

Perceived Effects of State-Mandated Testing Programs on Teaching and Learning:

Findings from Interviews with Educators in Low-, Medium-, and High-Stakes States

*National Board on Educational
Testing and Public Policy*

Marguerite Clarke, Arnold Shore, Kathleen Rhoades, Lisa
Abrams, Jing Miao, and Jie Li

Lynch School of Education

Boston College

January 2003



NBETPP

ACKNOWLEDGMENTS

We are grateful to The Atlantic Philanthropies Foundation for providing the funding that made this research possible. We also would like to thank our advisory board — Albert Beaton, Robert Hauser, Henry Levin, Audrey Qualls, and Daniel Stufflebeam — for their suggestions and feedback at various stages of the project. Others who were instrumental in the completion of this report include Amie Goldberg, Catherine Horn, John Madura, Miguel Ramos, and Stuart Yeh, who assisted with the interviewing and data analysis process; and Irwin Blumer, Kelvin Gregory, George Madaus, Joseph Pedulla, and Michael Russell, who provided ideas and advice. In order to preserve the confidentiality of our interviewees, we offer a general, but heartfelt, thanks to those who helped us gain access to districts and schools in Kansas, Michigan, and Massachusetts. Most of all, we are indebted to the hundreds of educators who generously contributed their time to answering our questions. While we acknowledge the help of many, responsibility for the content and conclusions of this report rests entirely with the authors.

Marguerite Clarke, Arnold Shore, Kathleen Rhoades, Lisa Abrams, Jing Miao, and Jie Li

November 2002

EXECUTIVE SUMMARY

Standards, tests, and accountability are the key policy components of standards-based reform in public education. The standards outline the expectations held for all students, the test provides a way to judge student performance against these standards, and the accountability component provides an incentive – in the form of stakes attached to the test results – for those involved to make the necessary changes in order to meet performance expectations.

The goal of this National Board study was to identify the effects of state-level standards-based reform on teaching and learning, paying particular attention to the state test and associated stakes. On-site interviews were conducted with 360 educators in three states (120 in each state) attaching different stakes to the test results. In Kansas, state test results were one of several pieces of information used to determine school accreditation, but had no official stakes for students. In Michigan, school accreditation was determined by student participation in, and performance on, the state test, and students received an endorsed diploma and were eligible for college tuition credit if they scored above a certain level on the eleventh-grade tests. In Massachusetts, school ratings (and potential takeover) were based on the percentage of students in different performance categories on the state test, and students – starting with the class of 2003 – had to pass the tenth-grade test in order to graduate from high school. Thus, as one moves from Kansas to Michigan to Massachusetts, the stakes for educators remain fairly constant (from moderate/high in Kansas to high in Michigan and Massachusetts), but the stakes for students increase dramatically (from low in Kansas to moderate in Michigan to high in Massachusetts).

Interviewees included elementary, middle, and high school teachers as well as school- and district-level administrators in the three states. Interviews were conducted between winter 2000 and fall 2001 and included the following broad topic areas:

- (1) The effects of the state standards on classroom practice
- (2) The effects of the state test on classroom practice
- (3) The effects of the state test on students

The main study findings are presented below, followed by policy recommendations (see Box 1 for a summary of recommendations). *Taken together, these findings suggest that stakes are a powerful lever for effecting change, but one whose effects are uncertain; and that a one-size-fits-all model of standards, tests, and accountability is unlikely to bring about the greatest motivation and learning for all students.*

Perceived Effects of State-Mandated Testing Programs on Teaching and Learning

Report Recommendations

Recommendation 1: States should invest in high-quality professional development for educators that is ongoing, related to the state standards, and tailored to their particular needs and contexts.

Recommendation 2: Educators should be supplied with high-quality classroom materials and other resources that are aligned with the state standards and support their integration into classroom instruction.

Recommendation 3: States need to work with schools and districts to ensure that local and state standards and tests are appropriately aligned.

Recommendation 4: States need to make sure that their standards and tests are aligned not only in terms of content, but also in terms of the cognitive skills required.

Recommendation 5: States should put in place ongoing monitoring and evaluation of their testing and accountability systems so that unintended negative effects can be identified, and resources and support appropriately targeted.

Recommendation 6: States should be flexible in the options available to students for demonstrating achievement so that all have a chance to be successful.

Recommendation 7: Test results should not be used to compare teachers and schools unless student demographics and school resources are equated and the latter are adequate to produce high student performance.

Recommendation 8: There is a need to make the teaching and learning process an integral part of standards-based reform and to recognize that testing should be in the service, rather than in control, of this process.

Perceived Effects of the State Standards on Classroom Practice

We found no clear overall relationship between the level of the stakes attached to the state test and the influence of the state standards on classroom practice. Instead, our findings suggest that other factors are at least as important, if not more so, in terms of encouraging educators to align classroom curricula with these standards. At the same time, as the stakes attached to the test results increased, the test seemed to become the medium through which the standards were interpreted. Massachusetts educators most often mentioned using the state test as the target for their teaching efforts (over two-thirds of these interviewees) while those in Kansas were least likely to mention this (one-fifth of these interviewees). Other findings in this area are outlined below.

⚙ Overall Impact on Classroom Practice

Between half- and three-quarters of the educators in each state expressed neutral to positive opinions about their state standards, mentioning that they encouraged greater curricular consistency across schools and increased the emphasis on problem solving and writing. Kansas and Massachusetts interviewees were the most positive in this regard. At the same time, a sizeable minority (between one-fifth and one-third) in each state expressed concerns about the negative effects of the standards on classroom practice, among them that they could lead to developmentally inappropriate material and pace, curriculum narrowing, and decreased flexibility. Massachusetts interviewees were the most likely to mention these concerns.

⚙ Factors Related to this Impact

In all three states, the extent to which the state standards affected classroom practice seemed to depend on a number of factors. These included (i) the perceived rigor, developmental appropriateness, and specificity of the standards; (ii) the degree of alignment with local standards and tests; (iii) the degree of alignment with the state test; (iv) the stakes attached to the state test; and (v) appropriate professional development opportunities and other resources (e.g., textbooks aligned with the standards). Depending on the interviewee, the relative importance of these factors varied. However, the rigor, developmental appropriateness, and specificity of the standards; their alignment with the state test; and the availability of professional development opportunities and other resources were regarded as important by most interviewees.

❁ School Type Differences

In all three states, elementary educators reported the greatest impact of the state standards on classroom practice. For example, elementary teachers were almost twice as likely as their high school counterparts to mention that the state standards had changed their classroom curriculum in positive ways. This pattern was similar in Kansas (two-thirds of elementary teachers versus one-third of high school teachers), Michigan (one-third versus one-fifth), and Massachusetts (half versus one-quarter). Middle school teachers fell somewhere in between, with two-fifths in Kansas, one-quarter in Michigan, and one-third in Massachusetts reporting a positive impact on their curriculum. At the same time, elementary teachers were the most likely to note that the standards were not developmentally appropriate for their students. The proportion of elementary teachers voicing this concern was similar in Kansas and Michigan (about one-fifth in each) and slightly higher in Massachusetts (one-quarter).

❁ District Type Differences

Educators in the rural districts appeared to find it hardest to align their local curriculum with the state standards. The most frequently mentioned concerns included a lack of curriculum materials, few professional development opportunities, and the potential loss of local identity as a result of aligning with the more context-free state standards. In addition, almost two-fifths of the rural educators in Kansas and almost half of those in Massachusetts felt that their state standards were not developmentally appropriate (this was less frequently mentioned in Michigan). Educators in other districts in Kansas and Massachusetts were about half as likely to mention this concern. Educators in the suburban districts, although still a minority, were the most likely to report that aligning with the state standards impoverished their curriculum. On the other hand, educators in the urban districts were the most likely to view the state standards as a chance to equalize curriculum quality with other districts, although attempts to align were impeded by local standards and testing requirements in Kansas and a lack of capacity in Michigan.

❁ Subject Area Differences

In all three states, educators had the most concerns about the social studies standards. These concerns included (i) too much content to be covered, (ii) developmental inappropriateness, (iii) an emphasis on facts rather than concepts, and (iv) a lack of alignment with the state test.

Perceived Effects of the State Test on Classroom Practice

Overall, Massachusetts educators reported the most test-related effects – both positive and negative – on curriculum and instruction. Michigan educators reported fewer effects and Kansas educators slightly fewer again. Since this is a qualitative study, we cannot test the significance of these differences in terms of their relationship to the stakes attached to the test results. However, we can infer that as the stakes increase, so too do the consequences for classroom practice, making it imperative that the test is aligned with the standards and is a valid and reliable measure of student learning. Other findings in this area include the following.

❁ Impact on the Curriculum

In all three states, educators reported that preparing for the state test involved varying degrees of removing, emphasizing, and adding curriculum content, with the removal of content being the most frequently reported activity. Compared with their peers in Kansas and Michigan, Massachusetts educators reported about twice the amount of activity in these areas. Perceived positive effects of these changes included the removal of unneeded content, a renewed emphasis on important content, and the addition of important topics previously not taught. Perceived negative effects included a narrowing of the curriculum, an overemphasis on certain topics at the expense of others, and an overcrowded curriculum. In all three states, about one in ten interviewees felt that the state test had no impact on what was taught.

❁ Impact on Instruction and Assessment

Interviewees in all three states reported that preparing for the state test had changed teachers' instructional and assessment strategies. Massachusetts educators reported about twice the number of changes as their peers in Kansas and Michigan. Perceived positive effects of these changes included a renewed emphasis on writing, critical thinking skills, discussion, and explanation. Perceived negative effects included reduced instructional creativity, increased preparation for tests, a focus on breadth rather than depth of content coverage, and a curricular sequence and pace that were inappropriate for some students. In all three states, only a minority of interviewees (one in seven in Kansas, one in five in Michigan, and one ten in Massachusetts) felt that the state test did not affect instructional or assessment strategies.

❁ School Type Differences

In all three states, elementary teachers reported the most test-related changes in what and how they taught, and were about half as likely as middle or high school teachers to say that the state test did not affect their classroom practice. In particular, they were the most likely to report removing topics from the curriculum to prepare for the test (something that many of them viewed negatively) and emphasizing topics that would be tested. The removal of topics from the curriculum tended to decrease from the elementary level (three-quarters of Kansas, one-third of Michigan, and four-fifths of Massachusetts elementary teachers) to the middle school (one-third, one-quarter, half), and high school (one-fifth, one-third, half) levels.

❁ District Type Differences

Educators in rural and large urban districts were the most likely to note that significant amounts of classroom time were spent preparing for the state test. In addition, rural educators reported more test-related changes in what was taught than did those in the other districts. Overall, suburban educators reported the fewest changes in response to the test. However, there was an indication that targeted kinds of test preparation occurred in the suburban districts.

❁ Subject Area Differences

Reported effects were different for tested versus non-tested grades and subject areas, with teachers in the former more likely to mention negative effects such as an overcrowded curriculum, rushed pace, and developmentally inappropriate practices. At the same time, teachers in non-tested grades reported adjusting their curriculum to make sure that students were exposed to content or skills that would be tested, either in another subject area or at a later grade level.

Perceived Effects of the State Test on Students

As the stakes for students increased, interviewees reported a more negative impact on students. Specifically, Massachusetts educators were three times as likely as those in Kansas to note that the state tests negatively affected students' perception of education, created stress for students, and were unfair to special populations. *At the same time, if the test results had no consequences for students, this was seen as problematic since, along with overtesting, it could reduce students' motivation.* Interviewees' suggestions in this area included reducing the number of tests students had to take and making the state test more meaningful in students' lives. The latter did not necessarily mean attaching high stakes to the results, but rather giving students feedback on how they performed and showing them how their performance related to their classroom work. Other findings in this area are discussed below.

⚙ Overall Impact on Students

In all three states, interviewees reported more negative than positive test-related effects on students, such as test-related stress, unfairness to special populations, and too much testing. Massachusetts interviewees were the most likely to note these negative effects, and Kansas interviewees the least likely. For example, while two-thirds of Massachusetts interviewees and two-fifths of Michigan interviewees reported that their students were experiencing test-related stress, only one-fifth of Kansas interviewees did so. Perceived positive effects noted by a minority – one-quarter or less – of the interviewees in all three states included increased student motivation to learn, and improved quality of education. Massachusetts interviewees were the most likely to note these effects.

⚙ Differential Impact on Special Education and Limited English Proficiency Students

While some interviewees felt that the state tests could help special education and Limited English Proficiency (LEP) students get extra help that might not otherwise be available, their impact on these students was seen as more negative than positive. Massachusetts interviewees were three times as likely (two-thirds versus about one-fifth in the other two states) to note the adverse impact of the state test on special education students, particularly in relation to the tenth-grade graduation test. Suggestions for how to reduce the negative effects on special education and LEP populations included the provision of multiple levels or forms of the test, allowing students several opportunities to take the test, improving testing accommodations, and introducing greater flexibility in how students could demonstrate their knowledge and skills.

❁ Validity and Utility of Test Scores

Interviewees had two main concerns about the validity of the test results. The first was that overtesting reduced students' motivation to exert effort on the state tests, thereby compromising the test's ability to measure what they had learned. Roughly one-third of Massachusetts educators and one-fifth of Kansas and Michigan educators identified this as a problem in the interpretation of test results. The second concern was that the test results were not a valid measure for comparing schools and districts since they were affected by out-of-school factors. Over half of the Massachusetts interviewees and one-third of the Kansas and Michigan interviewees mentioned this. As for utility, about one-fifth of the interviewees in each state noted that the results came back too late to be useful, while others said that they never received test results but would like to. Those who did receive results were divided as to their usefulness for enhancing instruction.

❁ School Type Differences

Across the three states, elementary educators were the most likely to note that the tests created stress for students, with roughly two-thirds of Massachusetts, three-quarters of Michigan, and one-third of Kansas elementary educators mentioning this. Elementary educators were particularly concerned by the developmental inappropriateness of what students at this level were being required to do.

❁ District Type Differences

In all three states, large urban districts were where a host of issues converged. For example, interviewees in these districts had to grapple with the problems of little parental involvement, overtesting, and the challenges facing the large proportion of at-risk students. State-specific findings emerged in Michigan and Massachusetts. In Michigan, educators in the large urban district were the least likely to note that the scholarship money attached to the eleventh-grade test provided an incentive for their students. This finding, along with data indicating that white, Asian, and wealthy students are the most likely to get these scholarships, suggests that the state's goal of increasing access to higher education through the program is not being realized. In Massachusetts, urban educators were most concerned about the potentially high failure rates and increased dropouts due to the tenth-grade graduation test. While results for the first cohort of students to face this requirement were not available at the time of these interviews, their subsequent release confirmed some of these fears, with pass rates for the urban districts in this study almost half that of the suburban district.

Policy Recommendations

These findings illustrate the complex linkages among standards, tests, accountability, and classroom practice, especially in the area of unintended negative consequences. *In particular, they show that increasing the stakes attached to the test results does not necessarily bring about improvements in teaching and learning, but can adversely affect the quality of classroom practice and have a negative impact on at-risk student populations.* While further research is needed to determine whether this pattern of findings holds for other states, some general policy implications can be discerned. These focus on five factors – capacity, coherence, consequences, context, and curriculum – that seemed to influence the relationship among standards, tests, accountability, and classroom practice in all three states. Capacity and coherence emerged as important factors in the ability of the state standards to influence classroom practice. Consequences and context emerged as important factors in the impact of the state test and associated accountability uses on teachers and students. Curriculum was an important consideration in both areas. These five factors highlight the need for policymakers to do more than mandate standards and test-based accountability if the intent of standards-based reform – high-quality teaching and high-level learning – is to make it to the classroom.

Capacity

The study findings suggest that one of the biggest obstacles to implementation of the state standards was lack of capacity. This mainly took the form of limited professional development opportunities and inadequate resources, especially in the rural and urban districts and for elementary educators. Since appropriate professional development, high-quality curriculum materials, and support for teachers and administrators are crucial to any effort to improve student outcomes, more attention needs to be devoted to these issues, particularly in low-performing schools. In this regard, *we recommend that states invest in high-quality professional development that is ongoing, related to the state standards, and tailored to educators' particular needs and contexts.* It should include training in classroom assessment techniques so that teachers can monitor and foster student learning throughout the school year and should provide educators with tools for interpreting and using state test results. In addition, *educators should be supplied with high-quality classroom materials and other resources that are aligned with the state standards and that support their integration into classroom instruction.* Resources should include clear descriptions of the standards as well as examples of student work that reaches the desired performance levels.

Coherence

Another obstacle to implementation of the state standards was the lack of alignment between standards and tests. This took two forms: misalignment between local and state standards and tests, and between state standards and state tests. The former was most evident in the urban districts in Kansas. The latter appeared in all three states, particularly in relation to social studies. Misalignment of either sort can lead to a lack of focus in the classroom curriculum, overtesting, and large amounts of time spent preparing for and taking tests at the expense of instruction. In order to avoid these drains on classroom time, and the associated stress on educators and students, two recommendations are offered. First, *states need to work with schools and districts to ensure that local and state standards and tests are appropriately aligned*. Depending on the state and the assessment purpose, this could mean using the same test for state, district, and school requirements or spreading the tests out across subject areas, grade levels, or times of the school year. Second, *states need to make sure that their standards and tests are aligned not only in terms of content, but also in terms of the cognitive skills required*. This is particularly important if stakes are to be attached to the test results, since the test is more likely to become the medium through which the standards are interpreted.

Consequences

The study findings showed a distinction between stakes and consequences. Specifically, while mandated rewards and sanctions may be directed at one level or group in the system, their impact can extend in unexpected and undesirable directions. The most striking example in this study was a consistently greater impact on both students and educators at the elementary level, regardless of the stakes attached to the test results. Some of these effects were positive, but others produced a classroom environment that was test-driven and unresponsive to students' needs. This finding is of particular concern in the current policy climate since the accountability requirements of the 2001 No Child Left Behind Act are placing an even greater testing burden on the early and middle grades. With this in mind, *we recommend regular monitoring and evaluation of state testing and accountability systems so that unintended negative effects can be identified, and resources and support appropriately targeted*. This kind of ongoing monitoring and evaluation can also be used to identify and reinforce unintended positive consequences.

Context

Another study finding was that some of the biggest differences are not between states, but within states. For example, the greater impact on special student populations, the tendency for urban districts to spend more time on test preparation, and the increased burden on the elementary curriculum highlight the complexities involved in implementing a one-size-fits-all reform in different contexts and with different populations. Given these contextual variations, there is a need to recognize the dangers involved in using one test to make highly consequential decisions about students or educators. This is of particular concern in Massachusetts, where the graduation test acts as gatekeeper to students' lives and career opportunities. It is also of concern in the use of test scores to compare and make decisions about schools and districts. Two recommendations emerge from these findings. First, and in line with guidelines provided by several national organizations (e.g., American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), *we recommend that these kinds of consequential decisions not be made on the basis of a single test, but that states should be flexible in the options available to students for demonstrating achievement so that all have a chance to be successful.* One way to do this is to move toward an accountability system that uses multiple measures of teaching and learning, some of which could be locally developed and tied in with local goals. *A second recommendation is that test results not be used to compare teachers and schools unless student demographics and school resources are equated and the latter are adequate to produce high student performance.*

Curriculum

Findings in all three states suggest that when capacity or coherence is lacking, when context and consequences are ignored, and when pressure to do well on the test is overwhelming, the test dictates the curriculum, and students' individual differences and needs are set aside. Since a test is limited in terms of the knowledge and skills that can be measured, safeguards against this eventuality are needed if the broader learning goals of standards-based reform are to be achieved. Thus, *there is a need to make the teaching and learning process an integral part of standards-based reform and to recognize that testing should be in the service, rather than in control, of this process.* This refocusing increases the chances of deep, rather than superficial, changes in student knowledge. It also requires a fundamental change in the nature of state testing programs (see Shepard, 2002), away from an emphasis on accountability and toward one on providing information, guidance, and support for instructional enhancement. The impediment to making these kinds of changes is not a lack of knowledge: we already know a lot about how children learn and how best to assess what they have learnt (e.g., Pellegrino, Chudowsky, & Glaser, 2001). Rather, what is needed is a change in mindset and the willpower to make them happen.