DOCUMENT RESUME

ED 474 867 TM 034 809

AUTHOR	Clarke, Marguerite; Shore, Arnold; Rhoades, Kathleen; Abrams, Lisa; Miao, Jing; Li, Jie
TITLE	Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from Interviews with Educators in Low-, Medium-, and High-Stakes States.
INSTITUTION	National Board on Educational Testing and Public Policy, Chestnut Hill, MA.
PUB DATE	2003-01-00
NOTE	107p.; Additional support from Atlantic Philanthropies.
PUB TYPE	Reports - Research (143)
EDRS PRICE	EDRS Price MF01/PC05 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Educational Change; Educational Practices; Elementary Secondary Education; *High Stakes Tests; Interviews; *State Standards; *Teachers; *Test Results
IDENTIFIERS	Kansas; Massachusetts; Michigan

ABSTRACT

The goal of this 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 (elementary, middle, and high school teachers) in 3 states (120 in each state) attaching different stakes to the test results. In Kansas, state test results were used to determine school accreditation but had no 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 11th grade tests. In Massachusetts, school ratings were based on the percentage of students in different performance categories and students, starting in 2003, had to pass the 10th grade test to graduate. No clear relationship was found between the level of the stakes attached to the state test and the influence of the state standards on classroom practice. 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. Taken together, findings suggest that stakes are a powerful level for effecting change, but one whose effects are uncertain. A one-size-fits-all model of standards, tests, and accountability in not likely to bring about the greatest motivation and learning for all students. Three appendixes contain a grid describing state testing programs, the interview protocol, and the methodology. (Contains 1 figure, 17 endnotes, and 40 references.) (SLD)

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

INTRODUCTORY VIGNETTE

Dan Green* teaches mathematics at a middle school in a small urban district. A veteran educator, he worked as an engineer for several years before becoming a teacher. Dan is in favor of standards-based educational reform and believes that his state's curriculum standards and associated tests foster critical thinking – a vital skill in the "real world." The interviewer asks him to elaborate.

> *I think we need to get away from the drill-and-kill method....Students need to be able to trouble shoot, they need to be able to problem solve in many different ways. What a lot of students have trouble with is the idea of math involving their having to read a lengthy problem, then conceptualize how to attack the problem, then write it up. That has created a lot of problems for them....I know for a fact that some engineers did not move up in companies because they could only do computation. They couldn't think on their feet, they couldn't say what the problem was, and they couldn't write down how to solve it.*

The interviewer then asks Dan about the accountability component of standards-based reform – who is held accountable for results, and how.

> *I like the idea of accountability. Unfortunately, I think a lot of the pressure for accountability has fallen at the feet of educators: superintendents, department heads, principals, teachers in the trenches....I think that a lot of other people have to step up to the plate: the students, the parents, the community....At the same time, the one thing that I really do not buy into is the idea that one test could be the basis for [determining] student graduation. That is very disturbing, that's very upsetting....I think there should be at least a three-tier evaluation process. [The first component] should be their grades, along with teacher evaluations. The second component could be the state test. And I think the third component could be portfolios – show your work. We do that in the engineering field.*

Like several educators we interviewed, Dan likes his state's standards. He agrees that he should be held accountable for helping students to reach them, but is troubled by the extent to which that burden has been placed on educators. At the same time, student accountability for learning – at least in the form of performance on a single test – presents its own problems. In this report, we explore the pros and cons of standards, tests, and accountability in three states, and through doing so, try to understand their impact on students, and on the classroom practices of Dan and other educators.

**Not his real name*

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

CONTENTS

IN MEMORIAM

Audrey Qualls

In early January 2003 we were saddened to learn of the death of our advisory board member, Audrey Qualls. Audrey was Professor of Measurement and Statistics at The University of Iowa. Throughout her career, she made valuable contributions to the fields of testing and measurement. She also made great contributions to this project. Her humor and perceptive remarks as well as her willingness to review and provide feedback on project materials helped guide us through much of our early work. Audrey was a wonderful scholar, intellect, and friend. She will be missed.

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.*

Box 1

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 over-crowded 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.

SECTION ONE

INTRODUCTION

A low-stakes test has no significant, tangible, or direct consequences attached to the results, with information alone assumed to be a sufficient incentive for people to act. The theory behind this policy is that a standardized test can reliably and validly measure student achievement; that politicians, educators, parents, and the public will then act on the information generated by the test; and that actions based on test results will improve educational quality and student achievement. In contrast, high-stakes policies assume that information alone is insufficient to motivate educators to teach well and students to perform to high standards. Hence, it is assumed, the promise of rewards or the threat of sanctions is needed to ensure change. (Heubert & Hauser, 1999, pp. 35-36)

The standards component of standards-based reform usually takes the form of state-approved documents that specify for each subject area what is to be taught and to what level.

The release of *A Nation at Risk* in 1983 triggered the call for world-class standards in U.S. education (National Commission on Excellence in Education, 1983). In the years that followed this warning of "a rising tide of mediocrity" in public education, one state after another began the move toward standards-based reform. This model for educational reform comprises three key policy components: rigorous standards in core subject areas, tests aligned with these standards, and accountability for results. The model has received strong backing from the business community as well as both sides of the political aisle because it is seen as a way to achieve excellence and equity in public education, improve U.S. performance on international assessments, and make the country a competitive force in the global economy.[i]

The standards component of standards-based reform usually takes the form of state-approved documents that specify for each subject area what is to be taught and to what level. The aim is to provide guidelines that teachers can use to create a challenging and high-quality curriculum for all children, regardless of where they attend school. At the time of this study, 48 states, in addition to the District of Columbia, had standards in the four core areas of mathematics, English, science, and social studies although the rigor and specificity of these standards varied considerably across states. Another state — Rhode Island — had standards in three of these subject areas while Iowa was the only state without state-level standards (Quality Counts, 2002).

Box 2

Testing Terminology

Test

A test is a set of questions or situations designed to permit an inference about what an examinee knows or can do in a given area. For example, by asking a sample of questions (drawn from all the material that has been taught), an algebra test is used to estimate how much algebra a student has learned. Most commonly used tests have the examinee select from a number of answers (e.g.,multiple-choice tests) or supply oral or written answers (e.g., structured interviews, essay questions). A test can also require the examinee to perform an act (e.g., read aloud from a book) or produce a product (e.g., compile a portfolio, write a book report). Because they are based on samples of behavior, tests are necessarily imprecise and scores should be interpreted carefully.

Standardized Test

A test is considered standardized when administration and scoring procedures are the same for all examinees (e.g., all seventh graders answering the same questions in the same amount of time on their state's mathematics test). Standardizing the process helps ensure that no test taker gains an unfair advantage on the test and that the test results can be interpreted in the same way for all students.

Accommodations

Under certain circumstances, the test content, format, or administration can be modified to accommodate test takers unable to take the test under standard conditions. Accommodations are intended to offset or "correct" for distortions in scores caused by a disability or limitation. Examples of testing accommodations include large-print versions of the test for students with visual disabilities and simplified language versions for students with limited English proficiency.

Reliability

In testing, reliability refers to the consistency of performance across different instances of measurement — for example, whether results are consistent across raters, times of measurement, or sets of test questions. A test needs to demonstrate a high degree of reliability before it is used to make decisions, particularly those with high stakes attached.

Validity

Validity refers to whether or not a test measures what it is supposed to measure and whether appropriate inferences can be drawn from test results. Validity is judged from many types of evidence. An acceptable level of validity must be demonstrated before a test is used to make decisions.

Sources: National Commission on Testing and Public Policy (1990); National Research Council (1997); U.S. Congress, Office of Technology Assessment (1992).

The second component of standards-based reform is the test. Its purpose is to provide an external measure of how well students have learned the content and skills specified in the standards.

The second component of standards-based reform is the test. Its purpose is to provide an external measure of how well students have learned the content and skills specified in the standards. Since it is not possible to test everything outlined in the standards documents, students are assessed on a subset of what they are supposed to have learned, and this information is used to infer how well they have mastered the broader subject area. Given the considerable body of research showing that testing drives much of what teachers do (e.g., Madaus, West, Harmon, Lomax, & Viator, 1992), one of the principles underlying the early development of these testing programs was to create "tests worth teaching to" (Resnick, 1996). This meant moving away from traditional multiple-choice tests that required students to select the right answer from the options given, and toward assessments that required students to demonstrate their knowledge and skills in novel situations, provide elaborated responses to questions, and explain the reasoning behind their answers. The expectation was that if teachers taught to these tests, they would be exposing students to the kinds of learning experiences that are at the heart of standards-based reform. Despite what some saw as the promising nature of this experiment, it proved difficult to implement on a large scale due to high costs, logistical issues, and concerns over the validity and reliability of the test results (e.g., Koretz, Barron, Mitchell, & Stecher, 1996). Most states now use a combination of multiple-choice, extended-response, and short-answer questions (Quality Counts, 2002) (see Box 3 for a description of how results are reported).

Box 3

Standards-Based Reporting

A distinguishing aspect of these testing programs is the way in which results are reported. Many states employ labels that describe a student's overall performance on the test in terms of specified performance levels. Since there is no firm mathematical procedure for choosing the cut-points between performance levels, committees are usually formed and judgment is used to decide where they should be set. For example, if the range of possible scores on a social studies test is 0 to 100, the committee may decide to designate all scores below 40 as *Below Basic*; scores between 41 and 60 as *Basic*; scores between 61 and 80 as *Proficient*; and scores of 81 and above as *Advanced*. Thus, if a student receives a score of 60 on the test, this will place her in the *Basic* category. If another student receives a score of 61, this will place her in the *Proficient* category. These labels are then used to report publicly on the extent to which students are meeting the state standards for a particular subject or grade level.

Since the choice of cut points is judgmental, it is frequently called into question (e.g., Shepard, Glaser, Linn & Bohrnstedt, 1993). While the committee in the above example chose to designate all scores between 61 and 80 as *Proficient*, another committee might have chosen all scores between 70 and 85. Because there are a number of approaches to standard setting (see Horn, Ramos, Blumer, & Madaus, 2000 for an overview), each committee might be able to present a defensible argument for why they chose their particular cut points.

In addition, the use of cut points and associated performance labels reduces the amount of information that is conveyed about student performance. As a result, large changes in student performance may go unrecognized (e.g., a score of 41 and a score of 60 are both considered *Basic*) and small ones may be magnified (e.g., because of a one-point difference the two students in the above example fell into different performance categories).

The third policy component of standards-based reform is accountability. This involves holding some member or members of the education system accountable for how well students have learned the content and skills laid out in the state standards. Since the state-mandated test is often the only measure used to gauge student learning and progress, test results are the most common method used for holding students — or when aggregated, teachers, schools, or districts — accountable. For example, at the time of this study, 18 states used their test results to make decisions about student promotion or graduation from high school, 17 used them to make decisions about school closure or reconstitution, and 30 publicly ranked or rated schools according to test performance (Quality Counts, 2002). The 2001 re-authorization of the Elementary and Secondary Education Act — also known as the No Child Left Behind (NCLB) Act — further increases the accountability uses of these tests by requiring all students in grades three through eight to reach the "proficiency" level on state reading and mathematics tests by 2014, and holding schools and school districts responsible for making adequate yearly progress toward these results (see Box 4 for a description of this Act).

The third policy component of standards-based reform is accountability. This involves holding some member or members of the education system accountable for how well students have learned the content and skills laid out in the state standards.

Box 4

No Child Left Behind Act

The main requirements of this federal mandate are:

Standards: States must put in place challenging content standards in reading and mathematics.

Tests: All students in grades 3-8 must be tested annually on these standards. Results must be broken out by student groups in terms of poverty, race and ethnicity, disability, and limited English proficiency.

Accountability: States must put in place annual statewide progress objectives ensuring that all groups of students reach proficiency on these tests within 12 years. School districts and schools that fail to make adequate yearly progress toward these goals will be subject to corrective action and restructuring measures aimed at getting them back on course. Schools that meet or exceed the annual progress objectives or close achievement gaps will be eligible for academic achievement awards.

States have until the 2005-06 school year to put these content standards and annual tests in place, and until 2014 to help all groups of students reach proficiency. States that do not comply risk losing some federal education funding (about 7 percent of their total budget).

Source: This description is adapted from a government fact sheet. It is available at: http://www.whitehouse.gov/news/releases/2002/01/print/20020108.html

Taken together, standards, tests, and accountability are seen as mutually reinforcing components of the overall push for excellence and equity in public education.

Taken together, standards, tests, and accountability are seen as mutually reinforcing components of the overall push for excellence and equity in public education. The standards outline the expectations that are held for all students, the test provides a way to judge student performance against these standards, and the accountability requirement 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 theory of action implied by this model[ii] is built on some key — but largely untested — assumptions, particularly in relation to the motivational power of the stakes attached to the test results. For example, it is assumed that teachers will pay more attention to the state standards in their daily instruction if their students have to take an aligned test; and that, if important decisions are based on the test results, the motivating power of the test will be increased. In addition, there is the assumption that the actions taken by educators or students in order to avoid sanctions or receive rewards will lead not only to improved scores on the state test, but also to improved teaching and learning.

While many studies have shown that state testing programs do have an impact on classroom practice (see Box 5 for a summary of findings), they are unclear as to how this varies according to *the kinds of stakes* attached to the test results. This is because most studies focus on individual states (e.g., Koretz, Mitchell, Barron, & Keith, 1996; Smith et al., 1997), or else do not look systematically across testing programs with different kinds of stakes attached to the test results. Thus, while there is much research on state testing programs, there is no firm basis for determining the precise mix of rewards and sanctions that will maximize the positive, and minimize the negative, effects on classroom practice.

The lack of clarity on this issue could be seen in the varied landscape of state testing programs at the time of this study. Some states held students accountable for the test results (e.g., Ohio), some held educators accountable (e.g., Kentucky), some held both accountable (e.g., Florida), and some held neither accountable (e.g., Maine). Within each of these groups, accountability could be further defined in differing ways. For example, some states held students accountable by requiring them to pass the state test in order to be promoted to the next grade (e.g., Delaware), others required students to pass the state test in order to receive their high school diploma (e.g., Nevada), and still others required students to do both (e.g., Louisiana). The accountability requirements of the NCLB Act reduce some of this variation by requiring all states to hold schools and school districts responsible for test results. At the same time, there is room for interpretation at the state level since the issue of stakes for students is not addressed. As states begin to work toward the NCLB goal of "proficiency for all," it is more important than ever that they consider the effects of the accountability uses of their test results on students and schools, and find ways to maximize the positive effects while minimizing the negative ones. *The goal of this National Board study was to better understand the effects of these different accountability uses by looking inside the black box of classroom practice.*

Box 5

Impact of State-Mandated Testing Programs on Teaching and Learning

Much of the research on state-mandated testing programs focuses on those that attach high stakes to the test results. In particular, these programs seem to attract attention when the stakes are for students. Consider, for example, the following headlines that appeared in Massachusetts newspapers around the time that scores for the 2001 administration of the state-mandated test, the MCAS, were released (students must pass the tenth grade test to be awarded a high school diploma):

"Boston students post record high MCAS scores" *(Boston Globe)*

"10th grade MCAS scores soar" *(Lowell Sun)*

"Business leaders laud MCAS result" *(Boston Globe)*

" MCAS score gains generate suspicion" *(Springfield Union-News)*

"School rankings represent hard work, wealth" *(Boston Globe)*

"Special ed student still struggling with MCAS" *(Metrowest Daily News)*

"Only one [student] passes [the] MCAS alternative [test]" *(Boston Globe)*

"Amherst [school district] may defy MCAS diploma rule" *(Springfield Union-News)*

"Thousands didn't take MCAS tests" *(Springfield Union-News)*

"MCAS racial gap widening" *(New Bedford Standard-Times)*

These banners reveal several of the issues that are fueling the debate over, and the research on, state-mandated testing programs. For instance, while some have ascribed improved scores to increased student learning (e.g., Grissmer, Flanagan, Kawata, & Williamson, 2000), others charge that there is a cost in real knowledge as students focus on learning what will be tested rather than the broader knowledge laid out in the state standards (e.g., Amrein & Berliner, 2002; Klein, Hamilton, McCaffrey, & Stecher, 2000). In addition, while proponents point to a reduced score gap between student groups on some state tests, others note the negative impact on minority, special education, and Limited English Proficiency students, particularly when promotion or graduation decisions are attached to the results (e.g., Orfield & Kornhaber, 2001). The strong relationship between test-based rankings of schools and students' socio-economic status also raises the question whether the scores reflect students' hard work or the increased learning opportunities that wealth affords.

Other issues have been raised in regard to the impact these tests have on teachers and schools. While the tests, especially when aligned with rigorous standards, can encourage educators to improve the quality of their curriculum and instruction, the pressure to improve scores can lead to teaching to the test (Madaus, et al., 1992) and to cheating scandals. Some have questioned the use of these tests to make highly consequential decisions about students (e.g., high school graduation) while teachers' judgment and school-based measures of student competency are ignored.

Overall, the research shows that these testing programs can have both positive (e.g., Bishop & Mane, 1999; Wolf, Borko, McIver, & Elliott, 1999) and negative (Jones et al. 1999; Smith, Edelsky, Draper, Rottenberg, & Cherland, 1991; Stecher et al., 2000) effects on teaching and learning (see Hamilton, Stecher, & Klein, 2002 or Mehrens, 1998 for a summary). Unclear is the mix of rewards and sanctions that will optimize the positive and minimize the negative effects.

The National Board Study

Goals

In 2000, the National Board on Educational Testing and Public Policy began a two-year study of state-mandated testing programs. The goal of this 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. Data were collected using mail surveys of teachers and on-site interviews with educators, with the former providing a national picture of teacher opinion and the latter an in-depth look at the testing programs of three states. The remainder of this report describes the interview portion of this study.

The goal of this 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.

Categorizing State-Mandated Testing Programs

Key to the study design was the inclusion of states with different kinds of stakes attached to the state test results. This required that state testing programs be categorized accordingly. When this was done, two general but overlapping groups emerged: (1) state testing programs with stakes for teachers, schools, or districts (hereafter referred to as educators), and (2) state testing programs with stakes for students. Each group could be further divided according to the severity of the stakes – i.e., high, moderate, or low. Box 6 contains the definitions used for each.

Box 6

Stakes Levels

Stakes for Students

Low Stakes: No consequences attached to the state test scores

High Stakes: Regulated or legislated sanctions or decisions of a highly consequential nature are based on the state test scores (e.g., promotion/retention, graduation)

Moderate Stakes: By default, all other test score uses (e.g., students may be given a certificate of mastery or other marker of success based on test performance)

Stakes for Teachers/Schools/Districts

Low Stakes: No consequences attached to the state test scores

High Stakes: Regulated or legislated sanctions or decisions of a highly consequential nature are based on the state test scores (e.g., accreditation, funds, receivership)

Moderate Stakes: By default, all other test score uses (e.g., ranked test scores for schools/districts available on the web or in local newspapers)

The product of this categorization process was a 3x3 grid of state testing programs, as shown in Appendix A, which was used as the organizing framework for the interview and survey aspects of the project.[iii] For the interview study, one state was selected from each cell in the top row of the grid, resulting in the following state testing program profiles (information on each is provided in Box 7):

- Low stakes for students and high stakes for educators (Kansas)
- Moderate stakes for students and high stakes for educators (Michigan)
- High stakes for both students and educators (Massachusetts)

These state testing program profiles are based on information found in state legislative documents in summer 2000.[iv] While this approach provided a common template for categorizing state testing programs, it did not always match the "on the ground" view, as was confirmed for us when we contacted officials in each state. Although officials in Massachusetts and Michigan agreed with our categorizations of their testing programs, officials in Kansas felt that their testing program was more moderate than high stakes for educators, since scores on the state test are only one of several pieces of information used to evaluate schools. Even with this shift for Kansas (i.e., to moderate stakes for educators), we were able to take a close look at differences in stakes for students while holding the stakes for educators fairly constant. This selection of states is of particular interest in the current policy climate, since the accountability model laid out by the NCLB Act requires that schools be held accountable for state test results, but provides some room for states to decide whether or how to hold students accountable.

This selection of states is of particular interest in the current policy climate, since the accountability model laid out by the NCLB Act requires that schools be held accountable for state test results, but provides some room for states to decide whether or how to hold students accountable.

Box 7

Kansas, Michigan, and Massachusetts State Testing Programs

Tested Subjects and Item Formats

At the time of this study, all three states had developed standards and tests in the core subject areas of mathematics, English, science, and social studies. Students were tested in each subject one or more times at the elementary, middle, and high school levels. A mix of multiple-choice and open-response questions was used on all of the Michigan and Massachusetts tests and one of the Kansas tests (the remainder used only multiple-choice questions). Tests were generally administered in the spring and results returned to schools by early summer or the beginning of the next school year. In all three states, aggregated results were publicly reported by student group (e.g., gender, ethnicity, Limited English Proficiency).

Stakes for Educators

In each state, the test results were used to hold schools accountable. In Kansas, results on the state test were used in combination with other information (e.g. graduation rate, attendance) to determine school accreditation. In Massachusetts, schools were held accountable for the percentage of students in the *Failing, Proficient*, and *Advanced* performance categories on the mathematics, English, and science tests. Schools that consistently failed to meet expectations could be deemed chronically underperforming and placed under new leadership. In Michigan, school accreditation was based on student participation in, and performance on, the state test. Due to criticisms of the heavy reliance on state test scores, Michigan adopted a new school accreditation policy in March 2002 (after we had finished data collection) that more closely resembles that of Kansas.

Stakes for Students

The official stakes attached to the test results for students differed considerably across the three states. In Kansas, no official consequences for students were attached to their performance on the Kansas Assessments. In Michigan, students who achieved a Level 1 or Level 2 on the eleventh-grade tests in mathematics, English (reading and writing), and science could receive an endorsed diploma and were eligible for up to $2,500 in tuition credits at an approved in- or out-of-state institution. Those who achieved a Level 3 on these tests could receive an endorsed diploma. In Massachusetts, students who were in the tenth grade at the time of this study were the first group to be required to pass (i.e., score in the *Needs Improvement* category or higher) the tenth-grade tests in mathematics and English in order to receive a high school diploma.

Accommodations and Alternative Assessments

Since the emphasis in all three states was to hold all students to a common set of academic standards, accommodations and alternative assessments were available to maximize student participation in the state testing programs. Federally legislated Individualized Education Plans (IEPs) and 504 plans determined the modifications available; limited English proficiency also was a consideration. Typically, accommodations involved changes in the time given for testing or the setting in which it was given. Where more substantial modifications were necessary, both Kansas and Massachusetts provided alternative forms of the test (e.g., students could submit portfolios of their work).

Testing Program	Kansas Assessment Program			Michigan Educational Assessment Program (MEAP)			Massachusetts Comprehensive Assessment System (MCAS)		
Year Testing Began	1996[1]			1989 (Reading), 1991 (Mathematics), 1996 (Science & Writing), 1999 (Social Studies)			1998		
Tested Subjects	Grade	Format	Standards Adopted[2]	Grade	Format	Standards Adopted[3]	Grade	Format	Standards Adopted
Mathematics	4, 7, 10	Multiple Choice (MC)	1990, 1993, 1999	4, 8, 11	MC, OR	1988, 1995	4, 6, 8, 10	MC, OR	1995, 2000
Reading and Writing/English	5, 8, 11	MC & Open Response (OR)	1996, 1998	4, 5, 7, 8, 11	MC, OR	1986, 1995 (Reading) 1985, 1995 (Writing)	3, 4, 7, 8, 10	MC, OR	1997, 2000
Science	4, 7, 10	MC	1993, 1995, 2001	5, 8, 11	MC, OR	1991, 1995	5, 8, 10	MC, OR	1995, 2001
Social Studies	6, 8, 11	MC	1999	5, 8, 11	MC, OR	1995	5, 8, 10	MC, OR	1997
Performance Levels	Unsatisfactory, Basic, Satisfactory, Proficient, Advanced			Level 4: Not Endorsed Level 3: At Basic Level Level 2: Met Michigan Standards Level 1: Exceeded Michigan Standards[4]			Failing, Needs Improvement, Proficient, Advanced		
Consequences for Districts, Schools, Teachers	Test results were one piece of information used to determine school accreditation (e.g. graduation rate, dropout rate, and professional judgement were also used).			School accreditation determined by student participation in, and performance on, the MEAP. Elementary schools were eligible for a Golden Apple monetary award if they met designated performance goals.			School performance ratings were based on the percent of students in the *Failing*, *Proficient*, and *Advanced* categories for each subject test. Schools not meeting performance expectations underwent review, and — if no improvement occurred — a change in leadership.		
Consequences for Students	No consequences for students.			Students scoring at Levels 1 or 2 on the eleventh grade tests were eligible for up to $2,500 in college tuition credit. Students scoring at Levels 1, 2, or 3 received an endorsed transcript			Students in the class of 2003 and beyond had to pass the tenth-grade English and math tests in order to graduate from high school.		
Accommodations and Alternative Assessments	Yes			Yes. Accommodations			Yes		

Sources:

Kansas: (1) http://www.ksde.org/assessment/assess_update2000.html (2) http://www.ksde.org/assessment/index.html (3) Quality counts. (2001, January 11). *Education Week, 20* (17). (4) C. Randall (personal communication, December 14, 2002).

Michigan: (1) http://treas-secure.state.mi.us/meritaward/meritindex.htm (2) http://www.meritaward.state.mi.us/mma/results/winter99.pdf (3) http://www.meritaward.state.mi.us/mma/design.htm (4) Quality counts. (2001, January 11). *Education Week, 20* (17).

Massachusetts: (1) http://www.doe.mass.edu/mcas/overview_faq.html (2) http://www.doe.mass.edu/Assess/ (3) http://www.doe.mass.edu/frameworks/current.html (4) http://www.doe.mass.edu/frameworks/archive.html (5) Quality counts. (2001, January 11). *Education Week, 20* (17).

Notes:

1 The state testing program began in 1991 with a pilot test in mathematics. Tests in other subject areas were added in subsequent years. 1996 represents the beginning of the second cycle of assessments (the testing program runs in cycles). However, it is the first cycle that began with tests in all subject areas. The state entered a third assessment cycle in the 1999-2000 academic year. Tests used in this cycle were based on the most recent versions of the state standards.

2 Multiple years indicate revisions to the standards.

3 Tests based on the 1995 standards were due for release in 2002.

4 Performance levels shown are for the high school tests. The following levels were used for the elementary and middle school tests: mathematics and reading (Satisfactory, Moderate, and Low); science (Proficient, Novice, Not Yet Novice); writing (Proficient, Not Yet Proficient); social studies (same as for the high school tests except for Level 4, which was termed "Apprentice").

Interviews with Educators

Also key to the National Board study design was a focus on educators since they are best positioned to witness the effects of testing policies on classroom practice. Classroom teachers in particular are an important voice in this conversation since they work at the intersection of policy and practice, and must turn a set of standards and test-related expectations into a set of educational practices. Approximately 360 tape-recorded interviews (120 per state) were conducted with educators at various grade levels, in multiple subject areas, and across several schools and districts in the three study states. Districts and schools were so chosen as to provide a representative socio-economic and demographic profile for each state as well as to illustrate the range of performance on the state test.[v] Interviewees were chosen using a purposive sampling technique to represent a variety of grade levels, subject areas, and teaching experience. The final interview profile in each state was as follows:

Also key to the National Board study design was a focus on educators since they are best positioned to witness the effects of testing policies on classroom practice.

- Four districts: large urban, small urban, suburban, and rural
- Up to six public schools in each district: two elementary, two middle, two high
- Six interviews in each school: the principal or assistant principal, some teachers at the tested grades or who teach a tested subject area, some teachers at the non-tested grades or who teach non-tested subjects, other faculty (e.g., special education teachers, counselors)
- Two interviews at the district level: the district superintendent, assistant superintendent, or director of testing

On-site interviews were conducted between winter 2000 and fall 2001 using a semi-structured interview protocol that covered the following topic areas (the full protocol is shown in Appendix B):

- Perceived effects of the state standards on classroom practice
- Perceived effects of the state test on classroom practice
- Perceived effects of the state test on students
- Perceived effects of the state test on the ways in which schools spend their time and money
- Perceived effects of the state test on the teaching profession

Interviews took between 30 minutes and two hours and were tape-recorded unless otherwise requested. The methodology used to code and analyze the interview data is outlined in Appendix C. Emergent themes were checked to see whether they held up across subject areas, grade levels, school types, district types, and states. Sub-themes (i.e., those mentioned by less than ten percent of interviewees) also were identified. Follow-up telephone interviews were conducted with a representative sample of 40 of the original interviewees to help resolve contradictions in the original interviews, confirm what seemed to be key themes, and obtain further information on seemingly significant findings that were mentioned by only a few respondents.

Overall findings from both sets of interviews are elaborated in the remainder of this report. The focus is on interviewee responses to the first three topics in the interview protocol, although findings that emerged in the other two areas are discussed where relevant. For each topic, overall findings are presented first and followed by a detailed discussion for each state. At the state level, findings are presented in terms of overall opinions (neutral, positive, and negative) and then by school-type (elementary, middle, high) and district-type (large urban, small urban, suburban, rural) themes. Differences in the opinions of teachers and administrators, new and veteran teachers, and those in high- and low-performing schools or districts are not systematically presented since the most striking and policy-relevant differences emerged at the school- and district-type levels.

Themes are presented in two ways: code frequencies and quotations. The former appear as proportions of interviewees who held a certain opinion. In research of this sort, these proportions are not altogether precise and are best taken as order-of-magnitude estimates, not exact amounts. Quotations were chosen to show the range and tendencies of responses to the topic. Since confidentiality was promised to all interviewees, any quotations used in this report are identified only by the state, district type, and (in the case of teachers and principals) school type in which the interviewee worked. Middle and high school teachers are further identified by their subject area specialization.

SECTION TWO

PERCEIVED EFFECTS OF THE STATE STANDARDS ON CLASSROOM PRACTICE

In this section, we report on interviewees' comments in regard to the first component of standards-based reform — the curriculum standards.

Standards can improve achievement by clearly defining what is to be taught and what kind of performance is expected. (Ravitch, 1995, p. 25)

The biggest problem is infusing the standards into daily and weekly classroom activity. We're like many other districts. We come up with check sheets and ask that our teachers check off when they've covered standards....I don't think that really means that teachers understand the standards or the bigger scope of things. (Kansas, Large Urban District, Deputy Superintendent)

In this section, we report on interviewees' comments in regard to the first component of standards-based reform — the curriculum standards. In all three states, standards are provided in the following curriculum areas: mathematics, English, science, and social studies. A *framework* document is usually provided for each subject area, and this includes the *standards* to be taught as well as *benchmarks* for what students should know and be able to do to demonstrate attainment of the standards (see Box 8 for the official terminology used in each state). Interviewees tended to use the terms *frameworks, standards,* and *benchmarks* interchangeably. Whichever term they used, the underlying purpose of these components is similar: to provide explicit guidelines for curriculum at various grade levels, and implicit guidelines for what is to be tested. How do educators view these guidelines? How do the guidelines influence classroom practice? What effect does the state test have on this influence? We explored these and other questions with teachers and administrators in the three states. Overall findings are described below and then elaborated on a state-by-state basis.

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 important to 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 be experiencing the most challenges in trying 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 were the most likely to report that aligning with the state standards impoverished their curriculum, although they were still a minority of these interviewees. 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.

We found no clear overall relationship between the level of the stakes attached to the state test and the influence of the standards on classroom practice. Instead, these findings suggest that other factors are at least as important, if not more so, in terms of encouraging educators to align classroom curricula with the 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).

We found no clear overall relationship between the level of the stakes attached to the state test and the influence of the standards on classroom practice.

Box 8

State-Specific Terminology

Kansas

Standard: A general statement of what a student should know and be able to do in a subject area. Standards are listed in the state curricular documents for each subject area.

Benchmark: A specific statement of what a student should know at a specific time.

Indicator: A specific statement of the knowledge or skills that a student demonstrates in order to meet a benchmark.

Michigan

Curriculum Framework: This document covers the state content standards and benchmarks for the subject areas of English, mathematics, science and social studies and is intended as a resource for helping schools design, implement, and assess their curricula.

Standard: A description of what students should know and be able to do in a particular content area.

Benchmarks: Learning objectives for each content area that further clarify the content standards.

Massachusetts

Framework: The overall document for a subject area, to be used for developing curriculum in that area.

Strands: The content areas in a subject area under which the learning standards are grouped.

Learning Standard: A statement of what students should know and be able to do in each strand area at the end of each grade span or course.

Below, we present a more detailed discussion of these findings on a state-by-state basis. States are presented in order of increasing stakes for students: from low (Kansas), to medium (Michigan), to high (Massachusetts). For each state, findings are first presented in terms of overall opinions and then broken down by school-type (elementary, middle, and high) and district-type (large urban, small urban, suburban, rural) differences. Since the emphasis is on the voices of educators — all 360 educators in three states — quotations are used liberally throughout.

Kansas

[The state standards have] been a huge improvement for public education. This is my sixteenth year as a principal, and things have changed dramatically in the last six to ten years....It really helps that everyone is reading off the same sheet of music. No matter what school you are in, or...what grade level, you have uniform expectations, and also for the most part a pretty uniform sequence of instruction. There's some flexibility, but the sequence is laid out fairly well....It particularly will help first-year teachers....For the more experienced teachers, it helps them to refocus. (Large Urban District, Middle School Principal)

You're trying to cover a lot of things and not doing any one thing well. We're just racing from multiplication to long division to fractions to decimals and they haven't done any of that very well. They're still coming to us in fourth grade not knowing their basic multiplication facts....They just aren't ready for this. (Suburban District, Elementary School, Fourth-Grade Teacher)

Overall Findings

The first quotation above reflects the views of two-thirds of the Kansas interviewees — that the state standards were having a neutral to positive impact on classroom practice. Reported effects fell into two connected areas: the linking of district- and school-level curricula to the state standards, and the changing or redefining of classroom work in response to the standards. Kansas educators' high level of receptivity to these changes was linked to two perceived benefits, voiced by more than half the interviewees: the state standards allowed for greater consistency in what was taught, and they helped teachers to focus on more important content while eliminating fluff. Typical comments included the following:

The standards help to add some consistency to the districts across Kansas. If you're in fifth grade here in our district and you move to [another town] the teachers are [going to] be teaching the same thing there that they're teaching here. So it makes for some consistency overall. (Small Urban District, Elementary School, Fifth-Grade Teacher)

The economic argument that runs through standards-based reform emerged in some interviewee comments, usually in reference to creating employable graduates. One teacher remarked:

Since we've aligned our teaching to the standards, it's not only helping the test results...we're [also] producing better students that are employable. They can go out and actually handle a job, write a memo. (Large Urban District, High School, Home Economics Teacher)

Despite the generally positive nature of the comments, some concerns emerged. For example, about one-fifth of the interviewees commented on the need for better alignment between the state standards and the state test.[i] In particular, interviewees emphasized the need for tighter alignment between the cognitive skills that students were required to demonstrate and those outlined in the state standards. The social studies and science standards and tests were singled out as most in need of this attention. A social studies teacher explained the importance of cognitive alignment as follows:

> *I am the department head, and I am trying to have the teachers in my department match the state standards with their curriculum. What I see going on...is that in all of the workshops and the in-services...we are being taught to teach kids to think broader, more globally, to synthesize information. Yet the [state] tests come right back to the same types of tests that we have always had; they are factually based tests. But we are no longer really teaching kids to learn facts. And I think that we're catching kids in the middle. (Large Urban District, High School, Social Studies Teacher)*

While this was a concern to a minority of the interviewees, it raises an important issue in terms of the potentially negative effects of misalignment on the quality of instruction, particularly in high-stakes environments where teachers may feel pressured to teach to the test.

School-Type Differences

The second quotation at the start of the Kansas section, from a fourth-grade teacher in the suburban district, illustrates another concern — the appropriateness of the amount and type of content in the standards. This issue played out differently at the elementary, middle, and high school levels, with elementary educators most concerned about developmental appropriateness, and middle and high school educators about the amount to be covered. For example, about one-fifth of the middle and high school educators noted that the large amount of content to be covered placed pressure on the pace of classroom work, resulting in decreased flexibility and students being left behind. The social studies and science standards were viewed as posing the greatest challenge in this regard. As one high school science teacher noted:

> *We go faster than probably some students are able, so we do lose some. The middle- to upper-level kids will survive....[Even with my current pace] I'm a unit behind the curriculum. So you either cover it slow enough and well enough that everyone gets it and you lose the end of the curriculum...or you cover the entire curriculum and go fast enough that you're going to lose some kids in the process. (Suburban District, High School, Chemistry Teacher)*

At the elementary level, the issue was framed more in terms of the developmental inappropriateness of some of the standards. About one-fifth of elementary educators voiced this concern, mainly in relation to the mathematics standards. A second-grade teacher explained:

> *There are some things that the kids are just not ready for. In second grade, [the] kids are just not mature enough to handle adding all the way to a*

dollar...you can't teach regrouping and that's essential for learning how to add money....You're really cramming in a lot at the very end [of the time preceding the assessment], to try to get all of those standards done. (Large Urban District, Elementary School, Second-Grade Teacher)

Still, most elementary educators in Kansas were positive about the impact of the state standards, with two-thirds noting overall positive effects on classroom practice.

District-Type Differences

In addition to concerns over alignment between state standards and tests, several Kansas interviewees commented on district-level factors that restricted their ability to align their curriculum with the state standards. For example, one of the goals of standards-based reform is to reduce disparities in curriculum quality between poorer (usually urban) and richer (usually suburban) districts. The assumption is that the more rigorous state standards will raise the curricula in the former to a level comparable to that in the latter. However, in Kansas, educators in the large and small urban districts were least likely to report that their local curricula were being linked to the state standards. This seemed to be due to the existence of district-level testing programs in the urban areas that were aligned with local rather than state standards. The urban educators, then, while generally enthusiastic about the state standards, were placed in the difficult position of trying to balance local and state requirements. As a teacher in the large urban district remarked:

I don't just worry about the Kansas standards, I also have to worry about district standards....They are not necessarily in conflict, but one [type of standard] may override the other. So a lot of the time, I feel like teaching to the Kansas standards but I really should be teaching to the district standards....Sometimes, there is this...battle. (Large Urban District, Middle School, Seventh-Grade Mathematics Teacher)

Other alignment issues surfaced in the rural district. While interviewees there were the most likely to report that local and classroom curricula were being linked to the state standards (half reported that local curricula were being linked and more than four-fifths that the classroom curriculum had been affected), this alignment process seemed to be accompanied by considerable growing pains. These were expressed in several ways. For example, almost two-fifths of the rural educators — four times as many as in any other district — indicated that the state standards were not developmentally appropriate for their students. One rural principal explained:

[The state standards and assessments] are a little too far advanced....Our English would be okay, our writing, reading, those scores tend to be right on, but everything in the sciences and social sciences and mathematics seems to be for a grade older. (Rural District, High School Principal)

Further problems were caused by an apparent lack of funding and access to training materials. In particular, many of the rural interviewees expressed a wish for more exemplars of the state standards and associated student work as well as guidance on accessing appropriate curriculum materials and professional development.[ii] A mathematics teacher remarked:

> *I usually look at [the standards] and think: What in the heck do they mean? Give me some useful examples. That way, I can go through my textbook, and say to my students, 'Hey guys, this problem is like what's on the state assessment.' Or if I see some examples...I can try to integrate them into homework assignments. But without examples, it's kind of hard to know what the state is looking for....[Without them], the standards aren't as useful as they could be. (Rural District, High School, Mathematics Teacher)*

At the same time, rural educators showed a high level of support for their state standards, and about three-quarters of them — twice the rate of educators in other districts — mentioned that the standards were having a positive impact on classroom curricula.

Michigan

> *I have to say that when the benchmarks came down it was a relief. We had something in our hands. Up until then teaching was textbook-driven, and in the elementary schools there were no textbooks, so it was creating everything yourself but not having a solid idea of where to go. So I like the idea of the benchmarks. I think our state has done a good job of them in science. They're pretty direct; you can look at them and see exactly what the kids have to know. (Suburban District, Middle School, Eighth-Grade Science Teacher)*

> *There's so many benchmarks to cover that probably our curriculum has become a little more superficial than it used to be. We lost our tenth-grade life science class and we really went into some depth in that....In adjusting our benchmarks to the [state standards and tests] we lost all of our anatomy....We also lost...comparing different organisms....[Students] also don't get the coverage of the cell and photosynthesis and respiration that they used to....Our district regards the [state test] as maximum knowledge and I would regard it as minimum knowledge. (Small Urban District, High School, Chemistry Teacher)*

Overall Findings

As in Kansas, Michigan educators noted two main effects of the state standards: the linking of district- and school-level curricula to the state standards, and the redefining of classroom work in response to the standards. Slightly more than half of Michigan interviewees viewed these effects in a mainly neutral to positive light. Perceived benefits included greater curricular consistency and — as illustrated by the first quotation above — a reduced burden on teachers in terms of devising classroom curriculum. The benefits of curricular consistency as they relate to the equity goals of standards-based reform are illustrated in the following comment:

> *One thing that [the benchmarks have] done, which was sorely needed, is to put a standardization over all of the schools and all of the districts so that you don't get children coming from, say [a poorer district], into our area who haven't had the same background and experiences....Students may not have had identical experiences but at least they have the same standards and learning benchmarks, and that has helped a lot. (Small Urban District, Elementary School, First-Grade Teacher)*

The mathematics standards were viewed particularly positively. This seemed to be due to their emphasis on problem solving and other higher-order thinking skills as well as the availability of aligned textbooks. In the following quotation, a teacher reflects on the changes he has seen in his students since these standards were introduced:

> *[T]he emphasis for over five years now is to get the students to read, comprehend what they read, explain how they got their answer....That's one of the major changes in our math program...a lot of problem-solving skills as opposed to doing more computation....The computational skills are still there and you have to use them, but you do more reasoning and mathematical thinking, making connections... This new way of teaching definitely increases the students' attention span. They have to sit still and really think. I notice when I look around, a lot of students who at one time didn't have that ability to sit still...now focus more. They sit there, and they're thinking and they're concentrating. (Large Urban District, Middle School, Sixth-Grade Mathematics Teacher)*

While comments about the standards were mainly neutral to positive, about one-third of Michigan educators voiced specific concerns about their impact on classroom practice. One of the main concerns was the loss of teacher flexibility due to the specific nature of the standards, particularly in areas like English where creativity and freedom of expression tend to be prized. An English teacher remarked:

> *I don't have any problem with the curriculum frameworks. I think they're good as a guide, but let's not take that guide and make it etched in stone....You want to be able to explore little different avenues and then go back to [the guide]. (Large Urban District, High School, English Teacher)*

Suggestions for how to deal with this loss of flexibility included paring down the number of standards to be taught or prioritizing the standards in each subject area.

In comparison with their peers in Kansas, Michigan interviewees were more likely to bring up the state test when discussing the impact of the standards. This is not surprising since the stakes attached to the test are higher in Michigan and the testing program has been around for longer, two factors that make it more likely to be on the minds of Michigan educators. This does not necessarily mean that the test was a more dominant influence on classroom practice. One teacher explained the relationship between the two as follows:

> *The state test just gives you a heightened awareness of how your students are going to be measured....What do they need to know, and what is it in my subject area that I should at least introduce them to before they take the test....It tells you that these benchmarks are important, so make sure they're in your instruction and the students are able to do those things. (Suburban District, Middle School, Social Studies Teacher)*

In particular, when good alignment was perceived among the standards, the test, and available resources such as textbooks, the standards seemed to be to the fore in educators' minds as influencing classroom practice.[iii]

School-Type Differences

Elementary educators reported twice as often as middle or high school educators that their school's curriculum was being aligned with the state standards (two-fifths of elementary versus one-fifth of middle and high school educators). They also were more likely to have changed their classroom curriculum in response to the state standards (three-quarters of elementary versus two-fifths of middle and high school teachers). External support for these educators seemed to be limited, as indicated by the following quotations:

> *I've been to lots and lots of conferences...to figure out how to write our curriculum so that we support the kids when they take the [state test]. That it follow the state benchmarks was the number one priority....So basically I just went to the website, downloaded the [state] benchmarks, and started from there for the curriculum for [kindergarten through fifth grade]. (Rural District, Elementary School, Fifth-Grade Social Studies Teacher)*

> *When I came to fourth grade we received new benchmark [documents] so...over the summer I redid the whole curriculum. I now have a yearlong curriculum that integrates all of the benchmarks, and...when they add new ones I fit them in. (Suburban District, Elementary School, Fourth-Grade Teacher)*

At the same time, elementary educators were the most positive about the impact of the state standards, with one-third noting overall positive effects on classroom practice (compared with one-quarter of middle school and one-fifth of high school educators).

As in Kansas, concerns about increased instructional pace due to the amount of content to be covered played out mainly at the middle and high school levels, and concerns about developmental appropriateness at the elementary level. It was often hard to disentangle whether problems with instructional pace were due to the standards themselves or the pressure of the state test (the Michigan Educational Assessment Program, or MEAP). For example, a high school history teacher answered a question about the effects of the social studies standards on classroom practice as follows:

> *There's pressure now to cover the content of the MEAP...to go from the Great Depression to the present. That's not always possible, but at least I see the necessity of going from there...to at least the Nixon administration....I'll be able to achieve that but a number of teachers won't because they've decided to lengthen their units and do more with literature....Their students will then be at a disadvantage when they take this test....So what [the MEAP] does in actuality is prohibit in-depth study of topics. (Small Urban District, High School, Tenth-Grade History Teacher)*

This comment also echoes a theme found in the Kansas interviews: the problematic nature of the social studies standards. In this regard, Michigan educators not only noted difficulties with trying to cover their content, but also noted that their lack of alignment with the state test made it still harder to know what to teach. The extent of concern over the social studies frameworks may also be because there is no natural hierarchy of knowledge in this subject area (unlike English and mathematics) that would ensure some natural redundancy and allow students and teachers to build more easily on what has already been taught. It also tends to be difficult to decide on the content to be taught, which may lead to frameworks that are overloaded with content or vaguely defined.

District-Type Differences

Other differences played out at the district level. Educators in the urban districts were least likely to report that local curricula were being aligned with the state standards (one-tenth in these districts compared with two-fifths in the suburban and rural districts). This pattern is similar to that seen in the Kansas interviews, but seems to be for different reasons. While many urban educators in Michigan felt that the state standards affected their curriculum and practice positively (one-third in the large urban and two-thirds in the small urban district), they often noted that efforts to orchestrate alignment at the district level foundered through lack of capacity. Asked about the impact of state educational reform efforts on classroom practice, a teacher in the large urban district remarked:

> *The reform efforts have demanded changes, and if you are a professional and you keep up on professional readings, you know how you can...respond to the changes. I think that [accommodating these reforms] requires more collaboration among staff to learn what's happening from grade level to grade level. I also think that reform has made demands on teaching and learning that have not yet been fully put into place because [the district's] ability to affect those changes to the required degree is lacking. (Large Urban District, Elementary/Middle School, English/Social Studies Fourth-Grade Teacher)*

Thus, while the equity goal at the heart of standards-based reform — to offer the same quality of education to children in rich and poor districts — was on the minds of many urban educators, they seemed to be lacking the support to make this happen.

While suburban and rural districts seemed to be doing more aligning of their local curricula with the state standards, their reasons were different. Educators in the rural district cited the mismatch between their previous curriculum and the state standards. Like their peers in Kansas, they hoped that aligning with the standards would improve their performance on the state test. Suburban educators talked more about having to winnow down the content usually taught in order to focus on the state standards, resulting — at times — in a less rigorous curriculum, particularly in science. The following quotations illustrate these trends in the two district types:

> *We just went through a year-long review...and redesign of our curriculum....The document we've used as our anchor point has been the state frameworks document [because] this district puts a lot of emphasis on our performance on the MEAP, and so we aligned ourselves with the frameworks document for that reason. When we open our classrooms in the fall it will be with classes that have as their backbone the frameworks document — the objectives of the state. (Rural District, High School, English Department Chairperson)*

> *It's had an impact on our curriculum in science...and that has been both good and bad; good in the sense that it caused us to teach some topics that we hadn't taught before that...were probably good for students to learn, topics such as weather — a little bit of astronomy, some earth science. At the same time...the curriculum is not as strong as it was, especially for the better student. The students that I deal with in [Advanced Placement] chemistry are not nearly as well prepared as they were in the past. And our scores are not as good as they were in the past on the [Advanced Placement] exam. (Suburban District, High School, Science Teacher)*

While the second quotation above highlights an instance when curriculum changes resulted in a less rigorous curriculum, this was not always the case. In fact, educators in the suburban district were generally positive about the effects of the state standards on classroom work. On the other hand, rural educators were the least positive of all those we interviewed in Michigan (only about one-tenth felt there had been a positive impact on classroom practice), and were the most likely to note that the test rather than the standards was the main influence on classroom practice.

Massachusetts

> *Part [of it is] convincing [everybody] that this is real. There is...the sense that it will go away. The frameworks will go away; if I just keep ignoring it, it will go away....We can't ignore it. I announced rather facetiously that we do have a state curriculum....It seemed like an obvious statement. It hit home for some people....A teacher asked me this afternoon, 'Are you telling me that I can't teach [what is] not in that curriculum?'....[I replied that if the students] learn everything in the state frameworks and you have time left over, okay, do what you want. (Rural District, High School Principal)*

Standards in themselves don't impact the quality of either teaching or learning....They're nice because they're a common target and send a message that we're all working toward that target, but they're not particularly meaningful to kids [or] useful as hooks to change instruction....What I attribute to the high stakes of [the state test] — the MCAS — is the urgency factor...this thing is coming down the train track and we've got to do something quick or we're going have kids potentially not graduate. That creates a sense of 'all hands on deck'; we've got the target, now we're trying to ratchet up the quality of instruction and [do that on a large] scale. (Large Urban District, Superintendent)

Overall Findings

Like their peers in Kansas and Michigan, Massachusetts educators noted two main effects of the state standards: the linking of district- and school-level curricula to the state standards, and the redefining of classroom work in response to the standards. About three-quarters of the Massachusetts interviewees — more than in the other two states — saw these effects in a neutral to positive light. For example, as in Kansas and Michigan, many saw the standardization of curricula throughout the state as a way of ensuring that all students were exposed to high-quality curricula and pedagogy.

Another perceived benefit, mentioned by about one-seventh of the interviewees, was that the state standards encouraged so-called curriculum spiraling (also mentioned in Kansas and Michigan, but to a more limited extent). This vertical alignment of curricula across grade levels allowed teachers to anticipate what students already knew, resulting in less repetition and a quicker pace of instruction. Many interviewees also were enthusiastic about pedagogical changes that were written into the Massachusetts frameworks. For example, a mathematics teacher remarked:

I think that [the framework's] use of technology, even the graphing calculators [is good]....I also like the fact that there's more group work, cooperation between students....Ten years ago, everything had to be neat, everything had to be the same, you couldn't [have students] talk to each other. [Students] either got it or didn't, and they used to get frustrated a lot. (Large Urban District, High School, Mathematics Teacher)

Others cited changes that benefited particular groups, such as students with special needs and students who learn more slowly. The number of references to the benefits for special needs children is worthy of note since they were rarely mentioned in Kansas and Michigan interviewee responses to questions about the state standards. A guidance counselor framed these benefits in terms of the higher expectations now held for these students and went on to explain:

We [now] have curriculum frameworks that every child must be exposed to. If you are a fourth grader and two years below grade level you still have those same curriculum frameworks....It could be that there are modifications and accommo-

> *dations so that you could access [the material] in a different way, but you have [the same] curriculum....The expectations are greater. The kids are exposed to all curriculum areas and content areas. They're not pulled out of the classroom during reading. They're not pulled out during math. They're in the classroom at all times. (Suburban District, Elementary School, Guidance Counselor)*

Despite the generally positive views held about the curriculum standards, several concerns were manifest. These included too much content to be covered in too short a time, an increase in developmentally inappropriate practices, and the misalignment of the standards and the state test. In comparison with interviewees in the other two states, larger numbers of Massachusetts interviewees brought up these negative aspects of the state standards.

The state test (the Massachusetts Comprehensive Assessment System, or MCAS) was always on the mind of Massachusetts educators, even as they spoke about the standards and their influence on classroom practice. About one-fifth of Massachusetts interviewees felt that the state standards and test were not aligned; about the same number felt that they were aligned. These differing opinions can be traced to the subject areas being commented on. Generally, the social studies standards and associated tests drew the most criticism for poor alignment, while the mathematics standards and tests were seen as having the best alignment.[iv]

In terms of which was driving classroom instruction, many believed that the state test was the greater force, while some felt that the impetus for change came from both the standards and the test. One principal summed up the difficulty many interviewees had in trying to tease out these issues:

> *I think [this test-related pressure] has improved the way children learn and the way teachers teach. I've seen a lot more constructivist, hands-on type of teaching. I've seen a lot more teachers really involving children in the learning process, and they would probably tell you that MCAS had nothing to do with it. All I know is that before [MCAS] they didn't teach that way, and now they do. You have to ask yourself, what made that happen? I think that happened...because of MCAS....[At the same time,] I find myself saying MCAS, but it isn't MCAS, it's the frameworks. (Small Urban District, Elementary School Principal)*

In comparison to their peers in the other two states Massachusetts educators most often mentioned using the state test as the target for their teaching efforts (over two-thirds of these interviewees versus one-third in Michigan and one-fifth in Kansas), suggesting that as the stakes attached to the test results increased, the state test was more likely to become the medium through which the state standards were interpreted.

School-Type Differences

Unlike in Kansas and Michigan, concerns about the speedier pace of instruction necessitated by trying to get through everything in the standards were heard at all levels of schooling, not just the middle and high school levels. Massachusetts educators found the social studies standards particularly problematic in this regard, as exemplified by the following quotations:

> *[Some of the standards] are right on, but some of them, for example...the social studies standards...are just too much. If you spent an entire year only doing social studies in fourth grade you might be able to get it [done]. Otherwise, it's very, very difficult, particularly if you don't have the material to do it. (Rural District, Elementary School, Fourth-Grade Teacher)*

> *It's like a shot in the dark because the state can take anything within that timespan [from] the Revolution to the Reconstruction, which is a tremendous amount of material, and ask any question [it wants]. Sometimes, when I'm feeling down about the test, [it seems like] buying a lottery ticket. You don't know what you'll get. (Small Urban District, Middle School, Eighth-Grade Social Studies Teacher)*

The faster pace of elementary instruction seemed to be linked to the large amount of content to be covered in the state standards. Since elementary teachers generally have to teach every subject area, and the elementary curriculum usually includes instruction in areas besides academic ones (e.g., social skills), this resulted in teachers scrambling to try to translate several sets of standards into daily lessons, and then fit everything else into the time allotted. Despite these time pressures, some elementary teachers refused to cut non-academic areas even though, as the teacher below remarks, "it's not something that they test."

> *I do a lot of work in my classroom on character building. It's really a big part of my [program] and is something I'll never cut no matter how much time I need....I do a lot of goal setting...and every Friday my class plays 15 minutes of games....It [fosters] a much better climate, the kids learn a lot better when they know how to play together....That will never be cut. And it's not something that they test. (Suburban District, Elementary School, Fourth-Grade Teacher)*

As in Kansas and Michigan, doubts about the developmental appropriateness of the standards were heard frequently at the elementary level, with one-quarter of these educators expressing concerns. However, unlike in Kansas and Michigan these concerns also often emerged at the upper levels, with one-quarter of middle school and one-seventh of high school educators reporting classroom changes they found developmentally inappropriate. At the high school level, interviewees wondered whether the lower-ability students could negotiate the state standards as well as the high achievers, hinting at multi-level testing, but with some fearing a return of tracking. Since students — starting with the class of 2003 — must pass the tenth-grade tests in order to graduate from high school, this was always on educators' minds as they discussed the issue.

> *My fear is that [the lower-scoring] group of kids is going to become a tracked group, essentially preparing to pass MCAS for the rest of their high school career. It's only logical....If I can't graduate until I pass this test, well, I'm going to do everything I can to get...ready to pass the test. So I can see...a system where you're going to have a whole bunch of remedial courses designed to get kids over that bar of the MCAS. (Rural District, Superintendent)*

Overall, elementary teachers reported far more changes in their classroom practice than did middle or high school teachers. While this pattern is similar to Kansas and Michigan, the intensity is greater. In fact, almost 100 percent of elementary teachers interviewed in Massachusetts said that the state standards had influenced their classroom practice and most felt it had done so in a neutral to positive way. At the same time, elementary teachers seemed to be carrying much of the burden involved in making these changes. The burden also was greater for new teachers, as suggested by the following comment:

> *The reform efforts have [had] a huge impact. As a fourth-year teacher with just three years of teaching in Massachusetts, coming to know and understand the frameworks and standards is a big deal. Yet is seems to be a self-study — not much guidance and clarity from administration. (Suburban District, Elementary School, Fourth-Grade Teacher)*

While middle and high school teachers in Massachusetts reported effects less frequently (about three-quarters at each level, with most feeling neutral to positive about them), they did so more often than their peers in Kansas and Michigan.

District-Type Differences

As in Kansas and Michigan, educators from the suburban and rural districts were about twice as likely as those from the urban districts to report that local standards were being linked to the state standards (two-fifths in the suburban and three-quarters in the rural versus one-third in the large urban and one-fifth in the small urban). While suburban educators were generally positive about the state standards, like their peers in Kansas and Michigan, a few (about one-fifth) noted that these alignment efforts could impoverish the curriculum. Rural educators had even more mixed feelings about the state standards. There was a strong local identity theme in their comments on this issue. One teacher explained:

> *Part of our [democratic values] is recognizing that each community is unique.... And we should have some control over what our students are learning. The state is taking that away and becoming Big Brother. (Rural District, High School, English Department Head)*

Here, the curriculum in place before the state standards were adopted seems to have been closely linked to the local context. Thus, in order to match it to the more context-free state standards, considerable realignment was needed. Some noted the benefits of these changes:

> *[The standards have] forced me to step back from what I'm doing and to look at whether I'm covering the entire curriculum. And I have found lapses....There are some things that I have avoided because I am afraid they will turn off the majority of kids....It's also been really positive because it's linked me to the [lower levels of schooling]. We have [several] feeder schools, and our kids have come with such varied backgrounds — every school teaching whatever it felt like at whatever level — [that] when we got them it was like a roller coaster. We had no idea where anybody was....Now we're beginning to have a dialogue, and the dialogue is helping us create [a] continuum. (Rural District, High School Principal)*

However, as in the other two states, there were growing pains. Almost half the educators in the rural district reported that the standards were not developmentally appropriate, and about one-quarter reported that they restricted teacher flexibility and encouraged teachers to skim over material. Overall, about half the rural educators in Massachusetts — almost three times as many as in the other districts — felt that the impact of the state standards on their curriculum had been negative.

As in Kansas and Michigan, the urban educators in Massachusetts were the least likely to mention that local curricula were being linked to the state standards. However, this seemed to be for reasons different from those given by their peers in Kansas and Michigan. A clue lies in the fact that the urban educators in Massachusetts were almost as likely as those in the suburban district to note that the state standards had affected classroom practice. This impact seemed to be linked to two factors: the Massachusetts standards have been around for some time and so have had a chance to filter into the classroom; and the state's 1993 Education Reform Act made resources available to urban educators so that they could learn about the standards and align their curriculum with them. In both the large and small urban districts, the benefits that the money tied to education reform had brought to the district were recognized.

> *One of the things that [education] reform has done is to help school systems reach their foundation budgets....In order to bring [our school district] up to the foundation budget, a lot of state money has come into the system for instructional resources, for teachers....The concept [of foundation budgets] in education reform was to ensure that all school districts have sufficient resources and an optimal level of funding from the state and from the local community. The concern had been that there were many communities [that, on the basis of] the property tax, were able to fund their school districts at a much higher level. The cities in particular were suffering from under-funding, so they established this optimal level, and over the seven years of education reform, [there was a push] to ensure that all communities met this optimal level of funding, which is a foundation, not a ceiling....The funding that's come in has helped enormously. (Large Urban District, Elementary School Principal)*

At the same time, there was a sense among urban educators that while the state standards had improved the quality of the curriculum, the state test had created a potential barrier for their students.

> *Yes, MCAS reflects frameworks. However, the frameworks themselves for an inner-city community are just pie in the sky, it's so unrealistic. Some of the standards – if we had them here for four years we wouldn't get to them. These children need skills that are going to help them cope in their future, at least to get themselves a job so that they can support themselves, and have some degree of success that is maybe better than what their parents had, and that's all we can hope for. We cannot work with more than we have, and yet some of the expectations of MCAS are far above what we could ever hope for our kids to do. (Large Urban District, Middle School, Eighth-Grade Teacher)*

Overall, educators in the three states were mainly neutral to positive about their state's standards even as they contended with the implications for classroom practice.

Overall, educators in the three states were mainly neutral to positive about their state's standards even as they contended with the implications for classroom practice. In the next section, we address their views on the state test and how this impacted on what and how they taught.

Box 9

Examples of Test Questions

Kansas

Sample question for the sixth-grade social studies test:

Which of the following is the best example of an American export?

A. The United States produces steel.
C. The United States is selling wheat to Russia.

B. Japan produces excellent cars.
D. The United States is buying salmon from Canada.

Answer: C

Michigan

Prompt for the 1999 fifth-grade writing test:

TOPIC:
Memories

PRE-TEST DIRECTIONS:
Talk about these questions with your group, making sure everyone gets to speak.

THINKING ABOUT THE TOPIC:
Can you think of funny or happy memories? Do you remember celebrating a holiday or going to a wedding, a festival, or a birthday party?

Can you think of any sad, frightening, or embarrassing memories? Do you remember saying goodbye to a friend, being involved in an emergency, or getting a bad haircut?

Do you remember any exciting moments? Do you have memories of cooking dinner by yourself? Riding on an airplane? Waiting for an announcement about making a team? Getting a part in a play?

TEST DIRECTIONS: WRITING ABOUT THE TOPIC:
Writers often write about past experiences. They often recall a favorite memory, an event like a celebration, or a time when they were happy, embarrassed, proud, or frightened. **Write about a memory**.

You might, for example, do **one** of the following:

- write about an exciting or funny time you remember very well **OR**
- explain why some memories become important and others do not **OR**
- write about a family memory you've heard over and over **OR**
- write about a memory that includes a person who is important to you **OR**
- write about the topic in your own way

You may use examples from real life, from what you read or watch, or from your imagination. Your writing will be read by interested adults.

Massachusetts

Question from the 2001 tenth-grade mathematics test:

At the first stop, 3/4 of the passengers on the bus got off and 8 people got on. A total of 16 passengers were left on the bus. Write an equation that can be solved to show how many passengers were on the bus before the first stop. Let *x* represent the number of passengers on the bus before the first stop. (You do **not** have to solve the equation.)

SECTION THREE

PERCEIVED EFFECTS OF THE STATE TEST ON CLASSROOM PRACTICE

Reading and writing, arithmetic and grammar do not constitute education, any more than a knife, fork, and spoon constitute a dinner. (John Lubbock, Astronomer/Mathematician)

We have done a lot of training and preparing [of] the kids since January. We did review packets. We had Friday math days just because of this one test....I felt I was a math teacher from January until spring break. We had to drop other curriculum areas because of this — spelling, writing....We couldn't drop science because we had a science assessment coming up at the same time. (Kansas, Suburban District, Elementary School, Fourth-Grade Teacher)

In this section, we report on interviewees' comments on the second component of standards-based reform — the state test.

In this section, we report on interviewees' comments on the second component of standards-based reform — the state test. In Kansas, the state test is referred to as the Kansas Assessments, or KSAs; in Michigan, it is the Michigan Educational Assessment Program, or MEAP; and in Massachusetts it is referred to as the Massachusetts Comprehensive Assessment System, or MCAS (see Box 9 for sample questions from each test). At the time of this study, the subject areas tested were the same in each state — mathematics, English, science, and social studies — with each subject tested at least once in elementary school, once in middle school, and once again in high school.

The test results were used for different purposes in each state. At the time of this study, they were one of several pieces of information used to determine school accreditation in Kansas, but had no official stakes for students. In Michigan, school accreditation was determined by student participation in, and performance on, the MEAP, while students could receive 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 were based on the percentage of students in different performance categories for the mathematics, English, and science tests, while students — beginning with the class of 2003 — had to pass the tenth-grade test in order to graduate from high school.

As was evident in the previous section, it can be difficult to disentangle the effects of the state test from those of the curriculum standards, since the latter provide the content on which the test is supposed to be based. It also is difficult to disentangle the effects of the test from the accountability system, since the latter uses the test results to hold educators and students accountable. Nonetheless, we asked our interviewees about the impact of the state test on classroom practice. For example, we asked them to describe how the state test affects what teachers include, exclude, or emphasize in the curriculum. We also asked them to describe how preparing for the test affects teachers' instructional and assessment strategies. How do educators view these tests and their impact on classroom practice? Can the tests themselves influence what and how teachers teach, particularly in relation to the state standards, or do they need the extra push of mandated consequences? Overall findings are described below.

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 as one moved from the elementary (three-quarters of Kansas, one-third of Michigan, and four-fifths of Massachusetts elementary teachers) to the middle (one-third, one-quarter, half), and high (one-fifth, one-third, half) school 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.

...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.

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. Below, we present a discussion of these findings on a state by state basis.

Kansas

[Guidelines for] the state test came out earlier this year. We had to chop the middle of our social studies curriculum in half so we could teach Kansas history, because the state decided to test Kansas history [a grade earlier than we usually teach it]. So our students have had to stop what they're learning in...Western hemisphere geography and cultures...to learn Kansas history instead. (Small Urban District, Middle School, Special Education Teacher)

I don't think that teachers have to eliminate topics because of the focus on what is tested. Teachers will stay focused on the curriculum...standards. They touch on all of those. Come test time maybe they concentrate a little more on those areas they know are going to be on the test [but] they try not to leave out anything. (Large Urban District, Elementary School, Special Education Teacher)

[The state score reports] mention things [the students] score poorly on and we try to improve scores in those areas — [for instance,] probability and statistics. [I used to think] 'Oh, it's second grade, come on!' and I would save that for the end of the year, but this year I plugged it in in February....And I know that third grade is doing the same thing, making sure they are addressing topics before [students] get to fourth grade so that they have plenty of practice. (Rural District, Elementary School, Second-Grade Teacher)

Overall Findings

These quotations illustrate that preparation for the Kansas Assessments led teachers to engage in varying degrees of *removing, emphasizing,* or *adding* curriculum content. The most frequently reported activity was the removal or de-emphasis of content not covered by the state test. Next most frequently reported was the emphasis of content that would be tested. Least often mentioned was the addition of content. In addition to these changes in *what* was taught, most Kansas interviewees felt that preparation for the state test had produced changes in *how* they taught. Overall, only one in ten felt that the test had no impact on what they taught and one in seven said it had no impact on how they taught. While it was evident that the state test was having a marked impact on what went on in Kansas classrooms, the amount of change was less than that reported in Michigan and Massachusetts.

Interviewees identified both positive and negative aspects of these test-related changes. Most of the positive comments focused on improvements to teachers' instructional strategies. For example, about one-fifth of the teachers noted that the emphasis on critical thinking skills on the test encouraged them to emphasize these skills in classroom instruction. Typical comments included the following:

> *My teaching strategies have changed — what I teach and the way I teach it. I teach very differently from a few years ago. I think that's for the better, I think that's the positive part of the testing. The tests [focus] on the kinds of thinking that students need to do. It's easy to change my teaching to address the test, because I do like the test. (Large Urban District, Middle School, Special Education Teacher)*

In particular, teachers noted that they were emphasizing writing skills more, and that the quality of their instruction in this area had improved due to the adoption of the six-trait analytical writing model (ideas/content, organization, voice, word choice, sentence fluency, and conventions). A principal explained:

> *Students have to have preparation in the six-trait writing model because that's how the state writing test is scored. If they aren't aware of how to develop their story around those six traits, then they won't do well.... That piece [of the state education reform effort] had a big impact all across the state. Teachers [wanted] some direction in how to teach writing skills...and that writing model was very good. (Small Urban District, Elementary School Principal)*

This principal also noted that since his teachers had adjusted their instruction to the skills measured by the state test, scores on other tests had gone up as well. The implication was that these instructional changes constituted more than just teaching to the state test since they produced learning that generalized to other measures.

In addition to the positive effects associated with preparation for the state test, educators mentioned that receiving students' test results helped them to understand which of the state standards students were not mastering, and thus allowed them to better tailor instruction.

An elementary school principal in the small urban district explained that his school had "an improvement team in each of the academic areas — reading, writing, and math. Those folks analyze the state test score data and develop a plan for how we're going to improve." Another educator described the process at her school as follows:

> *When we get [the state writing results]...I'll see where some of the weak areas are, and I'll hit that a little bit harder the next year. Of course, that takes more time out of your regular teaching time. Sometimes it seems like it's kind of a vicious circle....But on the writing, by seeing how well the kids have done, or where their weak areas have been on previous tests [and using that information to plan instruction], we have seen a gradual improvement of our writing skills across the district. (Rural District, Middle School, Communications and American History Teacher)*

Overall, however, Kansas educators mentioned using the state test results less frequently than did educators in Michigan or Massachusetts. In addition, almost one-fifth of them felt that the results came back too late to be useful.

Interviewees also commented on some negative consequences of preparing for the Kansas Assessments. Four interrelated concerns, each mentioned by about one-fifth of the interviewees, are discussed here. First, the effort required to cover the necessary content, especially close to test time, created a hurried pace of instruction, which some characterized as the test "driving" the curriculum. While noting that this could help focus teaching and remove "fluff," educators also felt that useful enrichment activities were being struck from the curriculum. The tensions involved are illustrated in the following quotations:

> *We can't go out of the box any more, we can't do the fluff any more. We are trying to meet those benchmarks and we just don't have the time. It doesn't help that the tests are in March, and for fifth graders the writing is in January. So you're cramming half the year, to teach everything, and it is just not possible....If you want to do a seasonal poem, for example, you feel guilty, because that's not really driving towards the standards they're going to be assessed on....And just from looking at my scores last year — the math assessment was in March, and I hadn't taught fractions yet — my kids didn't do as well in fractions. (Large Urban District, Elementary School Teacher)*

> *[Because of the need to cover content before the test] I feel that I don't get to do as many fun activities, like cooperative learning activities or projects....I can't fit too many of those in because it would take up too much time, which is a shame because it would help the kids learn about the concepts. So I feel as if this year I've done a lot more direct teaching than being able to do student-led learning [activities]. That's what really suffers. (Suburban District, Middle School Teacher)*

The latter comment is additionally interesting because the teacher's reference to becoming more teacher-centered in her instructional style is in contrast to the more student-centered approaches implied by the learning goals of standards-based reform.

A second concern mentioned by about one-fifth of the interviewees was that the state test forced them to focus on breadth more than on depth of coverage. This was partly attributed to the addition of instruction in reading and writing to subject areas such as mathematics and science, which meant that these skills competed with subject area content for classroom time. The following comment from a mathematics teacher exemplifies this issue:

> *As a math teacher I am required to work with [students] on reading and writing as it relates to mathematics, and that has certainly changed what we do in the classroom. It has put us in a big bind because our curriculum area has not [been reduced]....In fact, it's grown a lot — we've had to add statistics and probability, and things that weren't necessarily in...the curriculum [in previous years]. And now they're there...[but] the time we see our students has decreased immensely. I started teaching in 1972, and I now have 45 hours per year less with my students than I did at that time. (Large Urban District, High School, Mathematics Teacher)*

A third concern Kansas interviewees mentioned was that they felt conflicted between what they needed to teach their students so that they would do well on the state test versus what was developmentally appropriate. One aspect of this involved teachers having to adjust the sequence of the curriculum in order to accommodate the timing of the test, which could mean exposing students to content before they were ready. The following quotation illustrates this concern:

> *It is necessary to expose [students] to a lot more information because they will be tested on it....Your tradeoff [is between] instruction in the skills that they're lacking and instruction in things you know they are going to see on the test..And there are good effects from that exposure. They do become exposed to... not necessarily proficient at, but exposed to a wider range of information and topics, and some of that sticks...so they become a little broader in their knowledge base....But you are also taking time away from instructing those...skills that they don't have at all....that they never picked up along the way....Whether they never really understood vowel sounds or they never really understood subtraction, there are certain skills that [they] didn't pick up at the right time, and that's why they're in special education. (Small Urban District, Elementary School, Special Education Teacher)*

A fourth concern focused on the perceived impact on tested versus non-tested grades. At the time of this study, the subject areas covered by the Kansas Assessments were each tested at certain grades and not at others. Interviewees described this arrangement as resulting in a more crowded curriculum at the tested grades and a lighter load at the non-tested grades. Non-tested grades were considered to be "light" years because the perceived negative effects of test preparation (i.e., cramming, rushing content, teaching material that was developmentally inappropriate) did not significantly affect them. Teachers described the effects on their instruction as follows:

> *I'll [ask the other teachers], 'Do we have a writing assessment this year?' Not that I won't teach writing [if there is no test], but I won't feel that I have to teach [only*

> *the particular writing style that will be tested]. I can do a persuasive essay, I can do a narrative [essay], and so on. If I know there is a test and its going to be comparison/contrast or persuasive or whatever, that's what I focus on. That's all we do. (Small Urban District, High School, English Teacher).*

> *When I go back to fourth [grade] next year, I will probably do a lot of science and math integrated with everything, whereas this year I did a lot of reading and writing. The [state] test [that students will be] taking at the grade level you're teaching most definitely affects instruction all the way around. And you might shortchange science if you have to focus on reading. (Rural District, Elementary School, Fifth-Grade Teacher)*

These quotations illustrate that the aphorism "you get what you test" applies not only to the subject areas tested or not tested by the overall testing program, but also to those tested or not tested at a particular grade level.

School-Type Differences

Like the state standards, the state test had the most dramatic impact on the elementary curriculum. Elementary teachers were twice as likely as middle school teachers and three times as likely as high school teachers to say that they removed topics from the curriculum in order to prepare for the test (three-quarters of elementary, one-third of middle, and one-fifth of high school teachers mentioned doing this). They also were the most likely to note the negative effects of these omissions. At the same time, elementary teachers were more likely than middle or high school teachers to say that preparing for the test reinforced important skills and had improved the way they taught, with about one-third of elementary teachers mentioning positive effects in both of these areas. An elementary school principal reflected on these effects as follows:

> *The state assessment... is driving instruction. It is changing the way we teach. Before, we were pretty well textbook-oriented and knowledge-based. And now, the state assessment is telling us we have to teach more problem solving, more thinking skills. It's not necessarily bad, it's difficult in that we really have to train teachers to teach that higher-level thinking. [That is] completely different from, say, 25 years ago when I taught fourth grade. [Then] I opened the book and said, 'Now turn to page 18,' and 'Johnny, read.' Now teachers really have to put in a lot of preparation time, because they have to present the material and then do higher-level questioning and hands-on activities, which is good. (Rural District, Elementary School Principal)*

Elementary teachers also were the most likely to note the importance of familiarizing students with the format of the state test in order to make them more comfortable with the test-taking experience.

The state tests seemed to have the least impact on what or how high school teachers taught, with about one-fifth of them noting no effects in either area (compared with less than one-tenth at the elementary and middle school levels). However, a few high school teachers mentioned tailoring the curriculum to help prepare lower-level students for the test. For example, a high school mathematics teacher remarked:

> *I don't do much [test preparation] with my advanced classes, and I try to do more of it with my lower level. I justify that because three of the five classes I teach are at the college level....I think that's my first responsibility — [to] prepare them for the next college course. (Small Urban District, High School, Mathematics Teacher)*

As illustrated by this quotation, one of the main reasons for this lesser emphasis on the state test at the high school level was that these students are about to move on — either into the workplace or to higher education. Therefore, high school teachers' first priority was to prepare them for that stage in their lives.

District-Type Differences

There were some distinct differences in what educators in the four districts reported doing to prepare their students for the state test. Educators in the large urban and rural districts were the most likely to note that preparation for the test influenced their instructional and assessment strategies, with almost two-thirds in each district noting specific effects in these areas. In addition, teachers in the large urban district seemed to be the most frequent users of the scoring rubric from the state test in their classroom assessments. Echoing the district/state alignment issues discussed in Section Two, about one-fifth of the educators in this district mentioned that the district and state tests were either not aligned or in the process of being aligned. The most progress had been made in the area of writing, where the district and state tests had been combined into one assessment at the eighth grade.

Educators in the rural district were the most likely to say that they removed topics from the curriculum in order to prepare for the test, with about two-thirds of them reporting such changes (twice as many as in any other district). The removal of topics seemed to be linked to the misalignment between local and state curricula (as discussed in Section Two) since rural educators also were the most likely to mention that they added topics covered by the state test to the curriculum. In the following quotation, a rural principal reflects on some of the reasons behind these changes:

> *What the scores told us last year is that our curriculum is not aligned....That came screaming through and the high school math scores were just terrible. Last year [the high school] declared that everybody who came in starting next year would be taking algebra as a freshman, because the tenth-grade test is actually an algebra-II test. So with or without pre-algebra [students have to take] algebra I. The implications for our middle school are horrific, because we had three levels of math last year; this year we don't. That is an example of how performance on the state assessment determined a curriculum change [with strong] implications...down here at the middle school, and we were never consulted. [We were told] 'We're doing this, so you better do what you need to do to have them ready.' (Rural District, Middle School Principal)*

Educators in the suburban district talked about a more scaled-down version of curriculum changes, and these were more likely to be attributed to the state standards than the state test (one-seventh of the suburban educators said the test had no impact on what they taught and one quarter said it had no impact on how they taught). At the same time, suburban educators indicated that test preparation had become a part of classroom activity. The quotation from a suburban elementary teacher at the start of Section Three offers one example. Other comments included the following:

> *Starting in the fall, we really focus on this. Every day or two we'll throw a problem on the board when the kids come into class that is one of the state objectives...and just take...about 5 minutes to go over that problem. I know we've had some teachers that have gotten a copy of [another state's] test...and they've worked those problems....We have one teacher who does Friday quizzes. [He gives students] a topic for the week, maybe addition of fractions, and he'll just give a quick five-question quiz on Friday on [that topic]. (Suburban District, High School, Mathematics Department Chair)*

Overall, Kansas interviewees noted both positive and negative effects of the state test on what and how they taught. As will be seen below, Michigan interviewees had even more to say in these areas.

Michigan

> *When I teach, I put a lot of things into my lessons that will prepare students for the tests, and I remove a lot of the project-type activities such as writing plays, writing poetry, performance activities. Now we don't do a lot of that because we're concentrating on preparing for the tests. (Large Urban District, Middle School, Eighth-Grade English Teacher)*

> *We [must] prepare the kids to take their place in society. They don't understand checking accounts. They don't understand interest. They don't understand taxation or financing an automobile. We used to teach business math, we used to teach consumer math, and we used to teach some things that had real-world relevance. We eliminated all of that. (Small Urban District, High School, Mathematics/ Science Teacher)*

> *In reading...we do some comprehension [activities] and emphasize vocabulary words....Sometimes I feel stressed trying to review everything before the test because I want the students to be as prepared as they can be, giving them as many advantages as I can. (Rural District, Elementary School, Fourth-Grade Teacher)*

Overall Findings

As in Kansas, Michigan educators indicated that they were engaged in varying degrees of *removing, emphasizing,* or *adding* curricular content in order to prepare for the state test. The removal of content was the most frequently reported activity. The addition of content was least frequently reported. In addition to these changes in *what* was taught, the majority of Michigan interviewees felt that preparation for the state test had produced changes in *how* they taught. Overall, Michigan interviewees reported more changes in these areas than did those in Kansas. Michigan teachers also were more likely to note that the state test affected how they assessed their students (e.g., using multiple-choice and open-response questions that mirrored those on the state test, as well as giving students practice with released questions from that test). In all, only one in ten felt that the state test had no effect on what was taught and one in five said it had no effect on how things were taught.

As exemplified by the three quotations above, Michigan educators' reactions were mixed. For example, about one-fifth of the interviewees felt that these activities improved or reinforced important skills. As in Kansas, writing, literacy, and critical thinking were most often mentioned in this regard. A typical comment includes the following:

> *Our curriculum is set up around the MEAP....Looking at the lower levels and how writing was taught and what we're expecting now...there's a change now in what third graders are doing...away from grammar and mechanics...to actually writing....I think [these changes] are for the better. (Rural District, Elementary School, Fifth-Grade Teacher)*

Others noted positive effects that were related to the format of the MEAP tests. Specifically, interviewees mentioned that giving students similar types of questions in class could help them practice higher-level thinking skills.

At the same time interviewees noted negative effects such as an increased focus on breadth rather than depth of coverage in classroom instruction (mentioned by one-tenth of the interviewees). For example, one teacher remarked:

> *[I'm] trying to skim along very shallowly on the surface of things, not teaching toward mastery, but just getting students exposed...and hoping they'll remember the difference between Fahrenheit and Celsius for one test question. (Small Urban District, Elementary School, Fourth-Grade Teacher)*

As in Kansas, interviewees viewed the breadth-versus-depth problem as partly the result of having to teach literacy skills in addition to the subject area content that would be tested.

Others noted the distorting effect the tests had on the sequencing of the curriculum, again a theme that surfaced in the Kansas interviews. A teacher described the impact on her curriculum as follows:

> *Some of the things [on the test] are out of sequence for us....When you talk about math, the MEAP is given mid-year [and] after the Thanksgiving break I have to start reviewing some of those math areas that I won't [officially get to] until the end of the year....So for math, especially, it means [treating all of this material] out of context so that you can practice with [the students]. It takes a lot of time. (Suburban District, Elementary School, Fourth-Grade Teacher)*

Even an emphasis on writing and analytical skills could be viewed as having a negative impact on the classroom environment when it detracted from other desired classroom activities. In addition to the previously mentioned breadth-versus-depth problems, some viewed the emphasis on these skill areas as getting to the point of "drudgery." One English teacher remarked:

> *I would like to be doing other things, like working more with public speaking....We [once] did some things like demonstration speeches. [But] we kind of put that aside last year because we had to spend so much time on testing....My style of teaching just seems to be constantly, constantly, constantly reading, talking about main ideas, characters and all these things that are covered in these tests. It gets to be drudgery after a while. It's all you're doing. And writing skills...I do those things but I'm so conscious in the back of my mind that the test is coming, the test is coming. (Large Urban District, Middle School, English Teacher)*

As in Kansas, there were differences in terms of the perceived impact on teachers of tested versus non-tested grades or subject areas. Teachers at tested grades were most likely to report removing topics not covered by the test and emphasizing or adding topics that would be tested. Similar to their Kansas peers, they were the most likely to mention rushing content and teaching material that was developmentally inappropriate. At the same time, teachers of non-tested subjects or grades sometimes reported that they adjusted their curriculum to focus on content or skills that students would be tested on, either at a later grade level or in another subject area. A mix of incentives seemed to be operating here. In addition to wanting to help students perform better on the MEAP, these teachers sensed that their subject might be viewed as less essential if it did not aid achievement on the state test. For example, a business teacher remarked:

> *I know in our system here the testing average has been pretty low, and I think because more emphasis has been put on the test, it has forced us...to change some things....If there were going to be some cuts, the business department would probably be one of the areas to be cut. As far as the MEAP is concerned, we're not one of the content areas [tested], so...we had to...look at where the students are having trouble on the MEAP, and a lot of it is the [open-response questions where] they have to figure out how they get the answers. What we thought was, as a business department, when there's a science or math question...we could help by incorporating that and helping students figure out the response process, how you show your thinking and your work. (Large Urban District, High School, Business Teacher)*

While some educators mentioned using test results to help inform future instruction, others — particularly those at non-tested grades or of non-tested subject areas — said that they would like to receive results for individual students or classes, but did not know whether this information was available. Thus, there was a feeling of working in the dark, of not knowing whether strategies were working. As the business teacher quoted above noted: "The test results for the school are okay to see how we're doing as a whole, but [they don't] give the teachers the individual feedback to see [whether] their lessons are improved or having any effect." As in Kansas, about one-fifth of those who did receive test results said they came back too late to be useful.

School-Type Differences

As in Kansas, elementary educators in Michigan were the most likely to note the impact of the state test on classroom practice. Elementary teachers were twice as likely as middle and high school teachers to mention emphasizing topics that were covered by the state test (one-quarter versus about one-tenth at the other two levels). Teachers at the elementary level also were more likely to note that topics covered by the state test were added to the curriculum or taught close to the test administration. While teachers at all grade levels talked about teaching test-taking skills, elementary teachers were the most likely to say that they used commercial test preparation materials (almost one-fifth of these educators, compared with less than one-tenth at the other two levels), and that their teaching strategies changed after the test (almost one-fifth compared with less than one-tenth at the other two levels). As exemplified by the following quotations, elementary teachers' perceptions of these activities varied:

> *Basically, our curriculum is changing to the point that we're teaching to the test. All these things we have to get in before that test, we don't have time for the extras. We don't have time for spending a few extra days on a topic that the kids find interesting. We just don't have time. (Rural District, Elementary School, Fifth-Grade Teacher)*

> *As a teacher, I include more test-taking skills....I start, probably from the first week of school, [with] integrated, oral types of activities, and teaching high-level and low-level distractors....I think it's extremely important that they know those concepts because in Michigan, by the time they get to the [upper] grades, knowing how to handle testing documents is worth money to them. (Large Urban District, Elementary School, Fourth-Grade English/Social Studies Teacher)*

Elementary educators also voiced another theme: these tests changed the very nature of elementary education:

> *Well, you know kindergarten was never so academic. I mean...children learn a lot through play and socialization and they don't have the amount of time to develop that because [learning now is] so curriculum driven....There's not enough time. Five-year-olds take an enormous amount of time in their discovery process. It takes them a while to segue into the activity....Just when they're really excited and interested in what they're doing, boom, it's time to move onto something else because I've got to get everything in. (Large Urban District, Elementary School, Kindergarten Teacher)*

As suggested by these comments, elementary teachers in Michigan felt more negatively than positively about these changes. In fact, they were twice as likely as middle or high school teachers to report that preparing for the state test negatively affected the quality, pace, and developmental appropriateness of their teaching.

District-Type Differences

As in Kansas, educators in the large urban and rural districts most often mentioned engaging in test preparation activities. The intensive and time-consuming nature of these activities, which can begin on the first day of the school year, is illustrated in the following comments, one from a teacher in the rural district and the other from an assistant principal in the large urban district:

> *I would say pretty much the entire fall up until the MEAP test [is spent preparing students for the test]. A lot of kids come back after a long summer and they've forgotten...a lot of things....I'd say we spend a month just reviewing what they should have learned by now but maybe have forgotten....We spend three or four weeks preparing them for the MEAP test itself....The reading specialist and the math specialist come in...with specific ideas that will help on the MEAP test. [The principal] might come in for a pep talk or two. By the time the kids take the MEAP test, they know it's serious....They know they need to do their best. (Rural District, Elementary School, Fourth-Grade Teacher)*

> *Normally, [preparation for the test] should start at the beginning of each school year. Teachers should start incorporating test-taking strategies in their lessons. After we've assessed our previous test scores, teachers know what they need to work more on and they need to emphasize those particular skills in their lesson plans and then come up with strategies to present them to the children....So initially, it should start in August, when we come back to school. (Large Urban District, Elementary School, Assistant Principal)*

Teachers in both districts appeared to be the most frequent users of the state scoring rubrics in their classroom assessments (see Box 10 for an example of a scoring rubric from the fifth-grade MEAP writing test). Both also mentioned using the question formats (multiple-choice and open-ended) from the state test in their classroom assessments.

Educators in the large urban and rural districts were about twice as likely as those in the other two districts to note that this type of test preparation could improve or reinforce important skills. At the same time this opinion was voiced by a minority (one-quarter of the large-urban-district educators and one-third of the rural educators) and several acknowledged the deleterious effects of the test on their curriculum. Below, a science teacher in the large urban district compares the learning environment in his classroom before and after he had to prepare his students for the MEAP, and concludes that the former was better.

Box 10

Example of a Scoring Rubric

The following rubric accompanied the fifth grade MEAP writing prompt shown in Box 9:

Here is an explanation of what readers think about as they score your writing.

4 Central ideas may be clearly developed through details and examples. The writing may have a natural flow and a clear sense of wholeness (beginning, middle, end); the organization helps move the reader through the text. A clear and engaging voice is likely to be demonstrated through precise word choice and varied sentence structure. Skillful use of writing conventions contributes to the writing's effect.

3 A recognizable central idea is evident and adequately developed. The writing has a sense of wholeness (beginning, middle, and end) although it may lack details or have extraneous details which interfere with unity. Appropriate word choice and variable sentence structure contribute to the writing's effectiveness. There may be surface feature errors, but they don't interfere with understanding.

2 The writing shows a recognizable central idea, yet it may not be sustained or developed. There is an attempt at organization although ideas may not be well connected or developed; vocabulary may be limited or inappropriate to the task; sentence structure may be somewhat simple. Surface feature errors may make understanding difficult.

1 The writing may show little or no development of a central idea, or be too limited in length to demonstrate proficiency. There may be little direction or organization, but, nevertheless, an ability to get important words on paper is demonstrated. Vocabulary and sentence structure may be simple. Minimal control of surface features (such as spelling, grammar/usage, capitalization, punctuation, and/or indenting) may severely interfere with understanding.

Not ratable if:

off topic

illegible

written in a language other than English

blank/refused to respond

> *Before, I did stretch the top [students], but when I have the responsibility of these kids taking the MEAP, I have to make sure they understood this before they move on. If I didn't have the MEAP that teaching approach would change. I would probably shorten the review and move on, but because I know this is the curriculum and this stuff is going to be on the MEAP, I've got to be sure they know this....I would feel better if I could spend less time on revision. I'd feel that I was pushing them harder. Before, I really stretched the top...I had some poor readers of course and so I had this table where...my weak readers [were together] and the rest of the class was sort of surrounding them. I was always supplementing, always giving them a lot more....And that worked real well...and my [standardized test] scores...every year there was an increase. My kids did really well. (Large Urban District, Elementary School, Science Teacher)*

As in Kansas, educators in the suburban district were the most likely to say that the state test had no impact on how they taught. In fact, about one-third of these interviewees — twice as many as in the other districts — voiced this opinion. Some reasons given included an already high-quality curriculum, staff development opportunities, and a tradition of good educational practice at the school that removed the need for teaching to the test. One teacher remarked:

> *We really don't teach to the test. We may give students the opportunity to fill in answer boxes [so that they are familiar with the format of the state test]. I do know that there are some schools that push hard to teach to the test. We're really lucky because we do all sorts of staff development opportunities to help the staff feel comfortable with the curriculum. (Suburban District, Elementary School, Second-Grade Teacher)*

Another reason for the relatively lesser impact on the suburban curriculum emerged during an interview with a high school science teacher. This teacher mentioned that he did not focus on the MEAP because most of his students opted out of taking it. This apparent trend was corroborated by some of the large-urban-district interviewees, who also remarked that in their district parents were less likely to know about or exercise that option. Thus, lower-performing students in this district were more likely to take the state test and have their scores factored into the school average.

While suburban educators were the least likely to say that the state test affected how they taught, they did feel that it was affecting classroom practice in terms of class time devoted to test preparation and test administration. The overwhelming feeling was that this detracted from, rather than added to, the quality of teaching and learning since it encroached on time that could more usefully be spent on other things. One suburban educator remarked:

> *[Our district] is lukewarm about the whole MEAP in terms of its validity. As a district, we do fine on the MEAP...we tend to score at the top of the schools....But, we'd rather be focusing on other things...than proving to somebody that the kids [have] got the baseline information. It takes up a huge amount of*

time. It disrupts virtually two weeks of school. For some students it's traumatic. It's just a horrible thing to have to pay attention to when you're wanting to do more substantial kinds of things. (Suburban District, High School Principal)

Overall, Michigan interviewees reported more test-related impact on what and how they taught than did Kansas interviewees. One possible reason for this is that the Michigan testing program has been in place longer and thus has had more time to produce changes. Another is the stakes attached to the test results. As will be seen in the next section, Massachusetts interviewees reported similar types of effects on classroom practice as did their peers in Kansas and Michigan, but far more of them.

Massachusetts

I've gone to a more intensive approach to writing, developing essays, and open-response questions. The number of multiple-choice questions on my tests is now minimal. Last year, we spent one class period, usually Monday morning first period, working on open-response questions. After everyone had done one, we modeled what a good question would be, what a good response would be. This year, I'm stressing [to] the students [to make] an outline of what they want to say before they jump in and start writing, which was something I hadn't done before. (Small Urban District, Middle School, English Teacher)

In ninth grade [the students would] take a survival course, so we would actually...teach them how to build survival shelters and things....These were thematic courses designed to fit our uniqueness and our communities. We're very rural, so survival skills are important. That course is gone. We had 'Our Town' where we...actually integrated history and research techniques with the communities around here, so that when the kids moved out they really understood their roots. We had 'Troubled Times' for freshmen, which...helped students deal with adolescent issues, you know, drug abuse, sex, all those things. The course is gone....We had some wonderful, wonderful programs....These are the things that MCAS has killed. (Rural District, Middle/High School, English Department Chairperson)

Most students at this school are strongly encouraged to take one or more of three elective MCAS prep courses, either the full-semester course in ninth grade, or one of the two nine-week strategies courses in sophomore year, focusing on writing and math. I would say 50 to 60 percent take the ninth-grade course; 60 to 70 percent take either the writing or math strategies courses. (Small Urban District, High School, Mathematics Teacher)

Overall Findings

Massachusetts interviewees were the most likely to note that preparation for the state test was occurring in classroom lessons and that it became more specific or intensive as testing time approached. Part of this preparation involved the aforementioned activities of *removing, emphasizing,* or *adding* curriculum content. Writing was noted by many as an area that received special emphasis in the curriculum, although — as indicated by the first quotation above — this was tied in with preparing students to answer specific question types on the MCAS. In addition to these changes in *what* was taught, almost 100 percent of the Massachusetts interviewees — more than in the other two states — noted that preparation for the test influenced *how* things were taught. They were far more likely than their peers in the other two states to report that preparation for the state test affected classroom assessment. About one-fifth reported using the scoring rubrics from the state test in classroom assessments, one-third mentioned using open-response questions that matched the state test in format, and one-third mentioned using actual released questions from the test. Less than one in ten felt that the state test had no impact on what or how they taught.

One reason for the greater amount of test-related activity in Massachusetts was the stakes attached to the tenth-grade test results for students. Since students must pass the tenth-grade English and mathematics tests to receive a high school diploma, teachers at this level in particular felt pressured to give these areas more time and emphasis. One educator remarked:

> *During the school day we've extended the time on learning in both English language arts and mathematics because those are the focus of MCAS. Bottom line is you either pass those or you don't graduate, so we've extended the time on learning. (Large Urban District, High School Principal)*

Massachusetts interviewees perceived both positive and negative aspects of these curricular and instructional changes. For example, Massachusetts interviewees, similar to those in the other two states, noted that the teaching of literacy had gained new importance as a result of having to prepare students for the state test, and several felt that students' writing and critical thinking skills had improved as a result. The following comments illustrate these perceived improvements:

> *I think that the writing component of the MCAS math test has really improved math instruction. Where in any other career field [are you not] going to have to justify your work in words? If students can't explain what they found and how they found it to a boss in a report, what service am I doing them? Any time in life you need to be able to explain yourself...to justify what you have said, what you have done, and what you believe. So I think the writing component in all classes — because I really think the MCAS has deepened writing in all classes — is a terrific benefit of the test. (Large Urban District, Middle/High School, Mathematics Teacher)*

> *I think that the thing that's moving teachers is the realization that MCAS doesn't ask the [same] old rote question[s]. It asks thinking questions. If children are going to learn how to think, they need to do active learning. They need to be able to visualize a problem. They need to be able to think [about] how to research an answer. And you don't get [to that point] unless you're presented [with] a problem and you have more than one solution. (Small Urban District, Middle School Principal)*

Overall, about one-third of Massachusetts interviewees felt that preparing for the MCAS reinforced important skills. Running through some of the comments in this area were the perceived economic benefits of standards-based reform in terms of producing employable graduates. The first quotation in the above block is a case in point.

While interviewees perceived these test-related activities as useful, there were concerns. For instance, while writing activities had increased — and this was seen in a mainly positive light — some feared that students were being taught to write in limited, formulaic ways. An English teacher described the impact on her instruction as follows:

> *We do a lot of essays, just drafting. I sometimes don't even take it to final draft, just [concentrate on their] knowing how to respond to those essay direction words, formulating a thesis. (Rural District, Middle School, Eighth-Grade English Teacher)*

Almost one-fifth of Massachusetts interviewees — twice as many as in Kansas or Michigan — felt that the pace required in preparing for the test was developmentally inappropriate and that it forced them to focus more on breadth than on depth of coverage. One mathematics teacher remarked:

> *If I didn't have the MCAS hanging over my head, and if the kids didn't...I could teach [the] material more indepth. I could take these frameworks and really explain the connections [in the] content. The MCAS is such a constraint on my time that I can't always get my students to see the deep connections, even though I try to show them and approach the topic [in] different ways. The framework strands would absolutely support in-depth teaching; however, the test doesn't allow [that]. (Large Urban District, Middle/High School Mathematics Teacher)*

This comment is additionally interesting because it illustrates the interactions between the state standards and the state test. Specifically, the interviewee notes the constraint placed on his ability to implement the state standards due to the pressure to prepare his students for the state test. This is the more troubling because it prevents the teacher from capitalizing on the teaching and learning opportunities embedded in the state standards. It can also produce a one-size-fits-all approach to teaching. The quotation below is from a teacher who is struggling with this issue in her own teaching.

> *The frameworks are very good in that we now know specifically what it is we're supposed to be teaching at each grade level. And that is great to know...but the bad thing about it is, what do you do with the child who has not grasped the skills that were supposed to be taught in third grade and comes into fourth grade, and you know you are responsible for the whole curriculum taught [there]....Do you just put [these children] aside and go on with the fourth-grade curriculum? Or do you take the time to make sure the child has the third-grade curriculum cemented and has mastered the skills, which will slow up your fourth-grade curriculum? And you know there are high stakes involved, you have to complete this curriculum before the end of the year because you're going to be tested on it. So you're caught between a rock and a hard place. I feel lately in my teaching that I'm not teaching mastery to children, and [that] I'm doing them a disservice. (Large Urban District, Elementary School, Fourth-Grade Teacher)*

About one-third of those interviewed (twice as many as in Kansas or Michigan) noted that the removal or de-emphasis of content in order to prepare for the test had a negative impact on teaching and learning, particularly in terms of making it difficult to cater to students' individual strengths and interests (another aspect of the one-size-fits-all theme). As exemplified by the quotation below, concerns about the removal of content emerged at all levels of the system and in the suburban as well as other districts.

> *On the negative side, I think there's a serious danger that [the MCAS] can, by its nature, lead to a narrowing of the curriculum and [make] teachers...feel they have to put more time into just the high-priority subjects, when [these] may not be areas of greatest need for particular students. There may be students who are very much interested in the arts...and yet the arts, not being measured by MCAS or other high-stakes tests, are given short shrift by teachers. (Suburban Superintendent)*

As in Kansas and Michigan, there were mixed opinions on the usefulness of the test results. While some felt that they helped teachers to identify weak areas in the curriculum, thus helping to improve teaching and learning in these areas, others felt that the results were "useful, but only for test taking."The teacher who made this comment went on to explain:

> *We noticed that a lot of our kids did poorly on the open-response [questions], so we looked at those questions to see what they were asking kids to do. We decided that maybe before the test we should practice those questions a little more....Other than that, I don't try to let [the MCAS] affect my teaching at all or my thinking about my kids. (Large Urban District, Middle School, Fifth-Grade Teacher)*

In addition to these kinds of concerns, about one-fifth of the teachers noted that the test results came back too late to be useful.

School-Type Differences

Elementary teachers were twice as likely as middle or high school teachers (four-fifths versus two-fifths at each of the other two levels) to mention that they removed topics from the curriculum in order to prepare for the state test and that this had a negative impact on classroom practice. In addition, elementary teachers were more likely to note that the test reduced creativity in learning activities (two-fifths versus one-quarter at the other levels) and that preparing for the test created a developmentally inappropriate pace (one-fifth versus one-tenth at the other levels). While the following illustrative quotations are from teachers in urban schools, these concerns surfaced among elementary teachers in every district where we interviewed:

> *I feel as though I have so much to teach in a certain amount of time...[that] there are more things I'm responsible for in the classroom. So it seems [as if] all the fun things, all the things that were nice as an extra to get the point across, are put on the back burner because...I need to get a certain amount done, because now I'm responsible for their learning X amount in this amount of time. (Small Urban District, Elementary School, Sixth-Grade Mathematics and Science Teacher)*

> *The test is there from the moment you step into the classroom at the beginning of the year, till it's over at the end of the year. You wouldn't believe the level of fun and of real teaching that goes on right after the MCAS tests are taken. All of the pressure is off, you can now get into things that you really like to [do]. (Large Urban District, Elementary School Teacher)*

> *As for doing specific activities that are designed for MCAS, teaching kids how to write a five-paragraph essay in the fourth grade is strictly MCAS [and is] not related to anything else I do, curriculum-wise....And that I added in after the first year that I taught....We used to do long writings but I never formalized it into how you write an introduction, paragraph, [and so forth]. So I went back to how I was taught how to write in high school [and drew on] how I learned how to write to teach them how to write; and that was not something I was expecting...to teach fourth graders how to do. (Large Urban District, Elementary School, Fourth-Grade Teacher)*

Some additional themes in these quotations echo those heard in Kansas and Michigan. First, there is the greater burden on elementary teachers due to having to teach the standards in all subject areas as opposed to one or two. Second, there is the sense that these tests are changing the very nature of the elementary curriculum.

Elementary teachers also were the most likely to note that they used state-released questions from the test (two-fifths versus one-seventh at the middle and one-quarter at the high school levels) as well as the state scoring rubric (one-quarter versus one-seventh at the middle and one-fifth at the high school levels) in their classroom assessments. The scoring rubric was viewed particularly positively, mainly because it enhanced the teaching and learning process. A fourth-grade teacher remarked:

> *I have to honestly say that there is one good thing about MCAS, and that is the rubrics for writing because now...you can home in on where [the students really are], I like that part of it, I do like the rubrics....My [teaching] partner and I developed a children's rubric that we now use, which correlates with the state...ones, and now that's helpful, because you can give them a 1,2 or a 2, 1...and it's specific. They have the rubrics at home and they can take a look and say 'well, the reason why I got the 2 is that I listed things rather than expand them, I didn't use flavorful language, I didn't stick to my topic area.' That part of it is good. (Large Urban District, Elementary School, Fourth-Grade Teacher)*

While middle and high school teachers reported fewer MCAS-related effects on their curriculum, they reported far more than did their peers in Kansas or Michigan. These effects were differentially dispersed across subject areas, with teachers of English and mathematics reporting more impact than teachers of other subjects. This is understandable since tenth-grade students must pass the state test in both English and mathematics in order to receive a high school diploma. Thus, these teachers felt considerable pressure to make sure that their students were taught the content and skills that would be tested. A phenomenon that emerged only in Massachusetts was the addition of MCAS-preparation courses to the high school curriculum. A high school teacher described the course offerings at his school as follows:

> *Most students at this school are strongly encouraged to take one or more of three elective MCAS prep courses, either the full-semester course in ninth grade, or one of the two nine-week strategies courses in sophomore year, focusing on writing and math. I would say 50 to 60 percent take the ninth-grade course; 60 to 70 percent take either the writing or math strategies courses. (Small Urban District, High School, Mathematics Teacher)*

The social studies test seemed to have the least impact on classroom practice at the middle and high school levels. Teachers of this subject area said that they did not want to gear their classroom practice to this test because the nature of the questions would force them to emphasize memorization in their instruction. Instead, these teachers felt that the test should be redesigned to focus on critical thinking.

District-Type Differences

Like their peers in Kansas and Michigan, educators in the large urban and rural districts reported the greatest test-related impact on classroom practice. Educators in the large urban district tended to be the most positive about the impact on classroom practice, with almost half of them mentioning that preparation for the MCAS improved or reinforced important skills. The following illustrative comment is from a large-urban-district science teacher who described herself as "a student of the MCAS:"

> *I do a lot more...making them write, a lot more critical-thinking stuff, a lot more hands-on stuff....It has brought to my attention [that] I could be doing better....I have become a student of the MCAS test. I know what they ask...or I've gotten a feel for what they're asking so that I try to concentrate [on that]....I make my eighth-grade classroom almost an MCAS review year....[The students] come in at all different levels so during eighth grade they get exposed to a little bit of everything I started to see repeatedly on the test. So I have responded to the MCAS in trying to make sure that the vocabulary I use and the stuff I cover in class is in line with the test....That's actually been helpful. (Large Urban District, Middle School, Eighth-Grade Science Teacher)*

Educators in the large urban district talked about intensive test preparation activities, particularly at the high school level where the stakes for students are highest. For example, a mathematics teacher in this district explained:

> *We have school-wide testing once a week. We usually have the compare/contrast-type question, we have the open-ended question, and we have the multiple-choice. The whole school shuts down during that period. It's a two-prong thing; it gets our students test-wise, so to speak, and it also points in the direction that this test will hold them accountable. (Large Urban District, High School, Mathematics Teacher)*

Educators in the large urban district also talked about increased intensity in these test preparation activities as the test approached.

> *But right before the test, that's...almost four weeks before, [all] we're doing is practice. But yes, [even at the] beginning of the year, you constantly mention [that] this is the type of question [that is] on the MCAS...this is [like] the last question on the MCAS...these are always on the MCAS. You start that from day one in eighth grade, and even in the seventh grade. (Large Urban District, Middle School, Eighth-Grade Mathematics Teacher)*

These test preparation activities were not always viewed positively by those in the large urban district. For example, over one-third of these interviewees noted that preparation for the test forced them toward breadth rather than depth of coverage.

Rural educators tended to have the most negative attitudes about the impact of the state test on classroom practice. For instance, many of these educators were concerned about the loss of topics specific to their local area. The tenth-grade English teacher cited earlier (at the start of the Massachusetts section) described some of these losses, for example a survival course and courses on local history.

Rural educators also were more negative about the pedagogical changes produced by the MCAS. For example, over three-quarters of them felt that the test reduced creativity in teaching activities (compared with no more than one-fifth of the educators in the large urban, small urban, and suburban districts). They also were more likely to point out inconsistencies between the curriculum standards and the MCAS. One teacher remarked:

> *The math and science [frameworks] seem to have a lot of...life in them. As I read them, there are places where I can see how to turn that into an activity with my third graders. And the test...seems very flat [in comparison], particularly for young children....I feel like the test is just sort of...an end game. (Rural District, Elementary School, Third-Grade Teacher)*

Box 11

Teaching to the Test, Preparing Studen

In some follow-up interviews, we asked teachers and principals to clarify three terms that came up in the original interviews. These terms were (1) teaching to the test, (2) preparing students for the test, and (3) teaching to the standards or frameworks.

Teaching to the Test

This phrase had very negative connotations for most interviewees and was characterized as usually occurring among teachers of tested grades or subjects. Activities that matched this term included going over the actual test or questions from the test with students; using modified versions of test questions as practice in class; gearing everything in the classroom toward the test; matching released test questions to units in the state standards and then emphasizing those units in class; and taking older tests and giving them as practice. Many of these activities were seen as separate from the regular curriculum and detrimental to it. At the same time, there were differences in terms of which were viewed as poor educational practice versus outright cheating. For example, gearing everything in the classroom toward the test was seen as the former while having advance knowledge of the test questions and making sure that class work reflected them was cheating. All of these "teaching to the test" activities were characterized as taking time away from the regular curriculum, and none of the follow-up interviewees admitted to any of them. In contrast, educators from the initial round of interviews talked quite often about teaching to the test.

Preparing Students for the Test

Interviewees' responses were very consistent as to what constituted preparing students for the test. Mainly this meant familiarizing students with the format of the test and teaching them test-taking skills — e.g., making classroom tests look like the state test, and showing students how to deal with multi-mark questions (those with more than one correct answer), or how to make a good guess if they don't know the answer, or how to follow directions. The aim was to make students feel comfortable and mentally prepared to take the test. Some educators felt that preparing students for the test also involved emphasizing skills such as reading and writing so that students would better understand and answer the questions. It is noteworthy that preparing students for the test was seen as different from teaching to the test in that it was not detrimental to the regular curriculum. In particular, while teaching to the test was viewed as striving only for high test performance, preparing students for the test was believed to involve more transferable skills and knowledge. Still, the distinction between the two activities was not always clear among the follow-up interviewees, and often equally unclear among first-round interviewees.

Teaching to the Standards

Follow-up interviewees in all three states were most likely to say that they taught to the state standards, using them as guidelines for their classroom curriculum, with the content and skills to be covered broken down into units and daily lesson plans. Teaching to the standards was seen as different from teaching to the test and preparing students for the test; it was more curriculum-focused and produced student learning that was generalizable to other tasks and contexts. At the same time, many first- and second-round interviewees felt that the presence of the state test compromised their ability to teach to the standards. For example, teachers had to make time to familiarize students with the test format and to teach them test-taking skills. They also had to carve out review time to make sure that earlier content was remembered. These activities took time away from teaching the state standards. Below are examples of comments made by interviewees during the first round of interviews that exemplify "teaching to the test," "preparing for the test" and "teaching to the standards" activities.

Teaching to the Test

(1) Using modified versions of test questions, or taking released questions from older tests and giving them as practice

I actually include published test items in all my tests. I count them more heavily than I would some questions I made up, which I feel is kind of wrong, but I need to prepare them, that's my job. I don't want them to go in cold to the state test. (Massachusetts, Large Urban District, Middle School, Eighth-Grade Science Teacher)

(2) Gearing everything in the classroom toward the test

We did a lot of training and preparing [of] the kids since

r the Test, and Teaching to the Standards

January. We did review packets. We had Friday math days just because of this one test....I felt I was a math teacher from January until spring break. We had to drop other curriculum areas because of this. [We dropped] spelling, writing....We couldn't drop the science because we had a science assessment coming up at the same time. (Kansas, Suburban District, Elementary School, Fourth-Grade Teacher)

We've done away with silent reading, and that was here for years, and now we're doing MCAS prep instead...and the eighth-grade team is busy right now putting together a packet that they will give to all of the students and all of the teachers. So...everything is pretty much MCAS-driven, MCAS-driven, MCAS-driven. (Massachusetts, Suburban District, Seventh-Grade Social Studies Teacher)

Preparing Students for the Test

(1) Familiarizing students with the test format

Kansas [has a] multiple-mark [test] where there is more than one right answer....Once you find a right answer, you can't stop and move on, you have to keep looking....So [learning how to take this test] is something we work on. We definitely expose [students] to that, because it's a totally new concept to them. None of the textbooks are set up that way. (Kansas, Rural District, Elementary School, Third-Grade Teacher)

(2) Showing students how to approach answering questions

In the second year of the cycle we started teaching [the students] how to examine questions....We started giving them techniques. How many question marks do you see in the question? If you see 3 question marks you know you've got to have 3 answers, not one. Little techniques like that would help them. (Massachusetts, Small Urban District, Elementary School Principal)

(3) Showing students how to make a good guess

Back in January, when we were taking the MEAP, I would tell them, 'if you don't know, guess!' because they would leave questions blank....One girl raised her hand and said, 'I passed the math test last year, and I guessed on every one!' and, I praised her, 'you're a really good guesser, see, class, if you get right all the ones you knew, and some of the ones you guessed on — I mean, you have a 1 in four chance, and sometimes you can eliminate some — you should be fine'. (Michigan, Small Urban District, Middle School, Eighth-Grade Teacher)

Teaching to the Standards

(1) Using the standards as guidelines for units and daily lesson plans

As far as what I teach in the classroom, my starting point is the state standards and benchmarks....Part of my department-head job is getting everybody...on board with not pulling out their textbooks because the textbook companies are way behind [in terms of] having all the MEAP content covered in their textbooks. Rather than pull those out....we put together a grades K-2 and a grades 3-5 support folder [with] all the stuff I've [collected] in workshops. (Michigan, Rural District, Elementary School, Fifth-Grade Social Studies Teacher)

I really try to incorporate the frameworks into my units. It hasn't always been easy, but I try to...annotate exactly what standard it is that I am covering at that time. (Massachusetts, Suburban District, Elementary School Teacher)

The distinctions the follow-up interviewees made among these three terms, and their views on which contributed to student learning, match the research findings in this area. For example, Koretz, McCaffrey, and Hamilton (2001) differentiated among various types of educator responses to high-stakes tests in terms of their likely effects on test scores and student learning. Three groups of responses emerged: "those that are positive (i.e., they have beneficial effects on learning and lead to valid increases in scores), those that are negative (i.e., they lead to distortions of learning or inflated scores), and those whose impact is ambiguous (i.e., they can be positive or negative depending on the specific circumstances)" (as described in Hamilton, Stecher, & Klein, 2002, pp. 87-88). Positive responses include providing more instructional time, working harder to cover more material, and working more effectively. Ambiguous responses include reallocating classroom instructional time, aligning instruction with standards, and coaching students to do better by focusing instruction on incidental aspects of the test. The single example given for a negative response is cheating. Only the positive and ambiguous response groups emerged in our study.

Overall, there was a strong sense of local identity among educators in this rural district and thus an aversion to the more context-free state standards and tests.

While educators in the large urban and rural districts reported the greatest impact on what and how they taught, educators in every district seemed to be paying attention to the format of the state test and incorporating it, along with the scoring rubrics, into their classroom assessments. Released questions from the state test also were being used for practice. All of these activities served to familiarize students with the format of the test and provided an opportunity to teach test-taking skills or test"wiseness," as one of the interviewees termed it. While some viewed this as part of the learning process — i.e., moving students from skill acquisition to skill demonstration or application — others focused on teaching students how to take tests in order to raise their scores. The former viewpoint is seen in the first quotation below, from a suburban teacher; the latter runs through the second quotation, from a rural teacher.

> *I feel that I've always prepared my kids for this type of thing, whether they were taking an SAT writing [test] or doing preparation for college or mid-year exams or what have you. We've always [taught] the skills and then had [the students] apply them, and that's what the MCAS does. (Suburban District, High School, English Teacher)*

> *The panic is about getting our scores up, and the way that you do that is to figure what [students] need to do to get higher scores, and how you can get them to do that. So some of that is teaching test-taking skills, which may not be teaching either content or even...what we think of as processing skills, like how to get information. It's just how do you do well on tests....We're feeling that we need to teach kids how to take tests. (Rural District, Elementary School Teacher)*

In all three states, interviewees noted both positive and negative effects of the state test on classroom practice.

In all three states, interviewees noted both positive and negative effects of the state test on classroom practice. Whether an impact was viewed as positive or negative was related to its perceived effect on students' overall learning. Specifically, educators seemed to view test-related effects as positive if they resulted in a general improvement in students' knowledge in a particular area — i.e., not only in better test scores. Instructional strategies that produced improved scores on the state test, but did not increase overall student learning, were characterized as negative. In the next section, we address these educators views on the impact of the state test on their students.

SECTION FOUR

PERCEIVED EFFECTS OF THE STATE TEST ON STUDENTS

Tests, if used judicially, are instruments of guidance to good teachers and a warning signal to society when children of one race or economic class are not prepared to pass them. But tests and standards without equity, without equivalent resources for all children, are not instruments of change but merely clubs with which to bludgeon children we have cheated from the hour of their birth and to humiliate their teachers. (Excerpt from commencement address by Jonathan Kozol, Lesley University, May 20, 2002)

To think that every child is going to be able to perform at the same level at the same time could only be dreamed up by someone who has no idea what children are, because it's so totally unrealistic. That's not human....Not all adults are the same. Why should we expect all ten-year-olds to be the same? (Massachusetts, Rural District, Elementary School, Fourth-Grade Teacher)

In this section, we report on interviewees' comments in regard to the second and third components of standards-based reform — the state test and associated stakes for students. In particular, we try to tease out the relationship between impact on students and the accountability uses of the test results. As previously mentioned, at the time of this study, state test results were one of several pieces of information used to determine school accreditation in Kansas, 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 could receive 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 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, but the stakes for students increase dramatically. With this in mind, we asked interviewees to describe the extent to which the state test affected student motivation, learning, stress levels, and morale. We also asked them to discuss the suitability of the test for their students in terms of content, format, and the presence or absence of accommodations (e.g., simplified text or translations for students with limited English proficiency, separate test settings for students whose disabilities cause them to be easily distracted). The overall findings are described below.

In this section, we report on interviewees' comments in regard to the second and third components of standards-based reform — the state test and associated stakes for students.

Overall Impact on Students

In all three states, interviewees reported more negative than positive test-related effects on students. Perceived negative effects included 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 that the state test had increased student motivation to learn, and had improved the 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, they were mostly viewed as having a more negative than positive impact on these students. Massachusetts interviewees were three times as likely (two-thirds of them 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. In terms of 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, suggest that the state's goal of increasing access to higher education through the program is not being realized. In Massachusetts, urban educators were most concemed 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.

These findings confirm what the literature on motivation has long shown: attaching stakes to tests will have a differential effect on the activation and maintenance of motivation in students (Kellaghan, Madaus, & Raczek, 1996). They also raise serious equity issues in terms of the greater impact on elementary students and special populations in all three states, and highlight the need for appropriate supports and interventions for these groups. Above all, they cast into doubt the possibility of finding a one-size-fits-all model of standards, tests, and accountability that will optimize motivation and learning for all students. Below, these findings are discussed on a state by state basis.

These findings confirm what the literature on motivation has long shown: attaching stakes to tests will have a differential effect on the activation and maintenance of motivation in students (Kellaghan, Madaus, & Raczek, 1996).

Kansas

This year they had the state math and the state science [tests]. They had the state performance assessment, they had reading. They had four [tests], and I was the last one....So the students were just about [out of] it before they even started, which made me a little scared about how they were going to perform. I think it helped a little bit giving them points for participation on it, but I got a lot of negative feelings [about the test]. (Suburban District, Middle School, Seventh-Grade Science Teacher)

I fear we will develop a class of people who in past generations were good, solid citizens, who kept good jobs, had a family, and were responsible community members. Those kinds of people in the current generation [of schooling are] having their self-esteem so bashed by this testing, that I question whether or not they will become the kinds of citizens that we need. This isn't doing anything, in my opinion, but creating an educated class and a non-educated class, and we will have only ourselves to blame in twenty years when these people rise up and do whatever it is they are going to do. (Rural District, Middle School Principal)

The students really stress out over [these tests]. You know, I play that positive game with them all year, I always tell them they're the best fourth-grade class, or the best fifth-grade class, the smartest, the most hard-working, and we do practice tests and other things, so I try to reduce that anxiety. And I tell them this is showing what they know and how well I'm teaching, and I try to take the pressure off of them. I think that if the kids started to have the pressure of 'Gee I'm gonna have to go to summer school' — like in Indiana; if you don't pass the test, you have to go to summer school, then you take it again, and if you don't pass, they're looking at retention — that's an awful lot of pressure on a child at any age. (Rural District, Elementary School, Fifth-Grade Teacher)

Overall Findings

While no official consequences for students attach to the Kansas Assessments, interviewees still reported test-related effects on students. The most frequently mentioned were that students had to take too many tests (one-third of interviewees), the tests were unfair to special student populations (one-fifth of interviewees), and the tests created stress for students (one-fifth of interviewees).

Overtesting was the dominant theme in the Kansas interviews and was viewed as reducing students' motivation to do well on the state tests. This apathy was perceived as being due in part also to the lack of consequences for students from the test results. This did not mean the lack of mandated stakes, but that students rarely received feedback on their performance, and so did not see the connection between their classroom work and their performance on the test. Two illustrative comments are offered here, one from a mathematics teacher and the other from a social studies department chair:

> *The kids are fed up. They feel as if all they ever do is test. [District-level testing], state tests in all the [subject areas], national tests, classroom tests. They're tested out completely. I don't think they have a clue what the difference is half the time. I think they care up to a point and then just say, I'm done. I'm toast. I can't do any more. I think [that the] order in which these tests are given determines how well they do....I think the kids — most of them — have good intentions, but after a while they realize they're never going to see their scores....Where's their incentive to constantly perform at a high level on everything? (Small Urban District, High School, Mathematics Teacher)*

> *It would be nice if we had the results back so that we could hold the students accountable for them. Communication Arts gives theirs early enough in the year that they can hold [students]...responsible for the district assessment....Our kids [say] 'I didn't take that too seriously because it doesn't really matter.' Well, if we are organizing our entire curriculum around it, somehow the state should get their [act] together and be able to give those back to us so that we can hold the students accountable for them....It heightens [the students'] level of concern just a little bit if they know they are going to be held accountable. (Suburban District, High School, Social Studies Chair)*

At the same time, some interviewees noted that students were beginning to realize the importance of the tests, and that their school was providing public recognition for students who have done well. For example, one teacher remarked:

> *We sometimes will get [the test scores] back by the end of the year. I always go and ask for the scores. I want to see [them]. I want to see what the kids have done. The kids want to see, too. If we can get them back by the end of the year it is great. [How I use the information] — I think mainly it reinforces what I already knew. But it's kind of a validation of what they can do. They're really proud of it too, because they usually give 110 percent. I try to tell them that this is important, but that everything they do in this class they're doing for themselves, not for me or for a grade. (Suburban District, Middle School, Eighth-Grade English Teacher)*

As suggested by this comment, one problem with getting students to "buy into" the state test was the delay in getting back the test results, with one-fifth of these educators noting that the results came back too late to be useful.

Even though the Kansas Assessments have low stakes for students, about one-fifth of interviewees felt that they created undue stress for some students, particularly special education and LEP students. While the state provides accommodations as well as alternative assessments for students who need them, educators still felt that the validity of these

students' test results was questionable. A special education teacher explained:

> *[About half of my students with Individualized Education Plans take the state tests with modifications. About half don't because] they are higher-functioning and so we feel that they're okay. But the half that are taking modifications, I don't know how valid or helpful their taking the test really is. For example, I have one [student] who doesn't read but she can understand what's going on in the classroom. And we just gave the social studies assessment a couple of weeks ago, and I didn't think she had a good understanding — I had to read it to her — of what was being read. Maybe she understood the question, but I don't know that she had really grasped that information when I presented it in the classroom. I wanted to stop and throw it out the door, but we're supposed to give it so I had to go through the whole test. I felt in that particular situation it was a waste of her time....I don't like it when kids don't get anything out of what they're doing. (Suburban District, Elementary School, Sixth-Grade Special Education Teacher)*

Interviewees had several suggestions for how to reduce the adverse impact on these students, including the provision of multiple levels or forms of the test, allowing students several opportunities to take the test, and improving accommodations for LEP students.[i]

An additional concern was the number of tests that special education students had to take. Interviewees explained that district, state, and special education testing requirements meant that these students usually wound up having to take more tests than did those in regular classrooms. For instance, one teacher said:

> *Special education kids get really mad and they get hit worse [than the regular students] because we also have special education tests....Not counting the special education tests, I spent 51 class periods last year on formal assessments. That's way too many days. To me, the writing assessment should be the same as the social studies assessment so that could be done on one test. There should only be one [set of tests]. There shouldn't be district and state tests. I think that these people need to get their acts together and we need to have one test[ing program]. (Large Urban District, Middle School, Special Education Teacher)*

This teacher's suggestion — that the testing requirements should be reduced — was echoed by many we interviewed, in the context of testing not just special education students, but students in general.

Educators also mentioned concerns about students in the regular classroom who had not been diagnosed with a learning disability but had more problems than the average student. Since these students did not get testing accommodations, they were, as one principal described, "falling through the cracks." He went on to say:

> *I would recommend alternative forms of the test for kids at lower levels...because you want to show progress. There are alternative forms or modifications for kids in special ed, but those are not available for these [lower-level] kids [who are in regular classrooms and don't have an Individualized Education Plan]. I would say that these kids are falling through the cracks. These are kids who are almost [candidates for] special education, but they're not quite. (Large Urban District, Elementary School Principal)*

A classroom teacher explained how these issues play out in a testing situation:

> *I have two learning-disabled kids, so they can have the test read to them...but if I have a fourth grader who is actually reading on a second-grade level that hasn't been identified, he's stuck in there...and he can raise his hand and I can walk up to him and...tell him a word, but I can't tell him what it means or read a whole sentence for him, it's only if there's a word that he can't figure out — so it's difficult. (Rural District, Elementary School, Fourth-Grade Teacher)*

These findings suggest that much work needs to be done if the state test is to provide a genuine opportunity for all students to show what they know and can do. At the same time, some interviewees mentioned that the state test had improved the quality of education for special populations, particularly in terms of getting students extra help that might not otherwise be available. It also should be noted that these concerns are similar to those voiced by interviewees in the other two states.

One-third of those interviewed felt that the test results were influenced by factors over which the school had limited or no control. This could include out-of-school factors such as whether a child was read to at home or whether children's parents encouraged them to study and do well at school. It also could include the mood a child was in on the day the test was given. One teacher remarked:

> *I don't use [the state test results] to show well, it looks like I'm deficient here or whatever, because it depends on the kid, it depends on the day, it depends on the test, it depends on the year. (Suburban District, Middle School, Eighth-Grade English Teacher)*

As suggested by this comment, so many factors went into determining these scores that it was hard to view them as an accurate or complete depiction of student learning. In fact, half of those interviewed felt that the test results did not represent student achievement in a subject area. One suburban middle school principal described how teachers at his school used results from the state and other tests to identify students in need of extra help. At the same time, he noted the need to avoid self-fulfilling prophecies in terms of expectations for what these students could do, particularly since "there will be some kids who get flagged and they will be doing just fine." In other words, these educators were uncomfortable with using test scores as the single measure of student learning.

School-Type Differences

Elementary educators were twice as likely to note that the state tests created stress for their students (one-third versus one-seventh at each of the other two levels). This issue cut across all district types and, as indicated by the quotation below, all student ability levels.

> *I think we have extremely high expectations as a school district, but are they too high? When you've got kids who are crying during the test something's not right....This year we didn't have that as much because we spent the whole two and a half months getting them ready, but last year we literally had kids sobbing. Our top kids [were crying] because they like to do a good job and they weren't getting it. If these students who are pretty sharp aren't getting it, what about the others? (Suburban District, Elementary School, Fourth-Grade Teacher)*

Elementary and high school educators were twice as likely as middle school educators to note that these tests were unfair to special populations (one-quarter at the elementary and high school levels versus one-tenth at the middle school level). High school educators in particular felt that not all students should be held to the same standards, and that multiple levels of the test should be available. For example, a special education coordinator said:

> *My concern is that with these assessments, we are testing kids on levels that are way beyond their scope and they are very, very frustrated. This year has been extremely difficult with our sophomores...these are students who are receiving special services in the learning center, so their math and reading levels are way below grade level because they're severely learning-disabled. Of course, our curriculum in the learning center isn't the same as the curriculum in a regular classroom....By the time a kid gets to be a sophomore, the content is getting harder and harder, the reading level is getting higher, the math is getting higher, so we wanted to give them a more practical curriculum, but yet they had to be tested in an area that was way beyond them. I just thought that was a little bit unfair....Do I think they need to be tested? Oh, absolutely. I think there should be some kind of assessment, but one that's more appropriate for the curriculum they're learning. (Small Urban District, High School, Special Education Coordinator)*

Some of the concerns raised by high school educators were related to the language requirements of the tests. For instance, the reading level of questions on the mathematics test made it difficult for LEP students to respond to questions, calling into question the validity of the test results as a measure of these students' mathematics knowledge. A high school teacher explained:

> *In the previous math tests...we saw that there was Algebra II-level thinking...plus fairly high reading ability. We think a lot of our LEP kids might be able to do some of the math questions, but they can't really figure out what the question is asking because they're not that familiar with the language....and the result is that those students don't test well on the state assessment at all....Some of them speak no English, and they have to take the test....I really feel they're being thrown into the water and they're not ready to swim. (Large Urban District, High School, Mathematics Teacher)*

In general, interviewee comments on the impact on special populations highlighted the tensions involved in wanting all students to be included in the assessment, and yet not wanting them to take a test that did not match their curriculum and knowledge level.

District-Type Differences

Educators in all districts expressed similar concern over the stress the state tests caused for students as well as the effects on special populations. They also were in agreement that not all students should be held to the same standards and that multiple levels of the test should be available. At the same time, there were some differences. For example, educators in the large urban district were the most likely to note that the state test had changed the quality of education for students in positive ways (one-quarter versus less than one-tenth

in each of the other districts). As one teacher remarked:

> *I think the impact of the state test on students' education has been more positive [than negative] because we, in turn, have been able to focus, as teachers. We really use some of the ideas from the testing and they're really good ideas, and the areas we need to work on are helpful to know. (Large Urban District, Elementary School, Special Education Teacher)*

Educators in the large urban and suburban districts were the most concerned about the number of tests that students were being required to take (two-fifths in the former and one-third in the latter). The quotation from a seventh-grade science teacher at the start of the Kansas section exemplifies the suburban perspective. The following is a typical comment from a large-urban-district interviewee:

> *We do a lot of state assessments, we do a lot of local assessments. Especially for a fifth-grade teacher — they have to give the district writing assessment, they have to give the state writing assessment. They have to give the district reading assessment, they have to give a state reading assessment. It would be nice if we could somehow pare down the assessments in writing and reading. Give one assessment that would give us the information we need, so that we're not constantly having these kids take assessment after assessment, back to back. Fifth-grade teachers start in January, and they don't have a week without assessments until April....By the time you give the last assessment, those kids are burned out...and I don't know that they're doing their best on it....It's a high-stress situation for a lot of them. It would be nice if we had one writing, one reading, and one math [test] that the district and state would accept. (Large Urban District, Elementary School Teacher)*

While large-urban-district educators expressed these concerns primarily in relation to the lack of alignment between district and state testing requirements, suburban educators were more likely to note that the state tests took away from classroom time that could be spent on more substantive content.

Michigan

> *Some of the kids get very upset when they have to take the test. They will get stomachaches and headaches — sometimes those are the kids that might do really well. The kids that are not going to do well don't care, they just say 'oh well, so what, I will just fill in the bubbles and go on.' The kids get so sick of hearing the word MEAP, they just moan and say 'not again.' It is an albatross, I think, for the most part. Everything is geared to the MEAP, the students get to the point where they hate the word MEAP. (Small Urban District, Elementary School, First-Grade Teacher)*

> *The first year that my students, if they passed the MEAP test, [could get a] scholarship [for college] I shared with them how I'd had to work and I had to pay student loan payments for ten years...and how happy I would have been to have that scholarship money....So we try to motivate them through money, and it seems like such a good motivator for most of us....I think that for some [students] it helps them and for some it doesn't. [Students] whose parents dropped out...it's hard to make them care about the test. I don't really think the MEAP motivates the unmotivated students. (Small Urban District, High School, Science Teacher)*

Our kids have already taken nine tests by January. They also took three in January — a writing, a social studies, and a science test....We usually don't get results back until the last week of school. What can you do with them then? That's my question, and that's my criticism of it. People say, 'Well, you can always use [them] with your next year's group,' but that's a different group of children. They're not only going on to another grade, but to another school....In April, they took three more tests...that are mandated by the school district....We haven't gotten those results back, either....And that's not all yet. Last week, the eighth graders concluded taking four tests....That was [a district-designed test]. (Large Urban District, Middle School Principal)

Overall Findings

At the time of the interviews, the Michigan test had moderate stakes for students (i.e., an endorsed diploma and scholarship money attached to doing well on the high school test). When asked about the effects of the test on students, Michigan interviewees reported the same three effects as their colleagues in Kansas: that the test created stress for students (two-fifths of interviewees); that students were being asked to take too many tests (almost one-fifth); and that the test was unfair to special populations (almost one-fifth). However, Michigan interviewees also noted effects that were mentioned not at all in the Kansas interviews, or by only a handful of interviewees. These were that the test content was too difficult or ambiguous (one-quarter) and that the test provided students with incentives (one-quarter). The incentive effect seemed to be confined to high school, the level at which scholarship money could be won. As will be seen below, there was a complex interaction between the motivating power of these stakes and student characteristics, suggesting that the incentive did not function similarly for all students.

As in Kansas, one-third of those interviewed felt that the test results were influenced by factors over which the school had limited or no control. Almost half felt that they did not represent student achievement. Some educators mentioned flaws in the test as well as poor design features that contributed to this problem. For example, one principal remarked:

They change the test format every year. Sometimes the print is poor, this year they had the 'Reading for Information' section broken down into little groups....Children never read that way, children read left to right. So, it would be easy to confuse the children. Every single year, we have a new dilemma. (Suburban District, Elementary School Principal)

At the same time, educators were making use of the test scores to target areas for future instruction. A teacher described the process at her school as follows:

We take the present test scores — they show the teacher with each class, and everything broken down with percentile and raw scores....If you see some of the areas where the kids do really well, it could be the teacher [is] strong [at] teaching that topic. I like all parts of math, but some teachers might like geometry, so they explain geometry a little better. If I see [that my students'] overall results for integers are low, I need to go back and look over how I'm presenting integers....Either I could do a better job presenting ...my lessons on integers, or I need to look at the kids to see why they're not doing as well. (Large Urban District, Middle School, Sixth-Grade Mathematics Teacher)

While interviewees mentioned using the results to identify curricular weaknesses, they particularly wanted information that would help them identify students who needed more help and provide some indication of how to help them. In fact one-third of the interviewees mentioned that they would like this type of diagnostic information.

School-Type Differences

Elementary and middle school educators were the most likely to note that the test content was too difficult (one-quarter of each versus less than one-tenth at the high school level). In addition, elementary educators were twice as likely as educators at the middle and high school levels to mention that the testing caused students stress (almost three-quarters versus two-fifths and one-fifth respectively). The following are some illustrative comments, one from a third-grade teacher and the other from a middle school social studies teacher:

> *I note that children who do well in class in general, that are high achievers, are very anxious when it comes time for these tests. Very anxious — they get nervous, you know, but that's just their personality. And then at the other end of the spectrum you have those children who are just lackadaisical, and they're just going to go down and do a, b, c, d, a, b, and couldn't care less and finish in ten minutes....The brighter kids have more invested; absolutely, I believe that. (Large Urban District, Elementary School, Third-Grade Teacher)*

> *I don't think I should be teaching to the test....I should be able to teach my regular class and then they should be able to take what I teach and formulate a test but, it seems to me that it's the opposite...that they take something that maybe hasn't even been taught and they test it. For example, on the MEAP test in social studies, which is a very extensive test — it will curl your hair, they are testing the children on geography and economics. They're in eighth grade, mind you, but they're tested in eighth grade on something they were taught in seventh. They're tested on economics, which they were taught in sixth, they're tested on history, which they just got, and they're tested on government, which they never had. So the test, to me, is kind of unfair...it really puts the students at a disadvantage, and the teachers [as well] because we never know what's going to be on the test. (Large Urban District, Middle School, Seventh-Grade Social Studies Teacher)*

These quotations illustrate some additional themes that also came up in the Kansas interviews. For example, the elementary school teacher's comment about the greater anxiety experienced by the high achievers suggests that the motivational power of these tests is complex and can act in undesirable ways. The comment about lower-performing students' arbitrary choice of answers also raises questions about the motivational power of the test as well as the validity of the test results. In addition, the middle school teacher's wish — that she should be able to teach as she normally did, and that the test developers should be able to take what she taught and formulate a test — echoes Kansas interviewees' concerns over test content inappropriateness for certain student populations. It also highlights the need for tests that complement rather than constrain the curriculum.

Since scholarship money rides on student performance on the eleventh-grade test, it was not surprising that high school educators were the most likely to note that the state test offered students incentives (two-thirds versus one-tenth at the elementary and one-fifth at the middle school levels). At the same time, these educators were divided as to whether this had a positive or negative impact on student learning. In addition, a sizeable minority — about one-fifth of the educators at this level — felt that the stakes attached to the test results did not affect their students. These differing opinions on the motivational power of the test can be best discussed by organizing them according to district type of the interviewees. We turn to this below.

District-Type Differences

Interviewees in the large urban district were the least likely to mention that the state test provided their students with incentives. For example, less than one in ten educators in the large urban district felt that the scholarship money attached to the eleventh-grade test motivated their students, while more than one-third of the educators in the small urban, suburban, and rural districts felt that it encouraged their students to try harder. One reason given for the lower motivational power of the test in the large urban district was that students felt that their chances of scoring at the necessary level were slim.[ii] Others linked it to students not viewing the test as important in the short term. The following comment exemplifies some of the frustrations felt by educators who knew that the results for these students would be used to hold them — but no one else — accountable:

> *[Of all of the tests that students have to take], I think that the MEAP actually rates lowest, because the students don't see any consequences....Even the promise of the $2,500 doesn't come until high school. If kids were long-range in their thinking, they'd all do better, but they're short-sighted....[Attaching the diploma to the high school test] means something to kids who are going on to college, but if kids can do well on the ACT [college admission test], what do they care whether their diploma is endorsed or not?....It means a lot to me because it's a reflection on how well our school has done, and the state is attaching money or accolades to the scores. It doesn't mean that much to the child or his parents, except when the parents see the scores in the paper and [judge the school by them]. (Large Urban District, Middle School Principal)*

Another aspect of the motivation issue came up in interviewees' concerns about students having to take too many tests. Interviewees in the large urban district were particularly concerned about overtesting and the effects this had on student motivation and effort. The following comment exemplifies these concerns, which were raised by about half of these interviewees:

> *When you have a test in October...then maybe another test in December...then you [have] two weeks [for] the MEAP...then again in March, we have...a nationally normed test that we give. So there's weeks out of the curriculum....Then [there's] some other test that they give at the end of the year....Now add all that up. How many weeks are these children actually getting [teaching and content]? Look at all the time that's test-based as opposed to [learning]. And the children feel it; when they come into the room and say, 'oh no, not another test,' something's wrong; I think the children resent it and then they're not going to do their*

best....Education shouldn't have to be drudgery for children. Education is their job. Yet it's [as if] you have to bribe them: come on kids, do well, we have to do this. (Large Urban District, Elementary School, Third-Grade Teacher)

A theme among educators in both the large and small urban districts was that the content of the test was too difficult. One reason for this was the perceived inappropriateness of the test for special education students (this concern also came through in the other districts). Another was that the items on the test were not appropriate for urban students given their background and experiences, and that these students were being set up for failure. In the following excerpt from an interview with a high school social studies teacher, these issues are framed in terms of the different experiences that students in suburban versus urban districts bring to school:

It seems that most state-mandated tests are not aimed at an inner-city urban high school. They're aimed at the suburban white middle class. That causes some of the things that need to be covered on the test to be difficult to do....[The items] are not oriented towards the urban or inner-city student, language-wise, and a lot of times the essay questions or the reading comprehension questions are just not something that an inner-city student is going to be familiar with....There are obviously some biases in there. (Large Urban District, High School, Social Studies Teacher)

Perhaps not surprisingly, educators in the urban districts were the most likely to note that the state test had changed the quality of education for students in negative ways, with about one-fifth of them mentioning this.

Massachusetts

I know there's nothing wrong with education reforms [that are] trying to give kids the type of skills they need to succeed in [this] century....But this type of test, it's high stakes, and it will penalize those kids who are at the bottom....I think that there's not [enough] flexibility for...the types of students I've been teaching, and I'm not sure what'll come out of it in the end. I do know in a couple of years you're going to have a large number of dropouts — kids who drop out after the tenth grade, minority students, second-language learners — because they're going to feel they're failures: 'Oh, I'm never gonna be able to pass this test....why should I stay in school?' That's going to be a serious problem. It's tough enough to keep a lot of those kids in school beyond the tenth grade, even beyond the ninth grade....The test isn't going to help us with that. (Large Urban District, Middle School, English-as-a-Second-Language Teacher)

The largest negative aspect is the fear the students have of the test....The students will say to me...around graduation time after they've had the test, 'You know, I've gotten B's and B-pluses, and I've done well all year. Suppose I fail the test in high school, does that mean I don't go anywhere, does that mean that all of these years of work have been no good?' That's their big fear. (Small Urban District, Middle School, Eighth-Grade Social Studies Teacher)

There were kids in tears over it, and there have been for the last two years. Kids who didn't want to come to school. Kids that had stomachaches they never had before, who would just put down their pencils in frustration. (Rural District, Elementary School Principal)

Overall Findings

Given that receiving a high school diploma hinged on passing the tenth-grade test, it was not surprising that Massachusetts interviewees reported more impact on students than did interviewees in Kansas and Michigan. These effects included some that came up in the Kansas or Michigan interviews — e.g., that the tests were unfair to special populations, created stress for students, and had content that was too difficult or ambiguous, but they cropped up much more often in the Massachusetts interviews. For example, two-thirds of Massachusetts interviewees felt that the state test was unfair to special populations, compared with one-fifth of interviewees in Kansas and Michigan. And, while one-fifth of Kansas and two-fifths of Michigan interviewees felt that the test created stress for students, more than two-thirds of Massachusetts interviewees did so. Concerns over difficult or ambiguous test content were also heard more often in Massachusetts than in Michigan (one-third versus one-quarter; this theme did not emerge in Kansas).

Additional effects mentioned by Massachusetts interviewees that came up rarely, if at all, in the other two states included that the tests negatively affected students' perception of education (two-fifths of interviewees) and were too long (one-quarter). While interviewees mentioned more negative than positive effects on students, those who reported the latter tended to focus on the motivational power of the test (e.g., one-fifth felt that the tests encouraged students to learn). As in Kansas and Michigan, the test was not seen as having the same motivational power for all students. For example, one middle school teacher in the small urban district remarked that it seemed to motivate the high achievers, but not "the kids who really need it...that's what's frustrating."

School-Type Differences

Elementary educators were the most likely to report that the state test created stress for their students, with two-thirds of them mentioning this concern compared with about half of the middle and high school educators. Some of this stress was attributed to the inappropriateness of trying to hold every student to the same standard. As a fourth-grade teacher remarked:

> *To think that every child is going to be able to perform at the same level at the same time could only be dreamed up by someone who has no idea what children are, because it's so totally unrealistic. That's not human....Not all adults are the same. Why should we expect all ten-year-olds to be the same? (Rural District, Elementary School, Fourth-Grade Teacher)*

Special education students were seen as particularly affected, even when they could use accommodations. An elementary school teacher described the following incident:

> *We had one little boy who had an accommodation to work on the computer [for] the long compositions ...and what normally took children anywhere from two [to three] hours...took this poor child two days. And by... the second day, he wet his pants, he was so nervous and so upset. (Suburban District, Elementary School, Head Teacher)*

Others noted that while some students rose to the occasion others were crippled by stress:

> *For them to sit in the testing situation for that length of time, they're exhausted....They do very well at the beginning [but] they're so tired at the end that you see a dropoff in the test scores....I've seen kids break down: 'I'm not doing this.' Crying. Complete frustration. I've seen kids get sick to their stomach...a lot of physical things that are responses to stress. And I've also seen kids do the reverse. Kids who didn't do so well during the year who really shine on the test. (Large Urban District, Elementary School, Fourth-Grade Teacher)*

As in Kansas and Michigan, elementary educators in Massachusetts noted that high-performing students tended to be the most nervous about taking these tests. For example, an elementary teacher described a very bright student who had been in her class the previous year, but whose mother decided to move her to a private school because she was nervous about taking the MCAS. Since students in private schools do not have to take the state test, the girl's mother felt that it would be a less stressful environment for her. Reflecting on this parent's efforts to help her child, the teacher remarked:

> *Hey, if I had children, I don't think I'd put them through this right now, I'd put them through private education. Why am I going to do that to my child? Why are you going to [put] your tenth grader through a test, a grueling test that might affect whether or not they graduate, when you can put them in a private school and not have to put them through that at all? (Large Urban District, Elementary School, Fourth-Grade Teacher)*

Middle school educators were most likely to note problems with the test itself. Almost half of them commented that the test was too long, its content was too difficult or misleading, or it didn't suit all learning styles. The issue of student motivation also came up. Some felt that the graduation test was motivating middle school students to focus on their work:

> *The kids worry about the test, but they worry about everything here, much more than I ever thought. The kids have become obsessed with their homework, the quality of their work. I have been around for so long, and in the last five years I have seen something I have never seen before. I am so glad that I stuck it out to this last five years. The students talk about homework and compositions [as] you have never heard other kids speak in your life. These kids are so into their education. It is incredible, they are worried about the test because of graduation, but they are not obsessed [by] the test. (Large Urban District, Middle School Principal)*

Others felt that these students were not really affected by the impending stakes at the tenth grade. One teacher described the difference in attitudes displayed by eighth graders versus tenth graders in his school:

> *Since I teach both eighth and tenth grades I noticed a difference this year between how they view the MCAS. The tenth graders are just fearful...the eighth graders know that they are going to move on to the ninth grade even if they don't pass and don't have to worry about the eighth-grade MCAS again. They approach it with the attitude that they want to do well, but if they don't then it is not much of a problem. They almost look at it as practice for the tenth-grade MCAS. (Large Urban District, Middle/High School, Mathematics Teacher)*

The main theme at the high school level was demoralization rather than motivation of students, with over half of these interviewees noting that the tests have negatively affected students' perceptions of education, particularly those of special education students. While students who fail the tenth-grade test can retake it up to four times before the end of high school, interviewees spoke of students who had already convinced themselves that they would never pass the test and thus would have to drop out.[iii] Urban students, minority students, special education students and English-language learners were seen as particularly vulnerable. One teacher remarked:

> *Some of these kids are so upset, really...it doesn't motivate them. The legislators don't have a clue. Clearly, whoever thinks that has never been in the classroom. They get upset over it, and they don't try any harder; if anything it becomes a defeatist mentality, especially in a school like this, where they don't get motivated from a challenge, they back down....I have heard about dropping out, but I don't know [whether] they would do that anyway....they come out with that reply, 'I don't care, I'm dropping out anyway.' (Large Urban District, High School, Ninth-Grade English Teacher)*

At the same time, high school educators were the most likely to note that the stakes attached to the state test had increased student motivation to learn, with one-third of them noting this effect (versus one-fifth at each of the other two levels). While about one-fifth of the high school interviewees saw this increased student accountability in a mainly neutral to positive light, feeling that it served as a wake-up call for some students, most felt that the MCAS should not be used as the sole criterion for graduation. As indicated by the quotation below, concerns in this area focused on the gatekeeper role of the test for students who were already struggling.

> *That is the biggest of the injustices that I see because students will become disaffected...those borderline students whom we work so hard to keep and involve. We work so hard to fill the cracks, and I'm not sure society understands how much we do for the cracks. That we, in the long term, we prevent people from going to jail. We help prevent people from going on welfare. We help prevent people from having children and becoming a burden on society....MCAS is going to create, I believe, higher dropout rates, higher criminal activity, higher violence in schools, greater numbers of disaffected. (Rural District, High School, Tenth-Grade English Teacher)*

This lack of support for the gatekeeper role of the tenth-grade test was linked to interviewees' perceptions that the test results were influenced by non-educational factors and did not provide a complete picture of student achievement in a particular subject area (more than half of Massachusetts interviewees voiced both concerns). Thus, it was unfair to use the results to make such a highly consequential decision about a student's educational career. These feelings resonated with public opinion at the time, since the tenth-grade tests had been the subject of widely publicized student walkouts and other protests.

District-Type Differences

Educators in the large urban district were the most likely to raise the above concerns about the tenth grade test, with three-quarters of them noting that the test results were affected by non-educational factors and almost two-thirds mentioning that the content on the test was misleading or too difficult. One educator remarked that the tests were "almost prejudicial towards inner-city kids." Echoing concerns raised by the large-urban-district educators in Michigan, she said:

> *They are rated on the same scale as the more affluent communities, where the kids go to science camp and MCAS camp, whereas our kids are lucky if they get out of the inner-city project a day or two over the summer. It's almost prejudicial towards them because of economics. It's their unearned stigma, whereas in the more affluent areas, it's their unearned privilege...and I do think it is going to start to pit communities against each other, because even the teachers are starting to feel the crunch when accountability starts coming down....I was an MCAS tutor for four hours a week for fourteen weeks...but we can't make up for a lifetime of missed learning opportunities, we can't do that....So the kids are feeling defeated and a lot of the teachers are feeling defeated. (Large Urban District, Middle School, Eighth-Grade Language Arts Teacher)*

Urban educators were worried about the potential fallout from high failure rates on the tenth-grade test (see Box 12 for the results from the spring 2001 test).[iv] With this in mind, resources were being poured into preparing students. The superintendent in the large urban district described some of these efforts:

> *[The MCAS] has gotten the state to reallocate resources so we can pay for summer programs...which is something new in terms of services for urban kids. It has created after-school programs in every high school in this district for 90 minutes a day, four days a week. It's forced the district to reassess itself and address whether the high schools are organized in a way that will help us get to where we need to be....Pilot schools have started....Those things are all partly manifestations of the urgency that these failure rates have created. Will it get to the point that urban kids are going to perform as well as wealthy kids in the suburbs? That's a long way away. It would have to be supported by fundamental realignment of resources. (Large Urban District, Superintendent)*

In fact, while teacher professional development in all three states had been affected by the state test (i.e., more was being offered and this was mainly test related), Massachusetts seemed to be the only state where state and district resources were being allocated toward programs and materials that would prepare students for the state test.

Box 12

Spring 2001 Tenth Grade MCAS Test Results

Results for the spring 2001 MCAS test administration had not been released at the time these interviews were conducted. When results were released in fall 2001, the overall pass rates on the tenth-grade tests were better than expected — 82 percent of students had passed the English test and 75 percent had passed mathematics.[v] However, there were stark differences in the pass rates for students in urban versus suburban districts as well as for regular versus special student populations (Massachusetts Department of Education, 2001, 2002)[vi] Taking our four study districts as an example, while almost 100 percent of students in the suburban district and four-fifths of those in the rural district passed the test, only two-thirds of students in the small urban district and half of students in the large urban district did so.

Another area of concern among the large-urban-district educators was the lack of appropriate accommodations for special populations, which made it difficult for these students to show what they knew or were able to do. For example, one high school principal noted that since only a Spanish version of the test was available, there were no appropriate accommodations for the many Albanian and Vietnamese students in her school who also struggled with English as a second language. A middle school principal described the lack of appropriate accommodations for special education students as follows:

> *It's one thing if a kid fails by choice — and there are kids that choose to fail by not doing the work or whatever. But I have special ed kids in this building that have interesting learning styles who are going to have the most difficult time [trying to] pass [MCAS]....There are bright kids in this building who just don't test well, who have particular learning styles — and you notice that I say learning styles rather than learning disabilities — learning styles that don't fit that test. (Large Urban District, Middle School Principal)*

Educators in the small urban district were the most likely to note that the state tests had changed the quality of education for students in a neutral to positive way (two-fifths). They also were the most likely of those we interviewed to mention that the test had increased student accountability in a positive way, as evidenced by the below quotation. In trying to understand why small-urban-district interviewees differed in this way, we concluded that it is not so much a feature of small urban districts in general, or this small urban district in particular, but rather that we had tapped into one of the perceived rationales for high-stakes testing: these tests are a wake-up call for students. As a teacher in this district remarked:

> *I think you're going to see a rippling effect. I think you're going to see some change in [students'] attitudes over the next year when those scores come back. When the present-day sophomores take the exam and all of a sudden word gets back to brothers and sisters, neighbors, relatives and friends, that these scores do count for something, there's going to be that rippling effect. It's going to work its way down through the system. And I don't necessarily think that's a bad thing. I think it's a wake-up call. (Small Urban District, Middle School, Eighth-Grade Mathematics Teacher)*

Box 13

Vocational T

Vocational-technical (voc-tech) schools have a mission that is different from their traditional public school counterparts. While traditional high schools are places where students prepare for college, voc-tech schools are generally places where they prepare for a trade, for example as electricians, mechanics, or drafters. The different missions of the two produce different responses to state educational reforms. To better understand these differences, we interviewed eleven teachers and administrators at one voc-tech school in Massachusetts, using the same interview protocol as in our study. Since this was a small sample and all of the interviewees were from the same school, what we learned from them may not be characteristic of other voc-tech school personnel in other settings. To avoid overstating our observations we report themes that enjoyed unanimity of opinion among these eleven educators; and to put our observations in context we compared them with data from other sources. Three major areas of agreement were culled from the interview data. These are as described below.

Application of the State Standards and Test to Voc-Tech Students

While respondents tended to agree with the reform efforts in Massachusetts in principle, they did not approve of the standard application of the state curriculum frameworks, or of the MCAS, to their students whose academic attainments and requirements differ from those of the traditional high school student. As one teacher remarked:

> *[The Department of Education is] saying that students have to be accountable for such high standards, and I think [these] are unrealistic and totally unfair in the case of kids who have not even had algebra....Shouldn't they be pursuing those basic skills that they are actually going to use? [And that] a person who enters a trade...needs to know? (Vocational-Technical School, Culinary Arts Teacher)*

That is, while students who attend college will likely need to know algebra, those pursuing a skills-based occupation may not.

A second and more immediate concern involves the unequal pass rates of voc-tech and traditional high school students on the tenth grade test, which we have estimated to be 57 percent versus 78 percent.[vii] These figures suggest that a large percentage of voc-tech students face the possibility of finishing high school without a diploma. While some might argue that a resilient vocational student could earn a comfortable living without it, others view this outcome as an additional adversity for students who have already struggled in school. One interviewee put it strongly:

> *[I have a student who] won't be able to do well on the long composition. She won't do well on the content part. She's not a critical thinker. She doesn't understand abstract questions. [She's a] literal kid. She's [also] organized, [she's a] very nice person....She'll be a great baker. She follows directions well. She does all of her homework. She gets an A in my class. She tries hard. She participates in discussions. She does work over again when it doesn't meet a standard...[but she won't pass the MCAS]. [If she's denied a diploma] she won't be able to be a baker. Some [other] kid won't be able to be a diesel mechanic because [she] didn't get a diploma. They're not going to get hired....Taking away a person's ability to make money or have a living, that's not fair. If you want to have the test and then stamp MCAS on the diploma...that's all right, I can deal with that. (Vocational-Technical School, English Teacher)*

nical Schools

Voc-Tech Mission

Respondents felt that the standard application of the state standards and MCAS intruded on their voc-tech mission by making changes in their program that diminished its benefits. More than a third of those we talked with had removed voc-tech topics to include MCAS preparation courses. About half mentioned that the pace had been stepped up to the point where it forced them to rush through material. A similar amount felt that the test-related content being taught was too difficult for their students. The changed curriculum, quickened pace, and heightened expectations had left many students demoralized. As one teacher put it:

> *A lot of our students are a little bit behind, a little bit below level. So to expect them to do something like [the MCAS] at the tenth-grade level is very frustrating for them and that bothers me....For example, last year at the end of the tenth grade, I tried to give [students] questions to help them prepare for the type of question they would see on the test. And after they took the test one student...said to me that she felt so badly for the upcoming tenth graders....She said [there] were so many things on that test we've never seen....It was a very sad comment to make....I think it's very depressing for many of them. (Vocational-Technical School, Mathematics Teacher)*

The majority of these interviewees indicated that the overall quality of education at their school had deteriorated due to preparation for the state test.

Acknowledging Different Types of Achievement

Interviewees were unanimous in the assertion that students achieved at different levels and in different areas. With three-quarters of the interviewees indicating that the test was unfair, many wanted a graduation alternative that reflected the different academic goals and objectives for their students and school. One person said

> *I would have no problem with...a high school diploma that communicates some level of distinction or elite performance or college readiness. There [could also] be a range of mid-level certifications....[It should be] something other than a certificate of attendance...no, it would be a bona fide high school diploma. And perhaps this diploma might not get you directly into [a] university, but it should...get you into a community college. (Vocational-Technical School, English Teacher)*

In August 2002, the Massachusetts Department of Education decided to issue a state-sponsored certification of completion to students who completed their high school requirements but did not pass the tenth-grade MCAS.[viii] This certificate could be used to enter community college or the armed services, thereby providing some opportunity for advanced study for voc-tech students who fail MCAS.

Suburban educators were the most likely to mention that the state tests created stress for students (two-thirds noted this), particularly special education students. A guidance counselor at a school in the suburban district described the following incident:

> *Two years ago we had a Down Syndrome boy who was very articulate, who was reading. Comprehension was weak, but he felt so good about himself and his mom and dad had been so pleased that their child went through a year fully exposed to all of the curriculum areas in all of the classes. We would modify and accommodate as necessary. They wanted him to take the MCAS and we fully supported it. He could not take the math, because...he was in a simple math program that was significantly below grade level, but everything else he wanted to take. Well, this boy started writing notes to us signing his mother's name, saying, 'Please excuse [John]. He has to take a nap today. He's very tired. Please excuse him from taking the MCAS.' When I got the first note, I said '[John], what's going on?' 'Nothing.' 'Who wrote the note?' 'My mother, my mother.' We called mom and she said, 'Oh, my glory, it probably is too much,' but we had already made a commitment for him to take it. I actually then worked with him. We just did a section at a time. I tried to make a game of it...let's just see how we do....It was just too stressful. (Suburban District, Elementary School, Guidance Counselor)*

At the same time, suburban educators were the most likely to mention that the test increased student motivation to learn (one-quarter of interviewees in this district), although this was characterized as a case of further motivating the already motivated. Both rural and suburban educators were most likely to note that the tests have no impact on their students, but this was still a minority viewpoint in both districts. Rural educators also were the least likely to note that the tests had changed the quality of education or increased student motivation to learn.

Looking across the three states, these findings suggest that as the stakes for students increase, so too do the perceived negative effects on students.

Looking across the three states, these findings suggest that as the stakes for students increase, so too do the perceived negative effects 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. In the next section, we briefly review the findings from Sections Two through Four and outline some recommendations for policymakers.

SECTION FIVE

CONCLUSIONS AND POLICY IMPLICATIONS

The economic/business analogy seems to have shaped and propelled the drive for accountability in education during the last decade. Since there are no profits to serve as indicators of whether or not schools are doing a good job, test scores have been assigned that function instead. (Raywid, 1987, pp.764-765)

The worst thing happening right now is the testing. The testing is being used for the wrong purpose. To use a test to see where we could improve is one thing, but to use a test to compare schools, or to humiliate teachers, or to reduce funding – it's very destructive. Then you change from trying to improve instruction to the creation of stress and pressure and maybe even cheating. If it was used for a different purpose, I think it would be okay. (Michigan, Suburban District, Elementary School Principal)

The goal of this 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. The findings suggest that this reform is having a profound impact. It is sharpening the focus of teaching and causing some students to take their academic work more seriously. At the same time, it has drawbacks: an overcrowded curriculum, overanxious students, and perhaps worst, overtesting. The reform also has structural flaws, and these can prevent the spirit of this reform from making its way to the classroom.

One of those flaws is the uncertainty of stakes as a lever for producing change. The findings illustrate that the relationship between stakes and impact on classroom practice is mediated by several factors, including the school and district type in which teachers work, and whether they teach a tested or non-tested grade or subject area. The impact on students is also uneven. In all three states, the motivational power of the stakes attached to the test results varied, with high-achieving and suburban students most likely to be motivated and low-achieving and at-risk students most likely to be demoralized. Other researchers have reported similar findings (e.g., Cimbricz, 2002; Clarke, Abrams, & Madaus, 2001; Firestone, Mayrowetz, & Fairman, 1998; Grant, 2001; Madaus & Clarke, 2001). What this study adds to the body of literature in this area is a systematic look at how impact varies with the stakes attached to the test results. The results of this view are summarized below.

Findings that were consistent across stakes levels

- In all three states, educators noted positive, neutral, and negative effects of the state standards and tests on teaching and learning;
- The effects on educators were consistent, with elementary teachers as well as those in rural and large urban districts reporting the greatest impact on classroom practice, and suburban educators reporting the least;
- The reported effects on students were also consistent, with interviewees reporting a more negative than positive test-related impact on students, particularly elementary students, special populations, and students in urban districts.

Findings that varied across stakes levels

- As the stakes attached to the test results increased, the test seemed to become the medium through which the standards were interpreted;
- As the stakes increased, so too did the number of reported effects on classroom practice;
- As the stakes increased, interviewees reported a more negative impact on students, particularly elementary students, special populations, and students in urban districts.

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. 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.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10*(18). Retrieved March 28, 2002 from http://epaa.asu.edu/epaa/v10n18/.

Bishop, J., & Mane, F. (1999, Winter). *The New York state reform strategy: The incentive effects of minimum competency exams.* CEIC Review. Philadelphia: The National Center on Education in Inner Cities.

Bogdan, R., & Biklen, S. (1992). *Qualitative research for education: An introduction to theory and methods* (2nd ed.). Needham Heights, MA: Allyn and Bacon.

Cimbricz, S. (2002). State-mandated testing and teachers' beliefs and practice. *Education Policy Analysis Archives, 10*(2). Retrieved January 9, 2002 from http://epaa.asu.edu/epaa/v10n2.html

Clarke, M., Abrams, L., & Madaus, G. (2001). The effects and implications of high-stakes achievement tests for adolescents. In T. Urdan & F. Pajares (Eds.), *Adolescence and education: Vol. 1. General issues in the education of adolescents* (pp. 201-229). Greenwich, CT: Information Age Publishing.

Firestone, W. A., Mayrowetz, D., & Fairman, J. (1998, Summer). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis, 20*, 95-113.

Grant, S. G. (2001). An uncertain lever: Exploring the influence of state-level testing in New York state on teaching social studies. *Teachers College Record, 103*(3), 398-426.

Grissmer, D., Flanagan, A., Kawata, J., & Williamson, S. (2000). *Improving student achievement: What state NAEP test scores tell us.* Santa Monica, CA: RAND.

Hamilton, L., Stecher, B., & Klein, S. (Eds.). (2002). *Making sense of test-based accountability in education.* Santa Monica, CA: RAND.

Heubert, J., & Hauser, R. (Eds.). (1999). *High stakes: Testing for tracking, promotion, and graduation.* Washington, DC: National Academy Press.

Horn, C., Ramos, M., Blumer, I., & Madaus, G. (2000). *Cut scores: Results may vary.* Chestnut Hill, MA: National Board on Educational Testing and Public Policy.

Jones, G., Jones, B., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impact of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan, 81*(3), 199-203.

Kellaghan, T., Madaus, G., & Raczek, A. (1996). *The use of external examinations to improve student motivation.* Washington, DC: American Educational Research Association.

Klein, S., Hamilton, L., McCaffrey, D., & Stecher, B. (2000). *What do test scores in Texas tell us?* Santa Monica, CA: RAND.

Koretz, D., Barron, S., Mitchell, K., & Stecher, B. (1996). *Perceived effects of the Kentucky Instructional Results Information System* (KIRIS) (MR-792-PCT/FF). Santa Monica, CA: RAND.

Koretz, D., McCaffrey, D., & Hamilton, L. (2001). *Toward a framework for validating gains under high-stakes conditions* (CSE Technical Report 551). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Koretz, D., Mitchell, K., Barron, S., & Keith, S. (1996). *Final report: Perceived effects of the Maryland school performance assessment program* (CSE Technical Report 409). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Madaus, G., & Clarke, M. (2001). The adverse impact of high-stakes testing on minority students: Evidence from 100 years of test data. In G. Orfield & M. Kornhaber (Eds.), *Raising standards or raising barriers? Inequality and high-stakes testing in public education* (pp. 85-106). New York: The Century Foundation.

Madaus, G., West, M., Harmon, M., Lomax, R., & Viator, K. (1992). *The influence of testing on teaching math and science in grades 4-12.* Boston: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Massachusetts Department of Education. (2001). *Spring 2001 MCAS tests: State results by race/ethnicity and student status.* Retrieved May 30, 2002 from: http://www.doe.mass.edu/mcas/2001/results/re_ss.pdf

Massachusetts Department of Education. (2002). *Progress report on the class of 2003: Percentage of students who have earned a competency determination statewide and by district.* Malden, MA: Author.

Mehrens, W. (1998). Consequences of assessment: What is the evidence? *Education Policy Analysis Archives, 6*(13). Retrieved July 14, 1998 from http://epaa.asu.edu/epaa/v6n13.html

Miles, M., & Huberman, A. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* Washington, DC: U.S. Department of Education.

National Commission on Testing and Public Policy. (1990). *From gatekeeper to gateway: Transforming testing in America.* Chestnut Hill, MA: National Commission on Testing and Public Policy.

National Research Council. (1997). *Educating one and all: Students with disabilities and standards-based reform.* Washington, DC: National Academy Press.

Orfield, G., & Kornhaber, M. (Eds.). (2001). *Raising standards or raising barriers? Inequality and high-stakes testing in public education.* New York: The Century Foundation Press.

Pellegrino, J., Chudowsky, N., & Glaser R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.

Quality counts 2002 (2002, January 10). *Education Week, 21*(17).

Ravitch, D. (1995). *National standards in American education: A citizen's guide.* Washington, DC: Brookings Institution.

Raywid, M. A. (1987). Public choice, yes; vouchers, No! *Phi Delta Kappan, 68*(10), 762-769.

Resnick, L. (1996). *Performance puzzles: Issues in measuring capabilities and certifying accomplishments* (CSE Technical Report No. 415). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Shepard, L. (2002, January). *The contest between large-scale accountability testing and assessment in the service of learning: 1970-2001.* Paper prepared for The Spencer Foundation's 30th Anniversary Conference, "Traditions of Scholarship in Education," Chicago, IL.

Shepard, L., Glaser, R., Linn, R., & Bohrnstedt, G. (1993). *Setting performance standards for student achievement: A report of the National Academy of Education Panel on the Evaluation of the NAEP Trial State Assessment: An evaluation of the 1992 achievement levels.* Stanford, CA: National Academy of Education.

Smith, M., 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.

Smith, M., Noble, A., Heinecke, W., Seck, M., Parish, C., Cabay, M., Junker, S., Haag, S., Tayler, K., Safran, Y., Penley, Y., & Bradshaw, A. (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.

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.

U. S. Congress, Office of Technology Assessment. (1992). *Testing in American schools: Asking the right questions.* Washington, DC: U.S. Government Printing Office.

Wolf, S., Borko, H., McIver, M., & Elliott, R. (1999). *"No excuses": School reform efforts in exemplary schools of Kentucky* (CSE Technical Report 514). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

NOTES

Section 1

i Many calls for school reform assert that high-stakes testing will foster the economic competitiveness of the U.S. However, the empirical basis for this claim is weak. For a review of the research in this area see H. Levin (2001), High-stakes testing and economic productivity, in G. Orfield and M. Kornhaber (Eds.), *Raising standards or raising barriers: Inequality and high-stakes testing in public education*, (pp. 39-49), New York: The Century Foundation Press.

ii The theory of action underlying standards-based reform is further elaborated in R. Elmore and R. Rothman (Eds.), (1999), *Testing, teaching, and learning: A guide for states and school districts*, Washington, DC: National Academy Press.

iii The mail survey was based on five of the nine cells in the 3x3 grid (cells with one state or fewer were omitted from the design). Three of these five cells overlap with those used for the interview study. Thus, interview study findings could be cross-checked with those from the survey.

iv At the time of the project's inception some states had fully implemented testing programs, while in others the consequences attached to test results had yet to take effect. Therefore, a time dimension (fully implemented/not yet fully implemented) was also added to the grid.

v In order to preserve their anonymity, we do not provide demographics or other details for these districts and schools. In general, the urban districts in each state had the highest percentage of minority students, Limited English Proficiency students, and students receiving free/reduced price lunch (for that state). Surburban districts tended to have the lowest percentage of students in each of these categories (rural districts matched them on the percentage of minority students). Surburban districts tended to score above average on the state test, urban districts tended to score below average, and the rural districts varied.

Section 2

i At the time of this study, Kansas was the only one of the three study states whose standards had not undergone an external alignment review.

ii The Kansas Department of Education website has downloadable versions of each standards document as well as sample questions/tasks that students could be given to demonstrate attainment of these standards. The material also contains instructional suggestions for classroom teachers. During the writing of this report, the mathematics standards were being revised and only the draft document was available on the website; the version that was in place when this teacher was interviewed had been removed.

iii The research group Achieve conducted an evaluation of the state's standards and test and recommended clarifying the content standards and benchmarks in order to further facilitate alignment. The MI-CliMB (Clarifying Language in Michigan Benchmarks) project is addressing this issue (http://www.miclimb.net/).

iv The research group Achieve concluded that the degree of alignment between the state frameworks and tests in Massachusetts was very high. See Achieve, Inc., (2001), *Measuring up: A report on education standards and assessments for Massachusetts*, MA: Author.

Section 4

i LEP accommodations, including audio tapes in English, and a side-by-side Spanish/English version of the mathematics, science, and social studies assessments (for grades 10 and 11 only), were being developed and were supposed to be ready in time for the spring 2001 administration.

ii There are also data to show that white, Asian American, and female students, and those in wealthier communities, were awarded a disproportionate number of scholarships. See D. Heller and D. Shapiro, (2000), *High-stakes testing and state financial aid: Evidence from Michigan*, paper presented at the Annual Meeting of the Association for the Study of Higher Education, Sacramento, CA, November 16-19.

iii Students must obtain a score of 220 (answer 40 percent of the questions right) on each of the English and mathematics sections in order to pass the test (the score range is 200-280 on each). This score corresponds to the bottom of the *Needs Improvement* category on the test.

iv In a February 2001 release, the Massachusetts Association of School Superintendents outlined several principles that they believed should be incorporated in MCAS legislation. These included that "multiple measures including MCAS or an equivalent test should be used to determine a graduation requirement," that "special attention for graduation requirements must be given to special education students, vocational technical students, and students whose first language is not English," and that "the current criteria to pass MCAS in order to be admitted to a community college should be repealed [since] community colleges have historically served as a vehicle of mobility for those students (often poor and those who speak a language other than English) who develop skills more slowly."

v Some have accused the state of downplaying the impact of scoring changes on the 2001 exam results (state officials said the changes did not affect failure rates). Statewide, the percentage of tenth graders who passed the MCAS English test climbed from 66 percent in 2000 to 82 percent in 2001, and for mathematics from 55 percent to 75 percent.

vi As of May 2002, 75 percent of the original class of 2003 had earned their competency determination. However, while 76 percent of Asian and 83 percent of white students had passed the tests needed to graduate, only 47 percent of African American, 38 percent of Hispanic, and 57 percent of Native American students had done so. The discrepancies are even starker when broken down by student status. While 83 percent of regular students had passed the graduation test as of May 2002, only 42 percent of students with disabilities and 18 percent of LEP students had passed (Massachusetts Department of Education, 2002).

vii These pass rates were calculated by using the spring 2001 test/retest figures reported in Massachusetts Department of Education (2002). The difference reported here should be considered conservative because we could only compare students in voc-tech schools with students in all other high schools; however, many large high schools have voc-tech programs and these students were only counted as traditional students. In addition, we did not use the fall 2001 enrollment figures that, when used, significantly drop the pass rates for most schools, particularly those in urban areas, indicating that large numbers of students left those schools either because they moved or because they dropped out.

viii Massachusetts Department of Education Commissioner's "Back to School" Update, August 22, 2002.

APPENDIX A
3X3 GRID OF STATE TESTING PROGRAMS

CONSEQUENCES FOR STUDENTS

CONSEQUENCES FOR TEACHERS, SCHOOLS, AND DISTRICTS

	High	Moderate	Low
High	Alabama California* Delaware* Florida Georgia* Indiana* Louisiana Maryland* Massachusetts* Mississippi* Nevada New Jersey New Mexico New York North Carolina South Carolina Tennessee Texas Virginia*	Arkansas Connecticut Illinois Michigan Pennsylvania West Virginia	Colorado* Kansas Kentucky Missouri Oklahoma* Rhode Island Vermont*
Moderate	Arizona* Alaska* Ohio Minnesota Washington* Wisconsin*	Oregon	Hawaii Maine Montana Nebraska New Hampshire North Dakota South Dakota Utah* Wyoming
Low	Idaho*		Iowa

**Indicates that the program was not fully in place at the time of this study.*

APPENDIX B
INTERVIEW PROTOCOL

Note: This protocol was used for teacher interviews in Kansas. The same protocol, with appropriate modifications in terminology, was used for administrator interviews in Kansas and for both teacher and administrator interviews in Massachusetts and Michigan.

This conversation is an opportunity for you to share your views about the implementation of your state-mandated testing program — how it has affected both your thinking and your practices. This is a list of topics that I'd like to cover [hand out list].

If possible, I'd like to cover the topics in order, but feel free to raise any issues that occur to you.

[Explain how confidentiality will be preserved – no teacher, school, or district will be identified in our reports. Ask permission to tape.]

[Ask the main question for each section, then prompt as necessary. Skip follow-up questions that have already been answered (unless you want to ask for elaboration).]

1. In your opinion, how have educational reform efforts in Kansas affected teaching and learning in your classroom?

- **A.** More specifically, in what ways, if any, have the state's established curricular standards (i.e., the Kansas Curricular Standards) affected what you teach?
- **B.** In what ways have the curricular standards affected how you teach?
- **C.** What, in your opinion, is the function of the curricular standards? What purpose do they serve?
- **D.** What do you see as the relationship between the curricular standards and the Kansas Assessments?
- **E.** What, in your opinion, is the function of the Kansas Assessments?
  - Do they seem to serve that function?
  - Is that function or purpose an appropriate use of a test?

2. In what ways, if any, have the Kansas Assessments affected what you teach? (If none, go to B)

- **A.** What, if anything, have you added to or eliminated from your curriculum to prepare students for these tests?
- **B.** When, if at all, do you receive your students' test scores? What do you do with that information?
- **C.** The state argues that the Kansas Assessments are intended to reflect the curricular standards and measure attainment of those standards.
  - To what extent do you think that the Kansas Assessments adequately and accurately fulfill those purposes?

- More specifically, do you think that students' test scores on the Kansas Assessments accurately reflect how well they have mastered the curricular standards?
- Do your students' scores on the Kansas Assessments reflect actual learning?

3. In what ways have the Kansas Assessments affected how you teach?

A. Do you spend classroom time preparing your students specifically to take the Kansas Assessments?

- Explain what you do to prepare your students specifically for the test.
- Do you give students practice questions? Go over the test format with them? Teach them test-taking skills?
- Can you estimate how much time you spend on this preparation?
- Does the amount of preparation time vary throughout the year? How?
- How does your preparation of students for the Kansas Assessments vary, if at all, from what you normally do to prepare students for a classroom test? To prepare students for a commercially developed standardized test (e.g. SAT 9)?
- Why do you choose to spend time preparing your students for the test? What motivates you to allocate your time in this way: a desire to see your students do well? Fear of sanctions if your students don't do well? Other incentives?

B. In what ways, if at all, have the Kansas Assessments affected the way you assess your students?

- Are assessment results given to teachers? When?
- Are the assessment results helpful to you for assessing what students in your classroom know and can do? How?
- Do you think they are helpful for assessing:
  - school-level performance? How?
  - district-level performance? How?
  - state-level performance? How?
- What is being done to help you understand the results of the Kansas Assessments?

4. In what ways, if any, have the Kansas Assessments affected your students? Or how, if at all, are students in your class affected by the testing?

A. How have the state-mandated assessments affected students' motivation to learn, if at all?

B. Have the Kansas Assessments affected student morale? In what ways and for whom?

C. Do you think the Kansas Assessments are appropriately suited to your students in terms of:

- Their content?
- Their format?
- The presence or absence of specific accommodations?

D. How do your students do on the Kansas Assessments?

E. Have the Kansas Assessments affected the number of students who have been retained in grade? Have dropped out? Have been required to participate in summer school? Have the results been used to group students (i.e., tracking)? In what ways?

5. In what ways, if any, have the Kansas Assessments affected the ways in which your school spends its time and money?

A. In terms of money resources, has funding been shifted among departments in your school to accommodate test preparation? In what ways?

B. In terms of time resources, have courses been added or dropped from the schedule in order to prepare students for the Kansas Assessments? In what ways?

C. To the greatest extent possible, explain how you think the Kansas Assessments have affected the ways in which your DISTRICT spends its time and money.

D. In your opinion, should Kansas Assessment scores be used as part of the state's resource allocation decisions for districts? For schools? Why? If yes, how?

6. Based on your personal experiences with teachers at your school, what, if any, are the effects of the Kansas Assessments on the profession of teaching?

A. State-mandated tests are sometimes viewed as a means of ensuring teacher accountability. In your opinion, is this appropriate?

B. Do you believe that the Kansas Assessments are high-stakes tests for teachers? How?

C. In your view, what effect have the Kansas Assessments had (or will they have) on the public's perception of teachers? Is the public's perception accurate? How so?

D. Do you think that these state-mandated assessments have had an effect on teacher recruitment, retention, and retirement at your school? How? In which grades? Is there a particular reason for that?

7. Are there other questions or issues related to the state-mandated standards or testing program that you would like to discuss?

APPENDIX C
METHODOLOGY

Access

The process of gaining access began in Massachusetts in winter 2000 and in Kansas and Michigan in early 2001. A similar procedure was used in all states. First, a high-level administrator was contacted and help requested with gaining access to a previously identified set of districts. These districts were selected using U.S. Census and state data. Once at the district level, access was requested to six schools that varied in student demographics and performance on the state test. The final choice of schools was based on these criteria, suggestions made by the district superintendent, and the number of schools available in the district (e.g., a rural district might have only one high school). Each school principal was then contacted and on-site interviews were arranged.

The Interview Process

Interviewers were faculty, researchers, and doctoral students from the Lynch School of Education at Boston College. Most had experience as classroom teachers. Before conducting the interviews, the interviewers went through training and were involved in pilot testing of the protocol. Nine interviewers conducted 360 on-site interviews in Kansas, Michigan, and Massachusetts between winter 2000 and fall 2001. Interviews took between 30 minutes and two hours and were taped unless the interviewee requested otherwise (this rarely occurred). Confidentiality was promised to all interviewees. Follow-up telephone interviews with a subset of interviewees were conducted in late spring 2002. One of the main goals of these follow-ups was to clarify three terms that had come up in the original interviews: (1) teaching to the test, (2) preparing students for the test, (3) and teaching to the standards.

Data Summarizing, Coding, and Analysis

Since the cost of transcribing the taped interviews was prohibitive, a data summarizing process was developed to record the information gained from each interview. The researcher listened to the interview tape and extracted the chunks of conversation that pertained to the topic of this study. These were then recorded as either a "quotation" or a "point" on an Interview Summary Sheet. Quotations are the interviewees' statements or replies; points are summaries, in the researcher's words, of the interviewees' opinions. Interviewers were trained in this technique and required to write up interviews accordingly. An assessment of the interviewers' consistency and accuracy in extracting information from the taped interviews showed good results.

Three team members who had each conducted a substantial number of interviews used an iterative inductive process (Bogdan & Biklen, 1992) to construct a code list. Initial codes were generated by using the interview protocol questions as well as some of the interview write-ups. The list was further refined through input from the larger research group. The final code list contained 263 codes, organized into eight topic areas that mirrored the structure of the protocol. In order to avoid missing unusual opinions or insights expressed by individual interviewees, a "catch-all" code was included under each topic area on the code list. Team members were trained in how to code the interview write-ups using this system. A study of the consistency among coders on codes used and pieces of text that should be coded showed high levels of agreement. In addition to this consistency study, weekly debriefing meetings were held throughout the coding process and some interviews were double-coded to ensure consistency.

Interviews were coded and analyzed using a qualitative data analysis program called HyperRESEARCH (http://www.researchware.com/). This program allows for the coding and retrieval of pieces of interview text within and across interviews, frequency counts of code use within and across interviews, and theory building and hypothesis testing. Interviews were coded for demographic (e.g., district type, school type, grade level) as well as thematic information (interviewee perceptions and opinions). After coding, themes were identified first by frequency counts of codes and then by inductive analysis of the coded interview output. By using this approach, we were in agreement with Miles and Huberman's (1994, p.40) view that "numbers and words are both needed if we are to understand the world." This seemed particularly appropriate given the large number of interviews involved and the attendant difficulties of trying to discern patterns within and across interviews. Themes were checked in terms of whether they held up across subject areas, grade levels, school types, district types, and states. Sub-themes (those mentioned by less than ten percent of interviewees) were also identified, and some of them further checked for validity using a member-checking procedure. Specifically, we conducted follow-up telephone interviews with a representative sample of 40 of the original interviewees in May 2002.

NBETPP

About the National Board on Educational Testing and Public Policy

Created as an independent monitoring system for assessment in America, the National Board on Educational Testing and Public Policy is located in the Carolyn A. and Peter S. Lynch School of Education at Boston College. The National Board provides research-based test information for policy decision making, with special attention to groups historically underserved by the educational systems of our country. Specifically, the National Board

- Monitors testing programs, policies, and products
- Evaluates the benefits and costs of testing programs in operation
- Assesses the extent to which professional standards for test development and use are met in practice

This National Board publication is supported by a grant from The Atlantic Philanthropies Foundation.

The National Board on Educational Testing and Public Policy
Lynch School of Education, Boston College
Chestnut Hill, MA 02467

Telephone: (617)552-4521
Fax: (617)552-8419
Email: nbetpp@bc.edu

BOSTON COLLEGE

Visit our website at **nbetpp.bc.edu** for more articles, the latest educational news, and information on NBETPP.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

X	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").