

# A study of performance on the first year of the MCAS tests

Sponsored by  
the University of Massachusetts  
Donahue Institute

by  
Dr. Robert D. Gaudet,  
Senior Research Associate

January, 1999

---

"Education, then, beyond all other devices of human origin,  
is the great equalizer of the conditions of men..."

-- Horace Mann

Testing plays an important role in most of the contemporary school reform efforts in the United States. The Massachusetts education reform effort is no exception. Its testing vehicle is the Massachusetts Comprehensive Assessment System or, as it's commonly known, the MCAS.

The chief objective of the state's education reform initiative is to enable public school students to achieve a certain level of knowledge and skill. The Massachusetts Department of Education has established this level by setting out what students are expected to learn in each basic subject. School districts are supposed to see to it that their students learn what they're expected to learn. The purpose of the MCAS is to gauge periodically how students are doing as they try to achieve this level of knowledge and skill.

Each year, in every district in the state, the MCAS tests are given to public school students in grades four, eight, and ten. They cover such academic subjects as math, science, and literacy skills. The test scores are broken down by individual student, school, and district. The scores for individual students are available to their parents, teachers, principals, and superintendents. The scores for entire schools and districts are available to the general public.

With the MCAS, the state has, for the first time in its history, an evaluation mechanism that measures how much progress students are making toward well-defined goals. At the same time, individual schools and districts are urged to anticipate and complement the MCAS by developing their own parallel methods of assessing how their students are doing. Thus, the education reform effort uses assessment as a way to help all students move toward a high level of academic achievement.

Just as this effort views higher student achievement as its end, it views the improvement of the public schools as its chief means to achieve this end. What happens in school is by no means the only or even the leading influence on how pupils currently perform on standardized academic tests. However, what happens in school obviously is the only means that is currently within the control of the schools themselves. So it's the only means of reform that is at the disposal of the education improvement effort as it now exists.

## **Improving Our Schools**

Thus, the more the test scores can be used to inform decisions about how to alter what happens in school, the better the chance to make the schools more effective in helping their students to improve their performance on standardized academic tests like the MCAS. Properly used, the results can pinpoint which approaches to teaching and learning are working and which are not. The MCAS also includes an array of diagnostic tools that let teachers and administrators spot areas where students perform poorly, so that they can work with the students to mend the weaknesses.

Consequently, the essence of education reform in Massachusetts can be summed up in a few words: Better student performance, through more effective schools.

However, for the MCAS to fulfill its intended role in the current education reform effort, there at least two important conditions that have to be met.

FIRST, the tests, and other assessments, must be fair and accurate. They must measure what children have learned, rather than just their social or economic background. They must not be biased for, or against, any group of students.

SECOND, they must be used to make the public schools more effective. Thus, the scores should drive an ongoing analysis of what makes the school experience effective. They must provide teachers with a critical piece of information about the potential learning problems and possibilities of individual students. And the information must be used as a basis for helping all students to do better.

To meet the second condition, we must be able to use the MCAS scores as one tool to discern the effectiveness of our schools. We must be able to establish how effective they are today, and to track the rise or fall of their effectiveness in the future. Thus, finding ways to measure school effectiveness is essential to education reform.

## **Measuring Effectiveness**

influenced by two broad sets of factors: school factors and non-school factors. The first entail what happens in school, and thus what is within the control of the school district itself. The second entails conditions outside the schools, such as the demographic profile of the students and the community. As we look at a given district's average score on an MCAS test, we have to be able to discern how much of the score is tied to school factors, and how much of the score is explained by non-school factors.

How well do the school design and the curriculum promote learning for all? Are teachers top-notch professionals who have both the skills and commitment to teach all students? Are professional development activities rigorously aligned with efforts to increase student achievement? Is there strong, solid leadership in the school? Are there high expectations for all? Are parents full partners in their children's education? Are there adequate resources to do the job? These are all questions about school factors.<sup>1</sup>

In the research reported in this paper, non-school factors consist largely of the overlapping demographic conditions of family life and community life. We use six such conditions in a given school district: its median level of educational attainment, its median income level, its percentage of households above the poverty line, its percentage of single-parent families, its percentage of non-English-speaking households, and its level of private school enrollment. Statistical analysis shows that these factors form much of the non-school influence on how the state's students do on such standardized tests as the MCAS.<sup>2</sup>

As we all know, students in advantaged districts tend to get higher standardized test scores than students in disadvantaged districts. Thus, if a district's students get a high average score on an MCAS or other standardized tests, the test score by itself doesn't tell us how much of the score is explained by school factors and how much is explained by non-school factors. A high score might be tied more to advantaged demography than to what actually happens in the district's schools. The score by itself isn't a sound guide to how effective the school district is.

We cannot begin to zero in on just how effective the school district itself is unless we can distinguish between the respective influences of the two types of factors. Only then can we discern how effectively the district itself performs, and how much it contributes to its students' average performance on the MCAS.

The Effectiveness Index provides insight into this distinction, and consequently provides some measure of the school district's contribution to its students' performance. Thus, it supplies a piece of crucial insight as to which schools are more effective.

For a given district, the Effectiveness Index (EI) gauges the impact that school factors have on the average MCAS score. The greater the positive impact of the school factors, the higher the district's Effectiveness Index will be.

The Index is calculated in the following manner: For a given district, the six demographic factors are used as the basis for projecting a likely average score on the MCAS. The demographically-likely score is then compared to the average score that the students in the district actually received. The Effectiveness Index is the number that represents the difference between the likely score and the actual score.

If the number is negative —if the actual score is lower than the likely score —then this suggests that what is happening in the schools in the district is not enabling its students to perform beyond the demographic expectations for them. If the number is a positive number —if the actual score is higher than the likely score —then this suggests that what is happening in the schools is helping the district's students to surpass the demographic expectations for them. (For a fuller account of the development of the Effectiveness Index, please see Appendix B.)

### **What the Effectiveness Index Tells Us: Statewide Results**

We applied the Effectiveness Index to the MCAS scores of the 200 largest school districts in the state. These districts comprise 93 percent of the total population.

The demographic differences between the 200 largest districts explain 86% of the variation in the districts' overall average test scores - that is, their scores for all of their test-taking students for the nine MCAS tests combined. Thus, though demography isn't destiny in this case, it sets a strong tendency.

A simple way to depict the respective contributions that demography and the schools make to the average level of student performance on the MCAS is this:

$$\text{DEMOGRAPHY} + \text{School} = \text{Average Score}$$

Nonetheless, a number of districts achieved test scores that are significantly higher than their demography predicts.

### **Four Types of School Districts**

The Effectiveness Index lets us identify three types of school districts: effective, noteworthy, and ineffective.

An EFFECTIVE district meets two specifications:

- 1) Its Effectiveness Index is a positive number - that is, its actual

2) Its actual score is equal to or higher than the average MCAS score for the state as a whole.

Thus, Stoneham, whose demography places it in the middle of the state's demographic ladder, is an example of an effective district. Its actual score is substantially higher than its likely score. And its actual score is higher than the statewide average score. Indeed, its actual score ranks it 46th among the state's 200 largest school districts. Its demography would predict that its score would rank 111.

A NOTEWORTHY district fits the first specification but doesn't fit the second. Since its performance helps its students to go beyond their demography, it is still worthy of note. For what such a district is doing can hold useful lessons for districts that are demographically similar, but do not outscore their demography. And such a district is more likely to deliver a return on future public investment than an ineffective district is.

Here, Everett and Worcester are outstanding examples. Everett's overall score on all nine tests combined is much higher than its demography predicts. Worcester's scores on the grade four tests surpass its demographic prediction.

An INEFFECTIVE district has a negative index number, and its actual score that is less than the average MCAS score for the state as a whole.

For each of these three types of districts, the Effectiveness Index sets a baseline for improvement. If a district is ineffective, then its short-term goal should be to become effective. If a district is noteworthy, its short-term goal should be to get its test scores high enough to exceed the statewide average. If a district is ineffective, its short-term goal should be to raise its actual test scores further, so that it will become a fourth type of district-- a SUCCESSFUL district.

A SUCCESSFUL district, as presently defined, gets a 75% pass rate on each of the tested subjects in each of the three grades.

### **Making the Grade**

Currently, no successful districts exist, because none of the state's 200 largest districts meet the department's definition of a successful district. Presumably, some districts will improve their test scores by enough in the near future to do so.

However, a number of districts did come close to achieving success. On five of the nine MCAS tests, 75% of Harvard's students earned passing scores. Harvard got an overall score of 2208 on all nine of the MCAS tests combined, higher than any of the other 199 districts. In Medfield and Wellesley, 75% of the students earned a passing score on

None of these districts is among the 50 most demographically disadvantaged in the state.

This study uses the Effectiveness Index to identify, and then to rank, on each of several fronts, the 50 most effective districts and 10 most noteworthy districts. Thus, on pages TK through TK, there are four sets of rankings:

Overall performance: all three subjects and all three grades combined

Grade four performance: all three subjects combined

Grade eight performance: all three subjects combined

Grade ten performance: all three subjects combined

### **The Importance of Reading and Writing**

For each of the three grades, this study also ranks the districts that are effective or noteworthy in language arts --essentially, reading and writing.

This is because reading and writing are necessary conditions for doing well on the MCAS tests. This is true even of the tests in mathematics. Many of the problems on the mathematics's tests, particularly in grade eight and ten, are word problems. You cannot understand these problems if you cannot understand the words. In all subjects, moreover, many questions call for a written answer, as short as a sentence or two or as long as an essay of several paragraphs.

Finally, the children who took the grade four MCAS tests in May of 1998 entered first grade in 1994, the first full year that the Education Reform Act was in force. Thus, they are the children of this long-term education reform initiative. The entire span of their K-12 experience will be shaped by the requirements established by the reform act. How well these children do in their school careers will be the first full measure of the impact of the act.

The three sets of language arts rankings appear on pages TK through TK.

This study highlights these districts because they might have lessons to offer to other districts to help them to enhance their contribution to their students' future performance of the MCAS.

make to the average level of student performance on the MCAS tests in this fashion:

$$\text{Demography} + \text{SCHOOL} = \text{AVERAGE SCORE}$$

## **Middle Massachusetts**

In the demographic ranking of the 200 largest school districts, 100 districts are concentrated in the demographic middle of the state. These districts, with 2 million people, make up what might be called Middle Massachusetts. They may be well-suited to play a crucial role in the short-term future of education reform.

For the state as a whole, as we've seen, demographic differences between the 200 largest districts explain 86% of the variation in the districts' average overall test scores. All or much of the other 14% of the variation is probably explained by the differences in how the school districts themselves behave.

However, this 14% isn't spread out evenly across all 200 districts. Little of this variation is found in the most advantaged districts, where the relationship between demography and test scores is generally strong. Thus, these advantaged districts generally get high test scores. The same holds for the most disadvantaged districts. Here, the relationship between demography and test scores is also generally strong. The actual test scores of all these districts are well below the statewide average.

The pattern in Middle Massachusetts is different. Its districts exhibit a wide range of test scores -- even though their demography is relatively similar. Thus, much of the 14 points of variation is concentrated in the 100 Middle Mass districts.

This variation can be seen in these bar graphs. For each of the 20 districts closest to the demographic middle of the state -- thus, the districts that form the middle of Middle Massachusetts itself -- the tip of one bar represents its demography, and the tip of the other bar represents its MCAS test score.

[INSERT BAR GRAPH] [MAYBE YOU LABEL ONLY THE POSITIVE DISTRICTS BY NAME.]

Since the demographic variation is slight, but the variation in test scores is great, this pattern suggests that much of the variation is explained less by demography than by differences in what the schools of Middle Mass are doing.

Further, the test scores of Middle Mass districts with high positive numbers on the

Thus, the scores of a MiddleMass district like East Longmeadow are very nearly equal to the scores of Longmeadow, one of the more demographically advantaged districts in the state.

Some MiddleMass districts did particularly well on the grade four tests. The scores of Shrewsbury, Pembroke, and East Longmeadow were equal to or higher than the scores of such advantaged districts as Norwell, Cohasset, and Duxbury. To be sure, these districts had high scores, as one would expect. What's less expected is that districts farther down the demographic ladder did just as well.

So, if more MiddleMass districts become as effective as East Longmeadow, Shrewsbury, and Pembroke, then more MiddleMass districts will get test scores as high as the test scores of the advantaged districts.

Moreover, insofar as MiddleMass districts are demographically similar, what makes for effective schools in an effective districts in MiddleMass is more likely to make for effective schools in an ineffective district MiddleMassachusetts.

Thus, in the short run, Middle Massachusetts can be an especially fruitful place to seek, and expect to find, a relatively swift rise in MCAS test scores.

#### **Education Reform in Massachusetts**

The Education Reform Act of 1993 provides an opportunity to transform our schools. The MCAS can be the backbone of our effort to do so. It can assess the performance of districts, schools, and individual students, and it can inform the public about its schools. More importantly, MCAS's built-in diagnostics can help teachers to help all children learn better. Under the act, increased state funds provide substantial amounts of new money for districts to use for reform.

Massachusetts stands at a critical crossroads. The elements are in place for exciting statewide reform, but the barriers to change are substantial. This study captures the role that demography plays in student performance. Again, though demography is not destiny, it does establish a tendency. If we overlook the tendency of disadvantaged districts to produce low scores, then we will continue to consign the children of those districts to a future of unfulfilled potential.

Without a broad long-term effort to do what is needed to enhance performance by all, we can expect more of the same: polarization of academic performance that troubles even those whose children are fortunate enough to have been born into a situation that makes a powerful contribution to their academic success.

-----

#### **FOOTNOTES**

There is no standard accounting procedure for establishing PPE. For example, some systems might include teacher retirement costs, capital costs, federal funds, and long-term disability obligations in their per-pupil spending figure. Others might not. So comparisons across districts are difficult to make.

2. Other family and community conditions are crucial to student success, but are hard to observe and measure. One would have to monitor many families and communities closely over time to discern how family and community behavior affects school outcomes. How many books are read in the family? How much time is taken up by TV-watching? How do the community's adults treat children other than their own? Does the community mentor its young people? It's hard to get reliable answers to such questions. But we do know that the children of advantaged families and communities are more likely on average to have resources and support, and children of less advantaged situations are less likely to have them. So we use gross measures of such support as a proxy for answers to the more specific questions that are so hard to pursue.

For more information, please contact Robert Gaudet at the UMass Donahue Institute. Phone (617)

## **APPENDIX A: Community Listings**

THE LISTS: TOP 40 OVERALL EFFECTIVE; TOP 10 NOTEWORTHY; SAME FOR OVERALL 4TH GRADE, 8TH GRADE, AND 10TH GRADE; 4TH GRADE, 8TH GRADE, 10TH GRADE LANGUAGE ARTS

(TOTAL OF 7 LISTS)

## **APPENDIX B: Deriving the Effectiveness Index**

The Effectiveness Index (EI) is derived by comparing actual scores on standardized tests with scores as predicted by a model which factors in the role community characteristics play in educational outcomes.

The Community Effects Factor (CEF) model was developed in a doctoral dissertation (*Education Achievement Communities: A New Model for "Kind of Community" in Massachusetts Based on an Analysis of Community Characteristics Affecting Educational Outcomes*, May 1999, University of Massachusetts, Amherst). That work is the basis for determining school effectiveness. The model examines the relationship between selected demographic characteristics and educational outcomes. These characteristics include: average education level, average income, poverty rate, single-parent status, language spoken, and percentage of school-age population enrolled in private schools. These variables were chosen because they correlate with achievement and because the education literature identifies them as connected to academic performance.

In order to refine a better CEF model, it is first necessary to factor the impact of these demographic variables on each other. This can be done through a technique known as principal component analysis that is a statistical mechanism that reduces many variables to a few salient ones that have the most impact on an outcome. Once the factors have been identified, a regression analysis produces the equations that can be used to either build a kind-of-community model or to predict expected district performance on achievement tests. The degree to which a community's characteristics

demography.

The CEF, which is a measure of the demographic lift or drag of each community concerning educational achievement, is a good point of departure for analyzing school and school district effectiveness. The CEF identifies expected levels of performance based on community characteristics which, for better or worse, are very powerful indicators of educational achievement in Massachusetts. In this analysis, Weston is the most demographically advantaged community in the state in terms of educational outcomes (CEF = + 2.8), and Lawrence is the least advantaged (CEF = -4.8). The CEF has a strong relationship, or correlation, to test scores.

Correlation is a process that identifies the interdependence of one variable with another. Correlation simply shows "the extent to which two things typically run together." [*The Economist*, 6 Dec. 1997, p. 82]. Correlation is not equivalent to causation; it can only reveal tendencies between variables, not identify causes. Correlations simply demonstrate relationships. A perfect correlation would be 1.0. For example, the correlation between inches and feet is 1.0 because it is a perfect linear fit; 12 inches always equals one foot. Correlations in real world situations involving human behavior are never 1.0.

The correlation, or the connection, between spending (Per-Pupil Expenditure or PPE) and achievement in Massachusetts is .28, which is relatively low. While spending clearly matters, merely increasing spending levels has a relatively weak impact on results. Increasingly, many people are coming to the realization that how a system spends money is more important than how much money it spends. The achievement outcome accounted for by the community effects factor (CEF) is much stronger; that relationship is .86. This is not to say the community context, the CEF, is the most important determinant of school success, but it is a significant element that must be a major consideration in any plan to improve education in disadvantaged areas.

The Effectiveness Index was generated in the following manner:

- o Utilize the 1998 MCAS results as an outcome indicator for achievement in each of the state's most populous 200 communities. (NOTE: This model does not evaluate results in the 151 smallest communities of the state which comprise about 7% of the population.)

- o Utilize the CEF model to predict a score for each district. This predicted score is based solely on community characteristics as they affect educational outcomes.

- o Compare the actual to the predicted score. Systems whose actual scores are significantly higher than predicted scores and whose absolute scores are at or above state average are identified as effective. Systems with positive effectiveness indexes but

The basis for this model was developed as a doctoral dissertation. The following provides detail on the statistics behind the model which is used to predict expected scores based on demographics.

(Statistical info)

REV 2/3/99