

Overview of Current Assessment Practices

Margaret E. Goertz
Consortium for Policy Research in Education
Graduate School of Education
University of Pennsylvania

While NCLB has increasingly focused the attention of educators, policymakers and the public on testing, assessment has been used for nearly a century to measure and evaluate student and school performance. However, the range and purpose of testing has expanded, particularly in the last three decades. This paper is intended to frame the Best Practices in State Assessment workshops by: 1) situating state assessments in the broader context of how we use tests today; 2) reviewing very briefly what we know about the strengths and limitations of state and related assessments, such as interim assessments; and 3) using this overview to identify issues we need to consider as we design more innovative assessments and assessment systems.

I make three main points in this paper. First, policymakers and educators use a limited number of tests, particularly state assessments, for multiple purposes. The design of new assessments and assessment systems must recognize these multiple uses because they embody different theories of action and technical requirements. Second, “we treasure what we measure.” Test-based accountability has focused attention on standards, but particularly that subset of academic standards and content domain that is covered by the tests. This focus has had both negative and positive consequences for teaching. Third, it is not only about the test. Even if we build a more innovative assessment system, one that “measures what we treasure,” its impact will be limited unless we create a strong system of support for accessing and analyzing the resulting data, and build the capacity of teachers to interpret and act on these data.

Few Tests, Many Uses

Table 1 shows different ways in which the results of student assessments are currently used for students, teachers and schools. The rows represent the different types of use, the columns the different foci of assessment use. These are not discrete categories or cells, and limited space precludes a more complete description of these uses. But this display does enable one to visualize the multiple uses of one assessment (e.g., state assessment), as well as identify areas where we should be developing multiple or different forms of assessment.

Uses for students. We can look at how state assessment results have been used for students by working our way down the column titled “Students.” Historically, we have used standardized tests to place students into instructional groups (tracking within and across classrooms) and programs (e.g., remedial, advanced; special education; ESL, etc.). Teachers use external, curriculum-based, and their own assessments to make instructional decisions about individual students.

Test results are also used for low- and high-stakes student accountability. Scores from high school graduation tests may be placed on student transcripts, a low-stakes action as employers rarely look at student transcripts. A few states and school districts use state tests results as one factor in making promotion decisions for elementary and middle school students. Most critically, students in 24 states must pass a state test to graduate from high school. This policy becomes effective in 2 more states in 2012. Eleven states have or are in the process of moving from “comprehensive” to end-of-course exams for high school graduation. These new tests will have implications for school accountability as 24 states also use their high school exit exams for school accountability under NCLB (Center on Education Policy, 2009b).

Table 1: Assessment Uses

Uses	Student	Teacher	School
Diagnosis	Instructional decisions Placement Allocation of educational services	Professional development and support	Resource allocation Technical assistance
Inform teaching/ learning		Focus, align, redirect content and instructional strategies	Focus on, align curriculum to skills/content School improvement planning
Evaluation	Certification of individual achievement	Teacher preparation programs Teacher pay	Program evaluation
Public Reporting	Transcripts		Parent or community action
External Accountability	Promotion High school graduation	Renewal Tenure Pay	Sanctions and rewards

Uses for schools. Working our way up the column marked “Schools,” we see that the most visible use of state assessment data aggregated to the school level is for external accountability under NCLB and, in some states, separate state accountability policies. Schools identified as not making adequate yearly progress (AYP) are subject to sanctions. In some states (e.g., Arizona, Florida), schools may receive rewards for achieving, maintaining AYP and/or showing growth on state assessments. While most external accountability systems use cross-sectional (or cohort) measures, a growing number of states are incorporating longitudinal measures of student growth into their state accountability policies either in accordance with the NCLB Growth Model provision or through their own state formulas.

These same test data are reported publicly and may be used by parents and community members to press educators to make changes, or by parents to choose schools. These data may be used by school and district staff to evaluate programs in schools, by school staff to design and implement school improvement plans, and/or by school staff to focus their attention on, and align curriculum to, specific skills and content. Test results may also be used diagnostically to identify areas for technical assistance to schools, and to direct or reallocate resources to and within schools.

Uses for teachers. As discussed in the next section of this paper, teachers also use student test data at the classroom level to align curriculum and instruction with standards and tested skills and content, and to identify professional development needs. In addition, a few states and school districts have begun to use student test data in pay for performance programs. For example, Minnesota’s Q-Comp program, the Teacher Advancement Program (TAP) and the federal Teacher Incentive Fund include student performance as one component of performance-based compensation. Sixty percent of Florida’s Merit Award Program awards must be based on

student performance on state (and in non-tested grades and subjects, other) assessments. And, responding to financial incentives from the Race to the Top program, policymakers are considering how to use these data to inform teacher evaluations and tenure decisions and assess the effectiveness of teacher preparation programs.

Implications of multiple test uses for assessment design. These different uses of tests incorporate different theories of action and technical requirements that must be considered in the design of assessment systems. For example, the theory of action underlying the use of tests to change classroom instruction is that by aligning assessments with specific content and performance standards, and by providing diversification of testing formats, teachers will have the information they need to change their curriculum and instruction. A continuous improvement model posits that teachers will gather information about student learning on an ongoing basis and will interpret and use that evidence in ways that improve instruction. Thus assessments for instructional purposes need to be timely and relevant to classroom instruction and must generate continuous and actionable information. Ideally, assessment is embedded in instruction with continuous feedback to the teacher. Flexibility is important; reliability is not as critical.

In contrast, the use of tests for external accountability assumes that information alone is insufficient to motivate educators to teach well and students to perform to high standards; therefore, incentives and sanctions are needed to stimulate action. This use of tests requires standardization of content, administration and scoring because it involves making comparisons, whether to a benchmark or to other students, groups or schools. Because of the high stakes usually associated with external accountability, the tests must be valid, reliable and fair.

Although some of the structural and technical requirements differ by test use, tests for instructional and accountability purposes share some common requirements. For example, to

generate valid information about student learning, all tests should be aligned to content standards that reflect a model of learning or learning progressions. Tests must include valid measures of student growth to guide instruction and to inform accountability decisions, particularly for teachers.

We Treasure What We Measure

Researchers, policymakers and educators have identified both advantages and disadvantages to using external tests, such as state assessments, in instruction and to drive instructional change.

Strengths of the current system. The current test-driven accountability system focuses educators' attention on standards. By and large, educators believe in the intrinsic value of standards. They believe that state standards identify what their students should know and be able to do, that the standards are compatible with good educational practice, and that the public should hold students and educators to account for meeting certain outcomes. Teachers find standards helpful for guiding their own instruction and they align their curriculum and instruction with them, although they believe that standards include more content than they can cover in a year, and are, in some instances, too vague to give useful guidance. Teachers also find standards useful for bringing focus and consistency of instruction within and across schools (Kannapel et al., 2001; Johnson, Arumi, and Ott 2006; Massell et al., 2005; Stecher et al., 2008).

The legitimacy of state assessment systems is much lower, however, particularly among teachers. Teachers do not believe that state tests are necessarily a good measure of their students' mastery of content, and many raise concerns about the lack of alignment among standards, assessment and curriculum. But teachers, schools and districts pay attention to the data generated by assessments. Teachers report that they review assessment results to identify students who

need additional help, topics requiring more emphasis, and gaps in curriculum and instruction. Teachers also indicate that they search for more effective teaching methods and change some elements of their instructional practice in response to state assessments (Goertz & Massell, 2005; Kannapel et al., 2001; Stecher et al., 1998; Stecher et al., 2008).

Districts and schools are increasing their use of student data in school improvement planning, to change curriculum and instructional materials, and to focus professional development (Harris et al., 2005; Massell, 2001; Padilla et al., 2006; Stecher et al., 2008). Although teachers feel that their students are over-tested (Johnson, Arumi, & Ott 2006), teachers, schools and districts want more information on student achievement. Since the mid-1990s, districts and schools have supplemented state tests in order to measure continuous progress toward district and/or state standards, provide instructional feedback to teachers, offer student-level information for parents and teachers, reinforce constructivist teaching through performance assessments, and/or evaluate programs (Massell & Goertz, 2002). With expanded state testing under NCLB, districts have turned to interim or benchmark assessments to track student progress throughout the school year.

Stronger accountability has also focused educators' attention on traditionally underserved populations of students. While some educators still question whether all students can learn to high standards, their expectations for these students are considerably higher than in the past. And districts have responded to accountability press by providing assistance to schools, although not always the kinds of intensive support envisioned under NCLB (Center on Education Policy, 2007; Padilla et al., 2006; Stecher et al., 2008).

Weaknesses of the current system. While the current test-based accountability system has focused teachers' attention on standards, the breadth (and ambiguity) of standards, coupled

with accountability focused on test results, had led teachers, particularly those in low-performing schools, to pay more attention to what is assessed than to the broader set of standards (Herman, 2004). Research has shown that this focus had led to a narrowing of the curriculum to tested subjects, content and pedagogy; increased instructional time on tested subjects; and increased test preparation (Center on Education Policy, 2008; 2009a; Herman, 2004; Firestone, Schorr & Monfils, 2004; Shepard, 2005; Shepard & Dougherty, 2001; Stecher et al., 2008). Teachers are paying more attention to “bubble kids,” or children who are performing at just below the pass rates of mandated assessments (c.f., Booher-Jennings, 2005; Christman et al., 2009). And, although teachers have taken steps to align instruction with assessments and standards and to use data to plan instruction, there is little evidence that their daily instructional practices have changed (c.f., Firestone, Schorr, & Monfils, 2004; Herman, 2004; Stecher et al, 2008; Wilson & Floden, 2001).

Teachers’ focus on tested content is even more problematic given the limited alignment of state assessments with their standards, and the lack of rigor of these assessments. Building on their earlier work (Porter, 2002; Porter, Polikoff, & Smithson; 2009), Polikoff, Porter, and Smithson (under review) examined the alignment of state assessments with state content standards post-NCLB. Looking at 8 to 9 states where they had sufficient data, they found low to moderate with-in state assessment-to-standards alignment, averaging about 0.30 in mathematics, 0.20 in language arts and slightly over 0.20 in science in 4th and 8th grades. That is, large proportions of content standards were not covered on assessments. When they examined the cognitive demand of the assessments, the researchers found that *on average*, with the exception of 4th grade mathematics, the assessments were considerably more concentrated at lower levels

of cognitive demand than the states' standards. There was significant variation across states in their sample, however, generally due to variation in the cognitive demand level of the standards.

A final limitation of the current assessment system is a by-product of provisions of the NCLB Act. The NCLB requirements of testing every student annually in grades 3-8, coupled with the short turnaround times for scoring and reporting data required to identify schools that must provide choice and supplemental services, have led states to increase their use of multiple choice formats on state assessments. Assessments composed primarily of multiple choice items have less potential to provide information on students' understanding and cognition than those with open-ended items and/or extended writing and rubrics that require students to present evidence of their thinking. In addition, as teachers tend to model test pedagogy, they have less incentive to incorporate problem-solving or written communication in their instruction (Herman, 2004).

Interim assessments. The past ten years have witnessed an explosion in the use of interim assessments by school districts across the country in a desire to give teachers more frequent data on student performance to use for instructional planning. Stecher et al. (2008), for example, found that that 53% to 73% of study districts in California, Georgia and Pennsylvania used interim or progress tests in elementary and middle school mathematics (a focus of their study) in 2006. Districts are keeping their interim tests even under pressure to cut budgets (Sawchuk, 2009), and draft guidelines for the U.S. Department of Education's Race to the Top Assessment Program encourage school districts to develop formative or interim assessments as part of comprehensive state assessment systems.

Supporters of interim assessments argue that these tests will provide data on student understanding, and that teachers' analysis of this data will in turn lead to greater differentiation

of instruction and better teaching of content, leading to improved student learning. Much of the belief in the potential of interim assessments to improve student learning comes, however, from the growing body of research on formative assessment (c.f., Black & Wiliam, 1998; Crooks, 1988; Hattie & Timperley, 2007; Natriello, 1987). Very few studies exist on how interim assessments are actually used by individual teachers in classrooms, by principals, and by districts, or on their impact on student achievement.

Some recent studies surveyed teachers about their use of test data in instruction. Many of these teachers reported that interim test results helped them monitor student progress and identify skill gaps for their students, and led them to modify curriculum and instruction (c.f., Christman, et al., 2009; Clune & White, 2008; Stecher, et al., 2008). These studies, however, did not examine how individual teachers actually analyzed and used these data to inform their classroom practice. A recent exploratory study of how teachers used curriculum-based interim assessments in elementary mathematics in two districts found that teachers took similar actions, using data from interim assessments to identify weak content areas or struggling students within their classrooms, to determine whether and how students should be re-grouped during re-teaching, to decide when to move on to the next topic or concept, and as the basis for cross-classroom conversations (Goertz, Olah & Riggan, 2009). Teachers were much less likely to modify the ways in which they actually re-taught the content. The researchers also found that most items in the assessments did not provide actionable information on students' misunderstandings, a precursor to instructional improvement.

There are also few publicly available large-scale studies on the impact of interim assessments on student learning. Two such studies found small but non-significant effects of the use of interim exams in mathematics (Henderson, Petrosino, Guckenburg, & Hamilton, 2008)

and reading (Quint, Sepanik, & Smith, 2008). As Quint, Sepanik, and Smith (2008) note, however, the limited time frame of these studies (one to two years) may be insufficient to see effects of interim assessments on student learning.

It's Not Just the Test

Building a better assessment system is necessary but not sufficient to drive and support instructional improvement. Teachers must have access to and support for the use of test data, and they must have the capacity to interpret and act on these data.

A large body of research on data-driven decision-making has identified conditions for successful use of assessment data. The first is access to a quality and usable Information Management System (IMS) that focuses teachers' attention on content as well as items (c.f., Marsh et al., 2005; Supovitz, 2006; Wayman & Stringfield, 2006). Yet, a national survey conducted in 2006-07 found that only 72% of districts stored scores from district tests in electronic data systems and only 41% of teachers reported having electronic access to their students' performance on interim or diagnostic tests. Only 33% of teachers with access to an electronic student data system felt capable of forming data queries (U.S. Department of Education, 2009).

Other conditions affecting data use include: a) expectations that interim assessment results will be used to inform instruction (c.f., Firestone & Gonzalez, 2007; Goertz, Olah, & Riggan, 2009; Knapp, Copland, & Swinnerton, 2007); b) a culture of data use (c.f., Ikemoto & Marsh, 2007; Massell, 2001; Supovitz, 2006); c) time to re-teach content and skills to students (c.f., Goertz, Olah & Riggan, 2009); and d) leadership to create explicit norms and expectations regarding data use (Datnow, Park, & Wohlstetter, 2007; Kerr et al., 2006; Lachat & Smith, 2005; Supovitz, 2006; Wayman & Stringfield, 2006; Young, 2006), build and support principals' and

teachers' skills in data-based inquiry (Knapp, Copland, & Swinnerton, 2007; Young, 2006), promote norms of collaboration (Ikemoto & Marsh, 2007; Young, 2006), and support the concept of continual learning (Sharkey & Murnane, 2006; Knapp, Copland, & Swinnerton, 2007).

Using data for instructional improvement also requires tightening the connection between data and classroom practice and building teachers' instructional knowledge and skills. For example, Goertz, Olah & Riggan (2009) found that teachers' capacity to interpret assessment data played a major role in how they used the results of interim, and even formative, assessments. Many teachers focused on procedural rather than conceptual sources of student errors on test items, diagnoses that appeared to inform their instructional planning during re-teaching. Teachers who assessed for conceptual understanding did so across multiple test formats and were more likely to use instructional changes strategies than those who did not. Teachers' mathematical knowledge for teaching also appeared to contribute to teachers' instructional and assessment practices.

Yet, many teachers lack the knowledge, resources, and support to link assessment results to teaching (Datnow, Park, & Wohlstetter, 2007; Kerr et al., 2006; Young, 2006). A synopsis of RAND research, for example, found that while most teachers and principals reported having access to workshops on interpreting assessment results, few found them to be helpful. Educators preferred training on the use of assessment results in instructional planning, but this type of support "was less often available" (Marsh, Pane, & Hamilton, 2006, pp.7-8). Thus, professional development for assessment use must move beyond a focus on "point and click" (to locate and organize data) to a focus on interpreting data (e.g., diagnosing student error) and on connecting this evidence to specific instructional approaches and strategies. It is also important that teacher

professional development focus on teacher content knowledge, and on developing teachers' instructional repertoires and capacity to assess for students' mathematical learning.

Conclusion

We have learned that the substance and use of assessments focus teachers' attention on tested content. Teachers have aligned their curriculum and instruction to standards and assessments, and used assessment data for instructional planning. However, in many cases, this has led to a narrowing of the curriculum and emphasis on test preparation, and has not resulted in teachers' changing their day-to-day instructional strategies. Teachers, schools and school districts crave additional data to inform instruction, but teachers have limited access to student results and insufficient preparation and knowledge to interpret and act on these data. Support for better data use requires leadership that will create a culture of data use and continuous improvement, build principal and teacher capacity, and provide time for collaboration and re-teaching. We have also learned that no one test can or should be used for multiple purposes.

Therefore, as educators, policymakers and researchers consider the design of more innovative assessments and comprehensive assessment systems, they must address a set of design and implementation questions. These include:

1. What do we want to test and how?
2. For what purpose(s)?
3. What are the technical requirements for each use?
4. What kinds of information do we want to generate and for whom?
5. What is the role of a state test in a comprehensive assessment system?
6. What supports will educators need to use the test (and assessment system) to improve instruction?

References

- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-75.
- Boohers-Jennings, J. (2005). Below the bubble: 'Educational triage' and the Texas accountability system. *American Educational Research Journal* 42, 231-268.
- Center on Education Policy (2007). *Moving beyond identification: Assisting schools in improvement*. Washington, D.C.: Author.
- Center on Education Policy (2008). *Instructional time in elementary schools: A closer look at changes for specific subjects*. Washington, DC: Author.
- Center on Education Policy (2009a). *How state and federal accountability policies have influenced curriculum and instruction in three states*. Washington, DC: Author.
- Center on Education Policy (2009b). *State high school exit exams: Trends in test programs, alternative pathways, and pass rates*. Washington, DC: Author.
- Christman, J., Neild, R., Bulkley, K., Blanc, S., Liu, R., Mitchell, C., & Travers, E. (2009). *Making the most of interim assessment data. Lessons from Philadelphia*. Philadelphia, PA: Research for Action.
- Clune, W. H., & White, P. A. (2008). *Policy effectiveness of interim assessments in Providence Public Schools* (WCER Working Paper No. 2008-10). Madison, WI: University of Wisconsin-Madison, Wisconsin Center for Education Research.
- Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58, 438-481.
- Datnow, A., Park, V., & Wohlstetter, P. (2007). *Achieving with data: How high-performing school systems use data to improve instruction for elementary students*. Los Angeles, CA: Center on Educational Governance, University of Southern California Rossier School of Education.
- Firestone, W. A., & Gonzalez, R. A. (2007). Culture and processes affecting data use in school districts. In P. A. Moss (Ed.), *Evidence and decision making*. The 106th yearbook of the National Society for the Study of Education, Part I (pp. 132-154). Malden, MA: Blackwell Publishing.
- Firestone, W. A., Schorr, R. Y., & Monfils, L. F. (2004). *The ambiguity of teaching to the test*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Goertz, M. E., & Massell, D. (2005). Holding high hopes: How high schools respond to state accountability policies. *CPRE Policy Brief RB-42*. Philadelphia, Pa.: University of Pennsylvania, Consortium for Policy Research in Education.

- Goertz, M. E., Olah, L. N., & Riggan, M. (2009). *From testing to teaching: The use of interim assessments in classroom instruction*. Research Report #65. Philadelphia, Pa.: University of Pennsylvania, Consortium for Policy Research in Education.
- Harris, D. M., Prosky, M., Bach, A., Heilig, J. V., & Karen Hussar, K. (2005). Overview of Actions Taken by High Schools to Improve Instruction. In B. Gross & M. E. Goertz (Eds.). *Holding high hopes: How high schools respond to state accountability policies* (pp. 81-93). Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-11.
- Henderson, S., Petrosino, A., Guckenburg, S.; & Hamilton, S. (2008). *A second follow-up year for "measuring how benchmark assessments affect student achievement."* REL Technical Brief 2008-No. 002. Newton, MA: Regional Educational Laboratory Northeast & Islands.
- Herman, J. (2004). The effects of testing on instruction. In S. H. Fuhrman & R. F. Elmore, *Redesigning accountability systems for education* (pp. 141-166). NY: Teachers College Press.
- Ikemoto, G. S., & Marsh, J. A. (2007). Cutting through the "data-driven" mantra: Different conceptions of data-driven decision making. In P. A. Moss (Ed.) *Evidence and decision making*. The 106th yearbook of the National Society for the Study of Education, Part I (pp. 105-131). Malden, MA: Blackwell Publishing.
- Johnson, J., Arumi, A. M., & Ott, A. (2006). Is support for standards and testing fading? *Reality Check 2006*, no. 3. www.publicagenda.org/reports/reality-check-2006-issue-no-3 16 Jun 2009.
- Kannapel, P. J., Aagaard, L., Coe, P., & Reeves, C. (2001). The impact of standards and accountability on teaching and learning in Kentucky. In S. H. Fuhrman (Ed.) *From the capitol to the classroom: Standards-Based reform in the states*. The 100th Yearbook of the National Society for the Study of Education, Part 2 (pp. 242-62). Chicago: University of Chicago Press.
- Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education*, 112(4), 496-520.
- Knapp, M. S., Copland, M. A., & Swinnerton, J. A. (2007). Understanding the promise and dynamics of data-informed leadership. In P. A. Moss (Ed.) *Evidence and decision making*. The 106th yearbook of the National Society for the Study of Education, Part I (pp. 74-104). Malden, MA: Blackwell Publishing.
- Lachat, M. A., & Smith, S. (2005). Practices that support data use in urban high schools. *Journal of Education for Students Placed at Risk*, 10(3), 333-349.

- Marsh, J. A., Kerr, K. A., Ikemoto, G. S., Darilek, H., Suttorp, M., & Zimmer, R. W. (2005). *The role of districts in fostering instructional improvement* (MG-361-EDU). Santa Monica, CA: RAND.
- Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making in education*. Santa Monica, CA: RAND Corporation.
- Massell, D. (2001). The theory and practice of using data to build capacity: State and local strategies and their effects. In S. H. Fuhrman (Ed.), *From the capitol to the classroom: Standards-Based reform in the states*. The 100th Yearbook of the National Society for the Study of Education, Part 2 (pp. 148-169). Chicago: University of Chicago Press.
- Massell, D., & Goertz, M. E. (2002). District strategies for building capacity. In A. M. Hightower, M. S. Knapp, J. A. Marsh, & M. W. McLaughlin (Eds.), *School districts and instructional renewal* (pp. 43-60). New York: Teachers College Press.
- Massell, D., Goertz, M. E., Christensen, G., & Goldwasser, M. (2005). The press from above, the pull from below: High school responses to external accountability. In B. Gross & M. E. Goertz (Eds.), *Holding high hopes: How high schools respond to state accountability policies* (pp. 17-41). Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education.
- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, 22, 155-175.
- Padilla, C., Skolnik, H., Lopez-Torkos, A., Woodworth, K., Lash, A., Shields, P. A., Laguarda, K., & David, J. L. (2006). *Title I accountability and school improvement from 2001 to 2004*. Washington, D.C.: U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.
- Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, 31(7), 3-14.
- Porter, A. C., Polikoff, M. S., & Smithson, J. (2009). Is there a de facto national intended curriculum? Evidence from state content standards. *Educational Evaluation and Policy Analysis*, 31(3), 238-268.
- Polikoff, M., Porter, A. C., & Smithson, J. (under review). The role of state student achievement tests in standards-based reform.
- Quint, J., Sepanik, S., Smith, J., & MDRC (2008). *Using student data to improve teaching and learning: Findings from an evaluation of the Formative Assessments of Students Thinking in Reading (FAST-R) Program in Boston Elementary Schools* MDRC (ERIC Document Reproduction Service No. ED503919) Retrieved August 7, 2009, from ERIC database.
- Sawchuk, S. (May 13, 2009). Testing faces ups and downs amid recession. *Education Week*, 28.

- Sharkey, N. S., & Murnane, R. J. (2006). Tough choices in designing a formative assessment system. *American Journal of Education*, 112(4), 572-588.
- Shepard, L. A. (2005). Formative assessment: Caveat Emptor. *The future of assessment: Shaping teaching and learning*. Paper presented at the ETS Invitational Conference, New York, NY. .
- Shepard, L. A., & Dougherty, K. C. (1991). *Effects of high-stakes testing on instruction*. Chicago, Ill.: Spencer Foundation.
- Stecher, B., Barron, S., Kaganoff, T. & Goodwin, J. (1998). *The effects of standards-based assessment on classroom practices: Results of the 1996-97 RAND survey of Kentucky teachers of mathematics and writing*. Los Angeles: CA: Center for Research on Evaluation, Standards and Student Testing.
- Stecher, B. M., Epstein, S., Hamilton, L. S., Marsh, J. A., Robyn, A., McCombs, J. S., Russell, J., & Nafital, S. (2008). *Pain and gain: Implementing No Child Left Behind in three states, 2004–2006*. Santa Monica, CA: RAND Corporation.
- Supovitz, J. A. (2006). *The case for district-based reform: Leading, building, and sustaining school improvement*. Cambridge, MA: Harvard Education Press.
- U.S. Department of Education (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.
- Wayman, J. C., & Stringfield, S. (2006). Technology-supported involvement of entire faculties in examination of student data for instructional improvement. *American Journal of Education*, 112(4), 549-571.
- Wilson, S., & Floden, R. E. (2001). Hedging bets: Standards-based Reform in classrooms. In S. H. Fuhrman (Ed.), *From the capitol to the classroom: Standards-Based reform in the states*. The 100th Yearbook of the National Society for the Study of Education, Part 2 (pp. 193-216). Chicago: University of Chicago Press.
- Young, V. M. (2006). Teachers' use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education*, 112(4), 521-548.