsh, Ikemoto, Darilek, and Barney (2006) studied the strategies that three districts pursued to support teachers' and schools' use of data for instruction. The strategies they identified included "the development of interim assessments and technology/systems for housing, analyzing, and reporting data; the provision of professional development and/or technical assistance on how to interpret and use student test results; the revamping of school improvement planning processes; the encouragement of structured review of student work, and the use of an [Institute for Learning]-developed classroom observation protocol, the Learning Walk, to assess the quality of classroom instruction" (Kerr et al., 2006, p. 504). Overall, Kerr and colleagues found that the strategies that focused on data use in the two more effective case study districts were associated with more teachers and principals reporting access to multiple data sources, viewing data as "useful for guiding instruction in their classrooms" (p. 510), reporting "more frequent and extensive use of data" (p. 511), and receiving more support from their respective principals.

---

[8] The typical cycle includes defining the data necessary and appropriate to a particular problem, collecting the data, analyzing the data, drafting action plans as a consequence of the analysis, and following through with the changed behavior (cf. Nickols, 2000; O'Dell & Grayson, 1998; Streifer, 2001; Thorn, 2002). In extended conceptions, organizations evaluate the modified processes (with data) and revise them as needed.

In particular, the two case study districts that were most effective in promoting data use for instructional improvement invested in a school improvement planning (SIP) process in one case and created a system of interim assessments, analysis, and reporting in the other. The first district provided the SIP template and set the expectation that school coaches would support data analysis for the SIP and use accountability mechanisms such as SIP implementation visits. The second district required interim assessments three times during the school year in addition to other, more frequent formative assessments and enabled teachers to use the data with a new data management system. Nevertheless, a majority of teachers surveyed in that district reported that their classroom assessments were "more thorough and provided more timely information" (p. 509). These findings suggest that although districts may use many strategies in stimulating and supporting schools in using data, the data must have legitimacy with teachers, especially if they are to invest the time in learning to use new software to analyze and display data.

Coburn and Talbert's (2006) theory-building work offers a framework that illuminates varying understandings of evidence across a district. They identify differences in individuals' beliefs about what makes evidence valid, including the psychometric properties of the assessment, the degree of alignment with desired outcomes, the ability of the assessment to provide insight into students' thinking and reasoning, the degree to which assessment results reflect teacher judgment and therefore are authentic, and whether the evidence is based on multiple data sources. Coburn and Talbert (2006) also find four purposes for which individuals believe evidence should be used: meeting accountability demands, informing curricular programming and sometimes simply validating decisions already made, grouping students for instruction, and informing instructional practices. They find that these conceptions of evidence map to a district's hierarchy. District administrators are more likely to rely on the psychometric properties of assessments and links to desired outcomes, while teachers and those administrators working directly with teachers are more influenced by insights into student thinking and reasoning and teacher judgment as sources of validity. Personal history and experiences with prior reforms also condition individuals' frames of reference as new assessments are introduced and shape their conceptions of evidence. These conceptualizations provide a beginning framework for looking at how uses of data differ by organizational level in school districts.

Detailed research is scant concerning the types of decisions that district administrators may use data for, the kinds of data they desire, and how their practices change as a result of data use. The literature lacks specificity about district leaders' major responsibilities, the types of data that may influence their decisions, and the organizational conditions under which that influence occurs. In addition to understanding teachers' formative uses of assessment and other data within the context of their instructional practices, parallel research is needed for school and district leaders. Grounded in school leaders' work, what decisions do they make and how do they use data to make those decisions? What are the attendant conditions that facilitate or frustrate their attempts to use such data? A subtle but conceptually important difference with this approach, compared to the bulk of the research on how teachers use assessments, is that it places school leaders' and district leaders' work at the center. Arguably, supporting instructional improvement should be the most important goal of these leaders' jobs; however, they must also undertake myriad other activities for schools to run smoothly. We know little about the use of data that would improve their decisions, which at least indirectly support teachers' effectiveness in the classroom. In this line of inquiry, principals and district leaders would not be part of the environmental factors that enable teachers to use assessment and other data effectively; they would be the data users as well.

Organizational Supports and Barriers to Teachers' Formative Use of Assessment

In descriptions of teachers' assessment practices and effects of specific interventions designed to improve their assessment knowledge and techniques, few studies focused on the attendant organizational conditions. Nevertheless, in most of those studies, various organizational factors do emerge as barriers or facilitators. Studies looking at schoolwide data use tend to emphasize organizational conditions more so. Below, we discuss leadership and professional learning conditions that influence teachers' use of assessment and other data for instructional improvement.

Leadership for Formative Uses of Assessment

Leadership emerges as a crucial factor in various studies focusing on assessment reform initiatives. Indeed, in certain respects, using assessment data formatively may simply be a special case of school leadership: communicating expectations for using assessment data and managing the assessment program, for example. And when different assessment practices become a goal, the leadership in question might be a special case of change management; that is, it sets the vision for reformed practices, delineates a change process, and creates a supportive and nonthreatening environment. However, Halverson,Grigg, Pritchett, and Thomas (2005) go further, arguing that the current era of accountability pushes conceptions of school leadership "beyond the traditional categories of instructional, managerial, and transformational practice to a new, and more specific conception of creating accountable learning systems in schools" (Halverson et al., 2005, p. 5).

Leaders—defined broadly—are acknowledged as the prime movers in creating new school cultures around using data and changed practices. Copland (2002) asserts that leaders in key roles catalyzed change at schools embarking on an inquiry-based school reform effort. Supovitz and Klein (2003) similarly find that "virtually every example of innovative data use in [their] study came from the initiative and enterprise of an individual who had the vision and persistence to turn a powerful idea into action" (p. 36).

For schools advancing inquiry practices, leaders were effective if they developed distributed structures that built broad-based engagement across school faculties. Those structures included defining a new lead teacher position for someone already accepted as an informal leader who takes on "leadership functions typically associated with the principalship" (Copland, 2002, p. 12). Moreover, the involvement of a broadly representative group of teachers in the inquiry work created a greater sense of joint mission and a "shared feeling that the reform work is 'integral' to everything they do" (Copland, 2002, p. 13).

Principal leadership is crucial to setting expectations for school staffs to consider data as decision inputs (Wayman & Stringfield, 2006a; Young, 2006) and to creating supportive environments in which teachers can share the successes and failures associated with assessment results (Wayman & Stringfield, 2006a). To facilitate teachers' use of assessment and other data, leaders attempt to create a system and norms of learning. Young (2006) offers a set of agenda-setting activities specifically in the context of teachers' learning to use assessment and other data for instructional purposes. These leadership roles include modeling appropriate uses of specific data for teachers, providing the rationale (or theory of action) for using data, strategically aligning expertise and resources to support teachers' learning about how to use data, and deploying resources to cover a new range of data-related functions (Young, 2006). In addition to establishing and support teachers' use of data, leaders' responsibilities entail different types of activity that can be productively informed by data, such as "diagnosing or clarifying instructional or organizational

problems"; "weighing alternative courses of action"; "justifying chose courses of action"; "complying with external requests for information"; "informing daily practice"; and "managing meaning, culture, and motivation" (Knapp, Copland & Swinnerton, 2007, pp. 77). These leadership activities are encompassed in developing an organization-wide culture of continuous improvement.

Schools generally lack the capacity to use formative assessment feedback (Halverson et al., 2005). This gap argues for leadership attention to building capacity, as required for any other proposed reform. The data-related functions described in Young (2006) identify dimensions of school capacity that are not explicitly defined in the typical school organization. These dimensions include uploading and downloading data reports; interpreting data and teaching teachers about using data; providing professional development, materials, and other needed resources to foster data analysis; facilitating meetings so that teachers consistently focus on what they might do in their classrooms as a result of data analysis; and holding teachers accountable for instructional changes that they agreed to (Young, 2006). Along the same lines, Halverson et al. (2005) identify structures and leadership actions that aid in a formative feedback system: "re-purposing in-house expertise" (p.33) to support teachers in literacy instruction that relies on ongoing assessments of oral reading and providing structured and scheduled opportunities to work with a facilitator who is expert in interpreting literacy assessment results as well as in teaching literacy.

Studies also refer to the leaders' role as direct capacity to use data. They point out that school leaders expert in analyzing data are better able to facilitate teachers' discussions by injecting critical questions at the right time to guide teachers toward more accurate analyses, by appropriately delineating a problem, and by drawing specific instructional implications (Driscoll, 1999; Herman & Gribbons, 2001; Lachat & Smith, 2005; Young, 2006). In contrast, Supovitz & Klein (2003) report that only 19% of school leaders surveyed "felt that they had the technical skills to manipulate the data in order to use it to answer questions that they wanted to ask" (p. 38). Ironically, leaders are thus responsible for mobilizing the school to perform tasks that many may not be able to do themselves.

These articles offer a vision of leaders as change catalysts, capacity builders, and experts in the context of creating school inquiry and data use practices. They alternately describe distributed leadership systems that involve principals, teachers, and specialists with specific data-related functions, as well as the particular leverage vested in the principalship. But they do not address how leaders can develop the ability to fulfill these roles.

Professional Learning Conditions

At the heart of continuous improvement principles and processes is the notion of teachers as learners—learning about how to improve their instruction, be it which students to focus on, what specific students need, or how they (as teachers) can acquire and refine the instructional strategies that meet student needs. The school as a learning environment for teachers represents a crucial lens for viewing teachers' opportunities to develop assessment analysis and new instructional techniques and strategies in response. Research on high-quality professional development points to general attributes that improve teachers' learning experiences including intensity, subject-matter specificity, collaborative settings, and building on teachers' prior knowledge (Corcoran, Shields, & Zucker, 1998; Garet, Birman, Porter, Desimore, Herman, & Yoon, 1999; National Staff Development Council, 1995). In addition, research on school systems that support teacher learning underscores the potentially pivotal role coaches can play in situating professional development in teachers' classrooms and bringing expertise to new teachers in particular (Elmore & Burney, 1999; Hightower,

2002).[9] These learning supports are reflected in a number of interventions designed to improve teachers' use of assessment and other data.

The role of coaches. The role of coaches or facilitators in interventions designed to build teachers' assessment capacity centers on guiding conversation and modeling assessment and instruction in the classroom. These interactions with teachers create informal accountability for their learning and for their attempts to change assessment practices. Examples of informal accountability mechanisms appear in initiatives designed to give teachers new assessment techniques and action research models, which teachers use in reflecting on their practices (e.g., Clarke & McCallum, n.d.; Fennema et al., 1993; Fuchs, Fuchs, Karns, Hamlett, & Katzaroff, 1999; Torrance & Pryor, 2001). Other studies of districts pursuing data-use strategies found that experts in dedicated professional development roles were a key factor in helping teachers make connections between assessment results and instructional actions (Goertz et al., 2009; Means et al., 2009; Young, 2008).

External coaches can model concrete classroom techniques. For example, in one study, teachers struggled with how to observe students and classroom activities and how to record their observations systematically to afford more reliable and valid inferences for instruction. Discussions among teachers who lacked in-classroom coaching support were abstract and disconnected from the classroom while discussions among teachers who had such coaches were less abstract and more connected when focused on student observations (Borko et al., 1997). Another study concluded that pre-service teachers who had mentors expert in performance assessments received more substantive feedback than their peers who had less knowledgeable mentors. The less knowledgeable mentors did not provide in-depth, constructive feedback on the assessment tasks, and their student teachers found the overall experience "often frustrating and less rewarding" (Morrison & McDuffie, 2003, p. 22).

Studies seeking to change teachers' assessment practices have found that coaches and facilitators not only provide assessment expertise but also serve as a source of teacher accountability. They do not necessarily evaluate the teachers; rather, the interactions with teachers create professional accountability for changes in classroom practice. For example, in the Formative Assessment Project at King's College, teachers stated that when researchers or facilitators came "to watch them putting into practice the commitments they had made in their action plans[, that act] was a strong motivating factor in ensuring that they gave attention to developing formative assessment" (Lee & Wiliam, 2005; p. 278). In another case, coaches "ask[ed] carefully selected questions and constantly reframe[d] teachers' orientations" (Driscoll, 1999, p. 88). The coaches maintained teachers' focus on the main objective of the professional development, "observation of student work and reflection on what that work reveals about student understanding" (Driscoll, 1999, p. 88). In this study, teachers' monthly reports also provided opportunities for individual reflection, and the knowledge that they needed to submit a report served as a professional accountability mechanism.

As noted above, high-quality professional development needs to be grounded in specific subject matter. In developing teachers' capacity to use assessments, subject matter is a necessary though implicit ingredient. Teachers must know what they are teaching to know what they can reasonably assess. And they must know enough subject matter to glean information about students' misconceptions from their responses. Reflecting the central importance of subject matter, the interventions included in this review in general offered professional development for content knowledge and the learning process as well as the particulars of assessment, deepening teachers' capacity to use assessments formatively within their subject-matter contexts.

---

[9] However, the coaching role is also fraught with dilemmas associated with being one of neither teaching nor administration (Smylie, 1990; see Burney, Corcoran, & Lesnick (2003) for a review).

Such professional development is a function of the level of teacher experience. As discussed previously, the average novice teacher does not possess the expertise required to develop assessment practices that truly support teaching. Collaborative structures that allow novice teachers to learn from more expert colleagues cannot entirely substitute for the tacit knowledge underlying experts' judgment about whether and how student performance and work meet quality standards. But those structures can provide novices with access to exemplars in both instructional and assessment repertoires. We discuss collaboration as a factor in professional learning next.

Teacher community and collaboration. Sociological studies have long observed that isolation and privacy are traditional norms among teachers (Little, 1990; Lortie, 1975). Even though building learning communities among teachers has become a prevalent reform strategy (McLaughlin & Talbert, 2001), teachers typically did not collaborate on assessment in the course of their daily instruction-related work (Cizek et al., 1995/1996; McMillan, 2002). More recently, on a 2007 national survey, a majority of K-12 teachers reported that they used their respective student data system on their own (78%) and with colleagues or department teams (71%) (Means et al., 2009, p. 17). However, more than half (59%) did so as part of district-led activities, not necessarily as a common routine in conducting their work (p. 17). Classroom assessments usually still falls under the sole purview of individual teachers. Teachers may also feel that sharing assessment results may reflect adversely on their instruction; such conversations are thus likely to require firmly established professional trust. The history of entrenched isolationism that characterizes teachers' assessment practices may account for why some researchers found orchestrating collaboration focused on assessment difficult (Morrison & McDuffie, 2003).

Reflecting research on learning communities (Little, 2003; McLaughlin & Talbert, 2001) and communities of practice (Lave & Wenger, 1991; Wenger, 1998), the assessment intervention studies incorporated teacher collaboration in their overall theories of change. The studies reported that collaboration that focused on assessment practices allowed teachers to reflect on their own practices and to share ideas with colleagues, an act that in turn improved the effectiveness of the initiative. For example, broad-based collaborative meetings between participating teachers, researchers, and district administrators in the Formative Assessment Project at King's College led researchers to conclude, "The support of working as part of a professional learning community seems, from interviews with the teachers of the project, if not essential, then at least highly desirable, to make sure that the ideas take root" (Lee & Wiliam, 2005, p. 277). Borko et al. (1997) linked the degree of teacher change in using assessments to the degree of a teacher's embeddedness in a learning community, within which teachers studied benchmark books together to learn new instructional strategies based on diagnostic assessments. In this setting, the teachers experimented with new ideas and shared their struggles with new practices. Lyon and Leahy (2009) identified four processes that helped teachers learn about their assessment practices: "collaborative problem solving"; "customization of existing techniques and creation of new techniques"; "[s]hared examples of positive feedback from students, teachers, and administrators"; and "[c]ommitment to the group" (p. 3). These elements illustrate how teachers' learning about assessments is embedded in broader school norms of how teachers discuss their instructional practice and the degree to which they make public their struggles and areas needing improvement. Indeed, in a study of implementing assessment for learning in high schools, researchers found that pre-existing norms were difficult to change, and they concluded that the process of building a trusting team in which teachers felt safe discussing their practices was as challenging as improving their understanding of assessments for student learning (Weinbaum, 2009).

Other case studies also identified teacher collaboration as a mechanism by which teachers learned to analyze data and learned new instructional strategies to address the concerns raised in

their analysis of assessments (Diamond & Cooper, 2007; Halverson et al., 2005; Wayman & Stringfield, 2006a; Young, 2006). Indeed, assessment data are not formative unless teachers make use of the information for instructional practice or program design. Thus, to the extent that teachers' joint efforts underpin this critical step of bridging data analysis and instruction-related decisions, collaborative structures may be a key lever in changing how teachers develop and refine their repertoire. Channels for accessing others' instructional expertise are particularly valuable for novice teachers because they allow them to make sense of assessment results and to capture ideas about what to do about the results (Young, 2006).

Like other attempts to effect instructional change, collaboration is potentially an important vehicle for learning about new assessment practices—how to analyze various types of assessments and, importantly, how to brainstorm connections between assessment results and instructional strategies. However, as Little (2002) points out, collaboration in itself does not necessarily lead to improved instruction. What do teachers talk about? How do teachers represent their instruction, and how do they characterize their practices? What artifacts of instruction do they share with their peers? Given that the reflective collaboration that the literature espouses typically occurs outside of the classroom and depends on teachers' representations of their practices—however accurate or thorough—discussions to help teachers use assessment data may be circumscribed and limited in efficacy. As Driscoll (1999) notes, teachers' degree of instructional change varied widely despite the "blanket of support" (p. 101) provided by the discussion group-based professional development in which they participated. It is tempting to argue that collaboration is desirable—even necessary—to building systemwide capacity for incorporating assessment results into instructional decision-making. But we know little about how that time should be structured, whether norms of trust and learning can be developed simultaneously or must meet a threshold level for teachers to begin fruitful discussions, and what additional supports teachers need to leverage the time they spend together.

Time. Time is a frequently cited barrier in the implementation of many reforms. With respect to instructional change, time constitutes a resource in multiple ways: scheduled time to learn and collaborate outside of the classroom, instructional time to implement different kinds of assessments, some of which may be more time-consuming than others, and instructional time to act on data analysis; and the elapsed time over which instructional change occurs.

The studies designed to improve teachers' assessment practices embedded assessments that were more time-consuming than traditional paper-and-pencil tests. For that reason, competition for instructional time limited teachers' willingness to experiment with new assessment techniques (Borko et al., 1997; Hall & Hewitt-Gervais, 2000; Stiggins & Bridgeford, 1985) or inquiry work (Ingram et al., 2004). However, it is interesting to note that when teachers claim that instructional and assessment time have a zero-sum relationship, they see assessment as separate from—rather than as integral to—instruction.

The multidimensional nature of time implies different strategies. Some initiative designers accounted for professional development time and elapsed time for teachers to adjust practice by working with teachers over an extended period (e.g., 18 months for the Formative Assessment Project at King's College [Black & Harrison, 2001; Harrison, 2005; Lee & Wiliam, 2005], 2 years for the Formative Assessment Project [Clarke & McCallum, n.d.]). Others staged the learning process for teachers; for example, as Torrance and Pryor (2001) describe, researchers assisted teacher researchers in collecting data from their classrooms and reflecting on their practice for half a year and then supported teachers in operationalizing changes in their practice based on findings from the first phase. Conversely, Morrison and McDuffie (2003) blamed the one-semester training for their observed project's lack of success in developing pre-service teachers' skill in analyzing children's thinking or in developing inquiry-based instruction, and they concluded that one semester was too

short to provide teachers with such skills. In the more recent emphasis on implementing interim or benchmark assessments, schools or districts offer teachers collaboration time to learn about and engage in analyzing data (Goertz et al., 2009; Means et al., 2009). The amount of time available during teachers' workdays is probably insufficient, however. In a 2007 national survey, approximately a quarter of sampled K-12 teachers (23%) reported having time during the work day to analyze data, while a majority (59%) reported accessing data outside of their paid work day (Means et al., 2009, p. 27).

Across all of these studies, time is required for the foundational work needed to enable teachers to understand the role of assessments in instruction—creating success criteria tied to learning goals for specific instructional units, and shifting their orientation toward assessment as integral to instructional decision-making. Professional development time outside of the classroom is required, as is time for changing teachers' conscious practice, and time for instructional experimentation—some of which is inevitably inefficient. In addition, certain data-related functions that facilitate teachers' use of assessment and other data fall outside teachers' conceptions of their roles, as discussed above. Providing time (among other resources) linked to specific organizational roles to perform those functions may be necessary to support schoolwide use of assessments for instructional improvement.

Data Systems

Many recent studies on data-driven decision-making in educational organizations proceed from the vantage point of what information systems offer—the types of data, the aggregations and disaggregations, frequency and timeliness of the collected data, the reporting functions, and visual representations of the data (King & Amon, 2008; Means et al., 2009; Moody & Dede, 2008). Studies also examine facilitators and barriers to continuous improvement processes as applied to a range of decisions, programmatic as well as instructional (Wayman et al., 2007; Means et al., 2009; Moody & Dede, 2008). As technology enables organizations to access and manipulate large datasets, data systems are increasingly becoming an important dimension of organizational capacity.

Consistent with related literatures discussed above, studies focusing on data systems as the lens of reform underscore similar organizational factors influencing practitioners' use of data. For example, Wayman et al. (2007) examine a small, rural district that has separate data systems in place but is working towards becoming a "data-informed district." The evaluation found facilitating factors and barriers to effective data use that were tied to district culture, differing skills of teachers at all levels, and data system infrastructures. In a case study of an urban school district (Moody & Dede, 2008), researchers found "pockets of implementation"(p. 253) in the district but noted that lack of staff collaboration was hindering the effectiveness of the district's data system. In both districts, unification across the data systems and in staff perception of the usefulness of data systems would have aided in creating more effective use of available data.

Although many recent studies focus on what data systems offer and the implementation of data systems, the research that is currently lacking would focus on the types of data that are included in data systems and how those data systems help teachers make formative instructional decisions (the national study by Means et al., 2009, being a key exception). Despite this gap, we can attempt to draw some implications from the present research for thinking about teachers' formative uses of assessment based on how they use the data available to them in these information systems. Data systems tend to provide teachers with student demographics, attendance, and program participation (e.g., free or reduced-price lunch), class rosters, annual standardized test scores, and benchmark test scores. Means et al. (2009) reported that in a national survey, almost one-third of teachers (29%) considered the data available in their respective data systems limited in usefulness for deciding what

and how to teach. Roughly one-quarter (24%) experienced difficulties in finding the data they wanted and one-fifth lacked knowledge about how to work the system, further diminishing teachers' access to and usefulness of their local data (p.18). In a study conducted by Wayman et al. (2007), teachers used data accessed through the district data system to identify individual students for remediation, develop recommendations for tutoring, and tailor instruction to individual needs. For instructional use, teachers preferred using classroom assessments rather than state test data because they felt the classroom assessments provided the best picture of student learning. Even with access to different types of data and the ability to triangulate across data afforded by the data system, "it was also clear that teachers want more knowledge about student learning than they feel these data provide them" (Wayman et al., 2007, p. 24). Similarly, in another study, two districts provided teachers with interim assessment results for individual students with the assumption that such data would translate into actions in the classroom. They fell short of that goal, however, because the assessments had limited applicability to teachers' instruction (Goertz et al., 2009).

Research examining how school and district personnel use data systems point to a need for building teacher capacity not only in manipulating the system but in analyzing and interpreting the various forms of data and making connections to instruction (King & Amon, 2008). Teachers also need leadership support, timely data, and data that are relevant to curriculum and student learning (Wayman et al., 2007; Young, 2006). Although organization-wide data systems provide teachers with access to multiple forms of data, afford easy disaggregations, and display graphical representations of the data, leadership, capacity-building, and access to relevant data circumscribe teachers' actual use of the those data to inform instruction.

Conclusions

This review of teachers' formative assessment practices has illuminated the imprecise nature of the term. Formative assessment includes performance assessment and paper-and-pencil benchmark tests, which are not necessarily distinguished from summative assessments, as well as interactions with students, questioning, student responses, and student observations, which begin to describe instructional approaches. Two conceptual issues arose early in our literature search: what is *formative*, and what do we include as assessments? We follow others (Wiliam & Black, 1996; Wiliam & Leahy, 2006) in using *formative* to describe both the process and teachers' intended and actual use of data rather than the assessment itself. That is, formative use of assessment and other data entails in*form*ing instruction to improve it. Second, we cast a broad net to find what is considered assessment, including instructional efforts to ascertain students' understanding for teachers' use in making the next move in the classroom.

Key themes

Novice teachers generally feel underprepared to tackle assessment, once they are faced with the realities of classroom teaching. Although teachers use a variety of formats, studies examining teachers' assessment practices point to their reliance on informal assessment and nontest information to identify student needs (Cizek et al., 1995/1996; Goertz et al., 2009; Shavelson & Stern,1981; Stiggins & Bridgeford, 1985; Supovitz & Klein, 2003). Teachers value tests they design themselves more than external assessments, despite acknowledging the need to improve their own assessment practices. Teachers' ability to use assessments for instruction interacts with their knowledge of student learning and subject matter (Aschbacher & Alonzo, 2004; Duschl & Gitomer, 1997; Goertz et al., 2009; Johnston, Afflerbach, & Weiss, 1993); it also depends on their ability to manage classroom interactions, which provides a foundation for undertaking assessment- and data

collection-related tasks (Black et al., 2004; Borko et al., 1997). Teacher preparation programs are ill-matched to teachers' expressed needs for classroom assessment (Impara et al., 1993; Stiggins, 1991), and teachers report that in-service professional development that focuses on assessment has been infrequent (Herman & Dorr-Bremme, 1983). Because teacher preparation is a weak socializating force, and because professional isolation and privacy are traditionally dominant organizational norms for teachers (Little, 1990; Lortie, 1975), their assessment practices tend to be individual and learned on the job through trial and error. Under the goal of data-driven decision-making, districts are increasingly providing more in-service around accessing and analyzing data from state and interim or benchmark assessments (Goertz et al., 2009; Means et al., 2009; Young, 2008).

Studies attempting to provide teachers with new assessment techniques have emphasized the importance of stating clear learning goals (Clarke & McCallum, n.d.; Duschl & Gitomer, 1997; van Zee & Minstrell, 1993). They focus on student needs (Driscoll, 1999; Harrison, 2005; Torrance & Pryor, 2001) and seek to build a classroom learning environment in which students feel secure in expressing their ideas and in risking exposure of misconceptions, as well as cogent insights (Black & Harrison, 2001; Fennema et al., 1993; Harrison, 2005; van Zee & Minstrell, 1997). Most of these studies take the teacher as the unit of analysis. Little research has addressed school- and district-level uses of data. Although much has been written about how to foster schoolwide data-driven practices, those writings seek to build teacher capacity throughout a system and improve teachers' classroom decisions schoolwide. Research that examines the distinct uses of assessment for school leaders' and district administrators' functions is generally lacking, albeit increasing.

The studies identified leadership and its multifaceted roles as critical organizational conditions that support or frustrate teachers' uses of assessment for instructional improvement (Copland, 2002; Goertz et al., 2009; Halverson et al., 2005; Supovitz & Klein, 2003; Wayman & Stringfield, 2006a; Young 2006). Similarly, professional learning conditions including collaborative norms and structures emerged as facilitating factors (Borko, et al., 1997; Diamond & Cooper, 2007; Halverson et al., 2005; Lee & Wiliam, 2005; Lyon & Leahy, 2009; Wayman & Stringfield, 2006a; Young, 2006). The lack of access to expertise stands in the way of teachers' making changes to their instructional decision-making, and not surprisingly, lack of time is a barrier. Time, too, is multidimensional. Reform efforts need to address time in terms of professional development time, instructional or classroom time (Borko et al., 1997; Hall & Hewitt-Gervais, 2000; Stiggins & Bridgeford, 1985), and the elapsed time required to alter teachers' practices (Black & Harrison, 2001; Clarke & McCallum, n.d.; Harrison, 2005; Lee & Wiliam, 2005), as well as time linked to specific roles to carry out new, data-related functions that heretofore have not been anyone's job. An increasing number of studies focus on the affordances of data systems in stimulating data-driven decision-making, and they invariably conclude that organizational conditions such as teacher capacity-building and school leadership are necessary components to using data in continuous improvement activities (Means et al., 2009; Moody & Dede, 2008; Wayman et al., 2007). Thus, although data systems are a key dimension of organizational capacity, they cannot ensure change in data practices that are embedded in the overarching organizational norms of learning, collaboration, isolation or privacy.

Gaps in the Literature

The literature we reviewed delineate the various types of assessments teachers use. However, they do not provide detailed descriptions of the process teachers use to make sense of myriad assessment and information sources, especially in the current data-rich environment. Evidence indicates that teachers' beliefs and conceptions of teaching shape what they embrace and what they discard. Paradoxically, these filters may become more entrenched as teachers cope with the ever-

increasing abundance of data they are expected to make sense of. But we know little about how their beliefs might evolve with enhanced professional development, a greater sense of an adult learning community, and experience with using data for formative purposes.

Along the same lines, current policy discussions of data-driven decision-making assume that not only more data but virtually *all* data can be helpful to teachers. Yet, to take annualized tests as an example, teachers readily list those tests' limitations for instructional purposes. Which data are useful to teachers for instructional decision-making, and why are they useful? Which data are not useful, and why are they not?

New information systems set the stage for generating more—and more precise—data. The types of analyses and information these systems can afford needs to be analyzed separately from what teachers do with the data systems, and what teachers then do with the resulting information. Each of those points carries the potential for slippage—each of these linear steps do not automatically occur. What supports do teachers need to access information? What teacher questions can information in the data system address? What teacher questions require information outside of the data system? And what is the relative importance of those two sets of questions from the practitioners' perspective? Addressing a parallel set of questions for school leaders and district leaders regarding their work would also be useful.

A major research design issue lies in conceptualizing the length of decision cycles. At one extreme, teachers make decisions moment to moment—the period that many of the studies reviewed here addressed. Lesson plans give shape and texture to the day's activities; however, teachers constantly adjust those plans on the basis of students' attention and engagement and their apparent understanding, misunderstanding, and skills. Many unrelated social behaviors that enhance or detract from the lesson also influence teachers' responses. Is it possible to study how formal or informal data can influence this kind of almost-tacit decision-making? What are other cycles for research—daily planning? units of study? Teachers' daily, weekly, and monthly activities should form the basis for understanding data use. Identifying the social groupings in which teachers conduct activities (i.e., as individuals, with self-selected colleagues, in grade-level or departmental teams, schoolwide, or cross-school committees) may indicate when teachers use data and other instances in which we might reasonably expect them to do so. Weiss and Bucuvalas (1980) caution that in organizational life, decision-making points are not always planned or apparent; indeed, certain activities apparently occur without any explicit decision point. Organizational theory and related research warn against an overly rational view of the organization (DiMaggio & Powell, 1983; March & Simon, 1956; Meyer & Rowan, 1977; Scott, 2001; Weick, 1976). Recognizing when teachers, school leaders, and district leaders make decisions that lend themselves to formal data analysis, when they make decisions with informal data, and when decisions do not allow for acquiring additional information would help circumscribe what "data-driven" means in classroom, school, and district contexts.

Assessment specialists are also concerned about the validity of teacher-made assessments. In analyzing a special issue of the *American Journal of Education* that focused on data use, Wayman and Stringfield (2006b) found that case studies predominated and that little effort was devoted to evaluating how well teachers used data and how good those data were. Implicit in that discussion is a question about whether all practitioners' decisions and purposes require the same level of validity as that demanded of high-stakes assessments.

Many of the studies discussed here were conducted before the standards-based reform movement and the advent of the current high stakes accountability measures. Have accountability policies and the abundance of large-scale assessment data changed teachers' assessment practices? Taking a broad-based look of teachers' current assessment practices would be integral to a baseline

review of teacher work. Moreover, contemporary structural and cultural reforms (e.g., small learning communities, grade-level teams, literacy coaches, teacher collaboration, development of communities of practice and professional learning communities, instructional leadership) are frequently taken for granted. These ideas have shifted our views about how other reforms should be implemented— attention to an explicit theory of action or change, investment in capacity-building, use of collaboration as reform strategy, and focus on teaching and learning. And of course, holding schools accountable for student outcomes is the underlying premise of the argument for data-driven decision-making. To the extent that they form the context and sources of support for practitioners' work with data, these reforms have influenced approaches to researching teachers' formative uses of assessment data. A systematic look at the organizational structures (i.e., school and district levels of the system and external intermediaries) that support or hinder teachers' efforts to improve instruction by using assessment information is necessary. With such an organizational approach, is the issue of formative uses of data simply a special case of leadership, organizational change, and organizational capacity-building? Or are dimensions of organizational capacity distinct and essential to reforms in teachers' and schools' assessment work?

Using assessment and other data to improve instruction is a powerful proposition. It is, however, a rational outlook on teaching as the core technology of schooling. Our current knowledge about how teachers assess and how assessment and other data contribute to their instructional decision-making suggests both technical and normative limits to data-driven decision-making as a reform strategy. Building data management systems and developing expertise in individuals and capacity in organizations necessarily engage the nonrational aspects of the systems. Beliefs and conceptions about roles and norms of professional learning will alternately bolster or obstruct efforts to improve teachers' and schools' uses of assessment for instruction. Looking forward, a comprehensive research agenda will recognize these dual facets of data use as an organizational system and seek to understand both the potential for, and the limits of, such a system.

References

References marked with an asterisk indicate additional studies included in the literature review.

* Aschbacher, P. (1999). *Developing indicators of classroom practice to monitor and support school reform.* CSE Technical Report 513. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles.

Aschbacher, P., & Alonzo, A. (2004, April). *Using science notebooks to assess students' conceptual understanding.* Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.

* Assessment Reform Group. (1999). *Assessment for learning: Beyond the black box.* Cambridge, UK: University of Cambridge School of Education.

* Atkin, J., Coffey, H., Moorthy, S., Sato, M., & Thibeault, M. (2005). *Designing everyday assessment in the science classroom.* New York: Teachers College Press.

* Bandalos, D. L. (2004). Can a teacher-led state assessment system work? *Educational Measurement: Issues and Practice, 23*(2), 33–40.

* Black, P. (2000). Research and the development of educational assessment. *Oxford Review of Education, 26*(3 & 4), 407–419.

Black, P., & Harrison, C. (2001). Feedback in questioning and marking: The science teacher's role in formative assessment. *School Science Review, 82*(301), 55–61.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). *Working inside the black box: Assessment for learning in the classroom.* London: GL Assessment.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74.

Borko, H., Flory, M., & Cumbo, K. (1993). *Teachers' ideas and practices about assessment and instruction. A case study of the effects of alternative assessment in instruction, student learning and accountability practices.* CSE Technical Report 366. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Borko, H., Mayfield, V., Marion, S., Flexer, R., & Cumbo, K. (1997). *Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development.* CSE Technical Report 423. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

* Boston, C. (2002). *The concept of formative assessment.* College Park, MD: ERIC Clearinghouse on Assessment and Evaluation. (ERIC Document Reproduction Service No. ED470206)

* Brandom, A., Carmichael, P., & Marshall, B. (2005). Learning about assessment for learning: A framework for discourse about classroom practice. *Teacher Development, 9*(2), 201–218.

Brookhart, S. (2007). Expanding views about formative classroom assessment: A review of the literature. In J. McMillan (Ed.). *Formative classroom assessment* (pp. 43–62). New York: Teachers College Press.

Burney, D., Corcoran, T., & Lesnick, J. (2003). *Instructional coaching: A review of research.* Philadelphia: Consortium for Policy Research in Education.

* Calderhead, J. (1981). A psychological approach to research on teachers' classroom decision-making. *British Educational Research Journal, 7*(1), 51–57.

Cizek, G., Fitzgerald, S., & Rachor, R. (1995/1996). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment, 3*(2), 159–179.

Clarke, S., & McCallum, B. (n.d.-a). *Gillingham Partnership Formative Assessment Project 2000–2001: Interim Report on the second term of the project. Oral feedback and marking against learning intentions.* London: Institute of Education, University of London.

* Clarke, S., & McCallum, B. (n.d.-b). *Gillingham Partnership Formative Assessment Project 2000–2001: Final report on the third term of the project and final conclusions. Third term focus: Target setting.* London: Institute of Education, University of London.

Coburn, C. E., & Talbert, J. E. (2006). Conceptions of evidence use in school districts: Mapping the terrain. *American Journal of Education, 112*(4), 469–495.

* Coffey, J., Sato, M., & Thiebault, M. (2005). Classroom assessment-up close and personal. *Teacher Development, 9*(2), 169–184

Copland, M. (2002). *Leadership of inquiry: Building and sustaining capacity for school improvement in the Bay Area School Reform Collaborative.* Stanford, CA: Center for Research on the Context of Teaching, Stanford University.

Corcoran, T., Shields, P., & Zucker, A. (1998). *SSIs and professional development for teachers.* Menlo Park, CA: SRI International.

Cortada, J. W., & Woods, J. A., (Eds.). *The knowledge management yearbook, 2000–2001.* Boston: Butterworth Heinemann.

Datnow, A., Park, V., & Wohlstetter, P. (2007). *Achieving with data.* Los Angeles, California: Center on Educational Governance, University of Southern California.

Daws, N., & Singh, B. (1996). Formative assessment: To what extent is its potential to enhance pupils' science being realized? *School Science Review, 77*(281), 93–100.

Deming, W. E. (1982). *Quality, productivity, and competitive position*. Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Engineering Study.

Diamond, J., & Cooper, K. (2007). The uses of testing data in urban elementary schools: Some lessons from Chicago. In P. Moss (Ed.), *Evidence and decision making. 106th yearbook of the National Society for the Study of Education* (pp. 241–263). Malden, Massachusetts: Blackwell Publishing.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*, 147–170.

Driscoll, M. (1999). Crafting a sharper lens: Classroom assessment in mathematics. In M. Solomon (Ed.), *The diagnostic teacher. Constructing new approaches to professional development* (pp. 78–103). New York: Teachers College Press.

Duschl, R., & Gitomer, D. (1997). Strategies and challenges to changing the focus of assessment and instruction in science classrooms. *Educational Assessment, 4*(1), 37–73.

* Elawar, M., & Corno, L. (1985). A factorial experiment in teachers' written feedback on student homework: Changing teacher behaviour a little rather than a lot. *Journal of Educational Psychology, 77*(2), 162–173.

Elmore, R., & Burney, D. (1999). Investing in teacher learning. staff development and instructional improvement. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession* (pp. 263–291). San Francisco: Jossey-Bass.

Fennema, E., Franke, M., Carpenter, T., & Carey, D. (1993). Using children's mathematical knowledge in instruction. *American Educational Research Journal*, *30*(3), 555–583.

* Firestone, W., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis, 20*(2), 95–113.

Fleming, M., & Chambers, B. (1983). Teacher-made tests: Windows on the classroom. In W. E. Hathaway (Ed.), *Testing in the schools: New directions for testing and measurement* (pp. 29–39). San Francisco: Jossey-Bass.

Fuchs, E. (1999). *Quality in customer service* [electronic resource]. New York: McGraw-Hill. Retrieved July 7, 2003, from Stanford University library Web site : http://site.ebrary.com/lib/stanford/Doc?id=5002847.

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children, 53*, 199–208.

* Fuchs, L. S., Fuchs, D., Hamlett, C., & Stecker, P. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal, 28*(3), 617–641.

Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C., & Katzaroff, M. (1999). Mathematics performance assessment in the classroom: Effects on teacher planning and student problem solving. *American Educational Research Journal, 36*(3), 609–646.

Garet, M. S., Birman, B. F., Porter, A. C., Desimore, L., Herman, R., & Yoon, K. S. (1999). *Designing effective professional development: Lessons from the Eisenhower Program.* Washington, DC: U.S. Department of Education.

Goertz, M., Oláh, L., & Riggan, M. (2009, December). *Can interim assessments by used for instructional change?* CPRE Policy Briefs RB-51. Philadelphia: Consortium for Policy Research in Education.

Gore, A. (1999). *Quality in government services* [electronic resource]: Reinvention of the federal government. New York: McGraw-Hill. Retrieved July 7, 2003, from Stanford University library Web site: http://site.ebrary.com/lib/stanford/Doc?id=5003216

* Gullickson, A. (1984). Teacher perspectives of their instructional use of tests. *Journal of Educational Research, 77*(4), 244–248.

Gullickson, A. (1986). Teacher education and teacher-perceived needs in educational measurement and evaluation. *Journal of Educational Measurement, 23*(4), 347–354.

Hall, B., & Hewitt-Gervais, C. M. (2000). The application of student portfolios in primary/intermediate and self-contained/multi-age team classroom environments: Implications for instruction, learning, and assessment. *Applied Measurement in Education, 13*(2), 209–228.

Hall, K., Webber, B., Varley, S., Young, V., & Dorman, D. (1997). A study of teacher assessment at Key Stage 1. *Cambridge Journal of Education, 27*(1), 107–122.

* Hallinger, P., & Murphy, J. (1985). Assessing the instructional management behavior of principals. *Elementary School Journal, 86*(2), 217–247.

Halverson, R., Grigg, J., Pritchett, R., & Thomas, C. (2005). *The new instructional leadership: Creating data-driven instructional systems in schools.* WCER Working Paper No. 2005-10. Madison, WI: Wisconsin Center for Education Research (WCER).

* Hamilton, L. (2003). Assessment as a policy tool. *Review of Research in Education, 27*, 25–68.

Harrison, C. (2005). Teachers developing assessment for learning: mapping teacher change. *Teacher Development, 9*(2), 255–264.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

* Heritage, M. (2008). *Learning progressions: Supporting instruction and formative assessment.* Washington, DC: Council of Chief State School Officers. Retrieved March 28, 2010, from http://www.ccsso.org/content/PDFs/FAST%20Learning%20Progressions.pdf.

Herman, J., & Dorr-Bremme, D. (1983). Uses of testing in the schools: A national profile. In W. E. Hathaway (Ed.), *Testing in the schools: New directions for testing and measurement* (pp. 7–17). San Francisco: Jossey-Bass.

Herman, J., & Gribbons, B. (2001). *Lessons learned in using data to support school inquiry and continuous improvement: Final report to the Stuart Foundation.* CSE Technical Report 535. Los Angeles: Center for the Study of Evaluation (CSE), University of California Los Angeles.

Herman, J., Osmundson, E., Ayala, C., Schneider, S., & Timms, M. (2006). *The nature and impact of teachers' formative assessment practices.* CSE Technical report 703. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Hightower, A. M. (2002). *San Diego's big boom: District bureaucracy supports culture of learning.* Seattle, WA: Center for the Study of Teaching and Policy, University of Washington.

Impara, J. C., Plake, B. S., & Fager, J. J. (1993). Teachers' assessment background and attitudes towards testing. *Theory into Practice, 32*(2), 113–117.

Ingram, D., Louis, K. S., & Schroeder, R. G. (2004). Accountability policies and teacher decision-making: Barriers to the use of data to improve practice. *Teachers' College Record, 106*(6), 1258–1287.

Johnston, P. H., Afferblach, P, & Weiss, P. B. (1993). Teachers' assessment of the teaching and learning of literacy. *Educational Assessment, 1*(2), 91–117.

Juran, J. M. (1992). *Juran on quality by design: The new steps for planning quality into goods and services.* New York: Free Press.

Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to promote date use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education, 112*(4), 496–520.

King, S.P., & Amon, C. (2008). Assessment data: A tool for student and teacher growth. In E. D. Mandinach & M. Honey (Eds.), *Data-driven school improvement: Linking data and learning* (pp. 71–86). New York: Teachers College Press.

Knapp, M., Copland, M., & Swinnerton, J. (2007). Understanding the promise and dynamics of data-informed leadership. In P. Moss (Ed.), *Evidence and decision making. 106th yearbook of the National Society for the Study of Education* (pp. 74–104). Malden, Massachusetts: Blackwell Publishing.

* Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research, 47*(1), 211–232.

Lachat, M., & Smith, S. (2005). Practices that support data use in urban schools. *Journal of Education for Students Placed at Risk*, *10*(3), 333–349.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.

Lee, C., & Wiliam, D. (2005). Studying changes in the practice of two teachers developing assessment for learning. *Teacher Development, 9*(2), 265–283.

Leonard, J. F. (1996). *The new philosophy for K-12 education: A Deming framework for transforming America's schools.* Milwaukee, WI: ASQC Quality Press.

* Lieberman, A., & Miller, L. (Eds.). (2001). *Teachers caught in the action: Professional development that matters.* New York: Teachers College Press.

Little, J. W. (1990). The persistence of privacy: Autonomy and initiative in teachers' professional relations. *Teachers College Record, 91*(4), 509–536.

Little, J. W. (2002). Locating learning in teachers' communities of practice: Opening up problems of analysis in records of everyday work. *Teaching and Teacher Education, 18*(7), 917–946.

Little, J. W. (2003). Inside teacher community: Representations of classroom practice. *Teachers College Record*, *105*(6), 913–945.

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

Lyon. C., & Leahy, S. (2009). *Developing assessment for learning through teacher learning communities*. ETS RM-09-01. Princeton, NJ: Education Testing Service.

Lysakowski, R. S., & Walberg, H. J. (1982). Instructional effects of cues, participation, and corrective feedback: A quantitative synthesis. *American Educational Research Journal, 19*, 559–578.

Mann, D., & Shakeshaft, C. (2003, January). In God we trust; all others bring data. *School Business Affairs*, pp. 19–22.

March, J., & Simon, H. (1958). *Organizations*. New York: Wiley.

Marsh, J., Pane, J., & Hamilton, L. (2006). *Making sense of data-driven decision making in education*. Occasional paper. Santa Monica, CA: RAND Corporation.

McLaughlin, M. W., & Talbert, J. E. (2001). *Professional communities and the work of high school teaching*. Chicago: University of Chicago Press.

McLaughlin, M. W., & Mitra, D. (2003). *The cycle of inquiry as the engine of school reform: Lessons from the Bay Area School Reform Collaborative*. Stanford, CA: Center for Research on the Context of Teaching.

* McMillan, J. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practices, 19*(1), 20–32.

McMillan, J. (2002). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement, 21*(1), 34–43.

* McMillan, J. (2003). *The relationship between instructional and classroom assessment practices of elementary teachers and student scores on high-stakes tests*. Virginia Commonwealth University: Author. Clearinghouse on Assessment and Evaluation. (ERIC Document Reproduction Service No. ED 472 164)

McMillan, J. H., & Nash, S. (2000). *Teachers' classroom assessment and grading decision making*. Paper presented at the Annual Meeting of the National Council of Measurement in Education, New Orleans, LA.

McMunn, N., McColskey, W., & Butler, S. (2003–04). Building teacher capacity in classroom assessment to improve student learning. *International Journal of Educational Policy, Research, & Practice, 4*(4), 25–48.

Means, B., Padilla, C., DeBarger, A., & Bakia, M. (2009). *Implementing data-informed decision making in schools—teacher access, supports and use*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation, and Policy Development.

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology, 83,* 340–363.

Moody, L., & Dede, C. (2008). Models of data-based decision making: A case study of the Milwaukee Public Schools. In E.D. Mandinach & M. Honey (Eds.), *Data-driven school improvement: Linking data and learning* (pp. 233–254). New York: Teachers College Press.

Morrison, J., & McDuffie, A. (2003, April). *Preservice teachers' development and implementation of science performance assessment tasks*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

* Murphy, S., Bergamini, J., & Rooney, P. (1997). The impact of large-scale portfolio assessment programs on classroom practice: Case studies of the new standards field-trial portfolio. *Educational Assessment, 4*(4), 297–333.

National Staff Development Council. (1995). *National Staff Development Council's standards for staff development.* Oxford, OH: Author.

Neesom, A. (2000). *Report on teachers' perception of formative assessment.* Coventry, UK: Qualifications and Curriculum Authority.

Nickols, F. (2000). The knowledge in knowledge management. In J. W. Cortada & J. A. Woods (Eds.), *The knowledge management yearbook, 2000–2001* (pp. 12–21). Boston,: Butterworth Heinemann.

Noll, V. H. (1955). Requirements in educational measurement for prospective teachers. *School and Society, 82,* 88–90.

O'Dell, C., & Grayson, C. J. (1998). Part I: A framework for internal knowledge transfer. In *If only we knew what we know: The transfer of internal knowledge and best practice* (pp. 3–30). New York: Free Press.

* Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning. *Assessment in Education: Principles, Policy, and Practice, 5*, 85–102.

* Plake, B., Impara, J., & Fager, J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice, 12*(4), 10–12.

Redman, T. C. (1999). *Measurement, information, and decision making* [electronic resource]. New York: McGraw-Hill. Retrieved July 7, 2003, from Stanford University library Web site: http://site.ebrary.com/lib/stanford/Doc?id=5002834.

* Ruiz-Primo, M., & Furtak, E. (2004). *Informal formative assessment of students' understanding of scientific inquiry.* CSE Report 639. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

* Sadler, D. (1982). Evaluation criteria as control variables in the design of instructional systems. *Instructional Science, 11*, 265–271.

* Sadler, D. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119–144.

* Sato, M. (2003). Working with teachers in assessment-related professional development. In J. M. Atkin & J. E. Coffey (Eds.), *Everyday assessment in the science classroom* (pp. 109–120). Arlington, VA: NSTA Press.

* Schafer, W. D., & Lissitz, R. W. (1987). Measurement training for school personnel: Recommendations and reality. *Journal of Teacher Education, 38*(3), 57–63.

Schmoker, M., & Wilson, R. (1993). *Total quality education: Profiles of schools that demonstrate the power of Deming's management principles.* Bloomington, IA: Phi Delta Kappa Educational Foundation.

* Schunk, D. H., & Rice, J. M. (1991). Learning goals and progress feedback during reading comprehension instruction. *Journal of Reading Behavior, 23*, 351–364.

Scott, W. R. (2001). *Institutions and organizations* (2nd ed.). Thousand Oaks, CA: Sage.

Shavelson, R. (2003). *On the integration of formative assessment in teaching and learning with implications for teacher education.* Stanford, CA, and Manoa, HI: Stanford Education Assessment Laboratory and University of Hawaii Curriculum Research and Development Group.

Shavelson, R. J., & Stern, P. (1981). Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. *Review of Educational Research, 51*(4), 455–498.

* Shepard, L. (2000). *The role of classroom assessment in teaching and learning.* CSE Technical Report 517. Los Angeles: Center for the Study of Education, National Center for Research on Evaluation, Standards, and Student Testing.
Smylie, M. (1990). Teacher leadership: Tensions and ambiguities in organizational perspective. *Educational Administration Quarterly, 26*(3), 235–259.
Stiggins, R. (1991). Relevant classroom assessment training for teachers. *Educational Measurement: Issues and Practice, 10*(1), 7–12.
Stiggins, R., & Bridgeford, N. (1985). The ecology of classroom assessment. *Journal of Educational Measurement, 22*(4), 271–286.
* Stiggins, R., & Bridgeford, N. (1986). Classroom assessment: A key to effective education. *Educational Measurement: Issues and Practice, 5*(2), 5–17.
* Stiggins, R., & Conklin, N. (1992). *In teachers' hands. Investigating the practices of classroom assessment.* Albany, NY: State University of New York Press.
* Stiggins, R., Conklin, N., & Bridgeford, N. (1986). Classroom assessment: A key to effective education. *Educational Measurement, 5*(1), 5–17.
Stiggins, R., Griswold, M., & Wikelund, K. (1989). Measuring thinking skills through classroom assessment. *Journal of Educational Measurement, 26*(3), 233–246.
Stokes, L. (2001). Lessons from an inquiring school: Forms of inquiry and conditions for teacher learning. In A. Lieberman & L. Miller, (Eds.), *Teachers caught in the action: professional development that matters* (pp. 141–158). New York: Teachers College Press.
Streifer, P. (2001, April). The 'drill down' process. *The School Administrator,* pp. 16–19.
Supovitz, J., & Klein, V. (2003). *Mapping a course for improved student learning: How innovative schools systematically use student performance data to guide improvement.* Philadelphia: Consortium for Policy Research in Education.
Thorn, C. A. (2002, April). *Data use in the classroom: The challenges of implementing data-based decision-making at the school level.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA. Retrieved April 22, 2002, from http://www.wcer.wisc.edu/mps/AERA2002/data_use_in_the_classroom.html.
Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal, 27*(5), 615–631.
Tunstall, P., & Gipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal, 22*(4), 389–404.
* Vacc, N., & Bright, G. (1999). Elementary preservice teachers' changing beliefs and instructional use of children's mathematical thinking. *Journal for Research in Mathematics Education, 30*(1), 89–110.
van Zee, E., & Minstrell, J. (1997). Using questioning to guide student thinking. *Journal of the Learning Sciences, 6*(2), 227–269.
Wayman, J.C., Cho, V., & Johnston, M.T. (2007). *The data-informed district: A district-wide evaluation of data use in the Natrona County School District.* Austin: University of Texas.
Wayman, J. C., & Stringfield, S. (2006a). Technology-supported involvement of entire faculties in examination of student data for instructional improvement. *American Journal of Education, 112*(4), 549–571.
Wayman, J. C., & Stringfield, S. (2006b). Data use for school improvement: School practices and research perspectives. *American Journal of Education, 112*(4), 463–468.

Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly, 21*, 1–19.

Weinbaum, E. (2009). *Learning about assessment. An evaluation of a ten-state effort to build assessment capacity in high schools.* Research report RR-61. Philadelphia: Consortium for Policy Research in Education.

Weiss, C., & Bucuvalas, M. (1980). *Social science research and decision making.* New York: Columbia University Press.

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

Wiliam, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal, 22*(5), 537–548.

Wiliam, D., & Leahy, S. (2006, April). *A theoretical foundation for formative assessment.* Paper presented at the American Educational Research Association and the National Council on Measurement in Education, San Francisco, CA.

Wininger, S., & Norman, A. (2005). Teacher candidates' exposure to formative assessment in educational psychology textbooks: a content analysis. *Educational Assessment, 10*(1), 19–37.

* Wolf, D., Bixby, J., Glenn, III, J., & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. *Review of Research in Education, 17*, 31–74.

Young, V. M. (2006). Teachers' use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education, 112*(4), 521–548.

Young, V. M. (2008). Supporting teachers' use of data: The role of organization and policy. In E. B. Mandinach & M. Honey (Eds.), *Linking data and learning* (pp. 87–106). New York: Teachers College Press.

About the Authors

Viki M. Young
SRI International

Debbie H. Kim
Northwestern University

Email: viki.young@sri.com

Viki M. Young is a senior researcher in the Center for Education Policy at SRI International. Her research focuses on organizational supports for teachers' use of data, high school reform, and district and state policies supporting school improvement.

Debbie Kim is a doctoral student in Human Development and Social Policy at Northwestern University and previously was a research analyst in SRI International's Center for Education Policy. Her research interests include investigating people's understandings and subsequent implementation of education policy and the ways organizational factors may influence their actions.

education policy analysis archives

Volume 18 Number 19 August 20th, 2010 ISSN 1068–2341