

WORKING P A P E R

Using Test-Score Data in the Classroom

BRIAN M. STECHER AND LAURA S. HAMILTON

WR-375-EDU

April 2006

This product is part of the RAND Education working paper series. RAND working papers are intended to share researchers' latest findings and to solicit informal peer review. They have been approved for circulation by RAND Education but have not been formally edited or peer reviewed. Unless otherwise indicated, working papers can be quoted and cited without permission of the author, provided the source is clearly referred to as a working paper. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. **RAND**® is a registered trademark.

Preface

Teachers and principals have access to test-score data from a variety of sources, including state accountability tests, and the success of test based accountability policies, like No Child Left Behind (NCLB), in raising student achievement will be strongly influenced by their use of these data to guide school policy and classroom practice. As part of a larger study of NCLB implementation, we surveyed (and, in a small number of cases, interviewed) elementary and middle school principals and mathematics teachers in three states to obtain information on the formats in which they received test-score data, their opinions about the quality of the data, the ways in which they used these data, and conditions that hindered or promoted effective use. This paper provides a descriptive summary of these data. The findings should provide insights for improving the utility of test-score information at the classroom level, and for improving the validity of large-scale testing systems more generally.

This material is based on work supported by the National Science Foundation under Grant No. REC-0228295. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

This paper was presented as part of the related paper session, “Using Assessment Results to Improve District and School Practices,” at the annual meeting of the National Council on Measurement in Education, April 2006.

-

Contents

Preface.....	iii
Introduction.....	vii
Background.....	1
ISBA Study Methods.....	4
Results.....	6
How Do Teachers Use Test Results?.....	7
What Factors Influence Test Use?.....	14
Elementary Teachers.....	15
Middle School Teachers.....	15
Examining Variation in Data Use and its Correlates.....	28
Discussion.....	29
References.....	35

Tables

Table 1. Sample Sizes for Survey Responses	5
Table 2. Percent of Mathematics Teachers Engaged in Selected Assessment-Related Activities Often During the 2004-05 School Year	8
Table 3. Percent of Mathematics Teachers Engaged in Selected Assessment-Related Activities More Frequently in 2004-05 Than in 2003-04.....	9
Table 4. Percent of Mathematics Teachers Reporting They Engage in Practices “A Moderate Amount” or “A Great Deal” as a Result of Mathematics Assessments, 2004-05	10
Table 5. Percentage of Teachers Agreeing or Strongly Agreeing with Statements about the Use of State Tests, 2004-05	12
Table 6. Percentage of Principals Reporting that State Assessment Data is Moderately or Very Useful for Making Decisions.....	13
Table 7. Percentage of Teachers Reporting Availability and Usefulness of Mathematics Test Result Summaries, 2004-05	15
Table 8. Percentage of Principals Reporting Availability and Usefulness of Information and Assistance on 2003-04 State Test Results.....	16
Table 9. Percentage of Teachers Agreeing or Strongly Agreeing with Statements about the Features of the State Tests, 2004-05	17
Table 10. Percentage of Teachers Agreeing to Statements Regarding the Quality of the State Assessments, 2004-05.....	20
Table 11. Percentage of Teachers Reporting Availability and Usefulness of Resources and for Using Mathematics Test Results, 2004-05.....	22
Table 12. Percentage of Teachers Reporting that Professional Development had a Moderate or Greater Emphasis on Selected Features, 2004-05.....	23
Table 13. Percentage of Teachers Who Agree or Strongly Agree that Principal at Their School Engages in Selected Behaviors, 2004-05.....	24
Table 14. Percentage of Principals Reporting That Assistance for Data Use was Available and Very Useful, 2004-05	25
Table 15. Percent of Mathematics Teachers Reporting Progress Tests with Various Features, 2004-05.....	27

Introduction

Background

As a result of the No Child Left Behind Act (NCLB) of 2001, all states have implemented statewide testing programs in reading and mathematics for all students in grades 3 to 8 and for one high school grade. Results are provided for every student (for use by schools and parents) and aggregated results are used to evaluate schools and districts. Under NCLB, severe negative consequences can accrue to schools and districts that repeatedly fail to make Adequate Yearly Progress (AYP). Given the high stakes attached to performance on these tests, it is important that states make sure they are reliable and valid for these purposes. Most states have taken extensive steps to evaluate the technical quality of their tests for judging the performance of students.

However, for the most part states have not examined the effects of testing on educational practice. This is an important omission because one of the central assumptions of NCLB is that the provision of high-quality information will promote improved educational decision-making (Stecher, Hamilton, & Gonzalez, 2003). For example, while describing the benefits of NCLB in a recent speech, President Bush stated, “As well, we better figure out who needs help early before it's too late. I mean, one of the reasons you ask the question, can you read or write and add and subtract, is you want to diagnose early. You can't solve a problem unless you diagnose it. The importance of the No Child Left Behind Act is to find out who needs help” (<http://www.whitehouse.gov/news/releases/2004/05/20040511-5.html>). Large-scale testing programs such as those mandated by NCLB are generally intended to send signals to

-
practitioners (both in the form of score reports and in the form of institutional sanctions) about whether their practices are effective or not, and to provide evidence they can use to promote improved outcomes for students. When tests are used explicitly for the purpose of shaping institutional behavior, as they are in the context of NCLB, validation requires further effort than most states have expended. In particular, states should monitor the consequences, both anticipated and unanticipated, of their testing programs (AERA, APA & NCME, 1999, Standard 13.1).

Research points to some of the consequences that high-stakes tests may have on district and school policies and practices (Corbett & Wilson, 1988; Hamilton, 2003; Koretz, Barron, Mitchell, & Stecher, 1996; Romberg, Zarinia, & Williams, 1989; Shepard & Dougherty, 1991; Stecher, Barron, Chun, & Ross, 2000; Pedulla et al., 2003). However, relatively less is known about the use of this information for school and classroom decision-making or the features of the results (such as their perceived quality) that influence their utility for these purposes.

Effective use of test-score data for instructional decision-making is widely viewed as one of the most important actions districts and schools can take to promote improved student achievement (Council of Great City Schools, 2002; Massell, 2001). Past research provides guidance regarding the conditions necessary to support data use, such as high-quality principal leadership (Mason, 2002; Supovitz & Klein, 2003), adequate resources, including teacher planning time (Ingram et al., 2004), and the use of interim, classroom-based assessments to supplement annual state tests (Herman, 2005). Marsh et al. (2005) further delineate research findings regarding the use of data in educational decision-making.

In the context of NCLB, it is particularly important to understand what conditions at the state, district, school, and classroom levels support high-quality use of data from state

-
accountability tests as well as from other sources. Use of information from tests is one of the paths through which the standards-based accountability (SBA) systems that have been adopted under NCLB are likely to affect student achievement. SBA systems include state content and achievement standards, tests, performance targets, and consequences that together are intended to motivate educators while providing them with information to help them identify and make appropriate changes to their practices.

In this paper we present some preliminary information about teachers' use of test-score information in the context of NCLB. The data were collected as part of an ongoing study of NCLB implementation in three states. Although the study was not designed to provide a comprehensive picture of teachers' data use, the summary presented here may be considered a first step toward understanding the degree to which NCLB is fulfilling its promise to improve information accessibility and utility. The information reported here will eventually be incorporated into a broader set of analyses that is designed to illuminate relationships among actions and conditions at each level of the system and the links among these actions, conditions, and student achievement outcomes.

The results reported in this paper provide information relevant to the following questions.

- How do teachers and principals use SBA test-score data?
- What features of test score reporting are relevant to its use by principals and teachers?
- What assistance is available to help teachers and principals use data for instructional decision-making?

ISBA Study Methods

The Implementing Standards-Based Accountability (ISBA) project is designed to examine characteristics of three states' SBA systems and to identify conditions at each level of the system that are likely to promote desired outcomes while minimizing negative consequences (see Stecher, Hamilton, and Naftel, 2005 for additional details on the ISBA project). Data use is one of several sets of actions that the study is examining. The ISBA study is gathering data in three states—California, Georgia, and Pennsylvania—that have taken different approaches to meeting the NCLB accountability requirements and that vary in the characteristics of students they serve. It is a three-year study, with data collection in the spring of 2004, 2005, and 2006. In this paper we focus on the 2005 data, but eventually we will examine changes in responses over the three-year period. The study examines both mathematics and science instruction, but for the purposes of this paper we focus on mathematics.

In each state, we used a stratified random sampling procedure to select approximately 25 districts, including rural, suburban, and urban districts, and we randomly selected approximately 100 schools per state from within these districts. We excluded high schools, which do not have testing requirements as extensive as those imposed on elementary and middle schools.¹ Within the sampled schools, all principals and all regular teachers of mathematics and science in grades 3, 4, 5, 7, and 8 received surveys. Sampling weights and nonresponse weights were generated to allow for survey estimates that can be generalized to the entire state from which the responses were drawn. In addition to surveys administered to teachers, principals, and district superintendents, we conducted case study visits to schools in six districts (two in each state),

¹ We have separate elementary school and middle school samples of principals and teachers, but these designations refer to the grade levels rather than to the building configuration. In a K-8 school, for example, teachers in grades 3-5 are assigned to our elementary sample and teachers in grades 7 and 8 are assigned to our middle school sample.

-
which provided an opportunity to gather richer information about educators' responses.

Additional details regarding data collection and methods for analysis are available in Stecher and Naftel (2006).

We developed separate surveys for classroom teachers, principals, and superintendents. The teacher surveys addressed both mathematics and science instruction, with skip instructions that allowed teachers to answer only those questions relevant to the subjects they taught. The teacher and principal surveys were customized to each state, which enabled us to use the state-specific names given to the assessment programs and standards (e.g., the Pennsylvania Academic Standards and the Pennsylvania System of School Assessment). Principal and teacher survey response rates were between 80-90 percent in all three states. In Table 1 we provide the numbers of elementary principals and teachers with survey data; these numbers apply to the results tables that follow. The standard errors associated with the principal survey responses are larger than those for the teacher surveys because the sample sizes for principals are much smaller. In each of the tables that present results, we provide standard errors, which indicate the level of precision of each estimate.

Table 1.
Sample Sizes for Survey Responses

	California	Georgia	Pennsylvania
Elementary math teachers	456	630	574
Middle school math teachers	251	455	262
Elementary principals	42	49	54
Middle school principals	37	46	34

The three states differ in a number of ways in their implementation of standards-based accountability. For example, they had very different experience with accountability prior to

-
NCLB, and they have taken different paths to integrating their old systems with the new federal requirements. Their standards are organized differently, and at the time these data were collected, the three states were still testing students in different grades. In 2004-05, California and Georgia were testing reading and mathematics in all grades 3 through 8, but Pennsylvania was only testing in grades 3, 5 and 8. More extensive information about the differences among the states with respect to testing and accountability can be found in Stecher and Naftel (2006).

Results

Under NCLB the results of annual state tests carry high stakes for schools. The results are the primary element in determining whether schools have made adequate yearly progress (AYP) and are identified for improvement, corrective action or restructuring. Over three-quarters of the mathematics teachers in tested grades in all three states agreed or strongly agreed with the statement “I feel a great deal of pressure to improve students’ scores on the tests.” (The percentage of mathematics teachers who agreed was about 10 percent higher in Georgia and Pennsylvania than in California.) Consequently, we might expect to find that principals and teachers attended to state test results carefully, and reports from respondents in California, Georgia and Pennsylvania confirm those expectations. Survey respondents also indicated that results from both classroom tests and annual state tests were useful to them for local improvement purposes. But many features of the state tests limited their utility, and there were differences in many cases among the states as well as between elementary and middle school mathematics teachers. Below we present information on how mathematics teachers and

principals reported using test results, and then examine features of the tests and local conditions that may affect reported usage.

How Do Teachers Use Test Results?

It is important to remember that annual state tests are only one form of assessment, and that most teachers administer classroom tests on a regular basis, and these tests play an important role in instructional planning. Responses from mathematics teachers in the three ISBA states confirm the role played by classroom assessments. Roughly one-half of the elementary and middle school mathematics teachers in California, Georgia and Pennsylvania indicated that they used test results on a weekly basis to individualize instruction to the needs of students (see Table 2). A similar proportion indicated that they used test results at least weekly to identify topics that were not understood by all students and to plan for re-teaching. The proportion of Georgia teachers reporting frequent use of tests for these purposes was slightly higher than the other two states.

Teachers also reported that they use tests results for diagnosis and instructional planning more frequently in 2004-05 than they had the previous year. Table 3 shows that about one-quarter of the teachers in the three states increased the frequency of test use for these purposes. (Less than 7 percent of the teachers in any state reported decreasing the frequency of test use.) Again, the increase in test use was slightly higher in Georgia than in the other two states.

The survey also probed specifically about the impact of the annual state tests. Mathematics teachers reported a variety of changes in their instructional practices in response to the implementation of the state test. About two-thirds of teachers in all three states reported that they focused more on their state standards and they looked for more effective ways of teaching

as a result of the state assessment system (see Table 4). These practices embody the kinds of behaviors that accountability systems are supposed to encourage.

Table 2.
Percent of Mathematics Teachers Engaged in Selected Assessment-Related Activities Often During the 2004-05 School Year

	California		Georgia		Pennsylvania	
	Elementary	Middle	Elementary	Middle	Elementary	Middle
Review assessment results to identify <u>individual students</u> who need supplemental instruction	55 (3)	50 (7)	62 (2)	51 (3)	58 (3)	45 (5)
Review assessment results to identify <u>topics</u> requiring more or less emphasis in instruction	48 (4)	44 (3)	59 (2)	52 (3)	51 (2)	42 (5)
Re-teach topics because student performance on assignments or assessments did not meet expectations	48 (3)	44 (3)	57 (3)	53 (2)	43 (3)	40 (4)

Notes:

Choices included never, rarely (a few times a year), sometimes (once or twice a month), and often (once a week or more).

Standard errors shown in parentheses.

Most teachers also reported that they focused more on the topics emphasized on the test, on the styles and formats used in the tests, and on test taking skills as a result of the assessments. For example, one Pennsylvania elementary school teacher explained, “We are spending a lot more time actually teaching the format to the test, not necessarily the curriculum as we would like to teach it, but teaching toward the test.” These behaviors have been associated with inflated test scores (i.e., scores that overstate actual achievement in the area of interest) and they can reduce the validity of inferences drawn from the tests.

Table 3.
Percent of Mathematics Teachers Engaged in Selected Assessment-Related Activities More Frequently in 2004-05 Than in 2003-04

	California		Georgia		Pennsylvania	
	Elementary	Middle	Elementary	Middle	Elementary	Middle
Review assessment results to identify <u>individual students</u> who need supplemental instruction	21 (3)	17 (4)	27 (3)	26 (3)	21 (3)	15 (5)
Review assessment results to identify <u>topics</u> requiring more or less emphasis in instruction	20 (2)	19 (3)	26 (3)	28 (3)	20 (2)	19 (3)
Re-teach topics because student performance on assignments or assessments did not meet expectations	24 (3)	24 (3)	29 (2)	30 (4)	18 (3)	16 (4)

Notes:

Choices included more frequently, the same, and less frequently.
 Standard errors are shown in parentheses.

Our interviews revealed that many districts have implemented “pacing plans” that dictate the order and schedule for teaching the curriculum to ensure that the standards are all covered and students are prepared for the test. The pacing plans have a major impact on instruction, and not all teachers think they are beneficial. For example, a Pennsylvania elementary school teacher bemoaned the change she saw in her school.

I think now teachers are afraid to use their good teacher sense. So you may have taught your concept, but then you’re going to go back and you’re going to fill in the gaps where they’re necessary. You’re going to give them what they need on that level to build the skills that are going to get them back up to that. And some of them now, because they’re so focused on that darned calendar, and needing them to get it, that they’ve kind of given up what good teaching is. A lot of them see the calendar and they are just paralyzed by it. They’re like, ‘I have to teach this, I have to test them, the kids better have all gotten it.’ And those who even used to be able to think outside the box, can’t do it anymore, because they’re so scared about that darned calendar.

Table 4.
Percent of Mathematics Teachers Reporting They Engage in Practices “A Moderate Amount” or “A Great Deal” as a Result of Mathematics Assessments, 2004-05

	California		Georgia		Pennsylvania	
	Elementary	Middle	Elementary	Middle	Elementary	Middle
Focus more on standards	73 (4)	66 (4)	77 (3)	72 (3)	76 (3)	69 (3)
Search for more effective teaching methods	67 (3)	58 (3)	74 (4)	69 (3)	60 (4)	59 (7)
Focus more on topics emphasized in assessment	63 (5)	57 (5)	72 (3)	73 (3)	73 (4)	71 (7)
Emphasize assessment styles and formats of problems	55 (4)	44 (5)	78 (4)	71 (3)	74 (3)	62 (4)
Spend more time teaching test-taking strategies	53 (4)	45 (7)	56 (4)	44 (3)	51 (3)	39 (4)
Spend more time teaching content	52 (4)	45 (6)	58 (4)	53 (3)	53 (4)	46 (3)
Assign more homework	43 (4)	29 (4)	29 (4)	29 (3)	30 (4)	13 (4)
Focus more on students who are close to proficient	37 (5)	19 (3)	36 (2)	38 (4)	29 (3)	22 (5)
Offer more assistance outside of school for students who are not proficient	29 (3)	26 (3)	34 (4)	41 (3)	21 (3)	19 (5)

Notes:

Standard errors are shown in parentheses.

Response options included not at all, a small amount, a moderate amount, and a great deal.

Another set of questions asked about instructional behaviors, including specific uses of the state test results for improving classroom practice. Although all the options were endorsed by majorities of teachers, more teachers found the results useful for improving curriculum than for addressing individual student needs. Table 5 compares the responses of mathematics teachers regarding three potential uses of state test results: improving curriculum,

-
individualizing instruction, and identifying the need to enhance their own subject matter knowledge or teaching skills. A higher percentage of Georgia mathematics teachers, at both the elementary and secondary levels, agreed that their state tests were helpful for all three purposes than California or Pennsylvania teachers. Roughly 80 percent of Georgia mathematics teachers agreed or strongly agreed with all three statements. This probably reflects something about the nature of the tests, the score reports, and the culture of improvement one finds in Georgia schools. The survey results do not allow us to separate these influences (or others), but we hope that the larger study will help us understand the key features that appear to make the Georgia tests more useful. In California and Pennsylvania, mathematics teachers were more likely to find the test results helpful for identifying areas to strengthen their own content knowledge than for tailoring instruction for individual students. In California, elementary teachers were more likely to agree with each of the statements than middle school teachers.

Principals also reported that state test results were useful to them for a variety of purposes (see Table 6). A Georgia elementary school principal described her use of the annual Criterion Referenced Competency Test (CRCT) results to plan with teachers, “I spend time with them in the Fall. 'Let’s look back over the... CRCT. Let’s see the strengths; let’s see the weaknesses. What is it you feel like you need to work on as a grade-level goal? What materials do you need?’” Overall, about two-thirds of principals in California and Pennsylvania found the results moderately or extremely helpful for developing their school improvement plans, as did almost all principals in Georgia.

Table 5.
Percentage of Teachers Agreeing or Strongly Agreeing with
Statements about the Use of State Tests, 2004-05

	California		Georgia		Pennsylvania	
	Elementary	Middle	Elementary	Middle	Elementary	Middle
The test results helped me identify and correct gaps in curriculum and instruction	63 (4)	53 (4)	86 (2)	84 (2)	63 (4)	58 (4)
The individual student results helped me tailor instruction to individual student needs	54 (4)	35 (4)	84 (2)	78 (3)	40 (4)	50 (5)
The test results allowed me to identify areas where I need to strengthen my content knowledge or teaching skills	70 (4)	55 (4)	89 (2)	79 (2)	69 (4)	60 (5)

Notes:

Standard errors are shown in parentheses.

Response options included strongly disagree, disagree, agree, and strongly agree.

State test results were also widely seen as useful for individualizing instruction, changing curriculum and focusing teachers' professional development. For example, the principal at a California middle school described the way the school's administrative team analyzed the data at the beginning of the school year and then taught teachers how to analyze the data during professional development (called "AYP 101"). Each teacher is given a list of "target group" kids who are close to proficient (4-5 students) and given pointers on how to differentiate instruction for these students. For example, the math coach reported going over strategies on how to involve these students, like moving them closer to the front of the class. He called this their "AYP strategy." Test results were less often used for assigning students to teachers (perhaps because

they are often not received until after school has begun), and for identifying the strengths and weaknesses of individual teachers.

Table 6.
Percentage of Principals Reporting that State Assessment Data is Moderately or Very Useful for Making Decisions

	California	Georgia	Pennsylvania
Developing a school improvement plan	77 (8)	98 (1)	73 (7)
Identifying students who need additional instructional support	72 (9)	95 (8)	66 (7)
Making changes to curriculum and instructional materials	71 (9)	79 (5)	83 (5)
Focusing teacher professional development	71 (8)	85 (4)	78 (8)
Making decisions on how much time is spent on each subject	54 (10)	69 (8)	51 (9)
Identifying teacher strengths and weaknesses	49 (9)	74 (5)	41 (9)
Making decisions regarding student promotion or retention	45 (11)	77 (6)	27 (9)
Assigning students to teachers	21 (5)	57 (7)	16 (5)

Notes:

Standard errors shown in parentheses.

Response options included not useful, minimally useful, moderately useful, and very useful.

It is interesting to note that more principals in Georgia than in California or Pennsylvania reported that test results were useful for each purpose. It appears that the Georgia testing program is perceived to have greater utility by principals than the testing programs in the other two states, a finding that is consistent with differences in teachers' responses across states. In a separate question, over 90 percent of principals in all three states indicated that one of their strategies for school improvement was "increasing use of student achievement data for improving of instruction." About two-thirds (64 percent) of all principals in Georgia found the

-
test results to be very useful for this purpose, compared with about one-half in California (54 percent) and about one-third in Pennsylvania (32 percent). Over 80 percent of the principals in all three states agreed or strongly agreed with the statement that teachers in their schools reviewed the state test results and used them to tailor instruction.

What Factors Influence Test Use?

A number of factors are likely to contribute to the usefulness of test results for administrative and classroom purposes, including the form in which scores are reported, the timeliness of reporting, and the attitudes and beliefs of principals and teachers about test validity and appropriateness. Survey responses shed some light on a number of features that are likely to be related to the use of state test results.

The manner in which results are aggregated may affect their usefulness to teachers and principals. States must report the proficiency level of each student and must summarize these data for each significant student subgroup and for the school as a whole. In addition, they may provide sub-scores for specific topics, item-level analyses, etc. According to mathematics teachers in our study, all three states provided results to schools summarized by subgroup and by subtopic. Eighty percent of elementary school teachers and middle school teachers in California, Georgia and Pennsylvania reported having access to mathematics test results by subgroup and by subtopic. However, as is shown in Table 7, roughly twice as many teachers reported that the subtopic or skill summaries were moderately or very useful as reported that the subgroup scores were moderate or very useful. This was true for both elementary and middle school teachers. Again, Georgia mathematics teachers stood apart from the other two states. A higher proportion

of Georgia mathematics teachers than California or Pennsylvania mathematics teachers reported that each type of score was useful to them.

Table 7.
Percentage of Teachers Reporting Availability and Usefulness
of Mathematics Test Result Summaries, 2004-05

	California		Georgia		Pennsylvania	
	Available	Moderately or Very Useful	Available	Moderately or Very Useful	Available	Moderately or Very Useful
Elementary Teachers						
Mathematics test results summarized by student subgroup	86 (2)	36 (4)	88 (2)	51 (3)	81 (3)	27 (2)
Mathematics test results disaggregated by subtopic/skill	88 (2)	68 (3)	94 (1)	80 (2)	82 (3)	66 (3)
Middle School Teachers						
Mathematics test results summarized by student subgroup	89 (3)	30 (5)	82 (3)	57 (4)	79 (5)	27 (4)
Mathematics test results disaggregated by subtopic/skill	82 (2)	59 (4)	90 (3)	83 (2)	75 (6)	65 (7)

Notes:

Standard errors are shown in parentheses.

Response options included not available, available and not useful, available and minimally useful, available and moderately useful, and available and very useful.

Percentages in “moderately or very useful” columns are percentages of those reporting the resource was available.

Principals’ responses were similar to those from teachers (see Table 8). For example, almost all principals in all three states reported having access to test results summarized by subgroup and by subtopic. However, while a higher percentage of principals found subtopic reports to be useful than subgroup reports, the difference was not significant. It seems sensible that school administrators, who make decisions involving the whole school, might find subgroup reports relatively more useful than classroom teachers, who are only responsible for the students

they are assigned. Georgia principals were more likely to find both types of reports useful than principals in the other two states. In addition, principals were asked whether the state test results are reported with explicit links to content standards or lesson plans. Having either type of elaboration in the score reports would facilitate the task of using the results for improvement of curriculum or instruction. Over 80 percent of the principals in Georgia (85 percent) and Pennsylvania (80 percent), and over 70 percent of the principals in California (71 percent) agreed or strongly agreed that their state test results included such explicit links.

Table 8.
Percentage of Principals Reporting Availability and Usefulness
of Information and Assistance on 2003-04 State Test Results

	California		Georgia		Pennsylvania	
	Available	Very Useful	Available	Very Useful	Available	Very Useful
Test results summarized for each student subgroup	100	40 (9)	94 (3)	78 (6)	100	32 (8)
Test results summarized by subtopic or skill	89 (8)	55 (9)	96 (2)	84 (4)	100	43 (8)

Notes:

Standard errors shown in parentheses.

Response options included not available; not useful; minimally useful; moderately useful; and very useful.

Percent "very useful" is computed for those principals who had results available.

There are other factors that are likely to influence the use of test results, including when they are made available to teachers and how easily they are understood. There were significant differences among the states in both the timeliness of reporting and the clarity of results based on reports from mathematic teachers (see Table 9). About 70 percent of elementary and middle school mathematics teachers in Georgia reported receiving state test results in a timely manner. In California and Pennsylvania, a similar percentage of middle school teachers' reported timely receipt of test results, but the percentages were much lower among elementary school teachers.

The difference was most notable in Pennsylvania, where barely one-third of elementary school teachers (36 percent) reported receiving test results in a timely manner. Late results are much less helpful. As one elementary school teacher explained, they serve as “more of a verification, because by the time you look at them, you’ve got a feel for the kids, anyway.”

Table 9.
Percentage of Teachers Agreeing or Strongly Agreeing with
Statements about the Features of the State Tests, 2004-05

	California		Georgia		Pennsylvania	
	Elementary	Middle	Elementary	Middle	Elementary	Middle
I received the test results in a timely manner	58 (4)	70 (2)	71 (3)	69 (4)	35 (4)	68 (3)
The test results were clear and easy to understand	72 (3)	67 (4)	95 (1)	93 (2)	74 (3)	74 (3)

Notes:

Standard errors are shown in parentheses.

Response options included strongly disagree, disagree, agree, and strongly agree.

A similar pattern of responses was obtained from principals when asked about timeliness of state test reports. About one-half of Georgia principals (47 percent) agreed that information about the school’s performance was received in a timely manner, compared to about one-third of California principals (37 percent) and only one-fifth of Pennsylvania principals (19 percent).

Most teachers in all three states reported that the test results were clear and easy to understand. However, this sentiment was almost universal among mathematics teachers in Georgia, while it was only true for about 70 percent of mathematics teachers in the other two states. Responses from principals to a question about the clarity of test score reports were similar to responses from teachers. Ninety percent of Georgia principals and 86 percent of Pennsylvania principals agreed that the information they received about the school’s performance was clear and easy to understand. Only 75 percent of California principals agreed that their state test

reports were clear. Principals explained some of the shortcomings of the state reports in interviews. For example, a Pennsylvania elementary school principal explained:

All we're given is, they've put these beautiful graphs on that say that this amount of students are proficient and this many aren't and etcetera, which is great. You can look at it and you can say, 'Okay, they're not.' But it's not criterion-referenced and it's really difficult. You know that in reading concepts we're not doing as well as we should be. Well, that's great, but under reading concepts there are like, thirty skills or thirty objectives to cover. Which ones? What is it that we're not... it doesn't tell us anything.

In the future we will review the state test reports from the three states in greater detail to try to identify the differences that make some appear clearer than others.

Test use may also be explained, in part, by teachers' perceptions of test quality. A majority of teachers in Georgia agreed that the state test was a good measure of students' mastery of content standards, but that was not the case in California or Pennsylvania (see Table 10). Elementary school and middle school mathematics teachers' responses to this question were similar in all three states. This is consistent with previously reported test use. The surveys did not reveal exactly what was lacking in state tests, but teachers' responses provided some clues. For example, about one-quarter to one-third of teachers in each state felt the test contained considerable material that was not in their curriculum; a similar percentage also agreed that the test omitted considerable material that was in their curriculum. Thus, alignment between the test and the mathematics curriculum seems to be a concern in all three states. Questions about content alignment were greatest among middle school mathematics teachers in Pennsylvania. About 60 percent of the teachers in Georgia and Pennsylvania agreed that the state tests had adequate representation of mathematical reasoning and problem solving, but this was not the case in California, where only about 40 percent of teachers agreed. California mathematics teachers thought their state test contained relatively less coverage of higher-order mathematical

-
thinking than teachers in the other two states. California mathematics teachers were also much less likely than teachers in the other two states to report that they aligned their instruction with the state mathematics assessment. About one-half of the teachers in California agreed with this statement compared with over 80 percent of the teachers in Georgia and Pennsylvania. Finally, about two-thirds of the middle school teachers in California and Pennsylvania agreed that their tests were too difficult for the majority of their students. Taken together these results may explain some of the earlier results concerning the use of test results. Certainly, they point to concerns about the appropriateness of test content and the alignment of tests with curriculum in California and Pennsylvania.

Another potential explanation for some of the California reactions is the continuing use of the state's own accountability system as a component of the NCLB system. The state's academic performance index (API) is growth based, rewarding improvement at the bottom of the distribution more than at the top, whereas the AYP rewards proficient performance only. The combination of systems sends complex signals to schools. As one California middle school principal explained, "what you really have to do is focus on the middle and try to move the middle to the top for the AYP. And for the API you have to focus on the bottom and move the bottom to the middle because the API rewards improvements."

Principals' judgments about test quality were consistent with teachers' judgments. Over 71 percent of Georgia principals agreed or strongly agreed that the state test scores accurately reflected the achievement of their school's students. In comparison, 54 percent of principals in Pennsylvania agreed with this statement, and only 41 percent of principals in California agreed.

Table 10.
Percentage of Teachers Agreeing to Statements Regarding
the Quality of the State Assessments, 2004-05

	California		Georgia		Pennsylvania	
	Elementary	Middle	Elementary	Middle	Elementary	Middle
The mathematics assessment is a good measure of students' mastery of content standards	42 (4)	38 (4)	60 (3)	59 (3)	45 (3)	50 (5)
The mathematics assessment includes considerable content that is not in our curriculum	33 (3)	32 (4)	24 (2)	27 (3)	25 (4)	43 (6)
The mathematics assessment omits considerable content that is in our curriculum	35 (3)	30 (3)	26 (2)	37 (3)	27 (3)	49 (5)
The mathematics test adequately measures mathematical reasoning and problem solving	39 (4)	41 (3)	61 (3)	58 (3)	59 (3)	59 (4)
I have aligned my teaching with the mathematics assessment	52 (4)	51 (4)	82 (3)	77 (2)	87 (3)	86 (3)
The mathematics test is too difficult for the majority of my students	47 (5)	65 (4)	31 (3)	46 (3)	47 (4)	64 (5)

Notes:

Standard errors are shown in parentheses.

Only teachers who taught a tested grade are included in this table. While this includes all the teachers in our sample for California and Georgia, it only includes mathematics teachers who taught students in grades 3, 5, and 8 in Pennsylvania.

Response options included strongly disagree, disagree, agree, strongly agree, and I don't know.

Usefulness might also be a function of the support teachers receive for interpreting the results and using them to guide instruction. Elementary and middle school mathematics teachers in all three states had access to workshops in which test results were presented and explained

-
(see Table 11). However, teachers in Georgia were more likely to find these workshops to be useful than teachers in the other two states. Less than one-half of the elementary and middle school mathematics teachers in California and Pennsylvania found the workshops they attended to be useful, compared with 59 percent in Georgia. It may be the case that the workshops were better in Georgia or that the teachers were less well prepared, so they appreciated them more. It is also possible that the Georgia teachers' impressions were affected by their generally more positive views of the state testing program as a whole.

Specific training on using test results for instructional planning was more widely available in Georgia than in the other two states. Three-quarters of the elementary mathematics teachers in Georgia (and over 60 percent of the middle school mathematics teachers) had access to specific training in using test results for instructional planning or school improvement, and about 60 percent of those who participated in the training found it moderately or very useful. California teachers were somewhat less likely than Georgia teachers to find training availability and useful, and Pennsylvania teachers were somewhat less likely than California teachers.

Although a number of companies and organizations have developed software to help educators analyze test results, only about one-third of mathematics teachers in California and Pennsylvania had access to such software, and only about one-third of those who used the software found it to be useful. In Georgia, roughly half of the mathematics teachers had such software available, and about half of them reported that the software was useful.

Table 11.
Percentage of Teachers Reporting Availability and Usefulness of
Resources for Using Mathematics Test Results, 2004-05

	California		Georgia		Pennsylvania	
	Available	Moderately or Very Useful	Available	Moderately or Very Useful	Available	Moderately or Very Useful
Elementary Teachers						
Workshops or meetings where mathematics test results are presented and explained	71 (3)	39 (5)	77 (2)	59 (3)	72 (4)	38 (3)
Training on how to use mathematics test results for instructional planning or school improvement	59 (4)	53 (5)	76 (3)	60 (3)	47 (3)	48 (3)
Computer software or systems for re-analyzing mathematics test results	36 (4)	37(4)	53 (3)	50(3)	31 (3)	29(4)
Middle School Teachers						
Workshops or meetings where mathematics test results are presented and explained	66 (3)	38 (3)	70 (3)	51 (4)	70 (7)	30 (4)
Training on how to use mathematics test results for instructional planning or school improvement	43 (3)	49 (3)	64 (4)	60 (4)	51 (6)	40 (4)
Computer software or systems for re-analyzing mathematics test results	33 (7)	35 (8)	48 (5)	52 (4)	25 (5)	34 (8)

Notes:

Standard errors are shown in parentheses.

Response options included not available, available and not useful, available and minimally useful, available and moderately useful, and available and very useful.

Percentages in “moderately or very useful” columns are percentages of those reporting the resource was available.

Another factor that may have influenced teachers' use of assessments was the professional development they received. Table 12 summarizes teachers' reports about the extent to which the professional development they received emphasized two aspects of testing: preparing students to take tests and using test results. In all three states, elementary teachers were more likely than middle-school teachers to report that their professional development included a moderate or major emphasis on test preparation or using test results. Both themes were emphasized with a greater percentage of teachers in Georgia than in California. The responses from teachers in Georgia and Pennsylvania were similar in terms of preparing students to take tests, but a higher percentage of Georgia teachers reported that their professional development emphasized the use of test results. In general, these results are consistent with responses to the earlier question on support for use of test results.

Table 12.
Percentage of Teachers Reporting that Professional Development had a Moderate or Greater Emphasis on Selected Features, 2004-05

	California		Georgia		Pennsylvania	
	Elementary	Middle	Elementary	Middle	Elementary	Middle
Professional development emphasized preparing students to take the state test	47 (6)	28 (4)	74 (3)	56 (3)	67 (3)	58 (4)
Professional development emphasized interpreting and using reports of student test results	44 (6)	24 (3)	65 (3)	45 (4)	36 (4)	33 (7)

Note:
Standard errors are shown in parentheses.

The principal may also influence teachers' use of data through encouragement and action. Table 13 reports on the percentage of mathematics teachers who reported encouragement and

help from their principal to incorporate the state standards and use data from the state test to adapt the curriculum. Emphasis on the state standards was almost universal, according to teachers. However, assistance from principals in adapting curriculum was somewhat less common. About two-thirds of the teachers in California and Pennsylvania reported that their principal helped them improve curriculum through analysis of test results. In Georgia, the percentage was notably higher among elementary school teachers.

Table 13.
Percentage of Teachers Who Agree or Strongly Agree that Principal
at Their School Engages in Selected Behaviors, 2004-05

	California		Georgia		Pennsylvania	
	Elementary	Middle	Elementary	Middle	Elementary	Middle
The principal at my school encourages teachers to review the state content standards and incorporate them into their teaching	95 (2)	93 (2)	97 (1)	90 (2)	94 (1)	94 (2)
The principal at my school helps teachers adapt our curriculum based on an analysis of state test results	68 (4)	64 (4)	85 (2)	71 (4)	70 (3)	65 (6)

Notes:

Standard errors are shown in parentheses.

Response options included strongly disagree, disagree, agree, and strongly agree.

Similar to teachers, the vast majority of principals in the three states reported the availability of workshops, during which results were presented and explained, and training, during which uses for school improvement were discussed (see Table 14). However, principals in California and Pennsylvania were less likely than teachers in those states to report that the workshops and training were useful to them. Only in Georgia were principals' reports consistent with teachers' reports. Principals were also much more likely than teachers to have access to

computer software for analyzing test score data. In many schools such software is only available to administrators or is located on an administrative computer that is difficult for teachers to access. However, with the exception of those in Georgia, principals did not find such software to be useful. Fewer than 30 percent of the principals in California or Pennsylvania who had access to analysis software report that the software was useful. In Georgia, in contrast, two-thirds of principals had access and more than one-half found it useful. Further examination of the computerized systems in Georgia will be undertaken in the future.

Table 14.
Percentage of Principals Reporting That Assistance for
Data Use was Available and Very Useful, 2004-05

	California		Georgia		Pennsylvania	
	Available	Very Useful	Available	Very Useful	Available	Very Useful
Workshops or meetings where test results are presented and explained	78 (7)	29 (10)	85 (4)	54 (7)	87 (6)	30 (7)
Training on how to use test results for instructional planning or school improvement	84 (6)	37 (9)	87 (4)	67 (6)	82 (7)	33 (6)
Computer software or systems for re-analyzing test results	57 (8)	28 (10)	66 (3)	53 (8)	71 (7)	20 (6)

Notes:

Standard errors are shown in parentheses.

“Very useful” is as a percentage of principals who had assistance available.

Response options included not available, available and not useful, minimally useful, moderately useful, and very useful.

Teachers were also asked about the administration and use of “progress tests.” We defined progress tests (which are also called “benchmark tests” or “interim tests”) as required standardized tests that are administered periodically to monitor student progress toward meeting state standards. The use of progress tests has been growing in recent years. In 2004-05, more

-
than two-thirds of elementary mathematics teachers and more than one-half of middle school mathematics teachers in Georgia were required to administer progress tests (an increase of a four to seven percentage points since 2003-04). In California, 62 percent of elementary teachers and 31 percent of middle school teachers administered progress tests in 2004-05 (up from 56 percent and 31 percent the year previously). In Pennsylvania, 47 percent of elementary teachers and 50 percent of middle school mathematics teachers were required to administer progress tests in 2004-05 (up from 30 percent and 28 percent, respectively, the previous year).

Progress tests are one of the strategies districts are using to promote greater “data based decision-making” on the part of teachers. In theory, the advantage of progress tests is that they provide quick results that can be used to inform instructional planning at regular intervals during the school year. As one Pennsylvania elementary school teacher described the advantages of the progress test, “It’s quicker, a faster turnaround. I teach it, I test it, and I can see who gets it, who doesn’t get it, and I can go back in and work with that student immediately. It’s just more direct.” Overall, responses from teachers suggest that progress test results are helpful for instructional planning. As Table 15 illustrates, large majorities of teachers who use progress tests think they are a good measure of student mastery, good preparation for the states tests, and, perhaps most importantly, they help teachers identify and correct gaps in curriculum and instruction. In most cases, the percent of teachers agreeing with these statements about progress tests are greater than the percentage agreeing with similar statements about the annual state tests. It is also noteworthy that, in almost all cases, progress tests were “low stakes” tests, i.e., there were not direct consequences for teachers as a result of scores on these examinations. Some schools even organized interventions around progress testing, a sophisticated data-driven strategy.

As one Pennsylvania middle school teacher explained,

We have an ESR [Enrich, remediate, and supplement] team here and they're the ones that administer it [the progress test]. They will correct it. And we got feedback from them. That's what I use more than sitting down and doing PSSA. They gave us a sheet on every student, what they did, where they got their points in which section of the math and the reading, and so forth. And that's what I use more.

Table 15.
Percent of Mathematics Teachers who Agreed or Strongly Agreed
about Various Features of Progress Tests, 2004-05

	California		Georgia		Pennsylvania	
	Elementary	Middle	Elementary	Middle	Elementary	Middle
Progress tests are a good measure of student mastery of content standards	57 (5)	59 (5)	62 (4)	68 (4)	70 (3)	69 (5)
Progress tests are good preparation for the state mathematics assessment	54 (7)	66 (8)	72 (4)	81 (3)	75 (3)	81 (6)
Progress tests help identify and correct gaps in curriculum and instruction	76 (5)	76 (5)	76 (3)	82 (3)	84 (3)	86 (4)
There are consequences for teachers associated with performance on the tests	3 (1)	5 (5)	9 (3)	8 (3)	4 (2)	7 (3)

Notes:

Standard errors are shown in parentheses.

The percentages are percentages of the total number of teachers reporting that progress tests were required.

Response options were strongly disagree, disagree, agree, and strongly agree.

However, not all teachers gave positive assessments regarding progress tests. Unless the tests provided information in a useful format, they were not perceived as beneficial. A Georgia elementary school teacher described the limitation of their progress tests: "I never felt like I got real meaningful information to let me know whether or not the kids were making satisfactory

-
progress. What we got was a percentage of the questions that the kids got right in various categories, but we had no idea how that related to passing/failing [CRCT].”

Examining Variation in Data Use and its Correlates

The descriptive information presented so far provides some evidence regarding the ways teachers use data from standards-based accountability systems and how that use varies among three states. Many of the comparisons suggest a more positive climate for data use exists in one state than in the other two states. To date, the study has identified a number of factors that might explain these differences, including historical and political context as well as assessment and accountability policies. However, with only three state “cases,” the ISBA study will not be able to disentangle all the relationships fully, although further study will explore the state-to-state differences in greater detail.

On the other hand, the study was designed to analyze within-state variation and to provide a better understanding of what district and school actions facilitate use of data in the context of NCLB-style accountability. The ISBA project is addressing these questions through a series multivariate analyses that reflect the hierarchical nature of education and of NCLB accountability. These analyses are not complete, but initial results suggest some promising directions.

First, we are exploring the degree to which teachers’ use of data varies within and between schools and districts. If resources and guidance provided by principals or district staff have a strong influence on what teachers do, we might expect to see a high degree of consistency among teachers within a school or a district (depending on the level at which the intervention took place). Preliminary examination of sources of variability in our survey responses reveal a

-
lack of this type of consistency in most teacher responses related to data use. For example, school-level intraclass correlations among teachers' responses to the questions about test-score data usefulness were generally around 0.10, and district-level intra-class correlations were lower. This suggests that most of the variability in these responses occurs among teachers within the same school. That is, within a school, different teachers tend to have different impressions of the usefulness of data. Opinions about the quality of the mathematics test were even more variable within schools.

Second, we are constructing a series of multilevel models to examine relationships among data availability, perceived usefulness of data, classroom practices, supports for data use, and (eventually) student achievement. Preliminary, exploratory models provide some suggestive evidence that school and district actions that are designed to promote better data use do influence what teachers do in the classroom, but, consistent with the intraclass correlations discussed above, these influences tend to be small. For example, teachers' propensity to adapt their instruction to focus on tested content is positively associated with their principals' opinions regarding the usefulness of test results and with district-level actions designed to promote awareness of state content standards. We are currently refining and testing several models, with the goal of providing evidence regarding the links between data use, supports, and desired outcomes.

Discussion

Findings from the ISBA study suggest a number of factors that influence the use and utility of test results for decision-making. These factors include specific features of the tests, the

-
support provided to teachers, and some aspects of state policy. Although they are primarily descriptive, the results reported here help to quantify the extent to which some level of data-driven decision-making occurs in classrooms, and the extent to which this use is changing. Future analyses will probe deeper to examine relationships among these variables and other aspects of standards-based accountability.

To recap briefly, over one-half of teachers reported that they often used test results to identify topics that need review and to identify students who need attention. Furthermore, about one-quarter of teachers reported that their use of test score data for these purposes increased in the past year. Teachers also reported specific changes in classroom practice that occurred as a result of the annual state assessments. Large percentages of teachers reported that they were attending more to standards, looking for more effective methods, and redirecting their lessons to focus on content as a result of the state assessment. On the other hand, more than half of the teachers reported focusing specifically on the topics contained in the assessments, on the styles and formats used in the assessments, and on test preparation activities. These results illustrate the dual nature of data-driven decision-making. To the extent that teachers are using the information contained in test scores to better focus instruction on state standards and on gaps in student understanding, then attention to data is leading to positive outcomes. To the extent that teachers are using test results to focus on incidental features of tests (e.g., unusual styles or formats), on the narrow domain embodied in tests (rather than the broad domain described by standards), or to spend excessive time in test preparation activities, then the attention to data may be leading to negative outcomes. These latter behaviors may lead to inflated test scores and reduce the validity of the accountability systems.

Teachers' perceptions about the quality of the data they receive from state tests may influence their use of the data, so attention to data and report quality is important. In general, teachers reported that the test results were clear and were available in a timely manner, although there were state-to-state differences that warrant further examination. According to our surveys, the majority of teachers in two of the three states thought their state test was not a good measure of student performance. Although we did not probe more deeply, these reports indicate that teachers believe either the conditions, the format, or the content of the tests are lacking, leading to invalid measures of student performance. For example, one-quarter to one-third of teachers in all three states thought the test content was not aligned with their curriculum. Such perceptions present a serious obstacle to increased use of test results in the classroom. Although most teachers reported taking steps to align their teaching with the mathematics tests, such efforts will be incomplete if their available curriculum materials do not match the content of the tests. Misalignment limits teachers' ability to use test results for decision making in two ways; it limits the relevance of the information about performance teachers receive, and it restricts teachers' options for instructional intervention.

On the other hand, teachers reported that they found the test data clear and easy to understand, and that they were receiving a variety of support for using test data. The support included workshops and professional development to help them understand the data. Less common were efforts to tie data to lesson planning and access to software for customized analysis or reporting of results. Teachers and principals in one state found the workshops and training to be very useful, but the judgments of teachers and principals in the other states were less positive.

Many teachers also relied on progress tests to help in instructional planning. In fact, teachers who had access to data from progress tests were more likely to report that these data were used for instructional planning than were data from annual state testing. This finding is consistent with responses that annual state test results were more helpful for curriculum improvement than for improving specific instructional practices. These findings suggest that data-driven decision-making may be better served by periodic standardized progress testing or formative classroom assessment than by annual state tests.

The variation we observed among responses from teachers in different states suggests that state policy matters. In particular, conditions in Georgia seem to be more positive with respect to test score reporting and use than in the other two states we studied. We have not had time yet to fully explore the features of each state context to try to understand these differences. However, we plan to spend more time investigating Georgia's approach to standards-based accountability to learn why teachers and principals in that state responded more positively about so many features of their standards-based accountability system.

In addition, the high level of variation we observed within schools suggests that school-wide policies are not having uniform effects. This finding is consistent with most studies that find wide variation in classroom practice within schools, and it confirms that it remains a challenge to foster school-wide shifts in attitudes and practices. These results reinforce the sense that school- and district-level actions that are geared toward all teachers are having at best a small effect on teachers' responses related to data use. They also suggest that future analysis examine teacher-level influences, such as the individual exposure to data-related professional development and training.

-
Unfortunately, these initial analyses leave a number of important questions about test use unanswered. We hope that future analytic work as part of the ISBA project will provide more insights into the functioning of standards-based accountability systems and the factors related to the use of test results for improvement. It is also important to note that these results were drawn from a study that did not have data-driven decision-making as its focus. More insight into the use of data for instructional decision-making is likely to come from research designed with that purpose in mind.

References

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: Authors.
- Corbett, H. D., & Wilson, B. L. (1988). Raising the stakes in statewide mandatory minimum competency testing. In W. L. Boyd & C. T. Kerchner (Eds.), *The politics of excellence and choice in education: The 1987 Politics of Education Association yearbook* (pp. 27–39). New York: Falmer Press.
- Council of Great City Schools (2002). *Beating the odds II*. Washington, DC: Author.
- Hamilton, L. S. (2003). Assessment as a policy tool. *Review of Research in Education*, 27, 25–68.
- Herman, J. (2005). *Making accountability work to improve student learning* (CSE Technical Report 649). Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing.
- 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.
- Koretz, D., Barron, S., Mitchell, K., & Stecher, B. (1996). *The perceived effects of the Kentucky Instructional Results Information System (KIRIS)*. Santa Monica, CA: RAND.

- Marsh, J., Kerr, K., Ikemoto, G., Darilek, H., Suttorp, M.J., Zimmer, R., & Barney, H. (2005). *The role of districts in fostering instructional improvement: Lessons from three urban districts partnered with the Institute for Learning*. Santa Monica, CA: RAND.
- Mason, S. (2002). *Turning data into knowledge: Lessons from six Milwaukee public schools*. Madison, WI: Wisconsin Center for Education Research.
- 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*. Chicago: University of Chicago Press.
- Pedulla, J. J., Abrams, L. M., Madaus, G. F., Russell, M. K., Ramos, M. A., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Boston: National Board on Educational Testing and Public Policy.
- Romberg, T. A., Zarinia, E. A., & Williams, S. R. (1989). *The influence of mandated testing on mathematics instruction: Grade 8 teachers' perceptions*. Madison, WI: National Center for Research in Mathematical Science Education, University of Wisconsin–Madison.
- Shepard, L. A., & Dougherty, K. C. (1991, April). *Effects of high-stakes testing on instruction*. Paper presented at the annual meeting of the American Educational Research Association and National Council on Measurement in Education, Chicago.
- Stecher, B. M., Barron, S. I., Chun, T., & Ross, K. (2000). *The effects of the Washington state education reform on schools and classrooms* (CSE Tech. Rep. 525). Los Angeles: Center for Research on Evaluation, Standards, and Student Testing.
- Stecher, B. M., Hamilton, L. S., & Gonzelez, G. (2003). *Working smarter to leave no child behind*. Santa Monica, CA: RAND.

-
- Stecher, B. M., Hamilton, L. S., & Naftel, S. (2005). *Introduction to first-year findings from the Implementing Standards-Based Accountability (ISBA) Project*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal.
- Stecher, B. M., & Naftel, S. (2006). *Implementing standards based accountability (ISBA): Introduction to second year findings*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Supovitz, J. A., & Klein, V. (2003). *Mapping a course for improved student learning: How innovative schools systematically use student performance data to guide improvement*. Philadelphia, PA: Consortium for Policy Research in Education, University of Pennsylvania Graduate School of Education.