

Exceptional Children

<http://ecx.sagepub.com/>

The Political Dilemmas of Formative Assessment

Sherman Dorn

Exceptional Children 2010 76: 325

DOI: 10.1177/001440291007600305

The online version of this article can be found at:

<http://ecx.sagepub.com/content/76/3/325>

Published by:



<http://www.sagepublications.com>

On behalf of:

Council for Exceptional Children



Additional services and information for *Exceptional Children* can be found at:

Email Alerts: <http://ecx.sagepub.com/cgi/alerts>

Subscriptions: <http://ecx.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

>> [Version of Record](#) - Apr 1, 2010

[What is This?](#)

The Political Dilemmas of Formative Assessment

SHERMAN DORN

University of South Florida

ABSTRACT: *The literature base on using formative assessment for instructional and intervention decisions is formidable, but the history of the practice of formative assessment is spotty. Even with the pressures of high-stakes accountability, its definition is fuzzy, its adoption is inconsistent, and the prognosis for future use is questionable. A historical and organizational perspective explores plausible explanations for that inconsistency. These possible explanations include the standard literature in the research-to-practice gap and must also include the current policy environment surrounding educational accountability. A number of organizational and historical/cultural hypotheses suggest potential limits of structured formative assessment. From several perspectives, the practical question that may shape the future of special education is the identity of individuals at the school level who are responsible for coordinating the collection of formative-assessment data, for analysis, and for responses to the data.*

Respondents told us that teachers had to be willing to acknowledge both strengths and weaknesses and be willing to openly discuss these with colleagues. In addition, organizational cultures that viewed accountability as helpful rather than threatening enabled complex DDDM [data driven decision-making] processes. (Ikemoto & Marsh, 2007, p. 124)

An observer of American education might well think that the most underrated achievement of the No Child Left Behind Act (NCLB; 2002) is the expansion of formative assessment. Many states and districts have begun using formative phonological awareness and pre-/early reading assessments such as the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) as part of

Reading First contracts. Many districts encourage or require that students take tests that are supposed to be shorter or parallel equivalents to the state tests several times each school year. And in late summer 2007, Representative George Miller released draft language for the reauthorization of NCLB that would explicitly provide incentives for school systems to use and expand formative assessment (U.S. House Committee on Education and Labor, 2007).

To researchers who have pushed for the use of formative assessment in education, these events all look positive for students with disabilities. Several decades worth of research on formative assessment point in a similar direction: When used properly, formative assessment is one of the most powerful tools available to guide classroom decisions (e.g., Black & Wiliam, 1998). Not only is formative assessment powerful, it is also relatively inexpensive

and arguably a more cost-effective intervention than more widely discussed alternatives (Yeh, 2008). At the classroom level, one could argue that formative assessment is the type of “tinkering reform” that Tyack and Cuban (1995) advocated: It uses the existing structure of the classroom teacher and the curriculum rather than fighting against the existing structure.

The problem with this rosy perspective lies partly in the practical issues with formative assessment—the coverage of the curriculum at various grade levels by assessments that have the right technical characteristics, the support by district and state leaders often distracted by many other matters, appropriate professional development, intraschool dynamics, and broader teacher uses of assessment (e.g., Roehrig, Duggar, Moats, Glover, & Mincey, 2008; Sharkey & Murnane, 2006; Wiliam, 2006). The practical concerns are sometimes in the relationship between schools and other entities; for example, when local districts are more likely to contract with commercial testing firms, will expertise and resources be sufficient to develop formative assessment? Toch (2006) argued that the expansion of state testing programs and district attempts to test children several times a year was leading to an exhaustion of the commercial test industry’s capacity. But the barriers to the expansion of formative assessment extend beyond practical issues. Organizational, political, and cultural frictions have occurred with the development of formative assessment. Although *data-driven decision making* is a common education buzzword, formative assessment may conflict with the way that schools work, the shape of public discourse around education policy, and how a plurality of Americans think about tests.

Despite the language about “data-driven decision making,” is the public ready for a school system whose core values and practices revolve around decision making? As implied by the opening epigram, teachers’ willingness to ask hard questions of their own instruction can depend on the existence of a “trusting, data-driven culture” (Ikemoto & Marsh, 2007, p. 124). To gauge whether the optimism described at the beginning of this article is warranted, one must understand the range of responses to formative assessment and different frameworks for understanding that range.

FORMATIVE ASSESSMENT AND ITS CONTENTS

Formative assessment and its variations are the local, classroom version of formative evaluation (Scriven, 1967). Young and Kim (in press) point out that the term *formative assessment* is flexible, perhaps to the point of losing meaning. Those advocating formative assessment have used it to mean anything from informal or haphazard in-class judgment to frequent quantitative measures of specific skills. For the purposes of this article, *structured formative assessment* refers to the latter. Even in its more general sense, the spirit of formative decision making is that one can make adjustments to a classroom or a program in response to early feedback, and such adjustments can dramatically improve long-term outcomes. Depending on whether such formative data are coupled to guidance about instruction, one can see formative decision making either as an independent inductive loop for teachers (when teachers have considerable freedom in which to experiment in response to data) or as something more similar to a GPS receiver, where teachers receive explicit guidance when the path of structured formative assessment data vary from a predicted/preferred route. (Doug Fuchs introduced the inductive-experimental metaphor to me in 1993. Also see Mandinach, Honey, & Light, 2006, for a more generalized description of the range of attitudes toward data.) As subsequently argued, both metaphors are at odds with the organizational behavior of school systems and with our political debates about schooling, and the consequences for coordinating special education services may be profound.

Despite the language about “data-driven decision making,” is the public ready for a school system whose core values and practices revolve around decision making?

A substantial body of evidence suggests that when teachers respond to structured formative assessment—when they base decisions on whether children’s performances improve by a reasonable amount—children with low achievement can close a large portion of the achievement gap

(Deno, 1985, 2003; Fuchs, 2004). Stan Deno generally receives credit for curriculum-based measurement or progress monitoring as a practical, efficient use of frequent data probes in academic areas, especially with students labeled with specific learning disabilities. The research on curriculum-based measurement echoes research on formative uses of classroom assessment more broadly (Black & Wiliam, 1998). Curriculum-based measurement has two key characteristics: (a) the capacity for local adaptation and (b) the ability to use it in a low-stakes context. First, curriculum-based measurement exists within a local context, ideally creating tests (or probes) that sample from a curriculum range rather than from a narrow set of skills. In theory, teachers can create their own set of probes on the basis of the prevailing curriculum standards. In practice, creating probes with the right technical characteristics is something that few teachers have the skills to do efficiently in addition to the central part of their job. However, both commercial instruments and online tools are available (e.g., Wright, n.d.). The second characteristic of curriculum-based measurement is that it does not necessarily use tests for individual high-stakes purposes but instead uses test data to guide instructional decisions. This focus does not inherently preclude high-stakes use of test data, but it does not require it, either. These characteristics may derive from curriculum-based measurement's origins in federally funded research and development of classroom techniques.

Despite its origins in classroom-based research, debates about formative assessment currently occur at the level of policy, not just classroom decision making or supervision at the school and district level. At the most visible level, formative assessment is part of the discussion about the future of NCLB. The August 2007 discussion draft for NCLB's reauthorization included a new requirement of school improvement plans, a requirement that improvement plans include

[the] current use of (or lack of use of) formative assessments and data-based instructional decision making to determine how changes to such formative assessments and data-based instructional decision making could address causes for the school not making adequate

yearly progress. (U.S. House Committee on Education and Labor, 2007, p. 178)

The phrase *formative assessment* appeared in other parts of the draft and referred to professional development for schools not meeting adequate yearly progress and in several Title I grant programs proposed in the draft. Although debate over NCLB reauthorization is on hiatus on Capitol Hill, formative assessment may well return as a central issue in 2010, when both houses of Congress and the new administration are likely to bring up NCLB reauthorization for serious consideration.

Structured formative assessment is also crucial to the development of response-to-intervention (RTI) frameworks for general and special education intervention services (e.g., Busch & Reschly, 2007). The Fuchs, Stecker, and Fuchs article in this issue focuses on a crucial concern for RTI: the granularity of intervention tiers and the definition of special education. Underlying RTI is a core assumption that choosing the appropriate tier of services is a decision based on a feedback loop: See where the child is and is heading, intervene, see where the child is *now* heading, make a new decision, and so forth. It is the programmatic implementation of structured formative assessment in which the individual child's intervention intensity is the focus. The details of that system are crucial; the general understanding of formative assessment and its power makes possible the policy debate over RTI.

That general understanding exists because formative assessment appeals to disparate groups of people. In appealing to different interests, the different ideas of formative assessment comprise what historians of education reform recognize as a malleable concept, a flexible ideal that can appeal to people and groups with different interests and priorities (e.g., Cuban, 1990). Formative assessment appeals to those who like to see teachers as professionals and intellectuals, because decision making can lie in the hands of skilled teachers (e.g., Shepard, 2000). (That end of the spectrum coincides with the previously described inductive-experimental metaphor.) Formative assessment appeals to researchers and supervisors who have been working in the trenches and seeking ways to start productive conversations with

teachers about instruction (e.g., Brunner et al., 2005). Structured formative assessment appeals to those with small-scale technocratic temperaments who are hoping to find data that can guide instructional decision making on a frequent and firm basis. (That end of the spectrum coincides with the GPS-receiver metaphor.) Finally, structured formative assessment appeals to those with grand designs of systems for whom iterative feedback loops form the sine qua non of problem solving.

Although formative assessment is appealing in theory, its practice as well as its definition is inconsistent (e.g., Young & Kim, in press). Studies that describe data-driven decision making exist (Brunner et al., 2005; Chen, Heritage, & Lee, 2005; Firestone & Gonzalez, 2007; Ikemoto & Marsh, 2007; Lachat & Smith, 2005; Mandinach, Honey, et al., 2006; Mandinach, Rivas, Light, Heinze, & Honey, 2006; Wayman, 2005); however, no refereed study since the passage of the NCLB describes the nationwide use of structured formative-assessment implementation. Although when this manuscript was written, one might have suspected that Reading First promoted the extensive use of some form of formative assessment in thousands of schools, no documentation discussed the fidelity, regularity, and use of structured formative assessment. The history over the past few decades must therefore be an initial guide. One can cautiously conclude that implementation of structured formative assessment has depended on local initiative, often most successful and most institutionalized where some type of formative assessment was the focused project of key individuals, often student services or supervisors of special educators. This impression is consistent with the most recent general review of the literature (Stecker, Fuchs, & Fuchs, 2005), which also suggests that teachers by themselves will not uniformly respond to data by changing instruction: The strongest claims for improvement in carefully designed studies have occurred when structured formative assessment couples with strong guidance on instruction. The question for formative assessment in the past 5 years is whether top-down incentives have created any more faithful implementation than the intervention research conducted during the previous 20 years.

The limited evidence from the previous 6 years is not reassuring. With the debates over Reading First, little attention has focused on implementation issues. The most visible implementation research surrounding NCLB has instead been on the cascade of consequences for failure to meet adequate yearly progress.

From both supporters and opponents of formative assessment (there are some opponents) have come hints that significant numbers of educators misunderstand the purpose of formative assessment. One state technical assistance center warns principals not to use DIBELS scores for retaining students or evaluating teachers (Florida Center for Reading Research, 2005); this warning implies that significant numbers of principals involved in Reading First grants in the state were using DIBELS scores inappropriately. Goodman (2006) reports anecdotal evidence of teaching to the compartmentalized subtests of DIBELS and of students being retained on the basis of DIBELS scores. These reports from very different perspectives on teaching reading imply that formative assessment in the NCLB era is not guaranteed to proceed any better than in the pre-NCLB era. Also troubling is the disconnect between special education research and the growing body of literature that is self-conscious about data-driven decision making. Of several dozen recent papers on the subject, including a 2005 special issue of the *Journal of Education for Students Placed at Risk* and the 2007 *Yearbook of the National Society for the Study of Education*, none referenced the existing research in special education. Even when the topic was as specific as using technology to guide teacher interpretation and use of data (Chen et al., 2005; Mandinach, Rivas, et al., 2006; Wayman, 2005), the researchers did not use the existing literature in that area.

ORGANIZATIONAL, HISTORICAL, AND CULTURAL FRAMEWORKS

Given the disconnect between the special education research conducted over the past 20 years and the newer research on data-driven decision making, several years will probably pass before educators produce and publish valuable research on the uses of formative assessment as it has ex-

panded in the past decade. But some hypotheses are possible, and this section focuses on organizational explanations of varying implementation and on historical patterns that may provide useful perspective.

ORGANIZATIONAL FRAMEWORKS

One can start to think about implementation issues with the literature on organizational behavior of schools and reform implementation. Four potential explanations of varying implementation come from the literature on organizational behavior, pointing to predictions that range from the relatively optimistic and malleable to the more pessimistic. These explanations include looking at implementation as mediated by the presence or absence of entrepreneurial individuals or units within a school system, systemic gaps between tightly controlled research and quickly scaled-up programs, the behavior of educators as street-level bureaucrats, and the role of testing as part of a broader organizational repertoire that removes responsibility for learning from schools and places it onto children and their families. Each perspective presents plausible and competing ways to look at formative assessment.

Change-Agent Entrepreneurialism. On the optimistic end of these questions is the claim that individuals and organizations have the power to change organizational behavior. In this view, both organizations and individuals can operate in an entrepreneurial fashion within a bureaucracy or as external change agents sensitive to local contexts (e.g., McLaughlin, 1990). This view implies that key individuals can mediate organizational use of formative assessment while they keep data-driven decision making on the active agenda in key meetings, in decisions on professional development, and in changing the discourse within an organization. This claim also implies that organizational changes can have substantial effects on outcomes, a statement that may be true for formative assessment, which does not inherently require dramatic changes in school routines, at least to start collecting data. Moreover, from this perspective, variation in implementation is a positive good, a sign of mutual adaptation. But an entrepreneurial perspective also suggests that the power of individuals or smaller components may

weaken when the change must be significant—not just collecting data, but perhaps making substantial changes in organizational behavior in response to the formative assessment.

Bridging Gaps and Scaling Up. The scale of change is a focus of the research-to-practice gap and scaling-up literature (e.g., Carnine, 1997; DeStefano, Dailey, Berman, & McInerney, 2001; Fullan, 2002). The key focus in this literature is the difference between specialized, targeted programs and wholesale change within a school system or state. Different explanations in the literature concentrate varyingly on the difference between academic and K–12 cultures, the difference between a marginalized (and buffered) program and institutional change, the question of institutionalization and long-term commitments, parallels to changing standards of care and practice in other fields, and the proper leverage points for researchers to guide change in states and districts. The scaling-up perspective on formative assessment implies that variations inevitably occur in the implementation and use of formative assessment because institutional commitments vary and the understanding of formative assessment diffuses from a research community into a school culture. From the perspective of structured formative-assessment advocates, variations in scaling up represent a problem of fidelity. But from the perspective of a researcher, they are a golden opportunity to explore key factors in success.

Loose Coupling in Systems Change. Variation in implementation is not the focus of researchers who see schools as loosely coupled systems (Weick, 1976) or educators as street-level bureaucrats (Weatherley & Lipsky, 1977). Instead, this view of schools as messy systems implies that implementation faces inevitable limits. Local decision making in large organizations limits even the best ideas or most important principles. Teachers can and do respond to directives in diverse ways, in part because they can: Teachers are far too autonomous in practice to order not only procedures but also the correct use of those procedures. In the view of educators as street-level bureaucrats, teachers and principals must make choices among absolute priorities because they have limited resources, time, and energy. No school can educate all children perfectly in this view; and as a result, the day-to-day decisions

about implementing any reform water down every reform, including the use of data (e.g., Lachat & Smith, 2005). In addition, educators often face competing, inconsistent demands; and attempts to satisfy the inconsistencies constrain their choices.

Teachers and principals are often autonomous for other reasons. Education is labor-intensive, and many schools have few choices about whom to hire; alleged information flows out of schools at a pace that inhibits verification by outsiders; those outside school systems themselves have limited resources and limited time to pressure schools to behave differently; and they rarely agree on evaluations of schools. If this perspective on school reform is correct, structured formative assessment will have relatively limited fidelity. Less variation will occur than what a change-agent view or a scaling-up view predicts because of the inherent limits of change on educators at the local level. This local-bureaucracy view suggests that the easiest response is a superficial compliance with formative-assessment mandates, lip service to the value of data-driven decision making but lacking dramatic changes in instruction. That perspective may not be fatal for formative assessment; as noted in the discussion on entrepreneurialism, the initial steps in formative assessment do not require dramatic changes. But this perspective also suggests that implementing formative assessment may require small steps taken over several years rather than a sweeping and sudden change.

Displacing Responsibility. The most pessimistic view of formative-assessment implementation belongs to those who see school-system behavior as displacing responsibility away from schools and toward students (e.g., Calhoun, 1973; Dreeben, 1968). The displacement argument is that educators use assessment to label individuals, not only their achievement in one area and progress but also their character or other (presumably) innate qualities such as intelligence or characteristics of their families and communities. Formative assessment's promise is that testing can break the zero-sum discourse that blames either schools (and teachers) or students (and their families) for low achievement. An expansion of structured formative assessment could induce a shift away from testing students and toward test-

ing instruction. For subsequently explained reasons, this explanation does not fit well within the broader culture's understanding of testing. More concretely, the displacement hypothesis posits that longstanding organizational culture uses testing as one tool in a broader repertoire of displacement. In the same way that the person-first language of *students with special needs* proposed in England quickly became *special-needs children* (e.g., Tomlinson, 1982), school-system behavior in the United States possesses the capacity and the habit of displacing blame through a testing environment.

The different organizational perspectives on implementing structured formative assessment have consequences for the future of special education as an organizational entity within school systems. (A reviewer's helpful observation about formative-assessment implementation suggested this issue.) Both an entrepreneurial perspective and a scaling-up perspective might predict that educators can promote implementation by assigning a single employee in a school to have responsibility for coordinating formative assessment. The identity of that coordinator could vary and might have important consequences for RTI. Assigning a special educator as a schoolwide formative-assessment coordinator is consistent with preserving special education services as a third tier in RTI (see Fletcher & Vaughn, 2009, for a broader context on the RTI debate). Assigning a counselor as schoolwide formative-assessment coordinator is consistent with the way that many schools currently administer annual state-mandated testing but has very different consequences for RTI and special education.

HISTORICAL AND CULTURAL PERSPECTIVES

The perspectives outlined in the previous section focus on organizational responses to reform in an abstract sense. But one can also bring historical and cultural perspectives to bear on the process of education reform. In one case, with hybridization, a historical argument closely relates to one of the organizational perspectives, on the role of educators as street-level bureaucrats who must choose among goals that can never be completely attained. In the other case, with a test-preparation culture, a historical argument connects the behav-

ior of schools with a broader conflict about the purpose of schooling and a tension between the collective, public goals of schooling and the private goals that individual parents may have. The argument in the second case is that a mechanism to advance private goals (test preparation) has spilled over into generally acceptable behavior for schools.

Hybridization. Cuban (1993) argues that teachers have commonly hybridized classroom reform efforts, mixing older routines with new behaviors. Reform of instruction often has broadened teachers' repertoires rather than shifting teachers dramatically from older routines. Cuban notes that for most of the 20th century, elementary teachers were more willing to experiment in changing the structure of the classroom, in part because elementary teachers have not faced the same curriculum pressures to which academic teachers in high schools have responded. Historically, external expectations have mediated the extent to which teachers have changed in response to reforms. That mediating role has existed not just for teachers but for schools and school systems. The response of schools and systems to reform was the subject that Tyack and Cuban (1995) tackled. In many ways, they broadened Cuban's previous argument from the classroom to the school and school district. To them, both school-system organizations and American culture have established expectations of schools, what they call an implicit grammar that inhibits change. Metz (1990) similarly argued that high schools often operate under a common script even if the quality of education varies dramatically.

As a result of this grammar or script, schools have often changed only to the extent that the changes conform to broader expectations of what a school is and should be doing. The tendency of American schools to hybridize reform coincides with the view of educators as street-level bureaucrats, having to parse out the portions of reforms that are adaptable within the existing structure of an organization or an individual's work. This piecemeal adoption of reform can be significant, as with the expansion of rights for students with disabilities (Cuban, 1996). But Tyack and Cuban (1995) point out that school routines modify reforms as much as reforms change school practices. This history suggests that the expansion of forma-

tive assessment in schools will modify structured formative assessment as much as it might change the existing structure of schooling.

A Culture of Test Preparation. A darker view of American schooling starts from the assumption that the culture of test preparation dominates our society's view of testing, with test preparation defined operationally as superficial techniques designed primarily to raise test scores and not as substantively educational in themselves (Nichols & Berliner, 2007). Test preparation has existed for some time, most visibly in connection with admissions tests to colleges, as well as graduate and professional schools. A recognizable part of any bookstore is the section on test preparation for the SAT, ACT, LSAT, MCAT, GRE, GED, TOEFL, and AP examinations, as well as a variety of professional licensure examinations. Parallel versions exist for state K–12 tests, but they are the newcomers in the test-prep genre. Companies such as Princeton Review and Kaplan also earn millions of dollars annually through test-preparation workshops. Test preparation is a part of modern culture in the United States, and eliminating the most damaging aspects of high-stakes accountability would not eliminate the culture of test preparation.

Historian and sociologist David Labaree (1988, 1997) observed that the culture of test preparation comes from the private uses of public schooling. Parents believe that part of the job of schooling is to prepare their children for success in life. That private goal of parents parallels but is not identical to the human capital argument that investing in education makes people more productive and is socially useful. For millions, the purpose of education is reduced to earning a degree or succeeding in school without really learning, as Labaree argues. With this instrumentalist view of education, test preparation is a positive good because it achieves an aim with significant payoff, or at least avoidance of pain. The superficiality of test preparation demonstrates the extent to which the private mobility goals of education have displaced other goals.

Test preparation can include such classroom activities as drills focused on specific questions known or anticipated to be on a high-stakes test and the redesign of classroom activities around the formats of tests, such as teachers asking questions

or requiring responses in the format of the high-stakes tests; such activity outside the classroom as all-school assemblies before high-stakes tests; or even changing school breakfast and lunch programs (Smith, 1991). Test preparation can be short-term, such as the changes in school meals that Figlio and Winicki (2005) analyzed; and it can be long-term, such as when classroom routines change to suit the vagaries of specific tests. If test preparation dominates American views of schooling and achievement, the future of formative evaluation techniques is closely related to the fate of a test preparation culture. When the immediate consequence of an exam drives the prospective response of test preparation, attempts to use tests as feedback to reshape instruction face an uphill battle.

A "USER'S" PERSPECTIVE, SCALE, AND FIT

Educators often compartmentalize the discussion of formative assessment and data-driven decision making, even within a systems perspective, so that they view it as a component of schooling that they can analyze as a separate part. As McNeil, Coppola, Radigan, and Vasquez Heilig (2008) have noted, the lived experience of education reform does not conform to this compartmentalized perspective. The perspective of educators and students puts this component alongside the other routines, structures, and occasionally chaos of the entire school environment. This in situ reality of any reform effort is the true origin of the warnings given by the previous perspectives: Teachers have limited time, competing goals and distractions pull them, and they operate in environments whose values may conflict with the assumptions of either structured formative assessment or more informal implementations of formative assessment.

In some ways, formative assessment can fit comfortably within existing school routines, and clever researchers have figured out multiple ways to make structured formative assessment efficient and nondisruptive, with clear guidance for instructional decision making. Instead of focusing on a relatively lean fit, the discourse around data-driven decision making revolves around the condensation of and response to massive datasets, far from the work of structured formative assessment

that special educators have developed. Data-driven decision making is often discussed as if it is an activity of upper-level administrators in any system. In addition to the previously described general concerns, one may reasonably worry that the discourse of data analysis will distance teachers from the process, undercut the legitimacy of formative assessment, and reinforce the tendency to see all data as high-stakes and something to game (e.g., Thorn, Meyer, & Gamoran, 2007). The main effort for most schools should be in asking hard questions about instruction and asking them regularly. One may reasonably worry that in combination with our test-preparation culture, high-stakes accountability may distract from that effort as much as encourage it.

The main effort for most schools should be in asking hard questions about instruction and asking them regularly.

CONCLUSIONS

The gap between the special education research literature on formative assessment and the data-driven literature is both astonishing and worrisome. One of the aspirations for the disaggregation of data within the NCLB framework and the pushing of RTI toward general education has been a new resurgence of hope that general education will take the needs of underperforming children more seriously. One side of this concern is the practical aspect: Will general educators broaden their repertoire of instructional techniques and how they respond to information about children? But another side of this concern is the relationship between research communities. To see the same issues raised in the previous 3 years that have been raised and explored for 20 years is discouraging.

There is a more optimistic potential. One could design an accountability system that was based entirely on structured formative assessment. A state could require that teachers collect formative data weekly (or twice a week for children who need extra monitoring) and make monthly instructional decisions on the basis of the information. As New York's law suggests, administrators

could evaluate teachers on their use of structured formative assessment—did they base their instructional decisions on the information, and did they follow through on the decisions? In turn, upper-level administrators could evaluate principals and their supervisors on their supervision of structured formative assessment in schools. Any high-stakes decisions about teachers and administrators could be the endpoint of cumulative formative assessment; the basis for these decisions could be whether the educator responded to the feedback that her or his supervisor gave. Because the root of such a system would be solid research, there is at least some reason to believe that it would have a greater effect on student achievement than the current systems of accountability.

In reality, this shift in professional responsibilities would encounter both political and organizational resistance. Politically, defenders of high-stakes testing might well attack formative assessment as lowering standards and lacking teeth. Such an occurrence happened in New York in spring 2008, when the state teachers' unions pushed for a moratorium on using test scores alone to make tenure decisions. Although statutory language required that teachers demonstrate that they could use test data to make instructional decisions, opponents of the moratorium have generally described it as a firewall between test scores and personnel evaluation.

Even if formative assessment becomes acceptable politically, organizational barriers still hinder making the uses of structured formative assessment a professional standard. According to researchers and teachers, teachers perceive that frequent assessment for structured formative purposes is a great paperwork burden and that they would need additional assistance with the logistics of frequent assessment and a supportive culture (e.g., Hasbrouck, Woldbeck, Ihnot, & Parker, 1999; Roehrig et al., 2008). Such concerns may mask a more fundamental basis for resistance: The culture of schools and the school year are at odds with structured formative assessment. Teachers plan lessons and years by discrete units of time—a 2-week unit on addition with regrouping, a month on *Romeo and Juliet*, and so forth. The school year unfolds as a sequence of activities. Formative assessment is a different way to organize a year; it involves planning around

instructional decisions instead of discrete chunks of time. At best, responses to formative assessment would be a layer on top of existing curriculum. Would hybridization allow structured formative assessment to flourish, or would hybridization strip it of any productive use?

Formative assessment is a different way to organize a year; it involves planning around instructional decisions instead of discrete chunks of time.

In addition, the history of high-stakes testing creates two additional barriers to the effective use of structured formative assessment. Those barriers come from people who are opposed to most standardized testing and from educators who think they have accommodated testing appropriately. As described in Dorn (2007), thousands of parents and teachers possess lingering suspicions that any test with a quantifiable score threatens the integrity of education. Moreover, some educators believe that they are accommodating high-stakes testing by teaching to the test and using test scores for crucial decisions about student opportunities and teachers. Those individuals would and can undermine structured formative assessment. Some educators have socialized themselves too well and see every test as high-stakes and every test as something to which they should teach. But that view is wrong, and structured formative assessment can only be successful if students work hard on tests that have relatively low stakes. The fact that test preparation is an endemic feature of schooling today endangers the future of formative assessment.

Despite these obstacles, accountability centered on structured formative assessment, both for children with disabilities and their peers, is worth fighting for. School reform is a back-and-forth process, with steps forward and then some steps backward; and both often happen at the same time. The place of structured formative assessment in the reauthorization of NCLB has more consequences for the rest of us than may be apparent: It is a point at which the public must decide whether rationality or bottom-line results should be the primary way to judge educators.

The current requirement for adequate yearly progress is geared to an endpoint rather than the informed decision making of structured formative assessment. A “trusting, data-driven culture” that Ikemoto and Marsh (2007) advocate requires a rational-decision perspective rather than a bottom-line perspective.

In the near future, we need studies that indicate how educators understand and respond to different types of formative assessment when the assessment is conducted outside rigorous supervision and without munificent technical assistance and professional development. Brunner et al. (2005), in their case study of how teachers responded to summaries of student achievement called the Grow Reports, is an example of an early study of such responses. Brunner et al. reported that approximately one third of teachers used the reports regularly, generally at the beginning of the year (because the reports lagged the test by 6 months) and for general allocation of efforts (focusing on skills that teachers perceived were weak at the beginning of the year). Roehrig et al. (2008) is another example of such a study; in that study, the researchers observed a range of teacher attitudes and the importance of the broader school culture beyond a school’s reading coach. But more research is necessary on the use of structured formative assessment on the ground. We know little about what principals, general educators, and special educators do with that data, and we need answers to a set of interlocking questions. The first questions are, How do educators think about structured formative assessment? How many classroom educators understand not only in the abstract that they can use structured formative assessment to adjust instruction during the year (not possible in the environment studied by Brunner et al., 2005) but also see a responsibility on their own part to regularly set aside time to review data and revise instructional plans accordingly? Crucial to answering this question is the proportion of teachers who question the utility of the data or see teaching the test format as a more practical response than improved instruction. Any research in this area must focus on specific decision making points rather than whether educators articulate a general understanding, because fidelity is likely to be weakest when educators feel most pressured to change their own behavior.

The next question is, How do organizations respond to structured formative assessment data? If my fears are borne out, school districts and individual schools are more likely to ignore than to respond to structured formative assessment data, because the compulsions of preexisting routines are more powerful than data that call for intervention. For example, one might predict that in many districts with curriculum-pacing calendars, those calendars will not change even if data for a specific classroom or school suggest that reteaching is necessary (also see Erickson, 2007). Where data suggest that children are underperforming, the most common organizational intervention may be the expansion of existing behavior (e.g., spending more time with the same reading curriculum) rather than a shift in behavior that requires organizational change (such as rearranging a school’s schedule to accommodate teaching in small groups; see Diamond & Cooper, 2007, for a variant of this behavior).

One might also predict that individuals within a district can mediate this response. If the implementation literature applies to the use of structured formative assessment, the quality of responses will vary most clearly with the presence or absence of an institutional leader or internal entrepreneur who commits to structured formative assessment; ensures that resources are committed to professional development; follows through in making schools and teachers accountable for using data; identifies appropriate personnel at the school level to coordinate the collection of data, its analysis, and appropriate responses; and does not place a higher priority on programs that interfere with the use of structured formative assessment.

The next question is, How do individual teachers respond to formative assessment data? Teachers have responded in carefully designed research studies in the literature—not as well-designed as one would like, at least not without strong guidance; but the literature supports well-designed uses of structured formative assessment. With a dramatic expansion in what districts and states claim is data-driven use of formative assessment, many opportunities exist to study the use of formative assessment in schools under a range of conditions. We do not know whether NCLB has changed that behavior in schools, nor do we know what types of environments are more likely

to stimulate reasonable or productive responses to data. The presence of a key individual committed to structured formative assessment may mediate the responses of individual teachers as well as organizational responses, and it is the identity and supervision of individuals with coordination responsibility in schools where those questions may well be answered. On the other hand, in systems or components that put pressure on teachers without a key individual committed to formative assessment, teachers may be less likely to implement or respond to structured formative assessment in ways that are helpful to themselves or to students.

REFERENCES

- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7–73.
- Brunner, C., Fasca, C., Heinze, J., Honey, M., Light, D., Mandinach, E., & Wexler, D. (2005). Linking data and learning: The Grow Network study. *Journal of Education for Students Placed at Risk*, 10, 241–67.
- Busch, T. W., & Reschly, A. L. (2007). Progress monitoring in reading: Using curriculum-based measurement in a response-to-intervention model. *Assessment for Effective Intervention*, 32, 223–230.
- Calhoun, D. H. (1973). *The intelligence of a people*. Princeton, NJ: Princeton University Press.
- Carnine, D. (1997). Bridging the research-to-practice gap. *Exceptional Children*, 63, 513–521.
- Chen, E., Heritage, M., & Lee, J. (2005). Identifying and monitoring students: Learning needs with technology. *Journal of Education for Students Placed at Risk*, 10, 309–332.
- Cuban, L. (1990). Reforming again, again, and again. *Educational Researcher*, 19, 3–13.
- Cuban, L. (1993). *How teachers taught: Constancy and change in American classrooms, 1880–1990* (2nd ed.). New York: Teachers College Press.
- Cuban, L. (1996). Myths about changing schools and the case of special education. *Remedial and Special Education*, 17(2), 75–82.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52, 219–232.
- Deno, S. L. (2003). Developments in curriculum-based measurement. *Journal of Special Education*, 37, 184–192.
- DeStefano, L., Dailey, D., Berman, K., & McInernay, M. (2001.) *Synthesis of discussions about scaling up effective practices*. Washington, DC: Office of Special Education Programs.
- Diamond, J. B., & Cooper, K. (2007). The uses of testing data in urban elementary schools: Some lessons from Chicago. *Yearbook of the National Society for the Study of Education*, 106(1), 241–263.
- Dorn, S. (2007). *Accountability Frankenstein: Understanding and taming the monster*. Charlotte, NC: Information Age.
- Dreeben, R. (1968). *On what is learned in school*. Reading, MA: Addison-Wesley.
- Erickson, F. (2007). Some thoughts on “proximal” formative assessment of student learning. *Yearbook of the National Society for the Study of Education*, 106(1), 186–216.
- Figlio, D. N., & Winicki, J. (2005). Food for thought: The effects of school accountability plans on school nutrition. *Journal of Public Economics*, 89, 381–394.
- Firestone, W. A., & Gonzalez, R. A. (2007). Culture and processes affecting data use in school districts. *Yearbook of the National Society for the Study of Education*, 106(1), 132–154.
- Fletcher, J. M., & Vaughn, S. (2009). Response to intervention: Preventing and remediating academic difficulties. *Child Development Perspectives*, 3(1), 30–37.
- Florida Center for Reading Research. (2005). *Florida's Reading First assessment: Principal's guide*. Tallahassee, FL: Author. Retrieved from <http://www.fcrr.org/assessment/pdf/principlesguide.pdf>
- Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review*, 33, 188–192.
- Fullan, M. (2002). *Change forces with a vengeance*. New York: Routledge.
- Goodman, K. S. (Ed.). (2006). *Examining DIBELS: What it is and what it does*. Brandon, VT: Vermont Society for the Study of Education.
- Hasbrouck, J. E., Woldbeck, T., Ihnot, C., & Parker, R. I. (1999). One teacher's use of curriculum-based measurement: A changed opinion. *Learning Disabilities Research and Practice*, 14, 118–126.
- Ikemoto, G. S., & Marsh, J. A. (2007). Cutting through the data-driven mantra: Different conceptions of data-driven decision making. *Yearbook of the*

- National Society for the Study of Education*, 106(1), 105–131.
- Labaree, D. F. (1988). *The making of an American high school: The credentials market and the Central High School of Philadelphia, 1838–1939*. New Haven, CT: Yale University Press.
- Labaree, D. F. (1997). *How to succeed in school without really learning: The credentials race in American education*. New Haven, CT: Yale University Press.
- Lachat, M. A., & Smith, S. (2005). Practices that support data use in urban high schools. *Journal of Education for Students Placed at Risk*, 10, 333–349.
- Mandinach, E. B., Honey, M., & Light, D. (2006, April). *A theoretical framework for data-driven decision making*. EDC Center for Children and Technology, paper presented at the Annual Meeting of the American Educational Researchers Association (AERA), San Francisco.
- Mandinach, E. B., Rivas, L., Light, D., Heinze, C., & Honey, M. (2006, April). *The impact of data-driven decision making tools on educational practice: A systems analysis of six school districts*. EDC Center for Children and Technology, paper presented at the meeting of the American Educational Research Association, San Francisco.
- McLaughlin, M. W. (1990). The Rand Change Agent Study revisited: Macro perspectives and micro realities. *Educational Researcher*, 19(9), 11–16.
- McNeil, L. M., Coppola, E., Radigan, J., & Vasquez Heilig, J. (2008). Avoidable losses: High-stakes accountability and the dropout crisis. *Education Policy Analysis Archives*, 16(3). Retrieved from <http://epaa.asu.edu/epaa/v16n3/>
- Metz, M. H. (1990). Real school: A universal drama amid disparate experience. In D. Mitchell and M. E. Goertz (Eds.), *Education politics for the new century: The 20th anniversary yearbook of the Politics of Education Association* (pp. 75-91). Philadelphia: Falmer Press.
- Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, MA: Harvard Education Press.
- No Child Left Behind Act of 2001. (2002). Public Law No. 107–110.
- Roehrig, A. D., Dugger, S. W., Moats, L., Glover, M., & Mincey, B. (2008). When teachers work to use progress monitoring data to inform literacy instruction: Identifying potential supports and challenges. *Remedial and Special Education*, 29, 364–382. doi: 10.1177/0741932507314021
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39–83). Chicago, IL: Rand McNally.
- Sharkey, N. S., & Murnane, R. J. (2006). Tough choices in designing a formative assessment system. *American Journal of Education*, 112, 572–588.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14.
- Smith, M. L. (1991). Meanings of test preparation. *American Educational Research Journal*, 28, 521–542.
- Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools*, 42, 795–819.
- Thorn, C., Meyer, R. H., & Gamoran, A. (2007). Evidence and decision making in education systems. *Yearbook of the National Society for the Study of Education*, 106(1), 340–361.
- Toch, T. (2006). *Margins of error: The testing industry in the No Child Left Behind era*. Washington, DC: Education Sector. Retrieved from http://www.educationsector.org/research/research_show.htm?doc_id=346734
- Tomlinson, S. (1982). *A sociology of special education*. London: Routledge Kegan Paul.
- Tyack, D. B., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.
- U.S. House Committee on Education and Labor. (2007, August 28). *Miller-McKeon discussion draft*. Retrieved from <http://edlabor.house.gov/bills/MillerMcKeonNCLBDiscussionDraft.pdf>
- Wayman, J. C. (2005). Involving teachers in data-driven decision making: Using computer data systems to support teacher inquiry and reflection. *Journal of Education for Students Placed at Risk*, 10, 295–308.
- Weatherley, R., & Lipsky, M. (1977). Street-level bureaucrats and institutional innovation: Implementing special-education reform. *Harvard Educational Review*, 47, 171–197.
- Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 21, 1–19.
- William, D. (2006). Formative assessment: Getting the focus rights. *Educational Assessment*, 11, 283–289.
- Wright, J. (n.d.). *Math worksheet generator*. Syracuse: Intervention Central. Retrieved from <http://www.interventioncentral.org/htmldocs/tools/mathprobe/addsing.php>
- Yeh, S. S. (2008). The cost-effectiveness of comprehensive school reform and rapid assessment. *Education*

Policy Analysis Archives, 16(13). Retrieved from <http://epaa.asu.edu/epaa/v16n13/>

Young, V. M., & Kim, D. (in press). Using assessments for instructional improvement: A literature review. *Education Policy Analysis Archives*.

ABOUT THE AUTHOR

SHERMAN DORN, College of Education, University of South Florida, Tampa, Florida.

Correspondence concerning this article should be addressed to Sherman Dorn, College of Education, University of South Florida, 4202 E. Fowler Ave., Tampa, FL 33620-7750 (e-mail: sdorn@coedu.usf.edu).

Paper originally prepared for Changing Conceptions of Special Education, a University of Delaware Symposium for Scholars, Practitioners, and Policy Makers, April 24–25, 2008. The comments of Laura Eisenman, Ralph Ferretti, and the anonymous reviewers were formative in the best sense of the word, but errors of fact and interpretation are the responsibility of the author alone.

Manuscript submitted July 2008; accepted October 2009.

Exceptional Children Journal article available at <http://ecx.sagepub.com/content/76/3/325.short>.