



PROJECT MUSE®

Views from the Classroom: Teachers' Opinions of Statewide Testing Programs

Lisa M. Abrams, Joseph J. Pedulla, George F. Madaus

Theory Into Practice, Volume 42, Number 1, Winter 2003, pp. 18-29
(Article)

Published by Ohio State University College of Education



➔ For additional information about this article

<https://muse.jhu.edu/article/41406>

Lisa M. Abrams
Joseph J. Pedulla
George F. Madaus

Views from the Classroom: Teachers' Opinions of Statewide Testing Programs¹

This article discusses teachers' views on state-mandated testing programs. An overview of the literature is presented, as well as results from a nationwide survey of teachers. Findings from both suggest that high-stakes state-mandated testing programs can lead to instruction that contradicts teachers' views of sound educational practice. In particular, teachers frequently report that the pressure to raise test scores encourages them to emphasize instructional and assessment strategies that mirror the content and format of the state test, and to devote large amounts of classroom time to test preparation activities. The article concludes that serious reconsideration must be given to the use of high-stakes consequences in current statewide testing programs.

A GROWING BODY OF EVIDENCE suggests that high-stakes testing can be a driving force behind fundamental change within schools (Koretz, Linn, Dunbar, & Shepard, 1991; Madaus, 1988; McNeil, 2000; Smith, 1991). However, there is a difference of opinion as to whether this change is for better or for worse. For example, while some feel that the guarantee of rewards or the threat of

sanctions is essential for promoting quality teaching and encouraging higher student achievement, others have found that high-stakes tests limit the scope of classroom instruction and student learning in undesirable ways (Stecher & Barron, 1999; Stecher, Barron, Chun, & Ross, 2000). Regardless of one's position on this issue, it is impossible to deny that statewide testing policies influence classroom instruction and student learning. The question addressed by this article is: How do teachers perceive the effects of these testing programs, particularly in the area of teaching and learning?

The article is divided into three sections. The first section presents an overview of the literature on teachers' perceptions of state testing programs. The second section presents findings from a nationwide survey of teachers. Results from the latter confirm many of the findings in the literature, but also present new information on the interaction between impacts and the stakes attached to the state test results. The article concludes with a commentary on the need to reconsider the use of high-stakes consequences when developing and/or implementing statewide testing programs.

Overview of the Literature on Teachers' Perceptions of State Testing Programs

Numerous research studies have investigated the effects of state-mandated testing programs—particularly those with high stakes attached to the

Lisa M. Abrams is a research associate, Joseph J. Pedulla is an associate professor, and George F. Madaus is a professor, all in education at Boston College.

test results—on schools, teachers, and students. The majority of these studies gathered information from teachers and administrators by using surveys, interviews, classroom observations, and various combinations thereof. Most studies tend to focus on a single state. Given the varied nature of state testing programs in terms of the format, grade level, and subject areas tested, it probably is not surprising that research on the effects of these programs yields both positive and negative results. Generally, research on the effects of state testing programs, particularly those with high stakes attached, has focused on classroom practices, teachers, and students.

This section summarizes the findings from survey-based research conducted in various states. In all of the states that are mentioned, high stakes were attached to test results at the school and/or student level. For example, Kentucky, Vermont, and Washington use test results to hold schools accountable. In Maryland, North Carolina, Texas, and Virginia, test results are used to make highly consequential decisions at both the school and student levels. The review of current research on teachers' perceptions of state testing programs is organized around four main topic areas: (a) impact on classroom practices in terms of the content of instruction and the strategies used to deliver instruction, (b) the pressure to prepare students for the state test, (c) impact on teacher and student motivation and morale, and (d) views of accountability.

Impact on classroom practices

Much of the research on state testing programs addresses their effects on what is taught. A common finding is that teachers report giving greater attention to tested content areas. For example, of the 722 Virginia teachers surveyed by McMillan, Myran, and Workman (1999), more than 80% indicated that the state Standards of Learning (SOL) test had impacted their instruction, particularly with regard to the content focus of daily lessons. Overall teacher responses led the study authors to conclude that, "teachers are placing greater emphasis on covering the content of the SOL" (p. 10). Often, increased attention toward tested content has led to a decreased emphasis on nontested curricular areas. For example, a study in Arizona indicated that teachers did not place as much emphasis on

nontested subjects such as social studies and science (Smith, Edelsky, Draper, Rottenberg, & Cherland, 1991). In Kentucky, 87% of teachers surveyed agreed that the Kentucky Instructional Results Information System (KIRIS) had "caused some teachers to de-emphasize or neglect untested subject areas" (Koretz, Barron, Mitchell, & Stecher, 1996a, p. 41). Teachers in North Carolina reported similar results (Jones et al., 1999).

The impact of the state test on instructional strategies is less clear-cut. While teachers in North Carolina (Jones et al., 1999) reported mixed effects on instructional strategies, the majority of writing teachers surveyed in Kentucky indicated that the KIRIS writing portfolios had a positive impact on writing instruction (Stecher, Barron, Kaganoff, & Goodwin, 1998). Similarly, in a previous study in Kentucky, 80% of the fourth grade and eighth grade mathematics teachers surveyed reported increasing instructional emphasis on problem solving and writing as a result of the portfolio-based state test (Koretz et al., 1996a). However, in Virginia, McMillan, Myran, and Workman found the state test to have had a greater impact on the content and pace of instruction rather than on the "mode of instruction" (1999, p. 10). Perhaps these differences in research findings are a function of the format of the state test, since Virginia's tests are predominantly multiple choice while the state test in Kentucky at the time of the study was based on portfolios.

Pressure on teachers to improve student performance

Teachers have responded to the pressure to improve scores on the state test, particularly in high-stakes settings, by spending more classroom time preparing students specifically for the state test. In Maryland, 88% of teachers surveyed felt they were under "undue pressure" to improve student performance on the state test (Koretz, Mitchell, Barron, & Keith, 1996b). An even larger proportion of Kentucky teachers (98%) responded similarly when asked the same question (Koretz et al., 1996a).

An increased emphasis on test preparation is one of the possible outcomes of the pressure teachers feel to improve student performance. Of 470

elementary teachers surveyed in North Carolina, 80% indicated that “they spent more than 20% of their total instructional time practicing for the end-of-grade tests” (Jones et al., 1999, p. 201). Similarly, a survey of reading teachers in Texas revealed that on average teachers spent 8 to 10 hours per week preparing students for the Texas Assessment of Academic Skills (TAAS) (Hoffman, Assaf, & Paris, 2001). The most common test preparation activities reported by Texas teachers included demonstrating how to mark the answer sheet correctly, providing test-taking tips, teaching test-taking skills, teaching or reviewing topics that would be on the test, and using commercial test-preparation materials and tests from previous years for practice (Hoffman et al., 2001, p. 6).

One concern stemming from the reported emphasis on specific test preparation activities centers on the validity of the test scores as a measure of student achievement. Specific test preparation activities, coaching, and instruction geared toward the test can yield scores that are invalid (Haladyna, Nolen, & Haas, 1991; Koretz et al., 1991; Linn, 2000; Madaus, 1988). For example, one would expect that if student scores are improving on the state test from year to year, scores on other tests that measure the same content and/or skills should show similar improvement. When trends in student performance levels on similar standardized tests are not consistent, the accuracy of a particular test as an indicator of student achievement is questionable. For example 50% of Texas teachers surveyed did not think that the rise in TAAS scores “reflected increased learning and high quality teaching” (Hoffman et al., 2001, p. 8). Based on comments provided by the responding teachers, the authors concluded that teachers regarded improvement on the TAAS as a “direct result of teaching to the test” (Hoffman et al., 2001, p. 9). Consequently, student performance on a highly consequential test may not generalize to other measures of achievement. For example, several studies have compared student performance on the state test with performance on other standardized tests that assessed similar content knowledge and/or skills. Koretz and Barron (1998) found that score gains on the KIRIS mathematics test were substantially larger than score gains for Kentucky students on

the math portion of the National Assessment of Education Progress (NAEP), suggesting that improved performance on the KIRIS math test did not necessarily reflect broader gains in student knowledge. Klein, Hamilton, McCaffrey, and Stecher (2000) found similar results when they compared results on the TAAS to the performance of Texas students on NAEP.

Impact on teacher and student motivation and morale

While intended to motivate teachers and students to achieve optimal performance levels, the high-stakes nature of state testing programs can have quite the opposite effect. With regard to teachers, researchers have cautioned that placing a premium on student test performance can reduce instruction to test preparation, thus limiting the range of educational experiences to which students are exposed and minimizing the skill that teachers bring to their craft (McNeil, 2000; Smith, 1991). In other words, the implementation of the state test may, in effect, lead to a de-professionalization of teachers. Studies also indicate that high-stakes assessments increase stress and decrease morale among teachers. According to Jones et al. (1999) more than 77% of the teachers surveyed indicated decreases in morale, and 76% reported teaching was more stressful since the implementation of the North Carolina state-testing program. Similar results were found in Kentucky and Maryland. Over half of the Maryland teachers and about 75% of Kentucky educators indicated that morale had declined as a result of the state test (Koretz et al., 1996a; Koretz et al., 1996b). In addition, 85% of Texas teachers surveyed by Hoffman, Assaf, and Paris (2001) agreed with the statement “some of the best teachers are leaving the field because of the TAAS.”

Other studies have raised similar concerns about the impact these tests have on students. Increased levels of anxiety, stress, and fatigue are often seen among students participating in high-stakes testing programs. All three can have detrimental effects on student performance. In a survey of North Carolina educators, 61% reported that their students were more anxious as a result of the state test (Jones et al., 1999). Similarly, one third of

teachers surveyed in Kentucky indicated that student morale had declined in response to the KIRIS (Koretz et al., 1996a).

According to Kellaghan, Madaus, and Raczek (1996), the key questions regarding student morale and motivation are: Which students does the test motivate, and what does the test motivate them to do? While the rewards or sanctions attached to test results may spur many students to achieve and even excel, they may drive others out of school. If students do not believe that an opportunity for success exists, the external motivating forces of the rewards or sanctions will have a minimal effect (Kellaghan et al., 1996). Ultimately, those students who view passage of the test as an insurmountable barrier may give up and drop out of high school. In this regard, several empirical studies have shown that the use of high-stakes state-mandated tests is associated with increased student drop-out rates (e.g., Haney, 2000; Reardon, 1996).

Tests as a means of accountability

Not only do the results of state tests provide information about the progress of individual students, the results are often aggregated to evaluate school and/or district performance. In 2001, 18 states rewarded schools with financial incentives for high or improved test scores, and at least 20 attached sanctions for schools due to poor student performance on the state test ("Quality Counts," 2002). In terms of the latter, schools might not only lose accreditation if students performed poorly, they might also lose funding and even be taken over by the state.

The majority of research on state testing programs has focused on the effects on classroom practices and has reported on changes in the focus, content, and pedagogy of instruction. In addition, several studies have directly tapped into teachers' views concerning the ways tests are used for accountability purposes. In North Carolina, 76% of the teachers surveyed "believed that the accountability program would not improve the quality of education in their state" (Jones et al., 1999, p. 202). However, research conducted in Maine and Maryland suggests that perceptions of the stakes attached to the test results may vary among teachers in the same state (Firestone, Mayrowetz, & Fairman, 1998),

thus suggesting that they can have a differential impact on schools within the same state. In other words, the intended effect of the rewards and/or sanctions tied to test performance may be influenced by other factors.

Few question the need for high standards and some mechanism for measuring student progress toward those standards. The main focus of the debate surrounding state testing programs centers on the severity of the sanctions attached to the test results and whether indicators in addition to test results should be used to hold educators and/or students accountable (Linn, 2000). The next section of this article describes a nationwide survey of teachers that included questions on some of these issues. In particular, this survey, which was carried out by the National Board on Educational Testing and Public Policy,² attempted to address the interaction between the stakes attached to the state test results and perceived impacts on teaching and learning. The focus of the survey items and the process used to select teachers enabled us to look critically at the relationship between school and student levels of accountability.

National Survey of Teachers' Perceptions of the Impacts of State Testing Programs

Survey development

The survey conducted by the National Board on Educational Testing and Public Policy (Pedulla et al., 2003) sought to ascertain teachers' attitudes and opinions about state-mandated testing programs. To this end an 80-item survey was developed. These items presented teachers with a series of statements about their state testing program, classroom practice, and student learning, and provided four response options: *strongly agree*, *agree*, *disagree*, and *strongly disagree*. Items focused on how the state test impacted classroom instruction and assessment; feelings of pressure associated with improving student performance; test preparation; teacher and student motivation and morale; and school, teacher, and student accountability. The survey was based, in part, on other surveys used in Arizona (Smith, Nobel, Heinecke, et al., 1997), Maryland (Koretz et al., 1996b), Michigan (Urduan & Paris, 1994) and Texas (Haney, 2000), as well

as a National Science Foundation (NSF) study of the influence of testing on teaching math and science in grades 4-12 (Madaus, West, Harmon, Lomax, & Viator, 1992) and a study of the effects of standardized testing in the Irish education system (Kellaghan, Madaus, & Airasian, 1980). Former and current classroom teachers were involved in two field test administrations; their comments contributed to the refinement of the final survey items.

Sampling

Of particular interest was how teachers' attitudes differed depending on the nature of the consequences or stakes attached to their state test results. Because each state is charged with its own educational policy development and implementation, state testing programs tend to vary, not only in terms of content and format, but also in terms of the consequences attached to the test results. The first criterion involved in selecting teachers to participate in the study was based on the consequences of the state test. Each state was categorized according to the nature of the stakes attached to their test results.³

The state classification process produced two general categories of stakes: (a) consequences for districts, schools and/or teachers and (b) consequences for students. Within these two categories, the severity of the stakes attached to the test results was classified as *high*, *moderate*, or *low* for both the district, school and/or teacher level and student level of accountability. For districts, schools and/or teachers *high stakes* refers to state-regulated or legislated sanctions of significant consequence such as accreditation, financial rewards, or placing a school in receivership (Heubert & Hauser, 1999). The low-stakes category included states with testing programs that did not have any known consequences attached to test scores. If the stakes attached to the state test for districts, schools and/or teachers did not meet the criteria of either the high- or low-stakes definitions, states were placed in the moderate category. The moderate-stakes category included states that publicly disseminated test results (e.g., reported test results in local newspapers) (Shore, Pedulla, & Clarke, 2001).

The categorization process of student level consequences attached to test results was based on

a similar framework. High stakes for students referred to state-regulated or legislated sanctions that included the use of test scores to make decisions about grade promotion and/or high school graduation (Huebert & Hauser, 1999). The low-stakes classification was applied to states in which no observable consequences resulted for students based on state test performance. The moderate-stakes classification served a default function and was used to categorize states where the consequences for students did not meet the criteria for either the high- or low-stakes definitions.⁴ The classification of states was based on information found in state legislation, direct contact with state departments of education, their personnel, and web sites at the time the survey was administered (January 2001). For the purposes of this article, teachers' responses from two of the five⁵ categories of state testing programs are compared. Responses from teachers in states that have high stakes for districts, schools, and/or teachers *and* high stakes for students (High/High) are compared with those from teachers in states that have moderate stakes for districts, schools and/or teachers *and* low stakes for students (Moderate/Low). The states classified in each of the two categories are as follows:

High/High stakes: Alabama, California, Delaware, Florida, Georgia, Indiana, Louisiana, Maryland, Massachusetts, Mississippi, Nevada, New Jersey, New Mexico, New York, North Carolina, South Carolina, Tennessee, Texas, Virginia

Moderate/Low stakes: Hawaii, Maine, Montana, Nebraska, New Hampshire, North Dakota, South Dakota, Utah, Wyoming

In order to avoid unnecessarily complex language, the high/high group will be referred to as teachers from high-stakes states and the moderate/low group as teachers from low-stakes states. Comparisons among these two stakes-level categories capture the broad range of teachers' opinions on issues related to their state testing program, and most clearly highlight differences in teachers' views. Table 1⁶ presents the profile of teachers from high and low-stakes states who participated in the study.⁷ This table illustrates that teachers in high-stakes states are slightly more diverse in terms of their race/ethnicity than teachers in low-stakes states, but otherwise these are relatively similar

groups. Table 2 presents a summary of the National Board teacher survey results reported on in this article.

Survey Results

Impact on classroom instruction and assessment

The curriculum standards or frameworks established by states are intended to articulate high expectations for academic achievement and clear outcomes for students. Such curriculum standards have the consequence of establishing homogeneity of course content, thereby focusing classroom instruction and providing teachers with a clear purpose (Goertz, 2000). Regardless of stakes levels, the majority of teachers were positive about their state's content standards or frameworks. Fifty-eight percent of all responding teachers reported that their state-mandated test is based on a curriculum that all teachers should follow. Similarly, more than half of all teachers (55%) reported that if they teach to the state standards or frameworks, students will do well on the state test.

Results according to stakes levels indicate that state tests have a differential impact on what content gets emphasized and how students are assessed. Forty-three percent of teachers in high-stakes states, compared to only 17% of teachers in low-stakes

states,⁸ indicated that the time they spent on instruction in tested areas had increased a great deal. In order to spend more time on tested curriculum, some teachers were placing less emphasis on non-tested content. One-fourth of teachers from high-stakes states reported that instructional time dedicated to nontested areas had decreased a great deal, compared to only 9% of teachers in low-stakes states. In general, teachers in high-stakes states reported significant decreases in time spent on instruction in the fine arts, industrial/vocational education, field trips, class trips, enrichment assemblies, and class enrichment activities. Teachers in low-stakes states did not report decreases in these areas. Perhaps most disconcerting was the substantial proportion of teachers in both types of testing programs (76% of high-stakes teachers and 63% of low-stakes teachers) who reported that their state testing program has lead them to teach in ways that contradict their own notions of sound educational practice. These results suggest that regardless of the rewards and/or sanctions associated with test results, the implementation of state testing programs has changed teaching in ways that many teachers feel negatively impacts the quality of instruction students receive. At the very least, teachers are uncomfortable with the changes they

Table 1
Profile of Responding Teachers from High- and Low-Stakes States

Respondent Characteristics		Percent in high stakes	Percent in low stakes
Gender	Female	83	79
	Male	17	21
Age	Over 40	67	69
Race/Ethnicity	African American	9	1
	Asian/Pacific Islander	1	5
	Hispanic	6	1
	White	84	90
Years of teaching experience	1-12	39	38
	13-20	23	22
	Over 20	38	41
School type	Elementary	59	60
	Middle	20	16
	High	21	24

Source: Pedulla et al. (2003)

feel they need to make to their instruction to conform to the demands of the state testing program.

Not only do teachers in high-stakes states report that they are spending more time on tested content, but state tests, especially those with high-stakes attached, are also influencing the frequency and manner in which teachers assess their students. The results suggest that teachers are constructing their own classroom assessments to mirror the format and types of questions on the state test. For example, 51% of teachers in high-stakes states, as

compared to 29% of teachers in low-stakes states, reported their classroom tests were in the same format as the state test. In addition, almost twice as many teachers in high-stakes states reported using classroom tests comprised of multiple-choice questions on a weekly basis (31% vs. 17%). These results are consistent with the findings of previous research in this area (Corbett & Wilson, 1991; Herman & Golan, n.d.; Madaus, 1988; Mehrens, 1998; McMillan, Myran, & Workman, 1999; Stecher et al., 1998) and add to the growing body of evidence

Table 2
Percent Agreement by Stakes Level

NBETPP Survey Results ¹	High Stakes	Low Stakes
Impact on Classroom Instruction and Assessment		
Greatly increased time spent on instruction in tested areas	43	17
Greatly decreased time spent on instruction in nontested areas	25	9
Created classroom tests in the same format as the state-mandated test	51	29
Used multiple-choice classroom tests on a weekly basis	31	17
Reported the state testing programs has lead teachers to teach in ways that contradict notions of good educational practice	76	63
Pressure to Raise Scores and Prepare Students for the State Test		
Strongly agreed they felt pressure from their district superintendent to raise scores on the state test	57	41
Strongly agreed they felt pressure from their building principal to raise scores on the state test	41	17
Strongly agreed that there is so much pressure for high scores on the state-mandated test that they have little time to teach anything not on the test	41	18
Spent more than 30 hours per year preparing students specifically for the state test	44	10
Prepare students for the state-mandated test throughout the year	70	43
Used test preparation materials developed commercially or by the state	63	19
Used released items from the state test to prepare students	44	19
Taught test-taking skills to prepare students	85	67
Impact on Teacher and Student Motivation and Morale		
Reported teachers at their school want to transfer out of tested grades	38	18
Strongly believed students are extremely anxious about taking the state test	35	20
Perceived students are under intense pressure to perform well on the state test	80	49
State test has caused students in their district to drop out of high school	28	9
Views on Accountability		
Inappropriate to use test results to award school accreditation	66	77
Inappropriate to use test results to evaluate teachers/administrators	82	90
Inappropriate to use test results to award teachers/administrators financial bonuses	87	96
Appropriate to use test results to award high school diplomas	57	37
Inappropriate to use test results to promote or retain students in grade	59	76

Source: Pedulla et al. (2003)

1. Differences between high and low stakes percents for each item are statistically significant (alpha = .001).

that high-stakes tests can result in a narrowing of the curriculum by encouraging teachers to focus instruction on tested content and de-emphasize non-tested subject areas, while also encouraging them to develop classroom assessments that mirror the format of the state test.

Pressure to raise test scores and prepare students for the state test

Survey results suggest that pressure is brought to bear on teachers, particularly those in high-stakes states, to raise test scores. In comparison to teachers in low-stakes testing programs, a greater proportion of teachers in high-stakes environments reported feeling pressure from district superintendents, principals, and, to a lesser extent, parents to improve student performance on the state test. Even though teachers in both high- and low-stakes states indicated feeling more pressure from their district superintendent than their building principal to improve student performance on the state test, the pressure was most acute for teachers in high-stakes testing programs. Seventeen percent of teachers in states with low stakes for students strongly agreed that they felt pressure from their building principal to raise test scores. In contrast, more than twice that percent of teachers from high-stakes states (41%) reported feeling such pressure. In addition, 41% of teachers in states with high-stakes testing programs strongly agreed that there was so much pressure for high scores on the state-mandated test that teachers had little time to teach anything not on the test. By comparison, 18% of teachers in low-stakes states felt this same level of pressure to teach to the test.

The pressure to raise scores and improve student performance requires teachers to devote substantial amounts of instructional time to test preparation. Teachers in high-stakes states reported spending more class time preparing students for the state test than did their counterparts in low-stakes states. Specifically, four times as many teachers (44%) in high-stakes states reported spending more than 30 class hours per year preparing students for the state test (e.g., teaching test-taking skills) (10% of teachers in low-stakes states reported the same). In addition, 70% of teachers in high-stakes states, compared to 43% of those in

low-stakes states, indicated that they were preparing students for their state test throughout the school year, rather than just during the weeks prior to the test administration. Teachers in both types of testing programs employed a variety of strategies to prepare students. A substantial proportion of teachers in both high-stakes (85%) and low-stakes (67%) testing programs reported teaching test-taking skills to prepare students for the state test. However, far greater percentages of teachers from high-stakes states (63% vs. 19% in low-stakes states) used specific test preparation materials that had been developed commercially or by the state. This finding may be a function of the greater availability of supplementary materials in high-stakes states. Similarly, 44% of teachers in high-stakes states indicated using released items from the state test in their instruction and preparation for the test in comparison to only 19% of teachers from low-stakes states. Teachers in high-stakes testing environments felt significantly greater pressure to improve student test performance and employed more teaching behaviors geared specifically toward the state test.

Central to the current state accountability models is the need for steady increases in test scores as indicators of improved student achievement and, in turn, school effectiveness. The survey results show when compared to teachers in low-stakes states, teachers in states with high-stakes tests spent more time preparing students and were more likely to engage in practices that can corrupt the capacity of the state test to serve as an accurate measure of achievement. Using released items, commercially developed preparation material, and teaching test-taking skills can benefit students by familiarizing them with the item format, thus reducing test-related anxiety and stress. However, these preparation practices can also negatively affect the accuracy of the state test results by potentially raising scores without increasing the skill or knowledge level of students (Haladyna et al., 1991; Koretz et al., 1991; Linn, 2000; Madaus, 1988). Consistent with the research literature, these findings suggest that highly consequential tests encourage teachers to employ test preparation strategies that may result in improved test scores on the state test but may not represent an actual improvement in achievement.

Impact on teacher and student motivation and morale

Teachers reacted to the increased pressures created by high-stakes testing by teaching test-taking skills, modeling classroom assessments after the state test, and emphasizing content that is tested. These survey results suggest that teachers who reported feelings of pressure from either their district superintendent or building principal were also likely to work in schools with lower teacher morale. Almost half (45%) of all responding teachers indicated that teacher morale was low in their school. Comparing those results by the stakes attached to the state test results, 38% of teachers in high-stakes testing programs compared to 18% of teachers in low-stakes testing programs, reported that teachers in their school wanted to transfer out of the grades in which the state-mandated test is administered.

Not only can these highly pressured school environments have a negative impact on teachers, but they can also affect students negatively. Students can experience stress, anxiety, loss of self-efficacy, decreased motivation, and frustration resulting from pressures associated with high-stakes testing. Over one third (35%) of teachers from high-stakes states and 20% of teachers from low-stakes states strongly agreed that students were extremely anxious about taking the state test. However, far greater percentages of teachers from high-stakes states (80% compared to 49% of teachers in low-stakes states) perceived students to be under intense pressure to perform well. Almost one third (28%) of teachers from high-stakes testing programs reported that their state test had caused students in their district to drop out of high school, while only 9% of teachers in low-stakes states reported that their state test was having this impact on high school students. These findings add to the growing body of evidence suggesting high-stakes testing may negatively impact teacher and student morale and motivation; ultimately contributing to increased departures from the teaching profession and/or increased high school dropout rates (Haney, 2000; Reardon, 1996; Smith, 1991).

Views on Accountability

Teachers in both high- and low-stakes states rejected the notion that test scores should be used

to hold schools and teachers accountable, but responded more favorably when asked about student accountability. For example, 66% of teachers from high-stakes states and 77% of teachers from low-stakes states felt awarding school accreditation based on test results was inappropriate. Similarly, 82% of teachers from high-stakes states and 90% of teachers from low-stakes states felt it was inappropriate to evaluate teachers/administrators on the basis of student test results. Teachers in both types of testing programs overwhelmingly opposed using test results to award teachers/administrators financial bonuses. Eighty-seven percent of teachers in high-stakes states, compared to 96% of teachers in low-stakes states, held this opinion. In comparison to school accountability and especially teacher accountability, a greater percentage of teachers from both high- and low-stakes states supported using test results to hold students accountable. However, teachers in high-stakes states held a more favorable view toward test-based student accountability than their counterparts in low-stakes testing environments. For example, 57% of teachers in high-stakes states compared to 37% of teachers in low-stakes states indicated that using test scores to determine whether students should graduate from high school was appropriate. Teachers in both types of testing programs held less favorable views toward using test results to determine grade promotion. Fifty-nine percent of teachers in high-stakes states, compared to 76% of teachers in low-stakes states, reported it was inappropriate to use test results to promote or retain students in grade.

These results regarding teachers' views on student accountability seem contradictory in light of the reported negative impacts of high-stakes testing on classroom practices. Seven out of 10 teachers in the high-stakes states reported that the state testing program had lead them to teach in ways that violate standards of good educational practice. In addition, when the stakes are high, the survey results suggest that a substantial proportion of teachers will align their curriculum and assessments to mirror the state test and devote a sizeable amount of time toward preparing students specifically for the state test. Yet 57% of teachers from high-stakes states indicated that it was appropriate to use test results to award high school diplomas. These seemingly

contradictory perceptions toward the educational impact and accountability function of high-stakes testing present a complex paradox that is difficult to explain from these data.

Conclusion

This article provided an overview of teachers' perceptions of the impacts of state-mandated testing programs on teaching and learning. The National Board survey results build on the previous research in this area by providing a national picture of how teachers working in various types of testing programs and under different forms of accountability models perceive the impacts of their state test. The results point to serious concerns about the perceived effect of high-stakes tests on the quality of education and on teachers and students.

The results suggest that the state test, rather than the content standards, is the more powerful influence on teaching practices. While teachers reported generally positive views towards their states curricular standards, particularly troubling was the substantial majority of teachers in both high- and low-stakes states that reported the state test has led them to teach in ways that contradict their own notions of sound educational practice. In addition, teachers in high-stakes settings were far more likely to report having greatly increased the instructional time devoted to tested content at the expense of nontested content and enrichment activities than were teachers from low-stakes environments.

Also a source of concern is the substantial allocation of instructional time for specific test preparation. Teachers from high-stakes states reported spending far more time than did their counterparts in low-stakes states preparing students for the state test, teaching test-taking skills, and using test preparation materials and released items from the state. These types of test preparation activities may call into question the validity of state test scores, which were originally designed to provide an objective, accurate measure of achievement, thus rendering any decision based on test scores (e.g., award school accreditation or high school diplomas) questionable.

Not only are teachers' views regarding the state test's negative impact on the quality of education and the emphasis on specific test preparation

disconcerting, the perceived human impact of the state test is also worrisome. The results suggest that, especially in high-stakes states, both students and teachers experienced test-related pressure. Eight in 10 teachers in high-stakes states reported that students were under intense pressure to perform well on the state test. While pressure on teachers may materialize by placing greater emphasis on test preparation, it may also have significant professional costs. For example, almost twice as many teachers in high- versus low-stakes states indicated teachers at their school wanted to transfer out of grades in which the state test was administered. Similarly, teachers in high-stakes states were far more likely to report the state test has caused students in their district to drop out of high school. These results suggest there is a potential for a substantial human cost resulting from highly consequential testing programs, of which the effects on future opportunities, particularly for students, are profound.

While teachers are involved in the conversations that lead to education policy in many states, these results suggest it is imperative to expand their role. Teachers' views regarding the impact of state testing programs suggest that in high-stakes states especially, the intended policies are not realized at the classroom level. In addition, the National Board results, coupled with previous research, indicate that highly consequential testing policies can contribute to low morale, increased frustration, diminished student learning experiences, and restricted curricular options. It is becoming increasingly clear that the anticipated goals of state testing policies are at odds with the realities of their implementation and can lead to unintended negative impacts. These negative impacts are further exacerbated by high-stakes uses of test results. Consequently, it is essential that policy makers reconsider the highly consequential nature of state testing programs and refocus education policies to place greater emphasis on supporting and improving teaching and learning, rather than relying on a system of rewards and sanctions to spur change in classrooms.

Notes

1. This work was supported by the Atlantic Philanthropies Foundation. The findings and conclusions in this article are those of the authors and do not

necessarily reflect the views of the Atlantic Philanthropies Foundation.

2. Located in the Lynch School of Education at Boston College, the National Board on Educational Testing and Public Policy is an independent organization that both monitors testing in the United States and seeks to provide information about the use and impact of educational testing.
3. The stratified random sample was drawn by Market Data Retrieval in December 2000.
4. In an effort to achieve a nationally representative sample of teachers, the teacher selection process involved four criteria. After identifying the stakes level classification for each state, we further organized the sample by school type (elementary, middle, or high school) and by the subject taught for high school teachers. High school teachers who taught courses in English, math, science, social studies, and special education were randomly selected to participate in the study. Lastly, the sample was further organized by geographic location to ensure that teachers from urban and nonurban areas were proportionally represented.
5. The state classification process produced a nine-cell testing program matrix. However, based on the categorization, one cell remained empty and three cells contained only one state (Iowa, Oregon, and Idaho). Because it was cost-prohibitive to sample these cells at the same rate as the other five, these state testing program cells were excluded from the study. Teachers from states with high/high, high/moderate, high/low, moderate/high, and moderate/low stakes for districts, schools, teachers, and students were selected to participate in the survey.
6. Percentages may not total 100 due to rounding.
7. A total of 12,000 teachers were mailed surveys in January 2001. A prenotification letter and three follow-up mailings were employed; this resulted in 4,195 returned useable surveys and a response rate of 35%. Surveys were received from teachers in every state sampled. The demographics show that the overwhelming majority of the teachers were late middle-aged females with considerable teaching experience. Sixty-seven percent of the responding teachers were over 40 years old and 40% had over 20 years of teaching experience. Consistent with the national teaching force, 58% of those responding were elementary school teachers, while 20% taught in middle schools, and the remaining 22% were high school practitioners (U.S. Department of Education, 2002). The teachers who completed the National Board survey were comparable to the national teaching force in terms of their age, race/ethnicity, the type of school in which they taught (elementary, middle, or high school), and teaching experience.
8. All differences between high- and low-stakes percents are statistically significant ($\alpha = .001$).

References

- Corbett, H.D., & Wilson, B.L. (1991). *Testing, reform, and rebellion*. Norwood, NJ: Ablex.
- Firestone, W.A., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis*, 20(2), 95-117.
- Goertz, M.E. (2000, April). *Local accountability: The role of the district and school in monitoring policy, practice, and achievement*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.
- Haladyna, T.M., Nolen, S.B., & Hass, N.S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20(5), 2-7.
- Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives*, 8(41). Retrieved September 1, 2000, from <http://epaa.asu.edu/epaa/v8n41>
- Herman, J.L., & Golan, S. (n.d.). *Effects of standardized testing on teachers and learning: Another look*. (CSE Technical Report 334). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Heubert, J.P., & Hauser, R.M. (Eds.). (1999). *High-stakes: Testing for tracking, promotion and graduation*. Washington, DC: National Academy Press.
- Hoffman, J.V., Assaf, L.C., & Paris, S.G. (2001). High-stakes testing in reading: Today in Texas, tomorrow? *The Reading Teacher*, 54(5), 482-494.
- Jones, G., Jones, B., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impacts of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan*, 81(3), 199-203.
- Kellaghan, T., Madaus, G.F., & Airasian, P.W. (1980). *The effects of standardized testing*. Educational Research Centre: St. Patrick's College, Dublin, Ireland and Boston College: Chestnut Hill, MA.
- Kellaghan, T., Madaus, G.F., & Raczek, A. (1996). *The use of external examinations to improve student motivation*. Washington, DC: American Educational Research Association.
- Klein, S.P., Hamilton, L.S., McCaffrey, D.F., & Stecher, B.M. (2000). *What do test scores in Texas tell us?* (IP-202). Santa Monica, CA: RAND.
- Koretz, D., & Barron, S. (1998). *The validity gains on the Kentucky Instructional Results Information System (KIRIS)* (MR-1014-EDU). Santa Monica, CA: RAND.
- Koretz, D., Barron, S., Mitchell, K., & Stecher, B. (1996a). *The perceived effects of the Kentucky Instructional Results Information System (KIRIS)* (MR-792-PCT/FF). Santa Monica, CA: RAND.
- Koretz, D.M., Linn, R.L., Dunbar, S.B., & Shepard, L.A. (1991, April). *Effects of high-stakes testing on achievement: Preliminary findings about generalization*

- across tests. Paper presented at the annual meeting of the American Education Research Association and the National Council of Measurement in Education, Chicago.
- Koretz, D., Mitchell, K., Barron, S., & Keith, S. (1996b). *The perceived effects of the Maryland school performance assessment program* (CSE Technical Report No. 409). Los Angeles: Center for the Study of Evaluation, University of California.
- Linn, R.L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4-16.
- Madaus, G.F. (1988). The influence of testing on the curriculum. In L. Tanner (Ed.), *Critical issues in curriculum* (pp. 83-121). Chicago: University of Chicago Press.
- Madaus, G., West, M., Harmon, M., Lomax, R., & Viator, K. (1992). *The influence of testing on teaching math and science in grades 4-12*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
- McMillan, J.H., Myran, S., & Workman, D. (1999, April). *The impact of mandated statewide testing on teachers' classroom assessment and instructional practices*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.
- McNeil, L.M. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.
- Mehrens, W.A. (1998). Consequences of assessment: What is the evidence? *Education Policy Analysis Archives*, 6(13) (add page numbers). Retrieved August 14, 2000, from <http://epaa.asu.edu/epaa/v6n13.html>
- Pedulla, J., Abrams, L., Madaus, G., Russell, M., Ramos, M., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
- Quality counts 2002: Building blocks for success [Special Report]. (2002, January 17). Bethesda, MD: *Education Week*.
- Reardon, S.F. (1996, April). *Eighth grade minimum competency testing and early high school drop-out patterns*. Paper presented at the annual meeting of the American Educational Research Association, New York.
- Shore, A., Pedulla, J., & Clarke, M. (2001). The building blocks of state testing programs. *National Board on Educational Testing and Public Policy Statements*, 2(4). Boston College, MA.
- Smith, M.L., Nobel, A.J., Heinecke, W., Seck, M., Parish, C., Cabay, M., et al. (1997). *Reforming schools by reforming assessment: Consequences of the Arizona student assessment program (ASAP): Equity and teacher capacity building* (CSE Technical Report 425). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Smith, M.L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20(5), 8-11.
- Smith, M.L., Edelsky, C., Draper, K., Rottenberg, C., & Cherland, M. (1991). *The role of testing in elementary schools* (CSE Technical Report 321). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Stecher, B., & Barron, S. (1999). *Quadrennial mile-post accountability testing in Kentucky* (CSE Technical Report 505). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Stecher, B., Barron, S., Chun, T., & Ross, K. (2000). *The effects of the Washington state education reform on schools and classrooms* (CSE Technical Report 525). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Stecher, B., Barron, S., Kaganoff, T., & Goodwin, J. (1998). *The effect of standards-based assessment on classroom practices: Results of the 1996-97 RAND survey of Kentucky teachers of mathematics and writing* (CSE Technical Report 482). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Urdan, T.C., & Paris, S.G. (1994). Teachers' perceptions of standardized achievement tests. *Educational Policy*, 8(2), 137-156.
- U.S. Department of Education. (2002). *Digest of education statistics 2001*. Washington, DC: National Center for Education Statistics.