State Testing and Inquiry Based Science: Are They Complementary or Competing Reforms?

Joni Falk and Brian Drayton

TERC

This is a working-paper version
Final version published in:
Journal of Educational Change 5:4 345-387, 2004

Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

Abstract: The effect of district strategies for improving high-stakes test scores on science teachers' practice is explored in case studies of six middle schools in six Massachusetts districts. At each school, science teachers, curriculum coordinators, principals, and superintendents shared their strategies for raising scores, their attitudes towards the test, the changes that they were implementing in their curriculum and pedagogical approaches, and the effects that the test was having on staff and on students. Results from these case studies suggest that districts chose markedly different strategies for raising scores on high stakes tests, and that the approaches taken by districts influenced the nature of pedagogical and curriculum changes in the classroom. District strategies for raising scores that were complementary to the district's prior vision of science reform tended to cause less teacher resentment towards the test than strategies that departed from previously adopted goals. Differing effects on teachers in socio-economically "advantaged," "middle," and "challenged" districts are discussed.

Key Words: Accountability, Educational Reform, High Stake Tests, Inquiry, Middle Schools, Middle School Teachers, Science Education, Science Curriculum, Standardized Test

Abbreviations: ITBS: Iowa Tests of Basic Skills, MCAS: Massachusetts Comprehensive Assessment System, MAT7: Metropolitan Achievement Test, NRC: National Research Council, NSF: National Science Foundation, PALMS: Partnerships Advancing the of Learning of Mathematics and Science (this is the name of the Massachusetts SSI), SSI: State Systemic Initiatives

Introduction

The current popularity of high-stakes accountability in education follows several decades of innovation and experimentation in science education. Although a student-oriented, inquiry-based approach has not yet come to characterize the majority of science classrooms (Kohn, 2000; Weiss et al., 2001), it has generally become the "gold standard." Policy documents at every level, from district to nation, encourage this approach, and textbooks, however traditional, make at least a statement of its value for authentic science learning (Mass. Dept. of Ed., 1996; NRC, 1996; Worcester Public Schools, 1998).

It has been well-documented that the inquiry-based approach, carefully nurtured and steadfastly defended by the National Science Foundation (NSF), the National Research Council (NRC), and other agencies, is complex. It is very demanding on the teacher as well as the student, and requires deep changes in classroom, school, and district culture. Such an approach requires an increased emphasis on student reflection, on student ownership of the scientific questions probed, on conducting investigations, on data collection, and on data analysis. Inquiry pedagogy favors depth over breadth, and conceptual understanding over an emphasis on science facts and technical words to describe these facts (National Research Council, 2000, pg. xii).

This conception of science education was embedded in the Massachusetts science curriculum frameworks developed in 1996 and was at the core of PALMS, the Massachusetts State Systemic Initiative (SSI), which was the state's most visible intervention in science education in the mid 1990s. In 1998, however, the state implemented the high-stakes Massachusetts Comprehensive Assessment System (MCAS), which was intended to reinforce the curriculum frameworks. The MCAS exam, during the years 1998, 1999, and 2000, was administered to students in grades 4, 8, and 10 in each school district. The scores for each school and district were published in the newspapers.[1] As of 2003, students in Massachusetts are required to pass the MCAS in Math and English Language Arts as a high school

[1] Despite the availability of school scores to the public, we have preserved anonymity of the six districts reported here in our study; all names have been changed.

Final version published in: Journal of Educational Change 5: 345-387, 2004
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

graduation requirement. During this study, teachers and administrators in many districts were under the mistaken impression that students would need to pass the Science section as well; this requirement, however, has been put on hold.

Defenders of the test claim that it serves to hold teachers as well as students accountable for a core curriculum and promotes the inquiry-based science learning mandated in the state standards. Those who oppose the MCAS say that it demoralizes students and teachers; forces science back towards a fact driven, "mile wide and inch deep" curriculum; and stands as an obstacle to a more open-ended, inquiry approach to learning. This debate mirrors controversies taking place in states across the country (Haney, 2000; Heubert & Hauser, 1999; Jones et al., 1999; Kohn, 2000; Smith, Heinecke, & Noble, 1999), and for this reason Massachusetts is a valuable setting in which to understand the dynamics of this kind of reform.

This paper grows out of findings from the first two years of research in an NSF-funded project, "The Inquiry Based Classroom in Context" (REC-9804929). The main purpose of this research was to study how middle-school science teachers in districts involved for at least four years in the Massachusetts State Systemic Initiative had come to interpret inquiry in theory and practice. The project focused on 40 science teachers in six Massachusetts middle schools. At the same time that this research began, districts, schools, and teachers started to formulate an opinion of, and response to, the newly introduced MCAS. Our presence in these schools at this time allowed us to gain insight into the influence of the MCAS on the existing vision of science reform in these schools. Through interviews, classroom observations, and examination of student work a complex picture emerged of how district responses to the MCAS were effecting changes in curriculum and pedagogy. In addition, we heard from teachers, principals, and superintendents about the impact that they felt MCAS had on their science staff and on their students.

In this paper, we address the following research questions:

1. What approach, if any, has each district taken to respond to the pressures to boost MCAS scores?
2. To what degree have district factors (including articulated MCAS strategy, district vision of science reform, and district demographic factors) influenced individual teachers' responses to the MCAS?
3. What changes (if any) in curriculum or pedagogical approaches are being made in the case study schools as a result of the introduction of the MCAS exam?
4. Is MCAS seen to be complementary to or competing with an inquiry-based approach to science education?
5. How do teachers describe the effect that MCAS is having on science staff?

Study Framework: Taking a Contextualized View of Complex Reform

In our study of teachers' interpretation and implementation of inquiry-based science, we conceptualized the classroom's systemic context as in Figure 1. The state sets educational policy through frameworks, graduation requirements, teacher certification, and the allocation of resources in line with these major policy statements. In addition, Massachusetts conveyed a strong message about the direction of reform through the State Systemic Initiative (SSI) program. In order to transmit this message, regional centers were established by the SSI to provide teacher professional development around innovative math and science curricula to all Massachusetts districts.

Under the SSI program the district became an important unit of change and intervention. Districts interpret policy mandates in multiple ways, for example by allocating money for teacher in-

Final version published in: Journal of Educational Change 5: 345-387, 2004
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

service courses and curriculum centers, by supporting (or prodding) teachers through curriculum supervisors and cross-district curriculum meetings, and by staffing decisions. When we speak of "district culture," we mean both the multiple channels of communication and influence by which district leaders convey to teachers and administrators the current policy priorities, and the means available to implement them, as well as the other kinds of communication and consensus that take place at the district scale (Falk & Drayton, 2001; Spillane & Callaghan, 2000; Spillane, 1996).

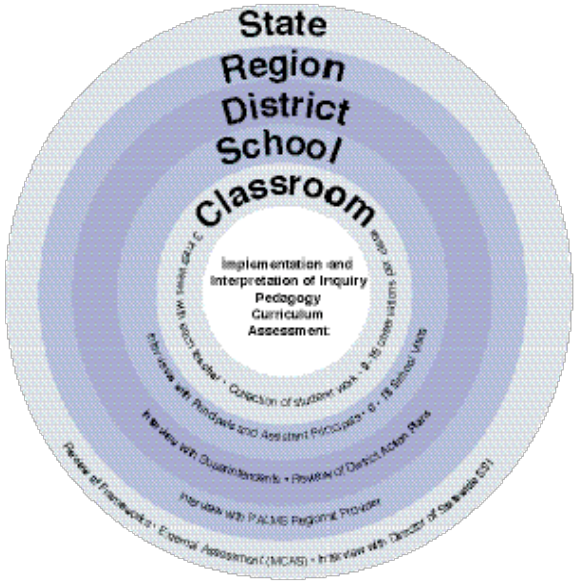

Figure 1. Data sources reflecting a contextualized view of reform.

Evidently, teachers' most intense cultural involvements are within the school and the classroom. In fact, one can say that although the classroom is the smallest unit spatially, it is the place where the most factors are at work. The teacher bears in mind the mandates and priorities of the school, district, and state leadership, but they also are engaged with the youth who are the school's reason for existence, and through the students the teacher encounters many aspects of the community within which the school is situated.

Thus our study is situated in a broad and dynamic view of classroom, school, and district culture. As we spoke with teachers and administrators and observed in their classrooms and schools, we were always seeing evidence of active interaction among the many levels of educational organization and culture. It is true that power relations — what is permitted, who is empowered to make decisions about hiring, resources, etc. — are key elements of a school or district culture (Sarason, 1996). However, in a complex reform which requires both logistical adjustments and considerable learning by practitioners, other factors are also important in shaping the atmosphere: people are aware of, and responsive to, the flow of information, resources, opinions, intentions, and objectives among their colleagues, superiors, subordinates, students and the wider community (Eisner, 1992; Hargreaves, 1999, 2001; Cuban, 1998; Maehr & Midgley, 1996).

Final version published in: Journal of Educational Change 5: 345-387, 2004
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

Review of the Literature

Research on High Stakes Testing and Classroom Effects

Studies have found that high-stakes testing has resulted in an emphasis on increased coverage of disparate topics, narrowing the range of instructional practices, test preparation, and increased use of drill and practice (Haney, 2000; McNeil, 2000; McNeil & Valanzuela, 2001; Smith, 1991; Smith & Rottenberg, 1991). Other studies suggest that high stakes tests exert pressure to teach within traditional subject domains, and discourage integrative pedagogies of the sort considered most desirable for middle-school education (Powell & Skoog, 2000, and see Firestone & Mayerowitz, 2000). Head-on preparation for the test, focusing on test-taking skills and drill in the "basics" at the cost of higher-order skills is a common response, especially in schools with a high proportion of low-achieving students (McNeil & Valanzuela, 2001). Studies have shown that teachers in high-minority classrooms, under pressure to improve student test performance, are more likely to teach test-taking skills, to increase emphasis on specific tested topics, to add topics extraneous to the curriculum just in case they might be tested, and to dedicate a large proportion of time to test preparation (Haberman, 2003; Madaus & Clarke, 2001; Settlage & Meadows, 2002; Cheng & Coutre, 2000). This response is common despite studies that have shown that a progressive, inquiry-oriented pedagogy is at least as good a preparation for proficiency tests as specific test-preparation (Newmann, Bryk, & Nagaoka, 2001).

Of further concern are large-scale studies that show that student gains on state tests (in almost all cases) do not generalize and are not transferable to other standardized tests, raising a question whether genuine learning has taken place (Amrein & Berliner, 2002; Koretz et al., 1991; Smith, 1991). In addition, Roderick and Engel (2001) have shown that the motivational response of low-achieving students varies considerably, suggesting that such tests do not provide the reliable incentive that proponents have claimed. In this connection, the social costs within the classroom are of concern (Stiggins, 1999). Some scholars contend that because student scores are used as criteria for sanctions on schools or teachers, as well as students, high stakes can have the effect of creating a sense of antagonism between teachers (who may feel their fate is unfairly determined by their students' scores) and students who can come to see the teachers as "cogs in a large accountability machine" (Settlage & Meadows, 2002).

In a more positive light, some studies have identified positive classroom effects such as emphasis on writing across the curriculum, use of graphic organizers, curriculum becoming more well defined, teachers having more accurate data on their students, and teachers' belief that student skills were improving (Abeille & Hurley, 2000). A study that examined the potential for authentic performance assessments to positively influence instructional practices in mathematics in Maine and Maryland found changes in curriculum but not in instruction. The authors concluded that the "effects of high stakes testing may be overrated by both advocates and opponents" (Firestone, Mayerowitz, & Fairman, 1998).

High Stakes Testing and Its Bearing on Science Education

Science education must deal with multiple fields of knowledge, and involves both many kinds of content and the acquisition of a complex set of reasoning and data collection skills (Drayton & Falk, 2001). In addition, recent science frameworks and standards have placed heavy emphasis on investigations, use of materials and scientific tools, analysis, and reflection, all of which impose a significant cost in teacher learning, and place demands on school structures such as scheduling and ability grouping which require time and experimentation to fully adjust (Drayton & Falk, 2001). Because of these characteristics of science education, a reform such as high stakes testing may well have a distinctive impact on efforts to reform science teaching and learning.

However, studies that focus on the effects that high-stakes examinations have on *science* classroom curriculum, pedagogy and assessment practices are still scant. Even the important recent studies by Pedulla et al. (2003) and Clarke et al. (2003), which explore differences in teacher attitudes, do

Final version published in: Journal of Educational Change 5: 345-387, 2004
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

not provide data of high enough resolution to understand effects on science education. Borman et al. (2000), and Kersaint et al. (2001) have made inroads in this direction by examining the relationship between standards based reform, constructivist pedagogy, and principal leadership. In one of the rare investigations of this topic, a study of the effects of high stakes tests on science instruction in British Columbia showed that in grade 12, the year of the exit exam in that province, teachers felt a significant negative impact on the range and flexibility of pedagogy they could employ (Wideen & O'Shea, 1997).

High-stakes Testing in the Context of Science Education Reform in Massachusetts

During most of the 1990s, the focus of science education reform at the district and school level in Massachusetts was on embedding inquiry-based science in classroom practice. This conception of science education was at the core of PALMS, the Massachusetts State Systemic Initiative for science and math reform. PALMS placed the emphasis on students' ownership of their learning, their engagement with problems and open-ended investigations, and their learning to reason with real-world data. The first Massachusetts Science Curriculum Frameworks (Mass. Dept. of Ed., 1996) were seen to be very much in harmony with the principles advocated by PALMS. Many districts responded by developing new curriculum frameworks, and aligned their curriculum materials and teacher professional development programs with the curricular changes.

The Commonwealth's Department of Education introduced the Massachusetts Comprehensive Assessment System (MCAS) exam in 1998 in order to hold students, teachers, schools and districts accountable to the Massachusetts frameworks. The MCAS included exams in Math, English, Social Studies and History, and Science and Technology/Engineering. The Commonwealth and individual districts have begun to use the student scores as indicators of school or teacher quality, and as part of a rating system that includes rewards and sanctions to schools and districts, thus increasing the stakes (Goertz & Duffy, 2001; Mass. Dept. Ed., 2000). At present, a student must have a passing score on both the English and Mathematics 10th grade tests to receive a diploma upon graduation. The Science test is not yet included in the graduation requirements, but a science graduation requirement is expected to be implemented within the next few years (the date not fixed; Mass. Dept of Education, 2003); hence, the stakes will continue to rise.

The science test combines both open-response (about 30% of the 8th grade questions) and multiple-choice questions (70%), and includes questions that bear on the scientific process, experimental design, and data analysis. Besides being asked to write clearly on scientific topics, students are expected to retain a wealth of facts that they have learned during the past several years. Districts which had spent much of the 1990s reshaping their science program to favor curricular depth over breadth and to incorporate student investigations, as advocated by PALMS and the NRC, were now reconsidering whether or not an inquiry-based approach would best prepare their students for the science MCAS exam.

Study Sites

The data for this study are derived from a research project that has been studying middle-school teachers' interpretation and implementation of the mandate for "inquiry" in the science classroom in 6 districts that have been involved with the State Systemic Initiative (SSI) for at least 4 years. The districts include two large urban, two small urban, and two suburban districts (Names of all districts and schools have been changed to protect confidentiality; demographic data are provided in Table 1).

Of our case study districts, the two suburban districts (one from "Middle MA," and one from "Advantaged MA") performed significantly above state average on the science MCAS. One small urban

Final version published in: Journal of Educational Change 5: 345-387, 2004 © 2004 Kluwer Academic Publishers
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

district ("Lower Middle MA") performed at the state mean, and three urban districts (all from the "15 Most Challenged MA districts") performed below the state mean. As shown in Figure 2, the scores decline in direct correlation with median income (percentage failure varies in inverse proportion to median household income).

The correlation between economic factors and MCAS scores should not be surprising. It has been well established by Gaudet (1999, 2000) that demographic characteristics that include average education level, average income, poverty rate, single-parent status, language spoken, and percentage of school-age population enrolled in private schools, accounted for 86% of the variation in score of the 1998 MCAS and 83% of the variation in the 2000 MCAS scores. This is consonant with other studies correlating student achievement with SES factors (Hoover, 2000; Robinson & Brandon, 1994; Rothstein, 1998).

Town (pseudonyms)	Approx. Median Income (from 1990 census)	% parents with BA or higher	District type	Gaudet community classification[2]
Harris	$49,000	39	Suburban district	Advantaged MA
South Allenville	$42,000	23	Suburban district	(middle) Middle MA
Allenville	$37,000	20	Small urban center	(lower) Middle MA
Wilcox	$36,000	21	Large Urban City	Challenged MA (one of 15 most challenged urban centers)
Stanton	$26,000	15	Large Urban City	Challenged MA (one of 15 most challenged urban centers)
Hendon	$23,000	15	Small Urban Center	Challenged MA (one of 3 Most Challenged in the state)

Table 1. *Median income, educational attainment, and demographic types of case study schools.*

[2] Gaudet (2000) has characterized Massachusetts districts with reference to important SES factors such as median income, average parental educational attainment, and linguistic diversity, as "Advantaged," Middle MA," "Challenged," and "15 Most Challenged."

Final version published in: Journal of Educational Change 5: 345-387, 2004
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

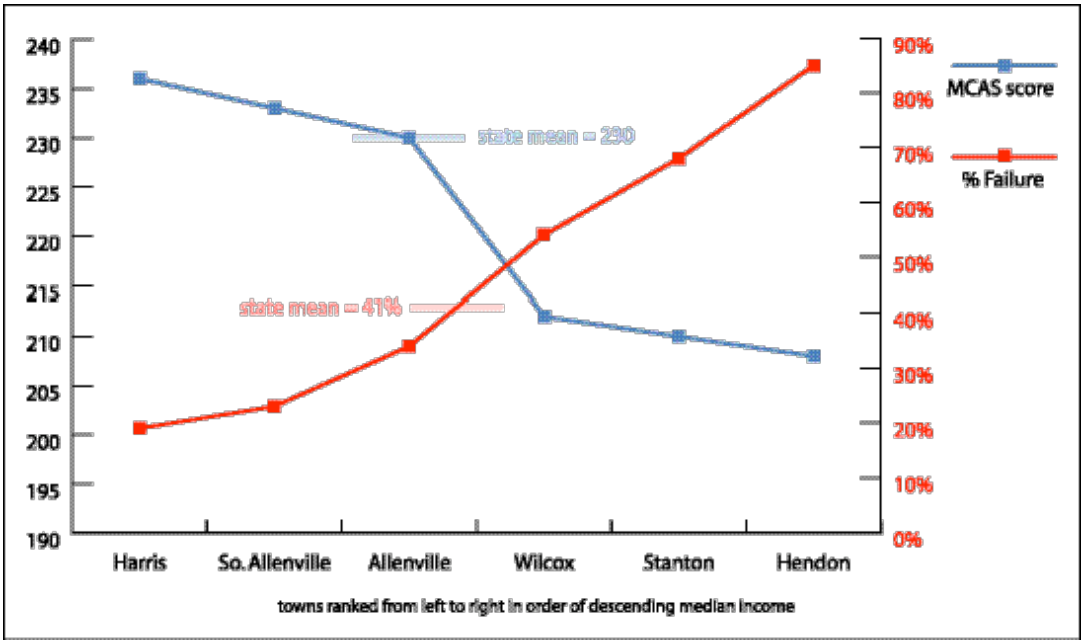

Figure 2. 2000 MCAS 8th grade science scores and % failure for case study districts.

Methodology

Data for this study was drawn from observations of teacher practice, interviews conducted with staff at every level of the system from Superintendent to classroom teacher, and artifacts collected from the state, district, school and classroom. These data enabled us to build a landscape in which to interpret the teachers' views, values, and practice.

Interviews and documentation.

In the six schools studied, we interviewed a total of 40 middle school teachers (of grades 6, 7, and 8). Each teacher was interviewed on three separate occasions, at the beginning, the middle, and the end of our observational period in their classrooms, for a total of 120 teacher interviews. In addition two newly hired, part-time teachers were interviewed once. We also collected documentation on textbooks or other curriculum materials used, and collected and discussed with the teachers samples of student work. In addition we interviewed all principals and assistant principals, science curriculum coordinators, regional providers of the SSI, and the superintendents of each district. All interviews were audio-taped and transcribed, and analyzed and coded by two researchers. Transcripts of interviews were coded for positive and negative reactions to the MCAS, perceived MCAS effect on students and on teachers, and MCAS effect on curriculum and pedagogical practices.

Observations.

We observed classrooms for 4-8 weeks in each school, for a total of 320 classroom periods. Each teacher was observed for an average of about 7.5 days. Classroom observers wrote field notes, and also recorded data on a variety of teacher and student variables on protocol sheets. Two observers were present for over 70 percent of the observations. Formal observation protocols (quantitative data) as well

Final version published in: Journal of Educational Change 5: 345-387, 2004
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

as observation field notes were analyzed by a team of researchers. Inter-observer reliability was over 80%.

Data analysis.

Our data analysis took a continuous comparison approach, so that as data were collected we developed hypotheses and theories which drove further analysis, and informed observations and interviews (Miles & Huberman, 1994). While formal research methods enabled us to look for generalized themes and responses, our dialogue with teachers, and their reflections on their practice, contributed crucially to the validity of the study's conclusions. As we gathered information and analyzed our findings, we were able to involve teachers in a dialogue about our work, our observations, and our understandings about their ideas, practice, and school culture. We also provided a formal time for review with teachers (Denzin & Lincoln, 1994; Patton, 1986; 1987; Stake, 1995).

The first round of data analysis took the form of narrative summaries and initial analyses of data from each school, drawing on and reconciling all data sources. Research team members prepared 10-20 page reports on each teacher, addressing four general areas, in each of which data were obtained relating to high-stakes testing in Massachusetts:

1. Interpretation of Inquiry
2. Conflicts and strategies (including the MCAS)
3. Personal, school-cultural and systemic influences
4. Coherence

On the basis of the qualitative narratives and the observational data, we developed two kinds of formal reports to use in member-checking with teachers: confidential reports for each teacher, and school reports on variations in practice and attitudes across the school, as well as cultural and demographic features. While the teacher report was only given to the teacher, the school report with aggregated school data (with individual teacher anonymity preserved) was given to teachers as well as curriculum coordinators, principals, and superintendents of each district. Teachers and administrators were asked for feedback in meetings with the investigators, and were also provided with forms for written feedback on the reports. These meetings and feedback forms provided a way to check the validity of our understandings of practice, attitudes, and process in each school and classroom.

Results

To explore the impact that the introduction of the MCAS was having on classroom curriculum and pedagogy, on inquiry-based science, and on students and teachers, and how that impact was mediated by district and school culture, we present short case studies of the six study districts. While there was variation between teachers' reactions to the MCAS exam and to the effect that it had on their teaching, both within a district and between school districts, we found a strong district effect. Hence, teachers within a district tended to have more similar responses to the test than teachers in different districts. The categories in which district-to-district differences were most notable were: teachers' assessment of the positive value of the test, the degree to which the test caused alterations in classroom curricular and pedagogical practices, teacher's evaluation of whether the MCAS was complementary or destructive to inquiry based science, and the degree to which the test affected teachers' morale and love of teaching.

There were certain topics on which there was broad agreement among almost all teachers regardless of school or district. These included teachers' reservations about the test's content, level of

Final version published in: Journal of Educational Change 5: 345-387, 2004
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

difficulty, and length as well as concerns about negative effects that failing would have on students' morale.

Areas of broad consensus across all six districts.

Teachers in all six districts felt that the test was too difficult. Comments such as "the test has unrealistic expectations for the kids of this age" and "some of the questions are ridiculously hard" were common. In addition teachers felt that it took a "shotgun approach to multiple topics" and required students to remember facts that they were taught two and three years earlier. In addition to the difficulty and breadth of coverage of science topics teachers from all districts commented that the reading level, especially the vocabulary and difficult phraseology, impeded student success.

Another obstacle that teachers across all districts frequently mentioned was that the "test is just too long for the students," "it [the testing period] goes on for two solid weeks, and by the time students get to the science section they have had it." In some schools other exams such as the MAT7 or ITBS follow or precede the MCAS, making the testing period even longer.

Teachers in all districts were concerned about the effect that the test was having on students' morale, especially for those who would receive a failing grade. Teachers reported noticing increased anxiety among students before the test. Some students erroneously believed that they would not graduate 8th grade if they failed. In one high performing school teachers commented that they had "A" and "B" students who did poorly on the test and they were worried about how the test would influence their students' self-esteem. Teachers at low performing schools were worried about the effect that the test would have on those who failed. There was particular concern expressed about students who had special needs, vocational students, and students who spoke English as a second language.

Case Studies

The six districts studied differed markedly in their attitudes toward the test and in the approach that they advocated to raising student scores. Likewise, the test had markedly different effects on staff from district to district. Schools that had higher baseline scores (Harris, South Allenville, and Allenville) had a clearly articulated strategy for raising test scores. The three lower-performing, challenged urban districts seemed still to be in the midst of figuring out a strategy.

All six districts had instituted after-school, summer school or weekend tutoring/remediation to address the test. This study, however, did not observe out-of-class remediation programs and its findings are limited to changes that were taking place within the classroom itself as a result of the MCAS. In this regard the factors that were considered in this study were changes in curricula and pedagogical approaches, the positive or negative values teachers felt these changes were having on their teaching, the effects on the implementation of inquiry-based science, and the effects teachers felt the test had on their own self-esteem.

Three Case Studies of Advantaged and Middle Income Districts: Harris, South Allenville and Allenville

Harris, South Allenville and Allenville are the demographically more advantaged districts in the study. Gaudet (see Table 1) classifies Harris as Advantaged and South Allenville and Allenville as Middle MA. Each of these districts had markedly different responses to the MCAS. Harris, which had never truly adopted an inquiry approach, concentrated on changing the scope and sequence of their science instruction to match the frameworks. South Allenville changed their prior emphasis on inquiry to

Final version published in: Journal of Educational Change 5: 345-387, 2004 © 2004 Kluwer Academic Publishers
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

a new emphasis on “back-mapping from the test.” In contrast, Allenville did not deviate from their inquiry reform efforts and instead deepend their committment to this reform.

School #1: Harris: primarily a reform of scope and sequence.

Harris is a suburban, fairly wealthy, unified school district; the town has a population of 14,428 with a median income of about $49,000. Gaudet considers it as part of Advantaged Massachusetts (see Table 1). Students in this study school generally score well on national assessments. Although this is a middle school, it is structured like a junior high, with each teacher teaching essentially the same material five times a day. The teachers are in teams, but the teams are large (one teacher per core subject area) and there is little flexibility in the schedule. The science faculty at the time of the study was composed of six science teachers, with a strong representation of veteran teachers. Years of service ranged from 3 to 34, with a mean of 25.7; of the six teachers, 5 had taught for 25 years or more; there was only one relatively young, new science teacher. There had been four principals over the previous six years; whether as a response to, or a cause of, this administrative turnover, these veteran teachers operated with a great deal of autonomy.

In general, Harris's teachers had high expectations for their science students in terms of learning and retention of factual material. Classrooms tended to be teacher-centered, relying to a large degree on content lecture (64% of class sessions observed). Students spent more time listening to lectures here than in the other two districts that performed at or above the state mean (South Allenville and Allenville). Students in Harris spent more time taking notes than did students in any of the other schools (see Figure 3). The lectures or "review discussions" were frequently quite lively and entertaining and the students often maintained a high level of engagement. Students here were less likely than students in either South Allenville or Allenville to be working in groups or pairs, or to be giving student presentations.

Although we observed hands-on investigations (35% of class sessions observed) they were usually brief in duration, and tasks and methods were usually defined by the teacher. Students were quizzed and tested frequently, usually with multiple-choice, matching, and short-answer questions. Despite the richness of the dialogue observed in some classrooms, in which teachers pressed their students to develop their reasoning, when it came to assessment Harris teachers most often relied heavily on factual or definitional quizzes.

The inquiry-based science reform, advocated by PALMS, was largely reinterpreted by this district as a call to adjust the curriculum's scope and sequence. The decision was made that life, physical, and earth science would each be taught for a third of the year during 6^{th}, 7^{th}, and 8^{th} grades where as previously a whole year was devoted to each subject. Teachers perceived the primary change as affecting the sequence of instruction, rather than a call to alter content or pedagogy to embed inquiry experiences. Although the district superintendent articulated that his vision of the reform was a pedagogical one that “primarily addresses increasing student involvement and engagement” he conceded that his principals did not share this vision: "They come out of the same mold; they didn't come out of my mold, they are scope and sequence."

Final version published in: Journal of Educational Change 5: 345-387, 2004 © 2004 Kluwer Academic Publishers
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

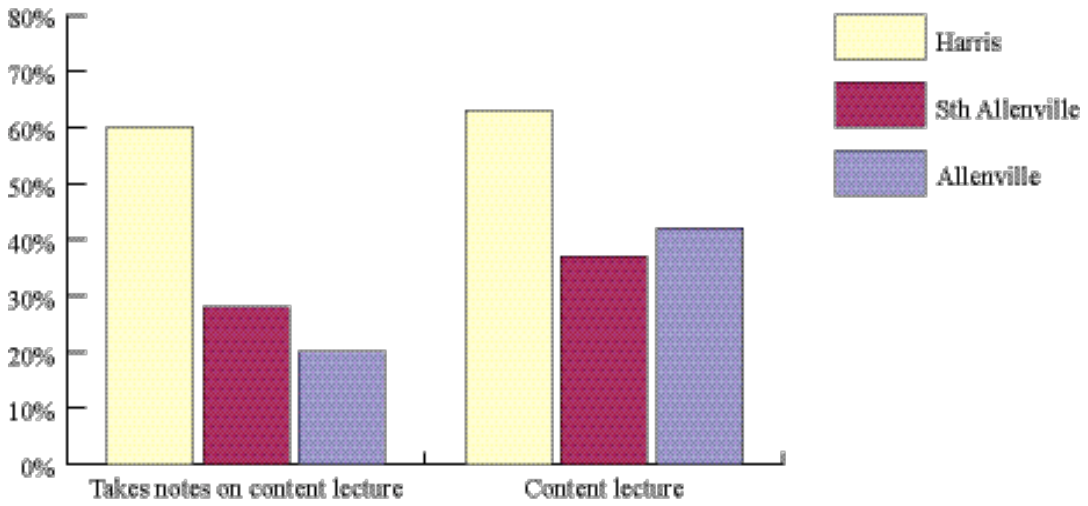

Figure 3. Proportion of classes in which content lecture or students taking notes were observed, in Harris, South Allenville, and Allenville.

The principal at the Harris study school spoke of the main goals of the reform as being "integration of science across the curriculum, increased use and integration of technology, and the application of science concepts to the real world." Student-centered investigations, data collection, analysis, and interpretation were not mentioned. The teachers generally shared the principal's understanding of the reform.

The Harris school strategy for addressing the MCAS revolved mostly around their preexisting plans to change the scope and sequence of the curriculum to insure that every subject (life, earth, and physical science including chemistry) is taught every year in grades 6, 7, and 8. This change was being put in place in order to align the curriculum to the district's interpretation of state standards as well as to the MCAS exam. Teachers in Harris maintained a strong sense of autonomy about specific topics that they chose to cover, and the texts and materials that they chose to use. Despite this, teachers articulated significant resistance to the curricular changes that were being implemented.

Harris's veteran teachers had developed a repertoire of activities, and a feeling of specialization in a particular science domain. Teachers stated that they were hesitant to move from being a "specialist to a generalist" and were also concerned that the "revised" curriculum was more fragmented. One example cited was the way that body systems are represented in the curriculum: the respiratory system is taught one year, the endocrine system the next and the reproductive system in yet another grade — there is no place in the scheme where the whole body as a system is presented.

The attitude of Harris administrators towards the MCAS was largely dismissive of its value and yet there was a commitment to maintain high scores. The superintendent did not mince words in expressing his skepticism about the MCAS:

> *The tests are merely necessary to satisfy the political requirements. 'We give you all this money, satisfy our appetite. Tell us we did it right'... I will go to any school, any kindergarten room and I will have the teachers line the kids up by socio-economic rank and I will tell you what the graduation rate is going to be for those kids and be 95% accurate.*

Yet a few moments later he added, "I got to tell you, every one of my principals has in their contract that their reimbursement is based on improving those scores." Yet despite this potential sanction, Harris staff did not appear to feel pressured or consumed with raising scores (in sharp contract to our next case study school, South Allenville).

Final version published in: Journal of Educational Change 5: 345-387, 2004 © 2004 Kluwer Academic Publishers
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

When Harris teachers reflected on the impact that the MCAS had had on their teaching they expressed a concern that in the future they may need to cut out labs and independent research to make room for MCAS preparation, but these changes had not happened yet. One teacher reported reviewing the rubrics of the essay section of the MCAS intended to define what answers were "unacceptable, acceptable or advanced." Several other teachers in Harris were observed occasionally reviewing an MCAS question with their students, and several teachers incorporated MCAS-like questions in their tests and quizzes.

Despite these accommodations for the MCAS, the test was not creating a cultural shift in Harris. The MCAS's short-answer, test-oriented approach did not feel like a departure from Harris teachers' past practice since Harris had a "testing culture" already. The vision of science in Harris was highly content-driven and had been cast as a "scope and sequence reform" rather than a reform of pedagogy or inquiry. Students had become well practiced in the art of multiple choice response. The culture of the MCAS was thus not foreign to this school, as Harris embodied a culture of high content coverage, lecture, quizzes and tests. In short, Harris teachers had never substantially accommodated to an inquiry-oriented, student-centered approach to science and therefore they did not feel that they were being asked to leave a desirable curricular and pedagogical reform behind in order to accommodate to the test.

Although the teachers thought that some of the MCAS questions were too hard, and inappropriate for eighth grade, teachers in Harris did not experience anxiety on their students' behalf that was evidenced in all our other case study schools. One teacher remarked "They know they have to take the test. I say to them you are responsible for your own education." Another teacher commented that the test served a useful purpose in that "the parents aren't complaining about the amount of work I'm requiring." The assistant principal was the only one who expressed reservations, saying she noted the fear in students' faces as they prepared to take the MCAS.

The MCAS appeared to have a very small impact on Harris' teachers' morale. This can be explained by a combination of factors. Harris is part of a unified school district with a largely decentralized method of operation. The district science coordinator was not perceived by the teachers as an effective presence, and the school had an exceptionally high turnover rate of principals. This resulted in teachers having a large degree of autonomy over their curricular and pedagogical choices. Hence the transmission of a mandate from the district to the school level was weak. Another factor that limited teacher compliance was that almost of all the science teachers at Harris had been there for decades. Teachers remarked that they had seen administrators, as well as other reforms, come and go, and that this reform would do the same. One teacher commented on his resistance to switching from physics to chemistry for the last third of the year:

> *I'm going to cheat a little until I get nailed, and they say "you know what – there seems to be a deficit in this area. The scores on the MCAS show you are not teaching this information." [Then] I'll present it [chemistry], but maybe I might try to de-emphasize it.*

As most of the teachers were nearing retirement they felt that they would not still be in the system once potentially damaging consequences of the MCAS exam took effect (e.g. graduation requirement, effect on teacher salaries).

Yet another factor is that the students in this fairly well-to-do district have historically performed relatively well on standardized tests so there was little fear that student performance would reflect poorly on the teachers. Last, the MCAS exam, which emphasizes retention of and facility with multiple content areas, is quite in tandem with the vision of science education at the school.

School # 2: South Allenville: A suburban school torn by two reforms.

South Allenville is a middle-class suburban district with a median income of $42,000. It is adjacent to the less wealthy Allenville urban center, which is also included in this study as case study #3.

Final version published in: Journal of Educational Change 5: 345-387, 2004 © 2004 Kluwer Academic Publishers
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

At the time of the study, there were eight science teachers on staff in South Allenville and their years of service ranged from 4 to 29 with a mean of 12.4 years.

The middle-school building is relatively new. The students are well behaved, and there is a general sense of order about the school. Each science classroom is large, and has desks in the front and a science laboratory in the back. There seems to be no shortage of science equipment, resources or state of the art computer labs.

Both the principal and curriculum coordinator of South Allenville were proponents of an inquiry-based approach and were well versed in the pedagogical and curricular demands of implementing such a curriculum. The curriculum coordinator was one of the first teacher leaders of the state systemic initiative, which provided the initial thrust to move science away from fact-filled lectures to a constructivist, inquiry-based approach.

The organizational structures in place in South Allenville are fairly traditional. While there are interdisciplinary teacher teams they are not small enough to afford flexibility in scheduling. The curriculum is currently linked to the Glencoe *Interactions* textbooks. The textbook series was chosen two years earlier because it seemed to be most closely aligned with the curriculum frameworks that integrate physical, life, and earth science within each year. Although the text does not promote long term investigations most of the teachers here had a strong commitment to students working in groups (47% of classrooms observed), and sharing strategies (56% of classrooms observed), and in integrating labs and hands-on activities (32% of classrooms observed).

Of all the schools observed, the MCAS test had the most profound impact on South Allenville. There was not a single day during our observation period where MCAS was not spoken about, either during our conversations with teachers or in their discourse with their students. Teachers reported that every faculty meeting is about the test. Every teacher in the school saw preparing students for the test as an essential element of their job. The curriculum coordinators took an active role in shaping the faculty's response. Over the summer before our observations, they had conducted a careful item analysis of the past year's test and revised what was to be covered in each grade based on this analysis. Commercial review books were given to all teachers so that they could review sample questions with students. Teachers incorporated "rapid-fire drill questions" and practiced open-ended questions from past tests in their classrooms.

Despite the energy that was expended on the MCAS, the assistant superintendent, principal, and curriculum coordinators expressed no particular liking for the test. The assistant superintendent mentioned that his school district had gone on record saying that having one measure of success is unfair to students. The principal was concerned that students who would have successfully gone to college in past years would now be discouraged or prevented by the MCAS from doing so. The curriculum coordinator, an advocate of inquiry, saw MCAS as a conflicting reform. Nonetheless all of them were committed to raising MCAS scores and saw increased MCAS scores as a measure both of the school's success and of their personal success. The curriculum coordinator expressed both his mandate and his ambivalence about it:

> *Right now the message is that it's important to hit all the area [s] that MCAS is asking for, so that the students will succeed on that test. What is good about that is that it gives us a more unified curriculum. ... The one good thing about the frameworks is that it makes everyone do the wrong thing together. That's a start. At least we're all marching to the same drum.*

The curriculum coordinator's ambivalence reflected his understanding that the approach that he was now espousing for raising student MCAS scores did not complement the previous inquiry-based science reform that he had worked hard to install. He explained:

> *In inquiry there are many dead ends…. Let's say the student frames a question. They say "I'd like to know about this and I think the reason is… "and then they test their hypothesis. That could*

be a dead end. ... You may have to back up and reframe the question, or come up with another possible solution, and then pursue that. If you're showing true inquiry that takes a lot of time. You'll have tremendous depth, but within a narrow range of actual content… whereas the frameworks, they're a mile wide and an inch deep."

How did the curriculum coordinator and principal reconcile themselves to abandoning depth for breadth? The curriculum coordinator explained that he saw this move as a temporary aberration, that once everyone is in alignment with the MCAS, inquiry-based science will be reestablished. Yet he was unsure when scores would be deemed high enough to do so.

The long-range view is this is something we have to get through initially, to realign the curriculum and to dictate to the teachers what they need to cover…. [then] we can relax about the MCAS test. This is just a hurdle we have to get over. This is not what it is going to look like in five years from now.

While it is not clear why this school felt under so much pressure to perform, it was clear that for the classroom teacher, the coordinator, and the principal, the MCAS had come to be the dominant measure of success. The curriculum coordinator said:

This is [MCAS results] how my success is measured, by my evaluators. I'm on a merit raise for next year and any increase in my salary is directly related to a rubric that has been developed. That puts pressure on us to have definite results.

The mandate to raise scores, shaped by assistant superintendent, principal and curriculum coordinator, exerted great pressure on the teachers in the school. The principal stated "One point can make the difference [for a student] between passing and failing and I tell the teacher not to underestimate their role in this," and indeed the teachers did not. When a sixth grade teacher described how her teaching had changed over the last three years she said, "The MCAS has completely changed what I teach and how I teach it and where my goals are—where I'm focused and where my end-product should be." When probed further about what her major science learning goals for her students this year, she replied:

That they're able to recognize the vocabulary in the MCAS and that they're able by process of elimination get rid of two answers that would be ridiculous and at least come down with a good chance of the correct one for the multiple choice, and then hopefully have a good chance on the open response, you know of getting threes and fours and the kids that are on the verge of failing, being able to push them up that step and just get them past it.

It is remarkable that these goals were expressed by a sixth-grade teacher whose students would not be tested on the MCAS until they were in eighth grade. Her comment points to the effectiveness of the transmission of a mandate from the district to principals, coordinators and to individual teachers. It also shows, in contrast to Harris, the high level of teacher compliance with this mandate.

The MCAS focus had a dramatic impact on the eight science teachers in this school. Their curriculum changed year by year as they added and subtracted topics in response to analyses of the previous year's MCAS. For example, two teachers reduced the time they had previously spent on the digestive and reproductive systems to make room for new topics. One teacher cut out genetics because she had heard it was not on the MCAS even though it was her favorite unit (and is included in the State Standards). Another teacher felt compelled to teach about two specific glacial features, drumlins and eskers, because she had heard that they were featured on the previous year's test (even though she did not know what they were and could not find information about them in her textbook). This adding and subtracting of topics diminished teachers' sense of mastery, and made some of them feel that the curricular topics had lost their interconnections. A common complaint voiced was that teachers were forced to cover more and more in less time. One teacher commented, "Last year we covered 13 chapters and now we are up to every single chapter." Another teacher commented, "Now we just do a little bit of

Final version published in: Journal of Educational Change 5: 345-387, 2004 © 2004 Kluwer Academic Publishers
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

everything, we don't do anything in depth anymore." Another remarked, "It [coverage] takes the wonder out of science. It just doesn't seem like there is enough time to allow kids to explore."

Teachers expressed their belief that increasing coverage of material and decreasing time (due to time needed for preparation for and administration of the test) inevitably led to changes in their pedagogy. Teachers remarked that they had eliminated interdisciplinary units that they had used in years past, cut down or eliminated multi-day investigations, and had fewer labs and explorations.

> *We cannot take the time to do the things that for kids and for me make learning fun. We have so many inter-disciplinary things that we've created over the years, and you just don't have the time to do them, because you are driven with that deadline of 'we've got to do this, we've got to do this, we've got to do this.'*

One teacher commented that she couldn't justify spending time on the students' acquiring lab techniques, which were not tested on the MCAS. Teachers had also added "rapid-fire questions" to their repertoire. These are short, multiple-choice, MCAS-type questions which teachers pass down from year to year so that, for example, seventh grade teachers can remind and drill students on facts they learned in sixth grade. Teachers were also observed having their class construct answers to open-ended MCAS questions. Such exercises, however, looked anything but open, with teachers specifying and practicing with their students the exact facts that would need to be included in each paragraph to get full score.

The impact of the MCAS on South Allenville's teachers' morale was palpable. Teachers expressed a lot of anxiety over the MCAS. One teacher said that she felt "partly to blame" when students performed poorly. Another teacher expressed worry about the possibility of not covering a topic that might be tested two years hence: "We don't want the superintendent to be able to say 'Oh, well, this was supposed to be covered in the sixth grade and this kid had Ms __ , so obviously she isn't doing her job." Others expressed a sense of frustration, and fear that they would not meet targets for coverage. Finally, teachers reported that this pressure was making it less fun to teach as well as for the students to learn.

School #3: Allenville: A middle class urban center with a coherent vision of inquiry.

Allenville is a small urban center with a median household income of approximately $37,000. Gaudet (2000) classifies it as "lower Middle MA," being neither a challenged nor an advantaged district. It is more urban and less privileged than neighboring South Allenville. Yet the Allenville school building is new, bright and spacious, providing many atrium ceilings with natural sunlight. The rooms are generally large for the class-size, and there are many interconnections between rooms, sometimes by doorways, sometimes by folding partitions which allow two rooms to become one. Each **s**cience class has both an area for desks and an area on the side with large lab tables with sinks. There are four to six computers in each classroom. The science faculty in grades 6, 7, and 8 is composed of ten science teachers, with a mix of young and veteran teachers who have from 4 to 36 years of teaching service, with a mean of 13.5 years.

This district has cultivated a coherent vision of inquiry (see Falk & Drayton, 2001), which is at the center of their understanding of science reform. The superintendent, in describing his vision stated:

> *It's a classroom where kids are doing real science, and that is they're investigating in a hands-on way. They're working collaboratively with other students. It's interdisciplinary in that they're using their math, they're using their writing skills, they're using reflection. The teacher is not up there lecturing, but the teacher has presented things for kids to investigate, and things that have a connection to their lives now, and future lives. And kids see a reason for doing it. They're interested in doing it. And we have plenty of supplies. And the teachers know the process, so they're not interfering with kids and giving kids answers too soon or ever. ... Kids are investigating and trying to solve a problem that either they have posed, or teachers pose for them. And, they're using the scientific method to come to some kind of theory that they can then prove*

Final version published in: Journal of Educational Change 5: 345-387, 2004 © 2004 Kluwer Academic Publishers
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

through inquiry, to getting data from other sources and testing it out in a lab, even in a small lab kind of experiment.

This is a vision shared across the district, articulated by the superintendent, supported by the principal, and transmitted through an effective curriculum coordinator. The school has a strong implementation of middle-school structure with flexible scheduling, small teacher teams, looping, a district-developed curriculum supporting short and long term investigations, and regular articulation meetings of the district's science teachers to discuss the curriculum grade-by-grade. The districts' vision of inquiry-based science is reflected in the teachers' pedagogical approach, in the curriculum materials that have been assembled, and in the multiple forms of assessment used. Practices observed related to an inquiry-based approach included students often working in small groups (63% of classes observed) or pairs (73%); frequent student engagement with data, with students revisiting and revising work (44%); and a strong focus on hands-on activities (38%). Of the six schools visited this one had the most consistent inquiry-based orientation (see Figures 4 and 5 below).

In startling contrast to South Allenville, Allenville administrators and teachers did not see the MCAS as a test that undermined the inquiry-based approach, which they had invested much energy installing over several years. A conscious decision was made to continue the reform that they had put in place and to resist the notion of increasing content coverage as a strategy to raise student scores. The strategy chosen by this district was to continue with the inquiry-based curriculum that they had adopted, to continue working on problem solving and short- and long-term investigations, and to increase an emphasis on writing across the curriculum. A professional development program was instituted to help teachers learn a new writing program. The superintendent explained why he believed that continuing with an inquiry-based approach would be helpful in the long run.

With the MCAS there's the content pressure, and I think a lot of people figure if I cover all this stuff… the kids are going to do better on the test. And, I think, maybe the answer is not to have a whole widespread coverage of content, but to do some things well. To look in the frameworks for the important area and do those well…We have an experience here in [Allenville] that often our open-ended questions are [scored] much higher than the state average, and yet some of our content [scores] in math and science, spelling, is lower by comparison. And the school committee says well, how can they do complex things if they can't do the basics? And it's hard to explain that. But, it's because they've had more experience doing complex things, and it's going to take them a while, but those basic skills, if they're using them in the context of solving real problems, eventually they're going to get better at them.

The curriculum coordinator shared the superintendent's resistance to a move towards increased coverage. Her role in the district's reform strategy was pivotal: she developed the curriculum kits, the resources, and alternative assessments, and facilitates "articulation meetings" for science teachers to discuss curriculum and pedagogy. When asked about the effect of MCAS she responded:

I don't care about MCAS and all that, that's incidental. I really think we're learning for life… To really learn and remember, the kids really need to be actively involved in doing and figuring out what's happening and why. That needs to be in place. … I really do believe that inquiry needs to be the focus…

Final version published in: Journal of Educational Change 5: 345-387, 2004 © 2004 Kluwer Academic Publishers
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

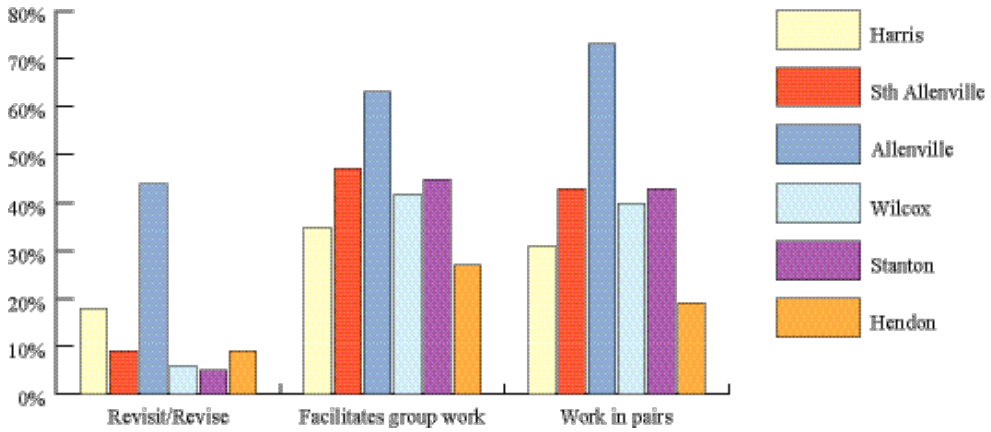

Figure 4. Frequency of collaborative and reflective practices in 6 MA study schools.

The principal in our study school in Allenville saw the MCAS as providing a lever to share best strategies across disciplines. In particular he was working on increasing writing skills across all subject areas and had a multidisciplinary team working on this.

Given the attitude of the leadership in Allenville it is not surprising that the science teachers were feeling relatively positive about the MCAS test. In fact some teachers were grateful that the test had "taken science off the back burner," "helped get a lot of materials into the classroom," "help[ed] students to take studies seriously," and "given direction."

Teachers' attitudes towards the test were not completely positive. Like their colleagues in South Allenville, Allenville teachers resented the time stolen for administering the test. They thought the test had "unrealistic expectations for kids at this age" and that it was unfair that the students would be tested on material they may not have learned. Nonetheless, teachers in Allenville did not feel pressured to cover more material for the MCAS. The time stolen was the time actually taking the test, not months of preparation towards it. The curriculum in place included lengthy hands-on investigations, data collection, and analysis, student presentations, journal writing, and performance assessments. None of the above has been changed to make room for MCAS preparation.

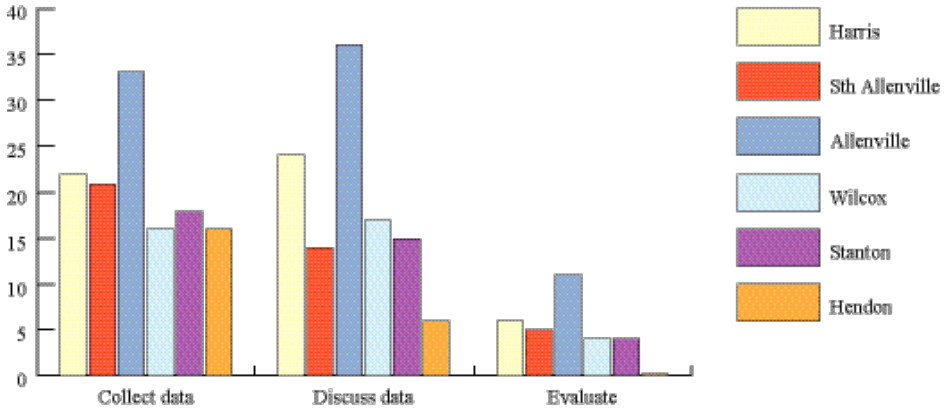

Figure 5. Frequency of collecting, discussing and evaluating data in 6 MA study schools.

The strategies that teachers reported using to prepare students for the MCAS include "increased writing across the curriculum," "more focus on thinking and open-ended questions," "practice reading

Final version published in: Journal of Educational Change 5: 345-387, 2004
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

directions," "using the MCAS rubrics with the students when grading papers," and requiring students to be "more detail oriented and to express their thinking in depth."

During our observations in these classrooms the Allenville teachers were not practicing specific MCAS questions. In contrast to the formulaic essay coaching observed in South Allenville, one teacher remarked, "I encourage them to trust that there is not only one right answer and that's why they are called open response." Only one teacher out of ten expressed the sentiment that teachers felt pressured by the test. None of the other teachers expressed negative effects on their teaching or on their self-esteem.

Three Highly Challenged Urban Districts: Wilcox, Stanton, and Hendon

We now examine the other three schools in our study, Wilcox, Stanton, and Hendon. Each of these schools is in one of the fifteen most demographically challenged districts in MA (Gaudet, 2000). In all three schools, the majority of students have failed the MCAS exam. In fact, in Hendon, the district with the lowest SES, 85% of their students failed the Science MCAS exam. All three of these schools have divided 6th, 7th, and 8th grades into segments so that life science, physical science and earth science are taught each year. They often used textbooks that are written at a simpler reading level to help their students.

These schools share some common challenges, as they serve students from very challenged demographic backgrounds. Teachers in all three of these schools face the dilemma of trying to be responsive to the state's mandates, while feeling that the demands mandated by the curriculum frameworks and implied by the MCAS may not be appropriate for their students. One teacher eloquently described her dilemma:

> *I feel like I'm going kind of slow, but if I speed up they won't get anything. I know I'm not going to get all the way but the way I'm thinking is, they cannot understand anything or they can understand the parts we get to.*

Similarly, another teacher said, "I feel guilty because I haven't covered it [all the material on the MCAS] or I feel guilty because they have been turned off of science." This tension between teaching to the needs of the child or teaching to the test is especially potent in these districts. The students have so many needs. Teachers saw that their students have significant deficits in behavior, reading level, and fluency in English, in addition to a lack of test taking experience. Thus, the MCAS nagged at superintendents, principals, curriculum coordinators, and teachers in these districts, yet there was no coherent philosophy or vision in place about how to address it. That is in part because the test felt so far out of reach, and the goal of getting most students to pass seemed unrealistic.

School #4: Wilcox: Large, urban, challenged district.

Wilcox is a very large, diverse, highly urban district with a population of approximately 170,000. Although the median income, $35,977, is only slightly below that of Allenville, it is a markedly more challenged environment, with 56% of students eligible for free/reduced lunch (state avg. 26%), 13% with limited English proficiency (state avg. 5%). Gaudet lists Wilcox as one of Massachusetts's 15 most challenged urban centers. In our case study school there were 780 seventh and eighth grade students, reflecting a significant minority population (9.1% African American, 28.3 % Hispanic, 9.9% Asian, .5% Native American and 51% Caucasian). The science faculty at the time of the study was composed of 4 veteran teachers and one new teacher. Years of service ranged from 1 to 35 years with a mean of 20 years experience.

The school is housed in an old but well-maintained building with large, clean and fairly well-equipped science classrooms usually having both a sink and one computer and a reasonable stock of lab equipment. In comparison to the two other inner-city districts to be discussed in this study (Stanton and

Final version published in: Journal of Educational Change 5: 345-387, 2004 © 2004 Kluwer Academic Publishers
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

Hendon), Wilcox students were less boisterous in hallways and classrooms, and more respectful of school property. While teachers noted the enormous challenges that many students face at home, they tended to be sympathetic to the students rather than critical of their behavior or abilities.

Wilcox had some school structures that promoted inquiry within the classroom. In particular Wilcox had block scheduling with some science classes extending over 80 minute periods. This allowed and perhaps encouraged some teachers to provide and include regular labs with ample time for student to collect and record data. This school also participated in an active partnership between the district and a local medical school. On a few occasions, we saw researchers from the medical school co-teaching labs with teachers.

While Wilcox had a science department, most planning time was done as part of interdisciplinary teams. Department meeting tended to focus on administrative matters rather than discussions about science content or pedagogy. So while each teacher was concerned with how to teach science in order to raise student scores, there was no unified instructional approach.

During the period of our observation we did note some classes where students spent a good deal of the period answering questions at the back of the text, but we did not see any MCAS questions (or questions similarly construed) given to students for practice. Teachers commented that they were increasing the time spent on writing, on having students write full sentences and beginning to write about topics in science. One novice teacher commented

> *I'm trying to get them towards giving them just a general topic and having them write about it, in science. I think that they can do that in English, but I think they are worried about it in science.*

While teachers reported adding emphasis on topics such as earthquakes, plate tectonics, and the composition of air (topics covered on recent MCAS tests) there was not the sense of careful backmapping to the test as in South Allenville. In general, the veteran teachers on the staff seemed relatively undisturbed by the MCAS test. One such teacher said, "my goal is process and if they are getting that they are on track." Another veteran staff member remarked that he "refused to get bent out of shape" over the test. While MCAS preparation seemed limited within Wilcox, there were hints that in the following year they would be bringing in commercial workbooks designed to improve MCAS skills.

School # 5: Staton: Highly challenged large urban district.

Stanton has a population of 157,000 and a median household income of approximately $27,000. The city is racially diverse: student enrollment is 28.5% African-American, 47 % Hispanic, and 22% white. Seventy-one percent of the students are eligible for free or reduced-price lunch.

The Stanton school is a challenging place to teach. The building is very old and shows wear. There are no working labs, little science equipment, strict limits on the amount of paper used for photocopying, and demanding teaching loads. The school had eight science faculty with a strong representation of veteran teachers. Years of service ranged from 4 to 31 with a mean of 15.8 years.

There did not seem to be a coherent vision of science education within Stanton. Our observations showed that, on average, science teaching at this school tended towards a teacher-centered approach. Content lectures were observed in 65% of classrooms while hands-on activities were observed in 26%. Discussions that sought predefined short answers (67%) were more prevalent than those seeking multiple strategies or brainstorming (25%). Most dialogue in the classroom was teacher directed. We observed students sharing strategies in 26% of classes.

Yet these figures by themselves do not paint an accurate portrait of science at Stanton, as there were marked variations in pedagogical style and teaching effectiveness among teachers at the school. While most teachers in Stanton seemed to spend a significant portion of each class on classroom management issues, and spent most of their class periods lecturing to largely inattentive students, two

Final version published in: Journal of Educational Change 5: 345-387, 2004 © 2004 Kluwer Academic Publishers
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

teachers at this school were inspiring in their ability to captivate students' attention, and to engage them in inquiry-oriented student investigations. In fact, the vast majority of the observations of hands-on activities, incidents of students working in teams, and of students sharing strategies that we made at Stanton occurred within these two classrooms.

Perhaps not surprisingly the MCAS exam had the strongest effect on these two inspiring teachers who had created inquiry-based classrooms despite the lack of labs and resources. These teachers had the most to lose. They worried that covering all the material tested on the MCAS would result in needing to change their pedagogical approach with their students. One of these teachers commented that there was no longer any time for him "to engage his students' athletic or artistic mind" as the MCAS only assesses only one type of intelligence. The other teacher commented that she will need to drop her "daily problem," which was motivating for her students, in order to increase material covered. She expressed resistance to complying with a mandated curriculum:

> *Teachers aren't enjoying what they are doing anymore. All of us think the best teacher was the teacher who loved what they were doing. A teacher who's just teaching, being told what to teach, is not contagious. And I think we're doing away with contagious learning, and it is going to be pounded into the kids.*

The lack of a coherent approach to science education in general was echoed in a lack of a coherent strategy towards the MCAS. While the superintendent stated that "it is clear that with exposure to constructivist thinking kids will do better on [MCAS] open-ended questions," the assistant superintendent did not share this perspective and spoke of the inherent tension between inquiry and MCAS, with "one stressing longitudinal investigation and the other exposure." Sample materials distributed to the teachers along with the district curriculum varied in quality, from those which could be used to support direct student inquiry, to those which focused primarily on following instructions and reaching foregone conclusions. At the time of the study, teachers were hearing no clear message coming from the leadership to the teachers as to whether to stay the course of inquiry or to increase breadth of coverage.

Some teachers have added specific curriculum units. Others have focused on vocabulary words, though it should be noted that these are not science vocabulary, but rather vocabulary that appears in test questions. Several teachers expressed the belief that they think that more hands-on experiences and labs might raise scores on the MCAS as their students would be better prepared for "thought experiments." Yet, they complain that there are no labs, no sinks, no resources and no materials available, so that doing more hands-on activities is very difficult.

School #6: Hendon: *Severly challenged small urban district.*

Hendon is a small urban district with a population of 44,000. It was rated by Gaudet as one of the three most demographically challenged districts in the state. The community has a very high proportion of immigrant residents; enrollment is 72% Hispanic, with a high proportion of the residents speaking Spanish as their first language, and 20% of students are considered "limited English proficient." The district has been seen to be in crisis for years, and at least one of its schools has seen intervention by the state Department of Education because of failure to improve quality. At the time of observation, our case study middle school had four science teachers on the staff, with from 2 to14 years of experience (mean = 8 yrs).

There was no clear mandate from superintendent, principal, or science coordinator about how to raise scores, though the district was getting considerable pressure from the state to make some improvement. Frequent leadership turnover had impaired the development of a coherent vision. An interview with the specialist for science teacher professional development revealed the range of challenges that she faced as she tried to engage teachers with both inquiry and the MCAS:

Final version published in: Journal of Educational Change 5: 345-387, 2004 © 2004 Kluwer Academic Publishers
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

The MCAS hurts [efforts at science pedagogy reform]… until people see that learning can happen, and that you can address these standards by doing inquiry… And then when you really do an inquiry-based activity, it could go on for weeks. And therefore you're not covering as many of those standards. … I have people who don't really understand science. So they're trying to use the textbook, and trying to follow it, because they don't really understand it. So to veer from it, you would have to know your science more. So that's difficult.

This teacher of teachers has advocated an inquiry-based approach within her district in the past. Yet the high rate of teacher turnover, combined with a lack of funds for teacher release time, has compromised her ability to develop this vision with her teachers, many of whom lack strong science backgrounds. While she believes, in principle, that inquiry based activities could be a rich approach to addressing the MCAS she must cater to teachers who have not received sufficient professional development and do not feel comfortable deviating from the textbook. In addition, the lack of decently equipped laboratories, science resources and materials presents further barriers to developing an inquiry based approach to address the MCAS.

With little guidance from district leadership, teachers were largely on their own in developing a response to the test. Teachers reported encouraging students to respond to open questions, and to have students generate their own answers rather than finding them in the book. In practice, this was rather frustrating. One teacher was observed asking her class to answer an open-ended question that was taken from last year's MCAS test. Out of a class of 20, only three attempted an answer. The rest of the class fidgeted, stared blankly at the page, or put their heads down in frustration. The distance between the current skills of these eighth graders and the difficulty of the question seemed unbridgeable. One teacher commented,

Most of these kids are not going to college. They get through high school and they are set. The MCAS is more so they can go to college and what is going to happen is that it is going to get them frustrated. The ones that are thinking of college will say "if this is college, forget it.

Another teacher worries about the effect that a failing score will have on the public, the parents, and students: "My fear is that the public and parents and even the kids are going to perceive themselves as not good scientists or not good students when I don't think that is accurate." While this test might be seen as a motivator in Middle MA, or in Advantaged districts, it is seen as a roadblock and a deterrent in this challenged city.

Discussion

In any given year the MCAS questions are the same in every district in MA; nonetheless, the effect that the MCAS has on teachers and students, and on teaching and on learning, varies significantly from district to district. In our study schools, we have seen several different district approaches, which ranged from changing the scope and sequence of the curriculum (Harris), to increasing coverage of material while decreasing investigations (South Allenville), to staying the course of inquiry (Allenville). The three highly challenged urban districts had not developed coherent strategies to address the MCAS. However, they expressed emerging plans to introduce commercial test preparation workbooks and to increase attention to test vocabulary and writing skills. The differences in strategies suggest that the test itself does not dictate change in instruction; rather it is the district's interpretation of and response to the test, in the light of previous pedagogical commitments (where these exist), that strongly influence the reactions of teachers within their classrooms.

It is interesting to compare the responses of South Allenville (school # 2) and Allenville (school #3). Both of these districts had invested significant administrative and professional development efforts to encourage an inquiry approach to science education before the introduction of the MCAS. Both

Final version published in: Journal of Educational Change 5: 345-387, 2004
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

districts had extremely active science curriculum coordinators who were effective proponents of inquiry-based science and were trained by the State Systemic Initiative to be leaders in this reform. Each of these coordinators had a significant impact on their schools. They each brought a sense of coherence to the pedagogical approach that was being used, and to the definition of what was taught within each grade and within each classroom. However, when the MCAS was introduced these two districts took markedly different approaches.

South Allenville adopted a "back-mapping from the test" strategy, which was supported by the superintendent, the middle-school principal, and the district's science coordinator, who played a key role in implementing this strategy. Along with this approach came stress to increase coverage, a shift of curricular content to topics appearing on the last MCAS, the development of "rapid fire" questions, and practice essays that were transformed into a "fill in the blank" format. All of these interventions ran directly counter to the previous reform that the school had been implementing in the previous 5 years. Inquiry-based science requires time for students to develop their own questions; time for data to be collected, discussed and analyzed; time for reflection; and time for students to share strategies and learn from each other (Drayton & Falk, 2001). The methods adopted for increasing MCAS scores in South Allenville diminished the time, the freedom and the leisure for scientific exploration.

The curriculum coordinator saw these interventions as only a temporary diversion from the inquiry-based approach that he had promoted during the past five years. He said, “This is just a hurdle we have to get over. This is not what it is going to look like in five years from now.” Yet this hope that teachers will focus on the MCAS for a few years and then return to an inquiry-oriented approach is a large gamble. Newly hired teachers will not have had the prior experiences of immersion in inquiry. More than this, school cultures are slowly built and they evolve organically over time. Once a culture of inquiry is lost, it may take a great deal of effort to regain it.

In contrast, the superintendent, middle-school principal and science coordinator in neighboring Allenville all share a commitment “to stay the course” of a program based on inquiry-based science. This vision is supported by structures within the school such as flexible scheduling, looping, small teams, and teacher articulation meetings to discuss approaches to curriculum (Falk & Drayton, 2000). This approach is evident in the classrooms observed where there were frequent opportunities for extended student investigations, and where students collected data, discussed and evaluated it. Rather than moving to cover more content in less time, students were encouraged to frequently revisit and revise their work.

The superintendent reported defending this policy to his school committee, which asked him, "How can they do complex things if they can't do the basics?" He explains that in the long run it is more effective to teach for depth than breadth and to be engaged with complex problem solving, rather than fact retention:

> *I think the answer is to look at the frameworks and focus on some important big areas, big questions, and … not worry about every single bit of content because you can't do it… Because you've got to show you understand by your writing, by your analysis of the problem, by the methods you're going to use. So, the more experience they have doing that in depth, the better they're going to do on the test.*

The difference in district leaders' response to the MCAS also influenced the teachers' attitudes towards the test. Teachers in South Allenville were frustrated as each year they were constantly changing what they taught in response to an analysis of test questions. As a result they were dropping favorite units and even topics specified in the frameworks that have not appeared on the test. In a sense the MCAS exam has replaced the frameworks as the master guide to what is to be taught. There is so much to be covered and not enough time to do it well. Such an approach impedes teachers’ abilities to teach to the “big ideas” and concepts in science. There is also no time to stimulate student curiosity and engagement and this has a dampening effect of both teaching and learning. As one teacher commented, “We cannot take the time to do the things that for the kids and for me make learning fun.” Even the thrust of

Final version published in: Journal of Educational Change 5: 345-387, 2004 © 2004 Kluwer Academic Publishers
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

professional development has moved away from inquiry and towards "increasing effective time management."

In contrast, most teachers in Allenville found that the test was helpful. They viewed the MCAS through the lens of their school culture of inquiry. The test had not destroyed this culture, but rather had prodded them to share best strategies, emphasize thinking and open-ended questions, and to come up with creative way to increase writing skills across all subjects through multidisciplinary teams. While teachers in Allenville still resented the time stolen from the classroom for administering the test, they did not experience a sea change in what they understood to be appropriate teaching and learning of science.

Finally, the differences in interpretation of the MCAS between South Allenville and Allenville account for differences in teacher morale. Staff in South Allenville bought into the idea that student test scores are the ultimate judge of effective teaching. Indeed, from the level of the principal to curriculum coordinator to teacher, there was the understanding that this is how their success is measured. In this atmosphere, a sixth grade teacher can come to fear that she will be blamed for not having covered a particular item that may be asked on the MCAS two years later, when her students reach eighth grade. In contrast, teachers in Allenville were encouraged to include a focus on big ideas, scientific habits of mind, and skills of investigation, articulation, and reasoning.

In contrast to Allenville and South Allenville, the Harris School presents yet another picture. Our observations showed that this school was even more test/quiz driven than South Allenville. Teachers lectured more and in general embodied a more traditional, teacher-centered approach. Yet the teachers here spent less time speaking with each other or with their students about the MCAS and they seemed less affected by it.

There are several factors that can explain this difference. First, the test was not interpreted in a fashion that required the teachers to change their pedagogical approach with their students. Second, Harris drew students from advantaged socio-economic backgrounds that have traditionally performed well on standardized tests so there was little pressure to raise scores. Third, the staff at Harris was "seasoned" (with a mean of 27.5 years of service), and most were close to retirement. This fact, coupled with high administrator turnover, allowed teachers to feel relatively free to close their doors and conduct their classrooms as they saw fit, with little motivation to comply with demands to change instruction in order to raise student scores. Teachers were not fearful of the negative future effects that the MCAS may have on teacher evaluations or salary as they knew they would soon no longer be in the system. Their sense of their efficacy about their teaching ability had been cultivated over decades and was not tied to student MCAS scores.

The three challenged urban schools (Wilcox, Stanton and Hendon, schools #4, 5, 6,) face a different set of challenges. They are just beginning to formulate an approach to improving MCAS performance. Practices that were mentioned by teachers (but not necessarily observed) included a move to increase writing assignments, to add topics on the test, and to add vocabulary drills of words that appear in test questions. In general, urban teachers were more likely to resist increased demand for coverage since they felt that if they went any faster they would lose their students completely. Schools with higher SES, such as South Allenville or Harris, seemed more subject to the pressure for coverage, as teachers believed that covering more would result in higher student scores, a view not shared in the urban districts. Hence the suburban schools were more likely than the challenged urban districts to increase coverage while decreasing the time previously allocated for student exploration. This finding adds a different perspective to studies that have concluded that schools with high proportions of low achieving students are more subject to head on preparation, drill, and coverage at the expense of higher level skills (Haberman, 2003; Madaus & Clarke, 2001; McNeil & Valanzuela, 2001).

Despite the fact that these three urban schools had not adopted coherent approaches that exerted demands for change in practice, teachers in these schools had a particularly negative attitude towards the test. They felt that the science test required a panoply of skills including reading comprehension, broad

Final version published in: Journal of Educational Change 5: 345-387, 2004 © 2004 Kluwer Academic Publishers
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

vocabulary, test-taking savvy, all of which their disadvantaged, (and many bilingual) students lacked, and which they were not in a position to remediate sufficiently to prevent widespread student failure on the MCAS. Teachers said that they felt that their duty was to teach at a speed and level that was appropriate to where their students were rather than to a test that they believed was far too difficult. Hence, they found themselves caught between two untenable positions, that of abandoning the test or abandoning their students.

The test also had a devastating effect on teacher morale. Teachers felt a sense of guilt at not being able to cover all the material that appeared on the test, even though they felt that they were proceeding at the correct pace to achieve their students' understanding. They felt that the test was being used to equate poor student performance with lack of competent teaching, despite studies that have shown a high correlation between test scores and SES/demographic indices (Gaudet, 2000). Several teachers reported feeling demoralized. One teacher commented, "We are pathetic." Some teachers expressed concern that their principal's salary would be tied to the results. A principal voiced a common fear that teachers' pay would be affected in the future. One teacher wryly commented that MCAS stands for "My Career at Stake."

Looking across all six districts, we return to the study question, Is the MCAS test beneficial or damaging to inquiry-based science instruction? In reply, it is important to acknowledge that, to an extent, "it depends." It depends first and foremost on the district's interpretation of the test. For Allenville the MCAS was seen as complementary to inquiry. It validated their emphasis on thinking skills and complex problem solving, and kept science off of "the back burner" when competing for resources with other subjects that are traditionally tested such as reading. It offered impetus to develop a writing program across the disciplines. In South Allenville the district leadership understood the test to be a competing reform. The most important value became the raising of student scores, rather than the raising of student understanding. Based on little evidence, the district decided that the best way to raise scores was to put an inquiry-based approach to science instruction on hold.

The effect of the MCAS also depends on whether the changes implemented by the district demand dramatic shifts in pedagogy. For example, teachers at South Allenville were asked to rethink pedagogies that took too much time, given the demands for coverage of tested materials, and hence the staff put a lot of the inquiry-based professional development that they had received in past years "into storage" while they waited for the MCAS storm to pass. Although Harris teachers did not need to change their pedagogical approach, one must consider the cost of veteran teachers shifting in their last years of teaching from one subject domain (e.g., biology) where they have passion and have developed expertise, to teaching all three (biology, chemistry, physical science) within one year, especially when there is little attempt made to integrate or make connections between these different content areas. As noted in other studies, such shifts, mandated abruptly to align practice with new policy frameworks, cause teachers to feel a loss of creativity and spontaneity, and "theft of their autonomy" (Hargreaves, 2003, pg. 91).

Last, much depends on whether the test is seen to be within shooting range as an appropriate target. While the consensus from all six schools was that the test was difficult, it was felt to be totally inappropriate in the urban inner city districts. Although MCAS advocates contend that the test will create a "level playing field" of expectations for student achievement, in the inner city schools it is seen to have the effect of closing doors, rather than opening them. Hence the effects that this test has on teachers and students in low socio economic districts where students traditionally do not perform well on standardized tests is different than the effect in high scoring, wealthier districts. Policy makers need to rethink the demands this test imposes so that they can provide a constructive challenge to a school like Hendon (where there is currently an 85% fail rate) rather than a frustrating barrier that seems impassible. Other studies have shown that such a barrier is more likely to yield an increased drop-out rate than increased motivation (Haney, 2000; Heubert & Hauser, 1999; McNeil, 2000).

Final version published in: Journal of Educational Change 5: 345-387, 2004 © 2004 Kluwer Academic Publishers
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

Having acknowledged the power of the district's interpretation and response to the MCAS in determining how the MCAS will influence their schools and inquiry based science, we did see some negative effects at work in all our study schools.

With the exception of Allenville, all schools experienced pressure for increased coverage. Yet, even in Allenville teachers reported having to cover material in less time since the MCAS was administered in May, and it was difficult to re-engage the students at the end of the year. Hence, the MCAS testing disrupted instruction in all districts. While this demand for increased coverage can be tempered by the district's vision, the MCAS emphasis on a broad range of topics promotes most districts to shift towards breadth over depth. Demand for increased coverage is the nemesis of the inquiry-oriented approach. In order to accommodate more content, teachers report doing fewer long-term investigations, labs, independent work, and student presentations. The test does not have an active performance component, so time spent on developing student investigations, lab techniques and skills seems unrelated to test performance — and consequently, without strong district leadership, these skills tend to become devalued.

All six of our case study schools had been reorganized around interdisciplinary teams of teachers as part of a middle school reform to create a more coherent experience for the young adolescent student and an environment in which all teachers in a student's life have opportunity to talk with each other. Interdisciplinary teams were also created in order to promote integration of curricula across domains. Teachers at each school mentioned that the MCAS made them cut interdisciplinary units and activities. Each subject domain was too pressed to cover the content that would be tested. This is an example of how one reform can inadvertently counteract another that is being implemented concurrently.

Finally, most of these schools had instituted heterogeneous classes so that students of different learning styles and different abilities could learn with and from each other. For some teachers the MCAS demands for "all students" is antithetical to the idea of respecting individual differences. As one teacher put it "they want everyone to develop at their own rate, but everyone has to be at the finish line together."

The effect on teachers' morale and attitudes towards teaching is a complex question that is worthy of further investigation. At the time of this study Massachusetts teachers were being offered professional development points to take courses offered at a teaching hospital on stress reduction techniques (e.g., visualization and diaphragmatic breathing exercises) to improve MCAS scores (Boston Globe, 4/1/01). This "Band-Aid" strategy for reducing the stress caused by high-stakes tests does not address a growing underlying problem. Our study schools have suggested several factors that tend to increase or prevent teacher stress. Teachers in schools where the remediation strategy to improve test scores was in sync with the overall vision of science education (as in either Allenville or Harris) were less likely to feel resentment to the test. Teachers in schools where the remediation strategy was a clear departure from the previously held pedagogical and curricular vision (as in South Allenville) were more likely to express a decrease in their love of teaching and their feeling of effectiveness in the classroom. Veteran teaching staff, such as in Harris where the mean years of service was 25.7 years, or in Wilcox where the mean was 20 years, felt less pressure to comply with mandates to change their instructional approach for the test. Teachers in highly challenged settings were often left demoralized as the gulf between test expectations and students' current skills seemed unbridgeable.

The kinds of stress experienced may also be related to the freedom teachers felt they had not to comply with a new school strategy to raise scores; as McNeil has shown, when teachers find new mandates burdensome or contrary to their educational philosophy, they may respond with various kinds of "resistance," which are in fact an important part of the dialectic of school change (McNeil, 1986; 2000). In our study schools, the potential for teachers to "opt out" or be "semi-compliant" seemed related to the clarity of the remediation strategy, the transmission of this strategy by superintendent, principal and curriculum coordinator, the proximity of teachers to their retirement, and teachers' perception of the effectiveness of a particular strategy for the specific children whom they served.

Final version published in: Journal of Educational Change 5: 345-387, 2004 © 2004 Kluwer Academic Publishers
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

This study has begun to look at the relationship between schools' remediation strategies for high stakes tests and their preexisting vision of science education reform. The relationship between these two is important in understanding teacher responses within their classroom. The story of the way the culture in these 6 districts shaped their initial responses to state testing makes a valuable contribution to our understanding of the forces that shape the translation of educational policy into practice (Darling-Hammond, 1990; Hargreaves et al., 2001; McNeil, 1986; Spillane & Callaghan, 2000). We see that in addition to factors such as community demographics, the availability of resources, and teacher attitudes, a district's previous history of innovation constitutes an interpretive frame that affects how new innovations and reforms are understood and acted upon. The study raises many questions about the nature of complementary or competing reforms, the role that a district can play in magnifying or ameliorating stress associated with high stakes tests and the effects of district strategies to improve MCAS scores on teachers' approaches to curriculum and pedagogy. Further study is needed on each of these topics.

References

Abeille, A., & Hurley, N. (2000). *How Massachusetts schools are using MCAS to change curriculum, assessment, and resource allocation*. Boston: Massachusetts Education Reform Review Commission (MERRC).

Amrein, A. L., & Berliner, D. C. (2002). High stakes testing, uncertainty, and student learning. *Educational Policy Analysis Archives, 10*(18).

Borman, K. M., Kromrey, J., Katzenmeyer, W., & Della Piana, G. (2000). *How do standards matter? Linking policy to practice in four cities implementing systemic reform.* Paper presented to annual meeting of American Educational Research Association, New Orleans, April 2000.

Cheng, L., & Coutre, J. (2000). Teachers' Work in the Global Culture of Performance. *The Alberta Journal of Educational Research, XLVI*(1), 65-74.

Clarke, M., Shore, A., Rhoades, K., Abrams, L., Miao, J., & Li, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: findings from interviews with educators in low-, medium-, and high-stakes states*. Boston: National Board on Educational Testing and Public Policy, Lynch School of Education, Boston College

Cuban, L. (1998). How schools change reforms. *Teachers College Record, 99*, 453-477.

Darling-Hammond, L. (1990). Instructional policy into practice: "The power of the bottom over the top." *Educational Evaluation and Policy Analysis, 12*, 233-241.

Denzin, N., & Lincoln, Y. (1994). *Handbook of Qualitative Research.* Thousand Oaks, CA: Sage.

Drayton, B., & Falk, J. (2001). Tell-tale signs of the inquiry-oriented classroom. *Bulletin of the National Assoc. of Secondary School Principals*, *85*(623), 24-34.

Eisner, E. (1992). Educational reform and the ecology of schooling. *Teachers College Record, 93*(4), 610-627.

Falk, J., & Drayton, B. (2000). *The Inquiry-based Classroom in Context: Bridging the Gap Between Teacher's Practice and Policy Mandates.* Paper presented to the American Educational Research Association, annual meeting, New Orleans April 2000.

Falk, J., & Drayton, B. (2001). Cultivating a culture of inquiry. *Hands On! 23*(2), p. 1.

Firestone, W. A., & Mayerowitz, D. (2000). Rethinking "high stakes": Lessons from the United States and England and Wales. *Teachers College Record, 102*, 724-749.

Firestone, W. A., Mayerowitz, D., & Fairman, J. (1998). Performance-based assessment and instructional change*:* The effects of testing in Maine and Maryland*. Educational Evaluation and Policy Analysis, 20*(2), 95-113.

Final version published in: Journal of Educational Change 5: 345-387, 2004
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

Gaudet, R. (1999). *Effective school districts in Massachusetts*. Boston: The Donahue Institute.
Gaudet, R. (2000). *Effective school districts in Massachusetts*. Boston: The Donahue Institute.
Goertz, M.E., & Duffy, M.C. (2001). *Assessment and accountability systems in the 50 states: 1999-2000*. CPRE Policy Briefs,RR-046. Philadelphia: Consortium for Policy Research in Education.
Haberman, M. (2003). The pedagogy of poverty versus good teaching. Eisenhower National Clearinghouse. Retrieved from http://www.enc.org/topics/equity/articles/document.shtm?input=ACQ-111376-1376.
Haney, W. (2000). The myth of the Texas miracle in education. *Educational Policy Analysis Archives, 8*(41).
Hargreaves, A. (1999). *Teaching in a box: Emotional geographies of teaching*. Presentation to Third International Conference on Teacher Education, Beyt Berl, Israel.
Hargreaves, A. (2001). Emotional geographies of teaching. *Teachers College Record,*103, 1056-1080
Hargreaves, A. (2003). *Teaching in the knowledge Society*. New York; Teachers College Press.
Hargreaves, A., Earl, L., Moore, S., & Manning, S. (2001). *Learning to Change: Teaching Beyond Subjects and Standards*. San Francisco: Jossey-Bass.
Heubert, J. P., & Hauser, R. M. (Eds.). (1999). *High Stakes: Testing for tracking, promotion, and graduation*. Washington, D. C.: National Academy Press.
Hoover, R. L. (2000). Forces and Factors Affecting Ohio Proficiency Test Performance, Retrieved from http://cc.ysu.edu/~rlhoover/OPT.
Jones, M., Gail, B. D., Jones, B., Hardin, L., Chapman, T., Yarbrough, T., & Davis, M. (1999). The impact of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan, 81*, 199-203.
Kersaint, G. K., Borman, M., Lee, R., & Boydston, T. L. (2001). Balancing the contradictions between accountability and systemic reform. *Journal of School Leadership, 11*, 217-240.
Kohn, A. (2000). *The case against standardized testing.* Portsmouth, NH: Heinemann.
Koretz. et al. (1991). State comparisons using NAEP: Large costs, disappointing benefits. *Educational Researcher, 20*, 19-21.
Madaus, G. F., & 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.* New York: The Century Foundation.
Maehr, M. L., & Midgley, C. (1996). *Transforming school cultures*. Boulder, CO: WestviewPress.
Massachusetts Deptartment of Education. (2003). http://www.doe.mass.edu/mcas/overview_faq.html#faq1
Massachusetts Department of Education. (2000). *Science and technology/engineering curriculum framework*. Malden, MA.
Massachusetts Department of Education. (1996). *Science and technology curriculum framework: Owning the questions in science and technology*. Malden, MA.
McNeil, L. (1986). *Contradictions of control: school culture and school knowledge*. New York: Routledge.
McNeil, L. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.
McNeil, L., & Valanzuela, A. (2001). The harmful impact of the TAAS system of testing in Texas: Beneath the accountability rhetoric. In G. Orfield, & M. L. Kornhaber (Eds). *Raising standards or raising barriers?* New York: Century Foundation Press, pp. 127-150.
Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis (2nd ed).* Thousand Oaks, CA: Sage Publications.
National Research Council. (1996). *National Science Education Standards.* Washington, D.C.: National Academy Press.

Final version published in: Journal of Educational Change 5: 345-387, 2004 © 2004 Kluwer Academic Publishers
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents

National Research Council. (2000). *Inquiry and the National Science Education Standards.* Washington, D.C.: National Academy Press.

Newmann, F. M., Bryk, A. S., & Nagaoka, J. K. (2001). *Authentic intellectual work and standardized tests: conflict or coexistence?* Chicago: Consortium on Chicago School Research.

Patton, M. (1986). *Utilization-focused evaluation* (2nd Ed.). Newbury Park, CA: Sage.

Patton, M. (1987). *How to use qualitative methods in evaluation.* Newbury Park, CA: Sage.

Pedulla, J. J., Abrams, L. M., Madaus, G. F., Russell, M. K., Ramos, M. A., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Boston: National Board on Educational Testing and Public Policy, Lynch School of Education, Boston College.

Powell, R. R., & Skoog, G. (2000). Middle level integrative curriculum reform: A study of teachers' epistemological theories. *Research in Middle Level Education Annual, 23*, 1-34.

Robinson, G. E., & Brandon, D. P. (1994). *NAEP Test Scores: Should They Be Used to Compare and Rank State Educational Quality?* Arlington, VA: Educational Research Service.

Roderick, M., & Engel, M. (2001). The grasshopper and the ant: Motivational responses of low-achieving students to high-stakes testing. *Educational Evaluation and Policy Analysis, 23*(3), 197-228.

Rothstein, R. (1998). *The Way We Were? The Myths and Realities of America's Student Achievement.* New York: Century Foundation Press.

Sarason, S. (1996). *Revisiting "The culture of the school and the problem of change."* New York: Teachers College Press.

Settlage, J., & Meadows, L. (2002). Standards-based reform and its unintended consequences: Implications for science education within America's urban schools. *Journal of Research in Science Teaching, 39*(2), 114-127.

Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher, 20*(5), 8-11.

Smith, M. L., & Rottenberg, C. (1991). Unintended consequences of external testing in elementary schools. *Educational Measurement: Issues and Practice, 10*(4), 7-11.

Smith, M. L., Heinecke, W., & Noble, A. J. (1999). State assessment becomes political spectacle. *Teachers College Record, 92*, 157-191.

Spillane, J. P. (1996). School districts matter: Local educational authorities and state instructional policy. *Educational Policy, 10*(1), 63-87.

Spillane, J. P., & Callaghan, K. A. (2000). Implementing state standards for science education: What district policy makers make of the hoopla. *Journal of Research in Science teaching, 37*, 401-425.

Stake, R. (1995). *The art of case study research.* Thousand Oaks, CA: Sage.

Stiggins, R. J. (1999). Assessment, student confidence, and school success. *Phi Delta Kappan, 81*, 191-198.

Weiss, I.R., Banilower, E.R., McMahon, K.C., & Smith, P. S. (2001). *Report of the 2000 National Survey of Science and Math Education*. Chapel Hill, NC: Horizon Research, Inc.

Wideen, F., & O'Shea, T. (1997). High Stakes Testing and the Teaching of Science. *Canadian Journal of Education, 22*(4), 427-444.

Worcester Public Schools. (1998). Science Curriculum for Worcester. Worcester, MA.

Final version published in: Journal of Educational Change 5: 345-387, 2004
Publication is available at http://www.kluweronline.com/issn/1389-2843/contents