Undergraduate Economic Review

Volume 6 | Issue 1 Article 8

2010

The Impact of High Stakes Standardized Testing on High and Low Achieving School Districts: The Case of the MCAS

Robert J. Havdala
Tufts University, robert.havdala@gmail.com

The Impact of High Stakes Standardized Testing on High and Low Achieving School Districts: The Case of the MCAS

Abstract

Ever since the passage of No Child Left Behind in 2002, high stakes testing has been at the forefront of education debate. Though these reforms are frequently meant to help underperforming schools, there may be unintended effects on high performing districts. This study examines the impact of the Massachusetts high stakes testing scheme (MCAS) by analyzing high school dropout rates, future plans of graduating seniors, and SAT results. I find evidence that high stakes accountability has had an insignificant, if not relatively negative, impact on low achieving districts and that Massachusetts students have not improved compared to national trends.

Keywords

accountability, high stakes testing, MCAS, high achieving, low achieving

Cover Page Footnote

The author would like to thank professor and thesis advisor Jeff Zabel, without whom this project would never have been possible.

1. Introduction

Since 1998, every student in a Massachusetts public school takes the state-administered Massachusetts Comprehensive Assessment System (MCAS) exam. For some students, the exam is the culmination of months of preparation, the determining factor as to whether they may graduate. But for many high achieving students, the two-week testing period is vacation: it's a break from classes. From an educator's perspective, this latter group is losing out on valuable class time if they all pass. So should we make them take it?

Backtrack to 1993, when the Massachusetts Supreme Judicial Court ruled in *McDuffy v. Robertson* that the state had not been doing enough to financially support certain schools. In the words of the court, "the reality is that children in the less affluent communities (or in the less affluent parts of them) are not receiving their constitutional entitlement of education as intended and mandated by the framers of the Constitution." The *Massachusetts Education Reform Act* promised to solve some of these problems. It was designed to reform and modernize Massachusetts public education by equalizing resources across districts, legalizing charter schools, and instituting a statewide exam system over a seven year period. This exam system became the MCAS.

Lawmakers argued that with MCAS results, the state could evaluate school progress towards meeting newly set learning expectations and adjust funding allocations accordingly.[1] Prior to 1998, the state used the Massachusetts Educational Assessment Program (MEAP) exam to evaluate success of schools around Massachusetts. It was administered biennially to both the 4th and 8th grades and either the 10th or 12th grade, depending on the year. However, schools had little incentive to focus on the exam as poor results led to few consequences. With the shift to MCAS in 1998, everything changed. By the spring of 2001, all 10th grade students in Massachusetts public schools were required to pass the English Language Arts (ELA) section and the math section in order to graduate. Schools faced repercussions for poor test results, including the threat of state takeover. And the stakes are only growing. In 2009, all 10th grade students were required to pass English and math sections, as well as one science section of their choosing, to graduate.[2]

The MCAS came at the same time when the federal government was adding its weight to testing too. *No Child Left Behind* (NCLB), passed in 2002, stipulated that states implement high-stakes exams, a qualification which MCAS

---

[1]The state recently took control of thirty underperforming high schools in Massachusetts after poor MCAS results and high dropout rates. Control was revoked from the district and given to the state's department of education.

[2] MCAS initially tested students in the 4th, 8th, and 10th grades in English and math to evaluate schools. Because of NCLB, the exam has expanded to the point that students must take MCAS exams every year from 3rd grade to 10th grade (except for 9th grade), in an increasing number of subjects.

fulfilled. On the national level, a number of papers have highlighted the consequences of NCLB and its impacts (Neal & Schanzenbach (2007), Reback (2007), and Loveless (2008), among others). The act has stirred much controversy, of which one debate is its effect on high achieving students. With its legal framework focused on improving underprivileged schools, it places little emphasis on those that consistently pass. It is possible, therefore, that these high achieving schools were adversely affected by laws intended to help the underprivileged. The same applies in Massachusetts. Though the MCAS may have helped the state identify struggling schools, there may very well have also been unintended consequences on high achieving students.

Instead of only focusing on underachieving schools, this paper will focus on the MCAS's effect on both low and high achieving districts. More specifically, I pose the question: has the MCAS had a detrimental effect on the academic achievement of high achieving districts in both an absolute sense and in a relative sense compared to low achieving districts? Given the emphasis on low performing school districts, it is distinctly possible that they have performed relatively better compared to high achieving schools since 1998 (the year MCAS was introduced) or 2001 (the year MCAS became high stakes). High stakes testing analyses over the past three decades on both the national and state level have produced conflicting results. With eight years of data since the shift to high stakes testing, it is an opportune time to examine the case of Massachusetts and examine whether the high stakes MCAS has had a positive or negative influence on the best and worst performing districts.

In Section 2, I consider previous research on high stakes testing, ranging from individual to national analyses. Section 3 describes the data set used for the study, while Section 4 develops the framework for analysis. This includes the metrics used for determining performance and the regressions run. The results section (Section 5) analyzes the outcomes of these regressions and determines whether the results are significant. In the conclusion (Section 6), the paper summarizes these findings in the context of previous studies on the national level. The appendices provide more detailed tabular results for the regressions.

2. Literature Review

This section cites a number of studies that address the effect of high stakes testing on student performance. Some specifically consider the impact on low achieving students versus high achieving students, while others consider the impacts more holistically. They vary widely in their results, and thus this section will be organized based on their findings. The first section will consist of studies that find the gap between low and high achieving students closing; the second will include studies which find the gap widening; the third section will contain studies that believe the gap remains unchanged; and the fourth will consider the impact on low income and urban students.

2.1 The gap is closing

One of the best known studies on high achieving students is by Tom Loveless (2008), who examined the impact of NCLB on high achieving students. He utilized national student-level data from the 4th and 8th grade National Assessment of Educational Progress (NAEP) exam, one of the nation's oldest exams and one that is administered to a random sampling of schools around the country.[3] He defined students at the 10th percentile as low achieving students and students at the 90th percentile as high achieving students, and tested the possibility that, since NCLB, the scores of high achieving students on the NAEP had slowed relative to those of lower achieving students. He analyzed these groups' NAEP scores over time, using 2002, the year that NCLB was passed, as the significant year in his regressions. His research confirmed his hypothesis, indicating over a year's worth of improvement of learning in low achieving students. Though high achieving students did not stop improving, their progress had slowed drastically since 2002.

Carnoy and Loeb (2002) found no difference in dropout rates after high stakes testing was implemented. They first created an index for the strength of accountability for each state and compared the index to student improvement on NAEP math tests between 1996 and 2000. In their model they used both 4th and 8th grade NAEP exams, and they included a term for "survival rate" (the proportion of students who reach the 12th grade) as well. The recursive regression model found no evidence of a relationship between accountability and high school completion rates or retention rates. However, they found that, on average, states which shifted to high accountability exams had greater improvements on the 8th grade NAEP math exam than states which did not implement high stakes testing. Carnoy and Loeb posited that schools with high achieving students may feel pressure and have a better ability to increase performance in response to greater external accountability.

Using individual level data in high stakes exams, Randall Reback (2007) examined the possibility that teachers overcommit resources to students immediately around the passing threshold. He focused on the Texas standardized testing scheme and found improvement in low achieving students. In contrast, high achieving students were unaffected. More specifically, students were particularly successful when their score was important to a school's accountability rating. This implies a short-term shift in resources towards low achieving students at the cost of high achieving students. To use Reback's own words, "Relatively high achieving students perform worse than usual if their own performance is irrelevant to the short-run accountability incentives."

2.2 The gap is widening

[3] One particular advantage of the exam is that it has no ceiling effect. In other words, scores are linear, making the difference between a 100 and 101 equal to the difference between a 200 and 201.

In contrast to Loveless, Carnoy and Loeb, and Reback, a wide array of studies have found an increasingly large gap between low and high achieving schools as a result of high stakes exams. Neal and Schanzenbach (2007) analyzed test scores in the Chicago school district from 2001 to 2002, a period when Chicago Public Schools shifted from a system of low stakes testing to a high stakes system. Though it was unclear whether high achieving students made any progress, low achieving students continued to lag far behind others. Only those students who were initially around the proficiency threshold had a significant improvement in scores. Such findings indicated the possibility that teachers focused their efforts on those students they felt could be pushed over the threshold, at the cost of those who were far above or far below (reflecting the threshold findings of Reback).[4]

In a study of 32 communities in metropolitan Boston, Bolon (2001) published a report on the 10th grade math MCAS. More specifically, he developed a model of performance on the exam in an attempt to be able to predict results on future exams. He found that covariates were immensely significant in his models—per capita community income, for example, accounted for 84% of the variance in performance, "by far the strongest factor in predicting tenth grade MCAS mathematics scores." Percentage of limited English proficiency students in each district was the second most significant influence, though its overall effect was small. Variables involving race were not found to be statistically significant, and school spending only had a weak association. Bolon concluded that social factors had the strongest impact, and that year-to-year changes were small, statistical uncertainties that could be explained by simple variations in the data. Furthermore, schools which succeeded were more likely to succeed in the future, whereas those who failed were more likely to continue failing. The increase in state funding from $1.3 billion to $3 billion for low-income schools seemed to have little impact. Overall, Bolon argued, the implementation of the MCAS actually widened the achievement gap further.

2.3 The gap is unchanged

In contrast to these previous studies, Jacob (2001) had differing results when he used data from the National Educational Longitudinal Survey (NELS) to evaluate the effect of exit exams on graduation rates and math or reading achievement. He found neither an appreciable effect of graduation tests on the probability that an average student graduates nor a positive impact on student achievement. In fact, the only students who seemed to be affected by the exams were the lower achieving students. In states where testing was present, the lowest decile saw small gains in learning, yet they also were 25% more likely to drop out of school than the bottom decile in states without high stakes tests. On average,

[4] Herein lies an adverse selection issue. Students who are far below the threshold might be encouraged to drop out by school officials so that district test scores are not lowered, and thus officials would not face repercussions.

students in both test states and non-test states made the same learning gains in high school, though those with testing schemes entered high school further behind, and these states also serve more disadvantaged populations.

Springer (2008) addressed the previously mentioned findings of Reback (2007) regarding teachers teaching to threshold students. With data from an internet test developed by the Northwest Education Association, Springer examined 3rd to 8th grade results in an unnamed northwest state. In this state with a new high stakes testing scheme, he did not find this threshold phenomenon, what he calls "educational triage." Instead, he found students at all levels improving under high stakes testing. Springer does leave open the possibility that schools shifted resources or studied to the test, but his results provide at least some evidence against the possibility of focusing on students at the threshold.

2.4 Effects on low income and urban students

Papay, Murnane, and Willett (2009) took a unique approach to the aforementioned threshold issue of Reback (2007) on the MCAS, choosing to analyze only urban school students around the passing threshold. Using individual level data from the 10th grade math MCAS in 2004, the authors attempted to predict whether a student would graduate based on their passage or failure of the exam. With a regression discontinuity design, their results indicated that there was no difference in dropout rate between those who barely passed and those who barely failed the exam, with one notable exception: students who were in urban and low-income school districts. Interestingly, these findings were not consistent with the ELA MCAS, which had no such effect. Papay, Murnane, and Willett provide the explanation that suburban districts have fewer students who fail, and can concentrate resources on struggling students. Urban, low income schools, however, have a greater number of failing students, have limited resources, and thus cannot afford to pay for one-on-one remediation as suburban schools can. This phenomenon is best exemplified with the following statistic: 80% of students in the 2006 cohort graduated on time, but only 57% of urban, low income students did. Still, the authors concede that the effect of high stakes exams is unclear: they could encourage a student to graduate who might not otherwise, or alternatively they could discourage a student to drop out who otherwise might have graduated. Papay, Murnane, and Willett sum up their results, writing, "Failure is clearly only one of the many factors that contribute to the dropout decision."

Warren, Jenkins, and Kulick (2006) undertook a long-term analysis of high school exit examinations in association with high school completion as well as GED rates between 1975 and 2002. In contrast with other studies, they account for more difficult high stakes exams as a separate group than simpler, minimum competency exams. They found an association between the presence of high stakes exam schemes and lower high school completion rates, as well as higher rates of GED test taking. This association was also more evident when poverty rates or proportions of racial and ethnic minorities increased. Additionally, in states where exams were deemed "more difficult," graduation rates were about

2.1 percentage points lower than schools with minimum competency high stakes exams.

2.5 Conclusion

The jury is still out as to whether high stakes testing provides adequate incentives for education systems to improve performance. From these studies, there really is no clear-cut answer as to whether high stakes tests close or widen the achievement gap. Thus, further research is necessary to investigate the true answer to this question. In the following section, the study addresses the methodological framework used to analyze the data, including a summary of the performance variables and covariates used in the models.

3. Data Description

The data set was compiled using information from the Massachusetts Department of Education (MADOE) and the US Department of Education (USDOE).[5] MEAP and MCAS test score data was gathered from the MADOE, while school characteristics were taken from both sources. Additional dropout rates, senior plans, and SAT results from the MADOE were also merged into the dataset. All data used are at the district level.[6] Statistics are reported annually between 1994 or 1995 and 2007 or 2008, depending on the performance statistic. In cases where only school data are reported and there is more than one high school in a district, weighted averages are calculated using enrollment as weights. Additionally, all vocational, charter, and agricultural schools are omitted from the data, since their funding structure and focus differ from traditional public high schools.[7] Any year listed refers to the academic year which started the autumn prior; for example, 2008 refers to the 2007-2008 school year. Finally, all figures and tables referenced in this section are located in Appendix 1.

3.1 High Achieving versus Low Achieving Districts

To determine which districts are high achieving and low achieving, I use 10th grade scores from the 1996 MEAP exam and 2000 MCAS exam. These two years were chosen because of the shift from the MEAP to the MCAS in 1998, and the shift from a low-stakes MCAS in 2000 to a high-stakes one in 2001. I refer to these years as "critical years" in the methodology.

The MEAP was administered to 4th, 8th, and 12th grade students in 1988, 1990, and 1992, and to 4th, 8th, and 10th grade students in 1994 and 1996. The test was made up mostly of math and English multiple-choice questions. Students designated as special needs or with limited English proficiency were not required to take the exam. Results were reported on a scale between 1000 and 1600, with the state average set at 1300. Individual results were reported in one of five levels.

[5] The dataset was initially used for Downes and Zabel (2007).

[6] Limitations in dropout rate data prevented any analysis on the school or individual level.

[7] Charter schools in particular were legalized in 1993, though most were not founded until 1998 or later.

To compare districts, I summed the mean math and English MEAP scores for each district.[8] MCAS results, on the other hand, are reported in one of four performance levels.[9] For the year 2000, I summed the total percentage of students from each district in the advanced and proficient category (required for graduation) for math and English, and grouped them on that metric.

Four different groupings were used in this study: two in 1996 and two in 2000. Within each year, high achieving districts were defined as those in either the top quartile or top decile of school districts within Massachusetts in that year's exam, while those in the lowest quartile or decile were defined as low achieving districts. These groupings were fairly robust regardless of the grouping definition.

3.2 Performance Variables

Three variables were used in this paper to measure district performance: high school dropout rates, senior plans, and district average SAT math and verbal scores. Means for these variables are summarized by year in Table 1.2.

One important goal for educators is to keep kids in school. The Massachusetts Education commission notes that students who never graduate with a high school diploma are much less likely to get a job and are more likely to be involved in crime (Lee, Shaefer, and Messner-Zidell 2007). As such, the dropout rate for students in grades 9-12 is a particularly good metric to gauge a district's ability to fulfill this objective. The adjusted rate was first used in 1993, the earliest point at which this report uses the data (see Appendix 2 for MADOE's calculation for dropout rate). Figure 1.1 depicts these annual Massachusetts rates by group. A time trend regression shows that the groups are, in fact, diverging over time (see Table 1.3), an indication that high achieving groups are performing better relative to low achieving groups. However, a separate regression indicated that there was no significant break or change from this trend in 1998 (see Table 1.4) or 2001. Additionally, some of the fluctuations in data can be attributed to the changes in the dropout calculation, also noted in Appendix 2.

The second performance variable is the results from an annual poll given to all seniors graduating from Massachusetts high schools. The methods of this survey are described by the MADOE as follows: "Data about the plans of high school graduates were obtained from the Year-End School Indicator Report, a survey of Massachusetts public schools conducted by the Department of Elementary and Secondary Education at the end of every school year. School officials report the number of graduating students by gender and race across nine categories of post-graduation plans." On this survey, students could be placed in

---

[8] All 1992 MEAP scores were fairly correlated to those from 1996 (r=.7588). Low achieving district scores had about the same correlation (r=.7451), though high achieving district scores were less correlated (r=.5732).

[9] A large number of special needs students are required to take the MCAS, not true for the MEAP, though exemptions are still made. Students with limited English proficiency, however, must take the test.

one of the following options: 4-year public college, 4-year private college, 2-year public college, 2-year private college, other post secondary, work, military, other, or unknown. Results were given in percentage terms. I summed together all responses for 4-year universities to create the statistic "four." The use of 4-year universities is motivated by the focus of this analysis on high achieving students.[10] These results are available back to the graduating class of 1995.[11] Statewide means of senior plans are shown in Figure 1.2. Overall, the percentage of students attending a 4-year college is steadily increasing, though upon inspection between achievement groups, the gap seems to be diverging. A time trend regression confirmed these results (see Table 1.3), although again there was no significant break in 1998 or 2001.

The final performance variable is the average score in each district from the math sections and the verbal sections of the SAT. Scores in each section are on a scale of 200 (the lowest) to 800 (the highest), are given in 10-point increments, and are distributed on a normal curve.[12] The exam is administered on a national basis, and though it is technically not mandatory or high stakes, many colleges and universities require SAT results for admission. Thus, much like senior plans, SAT results are also a good indicator of high achieving students, particularly those who will attend college. Scores from the two sections were summed together to create one total SAT score variable. Because CollegeBoard, the testing agency, has private ownership of data rights, SAT scores are sparsely available and were only found for the following seven years: 1995, 1997, 2000, 2001, 2005, 2006, and 2007. The mean scores for each district, by year, are depicted by group in Figure 1.3. Results from a time trend indicated no significant change in the score gap (see Table 1.3) and further analysis showed no distinct break in 1998 or 2001.

Conventionally, one might think that schools aim to have low dropout rates, high college plans, and high SAT scores. In theory, dropout rates should increase with a new standard (such as a high stakes test) that may inhibit graduation. But, as noted by Papay, Murnane, and Willett (2009), there is the equal possibility that students who pass could discover a newfound motivation to graduate. Thus, the effect on dropout rate could go either way. Furthermore, a high dropout rate could unintentionally have a positive effect on senior plan and SAT results. Since the poll is done in percentage terms in the final year of school, all students must have already passed the MCAS, biasing results upward for the entire district. Additionally, students who do not pass the MCAS are unlikely to

[10] Undoubtedly, I recognize that those who attended two-year schools or those who entered the military or the workforce are not "failures" by any means—and many may have continued onto a 4-year college afterwards.
[11] On occasion, a district listed more than 50% of students "unknown," a situation in which the district's results for that year were omitted.
[12] The exam format changed in 2006—see Appendix 2 for further details.

ever take the SAT, raising that mean district score. Following this logic, there is a selection bias: all else equal, districts with high dropout rates could be more likely to boast better senior plans and SAT results than those with lower dropout rates because underperforming students would have dropped out. But with the dropout rate effect ambiguous, senior plan results and SAT scores could go either way as well. These countervailing effects makes it virtually impossible to distinguish whether a high percentage of seniors attending a four year college and a high SAT score is actually an indicator of district success or failure.

Based on these statewide trends, there has been little change in the status of Massachusetts students over the last 10-15 years. In general, rates remain relatively constant and, arguably, perhaps improving slightly. And though there is some evidence of the gap between high and low achieving districts widening with regards to dropout rates and senior plans, no break or shift seems to occur in either 1998 or 2001 that could be attributable to MCAS.

Covariates

The covariates considered for this report are the percentage of low income students, the percentage of special education students, the total enrollment of secondary students in each district (which was divided by 1000), and the expenditures per pupil in each district (also divided by 1000). This last variable is based on the overall day program cost per student, across all grades (data for only secondary school students was not available).[13] To account for inflation in regressions with binary variables, the natural log of per pupil expenditure was utilized. Summary statistics are noted in Table 1.5.

4. Methodology

This section develops the analytical procedures used and predictions based on the introduction and literature review. The question posited is whether high achieving school districts were negatively impacted by the implementation of the MCAS exam relative to low achieving school districts. First, I categorize each district as either high achieving, low achieving, or in the middle as described in the previous section. Then, three general types of regressions were used. Each regression was run twelve times: once for each performance statistic (dropout rate, senior plans, and SAT scores), once for each critical year (1996 or 2000), and once for each grouping (quartile or decile). In all cases, districts not in the high or low achieving groupings (the middle 50% or 80% of districts) were excluded from the regressions.

4.1 Long term model

I first look at a long term model and consider the performance trends between 1994 (or 1995) and 2008. Here, I estimate the difference in the average change in achievement between the low achieving and high achieving districts.

---

[13] Reporting procedures for district expenditure and per pupil expenditures changed in 2005. However, these effects should be minimal across all districts given the relative nature of the study.

Included in these models are district characteristics that might otherwise influence levels of achievement. These characteristics consist of the percentage of low-income students in each district, the percentage of special education students in each district, the overall secondary school enrollment in each district, and per pupil expenditure in each district.

$$A_{dt} = \beta_0 + \beta_1 Y_t + \beta_2 X_{dt} + \beta_3 \text{LOWER} + \beta_4 \text{LOWER} \bullet Y_t + \varepsilon_{dt} \qquad \textbf{(1)}$$

where

t = 1994-2008 for dropout rate, 1995-2008 for senior plans, and {1995, 1997, 2000, 2001, 2005-2007} for SAT scores

A_{dt} = achievement in district d in year t

Y_t = 1 if year = t and 0 otherwise

X_{dt} = additional covariates

LOWER = -1 if in lowest achievement grouping, 0 otherwise

These regressions are weighted by enrollment. Of greatest importance in the above model is the coefficient β_4 for the interaction term $\text{LOWER} \bullet Y_t$. If the effects of the MCAS have aided low achieving districts relative to high achieving ones, β_4 should be negative for dropout rates and positive for senior plans and SAT results.

Additionally, I use a second regression to measure the average impact after the implementation of the exam instead of using individual yearly interaction terms. The regression is as follows:

$$A_{dt} = \beta_0 + \beta_1 Y_t + \beta_2 X_{dt} + \beta_3 \text{LOWER} + \beta_4 \text{LOWER} \bullet \text{DUMPOST} + \varepsilon_{dt} \qquad \textbf{(2)}$$

where

DUMPOST = 1 if year equal to or after critical year, 0 otherwise

β_4 is the significant term in this equation and should be negative for dropout rates and positive for senior plans and SAT results if the gap is closing.

4.2 Fixed Effects

The fixed effects model is very similar to the long term model, but includes an additional term to account for any unaccounted time invariant district-level covariates that are correlated with LOWER in the previous model. This is represented by u_d in the model:

$$A_{dt} = \beta_0 + \beta_1 Y_t + \beta_2 X_{dt} + \beta_3 \text{LOWER} \bullet Y_t + u_d + \varepsilon_{dt} \qquad \textbf{(3)}$$

where

u_d = district fixed effects

Note that the dummy variable LOWER is omitted from the model, unlike in the long term model, because the fixed effects capture all time invariant factors that affect the dependent variable. As in the long term model, the important term is the coefficient β_3 for the interaction term $\text{LOWER} \bullet Y_t$. Once again, if the effects of the MCAS have helped the low achievers relative to the high achievers, β_3 should be should be negative for dropout rates and positive for senior plans and SATs.

Much like in the long term model, a second fixed effects model helps calculate the mean effect after the critical year. To do so, the regression uses an interaction term made up of two dummy variables:

$$A_{dt} = \beta_0 + \beta_1 Y_t + \beta_2 X_{dt} + \beta_3 LOWER \bullet DUMPOST + u_d + \varepsilon_{dt} \quad \textbf{(4)}$$

If the gap between high achieving and low achieving districts is closing, β_3 should be should be negative for dropout rates and positive for senior plans and SATs.

4.3 Individual District Linear Trends

This last model accounts for the possibility that the differences in performance between high and low performing districts would have changed in the absence of MCAS reform. Thus, the linear trends model accounts for this possibility by including individual district trend lines in the regression, notated here by $u_d \bullet t$.

$$A_{dt} = \beta_0 + \beta_1 Y_t + \beta_2 X_{dt} + \beta_3 LOWER \bullet Y_t + u_d + u_d \bullet t + \varepsilon_{dt} \quad \textbf{(5)}$$

where

$u_d \bullet t$ = individual district trends (unobserved characteristics for each district multiplied by the time vector t)

By controlling for the trend of each district without MCAS, this model measures the actual impact of the MCAS in each year. Once again, the β_3 coefficient is important as a negative value for dropout rate or positive value for senior plans and SAT results indicates that the gap between high and low achievers is closing.

As with fixed effects, this model is repeated with an averaged interaction term evaluating correlation between performance and achievement group *only* after the critical year. The model is:

$$A_{dt} = \beta_0 + \beta_1 Y_t + \beta_2 X_{dt} + \beta_3 LOWER \bullet DUMPOST + u_d + u_d \bullet t + \varepsilon_{dt} \quad \textbf{(6)}$$

β_3 should be negative for dropout rate and positive senior plans and SATs if the gap between high and low achievers is closing.

5. Results

Over the fifteen year period, I consider 226 districts in my analysis.[14] This section is divided into two sections: a more qualitative analysis of Massachusetts performance compared to other states, as well as relative analyses within the state, which are described using models in the previous section. Figures and tabular results can be found in Appendix 3.

5.1 State-by-state analysis

National-level data are as to whether Massachusetts has been improving over the last 15 years. On the national scale, Massachusetts surpasses most states in student academic performance. Massachusetts consistently scores among the best in the country on the NAEP exam (see Figures 3.1a and 3.1b for 8th grade exam results), with the state's relative performance remaining relatively constant

[14] A few districts are occasionally omitted each year because of reporting errors from the MADOE.

over the last 10-15 years—unchanged around the implementation of the MCAS. These results are consistent with international test findings that have shown that Massachusetts students outrank every nation except for Singapore (Meier 2002). When considering dropout rates, dropout rates have been steadily anchored around 3.5%, roughly in the middle of other states in New England. This may be partially due to the fact that other states in New England do not have cities as large as those in Massachusetts nor do they have comparable urban communities (see Figure 3.2). The national dropout rate, however, has recently been improving while the Massachusetts rate has remained stagnant (see Figure 3.3). In terms of the percentage of seniors going to college, the trajectories of both Massachusetts and the country are increasing at roughly the same rate, indicating little change around MCAS implementation (see Figure 3.4). Lastly, SAT scores in Massachusetts have been improving drastically compared to the rest of the nation since 1998 (see Figure 3.5). Given the variegated nature of these trends, there is no definitive answer that Massachusetts students have improved or worsened compared to the rest of the nation since the implementation of the MCAS.

5.2 Within-state analysis: 5.2a Long term model (1 & 2)

In examining the effect of the MCAS over time, the long term regressions consider the years prior to the MCAS (1994 or 1995) up until the most recent data (2007 or 2008). The results are given in Tables 3.1a-d. None of these regressions yield results that indicate significant changes in gaps. Only the senior plans regression with year 2000 groupings at the 25th and 75th percentile have significant interaction terms in 1995, 1996, 1999, 2003, and 2004—but jointly they are insignificant (F=1.29, p=.2122). The changes noticed within these models are a result almost entirely of the covariates. All covariates are significant at the = .01 level, except for percent special education in all models and per pupil expenditure in dropout regressions using decile groupings (also insignificant). The mean model long-term regression results are found in the first three columns of Table 3.4. Much like the previous models, the results indicate no significant effects. Again, most of the changes in the model are attributable to the covariates, which are significant in almost every case.

5.2b Fixed Effects (3 & 4)

I re-run the two long term regressions from the previous section but now account for potentially unobserved effects; the results are in Table 3.2, and are drastically different. No longer do the covariates explain dropout rates and senior plans, but all of the senior plans models have significant yearly interaction terms. In the 1996 25th/75th model, only 1998, 2002, 2004, 2005, and 2007 are insignificant, and jointly they are significant as well (p=.0033). The 2000 25th/75th model have similar results (2002, 2004, 2005, and 2007 are insignificant, jointly p=.0019), as with the 2000 10th/90th model (2001-2007 are insignificant, p=.0083). Only in the 1996 10th/90th model are the terms not jointly significant (1997, 1998, 2002, and 2005-7 are insignificant, jointly p=.1219), All of these significant coefficients are negative, indicating that the gap between low and high achieving districts has increased. In other words, high achieving districts are more likely to have students planning to go to a four year college in 2008 than earlier—roughly 4%, depending on the year and model. However, when taking an average of the interaction terms of the years after the MCAS, only the models with groupings in 2000 are significant (results in Table 3.4).

The fixed effects models using dropout rate as the dependent variable demonstrate similar results. The interaction terms in 1996 and 1997 are significant in the decile models, but were not enough to make a significant impact in the yearly interaction regressions. In the mean dropout rate models, however, the mean term coefficient is always significant (p=.0402; p=.0012; p=.0076; and p=.0279 respectively), also indicating an increasing gap. This indicates, once again, that the gap between high and low achieving districts is growing, implied by the fact that dropout rates have been decreasing at a faster rate for high achieving districts than low achieving districts since 1998. Though these results are not reflected in every yearly interaction term model, the mean model findings provide some support for the possibility that the gap is widening.

In contrast with senior plans and dropout rate, the SAT model has few significant results. Though there were significant results in 2001 for the 1996 $10^{th}/90^{th}$ groupings and 2000 $25^{th}/75^{th}$ groupings, none are significant in mean models.

Thus, these results seem to indicate that there is some evidence for an effect of the MCAS and, if anything, imply high achieving districts have done relatively better than low achieving districts.

5.2c Individual District Trends (5 & 6)

This final section runs the same regressions that were run in the previous fixed effects section, but also includes the trend lines for every district in the model, thus accounting for any potential trend prior to the MCAS in 1998. Results can be found in Table 3.3. The variable for 1994 is omitted in the dropout rate models and 1995 is omitted in the senior plans models and SAT models due to collinearity. Additionally, a number of individual district trends are omitted due to collinearity. However, no observations are lost in any regression.

No district trends regressions support the findings noted in the previous fixed effects models. Only the SAT models grouped in 2001 are jointly significant (p=.0007 and p=.0152) and those regressions only have a significant yearly interaction term in 2001. When running mean model regressions that average the overall effect after the MCAS, none are found to be significant (see Table 3.4). This implies that, though there seems to be some evidence indicating an increasing gap in the fixed effects models, that change in performance should be attributed pre-existing trends of each district.

5.3 Additional Tests

To test all possible explanations, these regressions were re-run with the following variations: critical years other than 1998 and 2001, excluding years after 2002 in mean models, and senior plans including seniors intending to attend either a 2- or 4-year college (rather than only a 4-year college).[15] As before, the results are mostly inconclusive. If anything, the results indicate a widening of the gap between high and low achieving districts in the fixed effects and district trends models, though these findings are sporadic and inconsistent.

6. Discussion

None of the long term models indicate any statistically significant effects of high-stakes testing on gaps. When fixed effects are included, some significant results become apparent: all of the mean models for dropout rate indicate a significant increase in the gap between high and low achieving districts, and a number of senior plans models find the gap increasing in terms of seniors going onto a four-year college. However, these effects disappear when prior trends are included, perhaps due to the large number of parameters included in the models. There are never any apparent effects on district SAT scores either. But when considering additional variations of these regressions, a few dropout and senior plans regressions show evidence of a widening gap.

Based on evidence from the national scale, it seems unlikely that the MCAS has had an absolute impact on either group. On NAEP exams, Massachusetts remains high above the national average. The percentage of seniors going to college remains constant. And though dropout rates have worsened relative to the national average, SAT scores are improving relative to the average—indicating an ambiguous effect overall (note that these last three metrics do not account for high and low achieving groups, a factor necessary to determine the actual gap). Additionally, around the years when the MCAS was implemented, there was no evidence of any break in trends. All in all, there is no clear evidence that the MCAS has either hurt or helped Massachusetts students in terms of dropout rate, percent of high school seniors attending a four year college, or SAT scores.

Thus, at the very best, the MCAS has had a negligible impact on both high and low achieving school districts in Massachusetts; at the worst, the MCAS has actually widened this gap. Admittedly, there is no counterfactual to know what the state would have looked like had there been no MCAS. Yet this study's results bring into question the true utility of the high stakes testing. Conventional wisdom has it that standardized testing would aid low achieving schools at the cost of high achieving ones, but these results indicate otherwise. In fact, they indicate a marginal effect at best. Though the MCAS's initial purpose under the *Massachusetts Education Reform Act* was to evaluate progress towards meeting

[15] Other critical years are considered because class graduation years do not necessarily line up with implementation dates (i.e. a student who takes the exam in 2001 is not scheduled to graduate until 2003). I exclude years after 2002 to account for potential effects of NCLB. Lastly, I include 2-year colleges in senior plans to offset confounds from 4-year college costs. None of these additional test results are available in the appendix.

certain curricula goals, *No Child Left Behind* fully expected these high stakes exams to close the this gap. In the latter sense, it appears that the MCAS has failed.

It is important to note that these results refer only to three performance variables, yet combined they provide a fairly holistic view of school populations. Dropout rates are frequently used to gauge the performance of low achieving students in each district. Senior plans and SAT results, on the other hand, evaluate the performance of high achievers much better than dropout rate. However, these metrics do have their limitations: senior plans are just that—plans—and may not necessarily translate into actual action. Furthermore, SAT exams are designed to evaluate an individual's basic knowledge of subjects, which may not be variable and thus not accurately reflect district performance. The same applies to the other metrics; it is possible that by 10^{th} grade, student performance is already set and unrelated to the quality of education provided by high school districts. Even so, similar concerns arise with most any such performance statistic.

The findings of this study fall into line with a growing body of research (such as Jacob (2001)) that suggests accountability has little impact on student performance. Further research in other geographic areas is necessary to determine whether the case in Massachusetts is an isolated one. Additionally, future studies should test students who have experienced the total, overall impact of NCLB throughout their educational experience, since the act was intended to affect all grades rather than just high school exit exams. Still, it seems as though the benefits of high stakes accountability may be fewer than previously considered. Though the exams do provide a tool to adjust state finances as appropriate, this study provides some evidence that they may be hurting low achieving districts. At the very least, based on these results, education reformers may want to give pause before continuing with plans to further implement high stakes testing.

Appendices

Appendix 1: Data Description Figures and Tables

Table 1.1: Variable Names and Definitions

Variable Name	Definition
dropout_rate	Four year district dropout rate
four	Senior plans
sat	Mean district SAT score (combined math and verbal)
critical year	1998 or 2001, depending on the regression
dumpost	= 1 if the year is after the critical year, 0 otherwise
lower	= 1 if district is defined as "low achieving", 0 otherwise
percent_low_income	Percentage of students considered "low income," as defined by the MADOE
percent_special_ed	Percentage of students in district considered as special

	education
enrollment	district student population, considering students in grades 9-12
ppexpend	Natural log of dollars spent per student within each district

Table 1.2: Mean performance variables by 1996 groupings (weighted by enrollment)

Dropout Rate					
Year	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
1994	7.609421	5.926844	3.864032	1.208415	.8468851
1995	7.429772	5.6422	3.624792	1.316532	.8753113
1996	6.247061	5.153229	3.425415	1.092398	.7758167
1997	6.539105	5.266373	3.537774	1.231741	.7566619
1998	7.02547	5.517542	3.552372	1.256589	.715624
1999	7.440297	5.8679	3.652625	1.111761	.6935771
2000	7.109153	5.648982	3.4924	1.106651	.6238931
2001	7.037298	5.6417	3.529432	1.035199	.6248069
2002	6.069376	4.927438	3.139517	0.9337584	.5389764
2003	6.668582	5.423087	3.431702	1.010492	.615037
2004	7.302666	6.03224	3.77084	1.058741	.6993818
2005	7.677438	6.175788	3.767443	0.9820527	.6783238
2006	7.410342	5.818034	3.462651	0.8578059	.5469732
2007	7.707436	6.166741	3.774623	1.1066	.6534988
2008	7.083289	5.611858	3.380825	0.8610518	.6029412
Senior Plans					
Year	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
1995	39.79091	44.44349	55.75297	75.62802	81.98032
1996	40.58042	45.75355	57.00801	76.22893	82.911
1997	39.318	44.53621	56.45624	76.76264	83.22346
1998	37.31496	42.84183	56.04421	76.96205	83.64376
1999	38.14162	43.41039	55.58307	76.33807	84.09256
2000	38.13686	43.57513	56.15636	76.00074	83.02211
2001	36.52235	42.60561	56.42805	77.15539	83.12391
2002	35.85154	42.07804	56.21693	77.24147	83.50712
2003	37.89574	44.44009	57.9318	77.96178	83.88916
2004	41.47276	45.86176	58.43731	78.63304	84.26327
2005	39.26914	44.92835	59.37812	79.73645	86.06256
2006	39.76345	45.23758	59.14301	79.91664	86.49944

2007	40.48618	45.64841	59.59997	80.88523	87.12196
2008	40.22649	44.77421	59.33587	81.78014	88.02857
SAT					
Year	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile
1995	874.518	915.8651	982.2497	1082.841	1131.968
1997	886.3007	923.5853	989.1631	1090.378	1140.634
2000	892.0295	930.7892	997.6647	1103.783	1156.943
2001	892.8858	931.334	999.2171	1101.79	1154.009
2005	916.8051	950.6796	1022.066	1129.913	1186.604
2006	904.4927	938.8459	1009.961	1120.543	1181.004
2007	902.4286	936.7442	1008.17	1116.539	1177.31

Figure 1.1: Dropout rate over time, weighted by enrollment (using 1996 25th/75th percentile groupings)

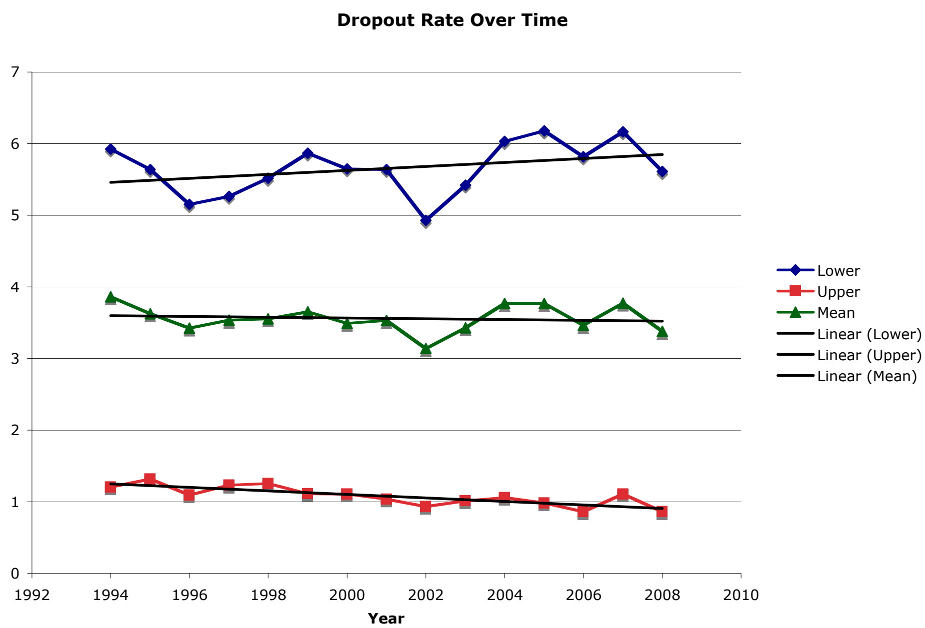


Figure 1.2: Percentage of high school seniors attending college in MA, weighted by enrollment

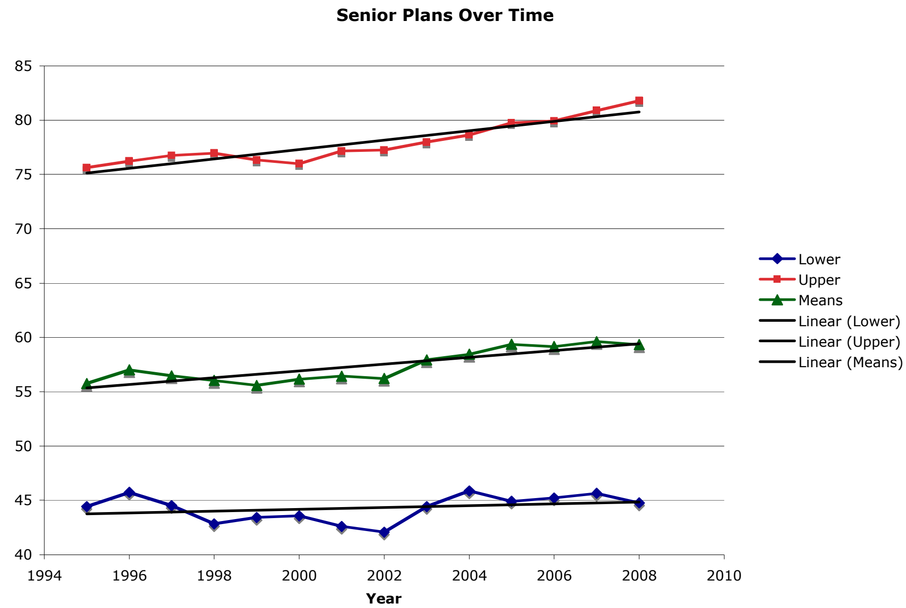
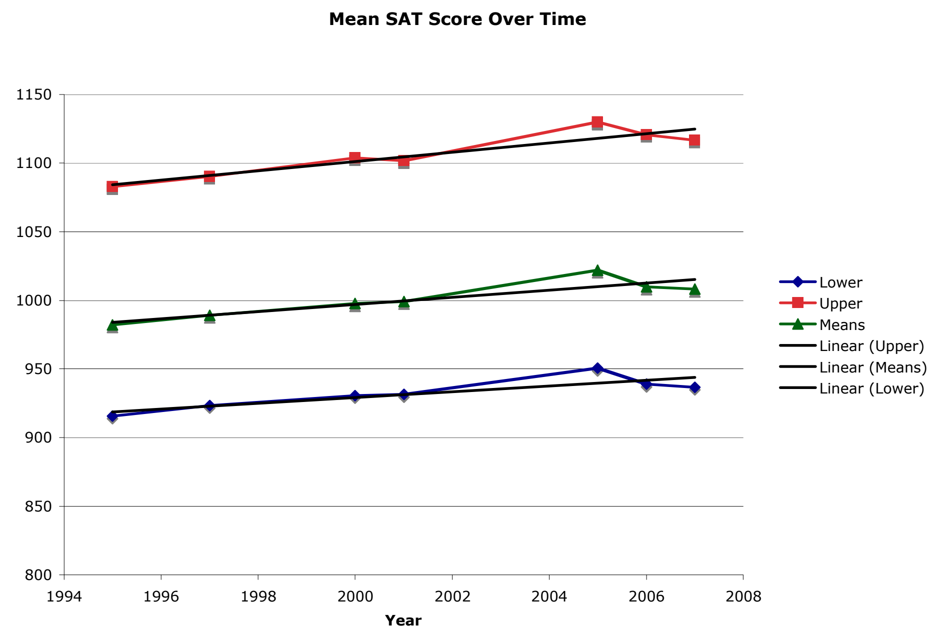


Figure 1.3: Massachusetts SAT Results (weighted by enrollment)



*Note: the above graphs use national data from the NCES, CollegeBoard and the BLS rather than the MADOE, and thus includes vocational and charter schools. Additionally, NAEP data from MA in 2002 was omitted in the national report.

Table 1.3: Time trend regression results (1996 25/75 groupings)

	(1)	(2)	(3)
VARIABLES	dropout_rate	sr_plans	sat
year	-0.018	0.420**	2.638**
	(0.012)	(0.080)	(0.532)
lower	-92.193*	627.000*	1,032.476
	(37.672)	(243.647)	(1,610.896)
yearlower	0.048*	-0.325**	-0.577
	(0.019)	(0.122)	(0.805)
Observations	3353	3133	1570
Adj R-squared	0.391	0.427	0.461
SER	2.321	13.52	66.62
Standard errors in parentheses ** p<0.01, * p<0.05			

Table 1.4: Regression with binary variable indicating post-critical year (1996 25/75 groupings)

	(1)	(2)	(3)
VARIABLES	dropout_rate	sr_plans	sat
year	-0.004	0.490**	2.606
	(0.028)	(0.167)	(1.463)
dumpost	0.040	-2.644	4.310
	(0.278)	(1.706)	(14.181)
Observations	1838	1716	859
Adj R-squared	-0.001	0.004	0.012
SER	3.305	19.75	104.6
Standard errors in parentheses ** p<0.01, * p<0.05			

Table 1.5: Means and standard deviations for covariates (weighted by enrollment)

Variable	Mean	Standard Deviation
Percent Low Income	26.03209	24.16897
Percent Special Education	16.39057	2.961837
Enrollment	2.886367	4.506382
Per Pupil Expenditures	2.045017	.3133035

Appendix 2: Changes in Performance Statistic Calculations

Since 1993, high school dropout rate has been calculated by the MADOE in the following manner:

$$\text{Annual Dropout Rate in year t} = \frac{\text{(Dropouts - Returned Dropouts) = Final Dropout Count}}{\text{Enrollment Grades 9 - 12 on October 1 in year t - 1}} \text{ x 100}$$

Note that the dropout rate adjusts for those who return to school.

There have been minor adjustments to the dropout rate calculation that have biased dropout rates in 2002, 2006, 2007, and 2008.

- In 2002, summer dropouts were omitted from the model, lowering the overall dropout rate (this was fixed in 2003).
- In 2005-6, the MADOE started working with the GED testing service, an exam which provides an equivalency for a high school diploma. Now, individuals who receive GED's are no longer considered dropouts (lowering dropout rates).
- In 2006-7, the MADOE began to consider students who claim to be transferring districts but never reenroll as dropouts (raising rates).

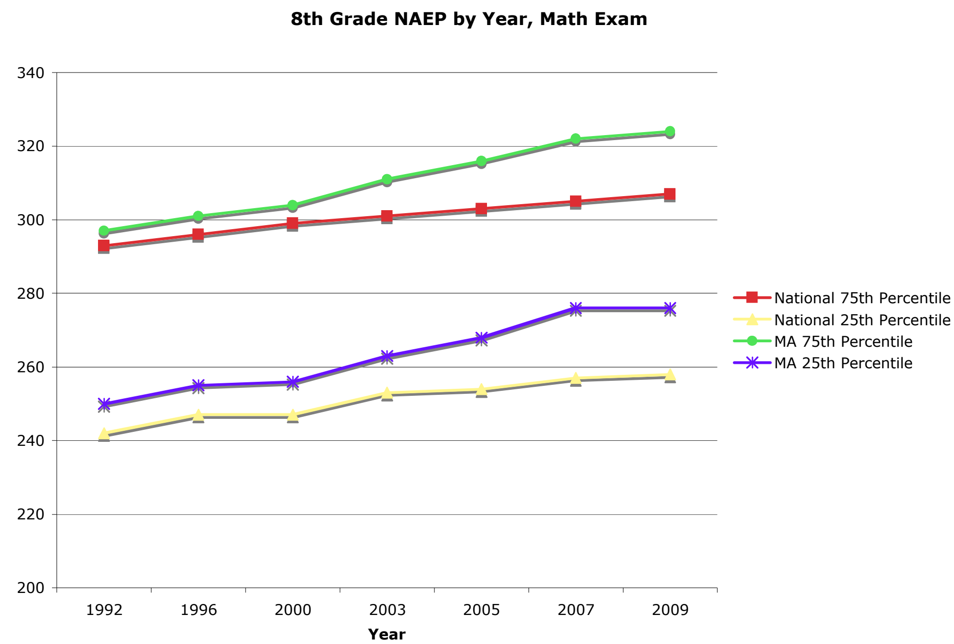
Though these changes do bias results, all regressions were relative and thus affected both high and low achieving districts equally. This is one reason why year dummies were included in the model to pick up year-to-year variations in the dropout rates.

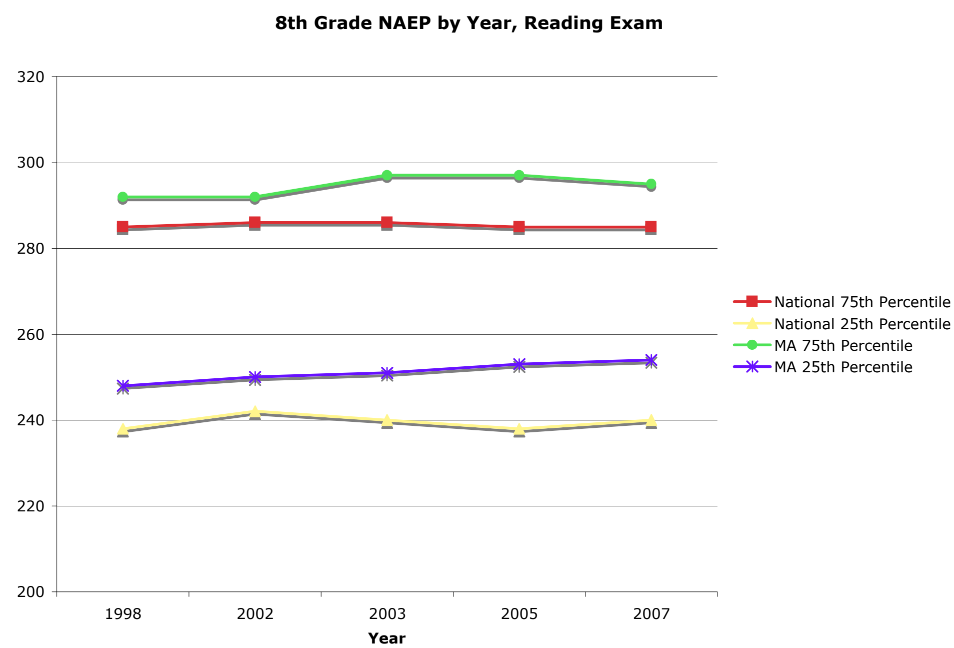
Furthermore, dropout rate regressions were run omitting these years. Results were almost identical to those cited in the report. The only notable differences were in the fixed effects models: both 1996 models had significant yearly interaction terms, and the 2000 10/90 mean model had a significant positive coefficient. Again, however, these effects disappeared when districts trends were included, and thus the conclusions in the paper remain the same.

Additionally, the SAT changed format, which reflected in the results reported for 2006. Specifically, the verbal section no longer contained analogies and antonyms but more sentence completions and reading comprehension questions. However, as scores are still scaled according to a normal distribution, there should have been no confounding effect of the new exam.

Appendix 3: Absolute Analysis and Regression Results

Figures 3.1a & 3.1b: 8th Grade Math and Reading NAEP Exam over time





*Note that the X axis is not to scale in the above two figures

Figure 3.2: Dropout rates around New England

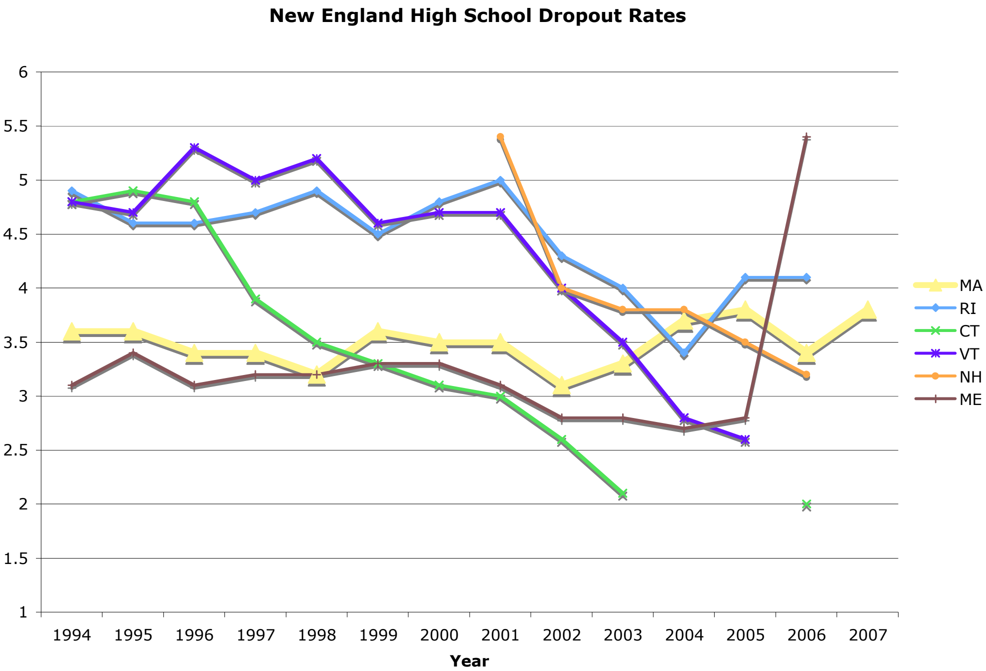


Figure 3.3: National dropout rate comparison

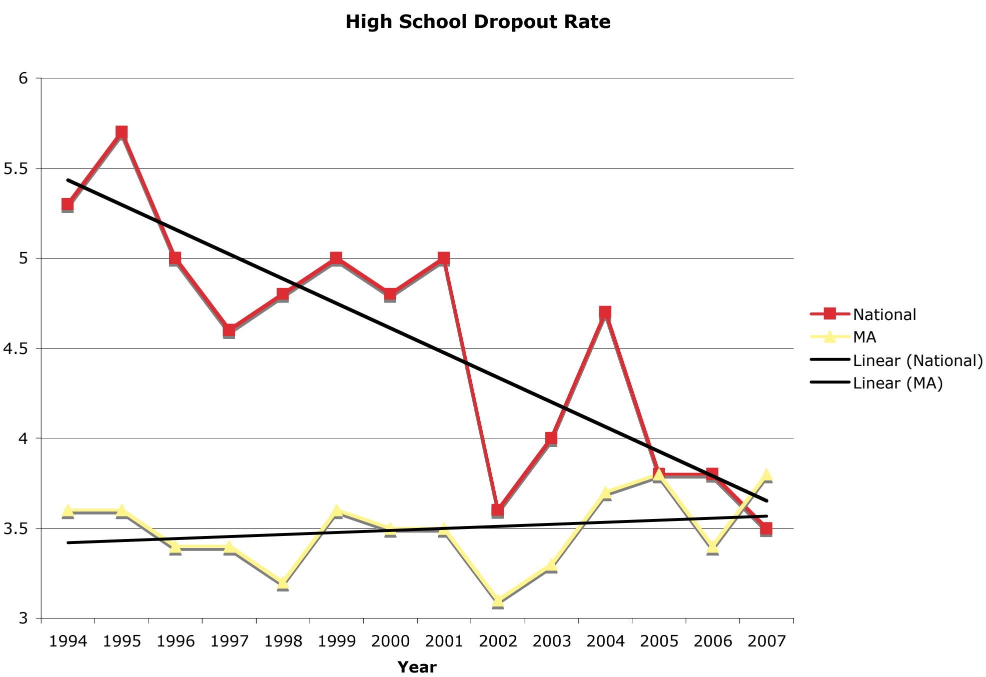


Figure 3.4: Percentage of high school seniors attending college over time

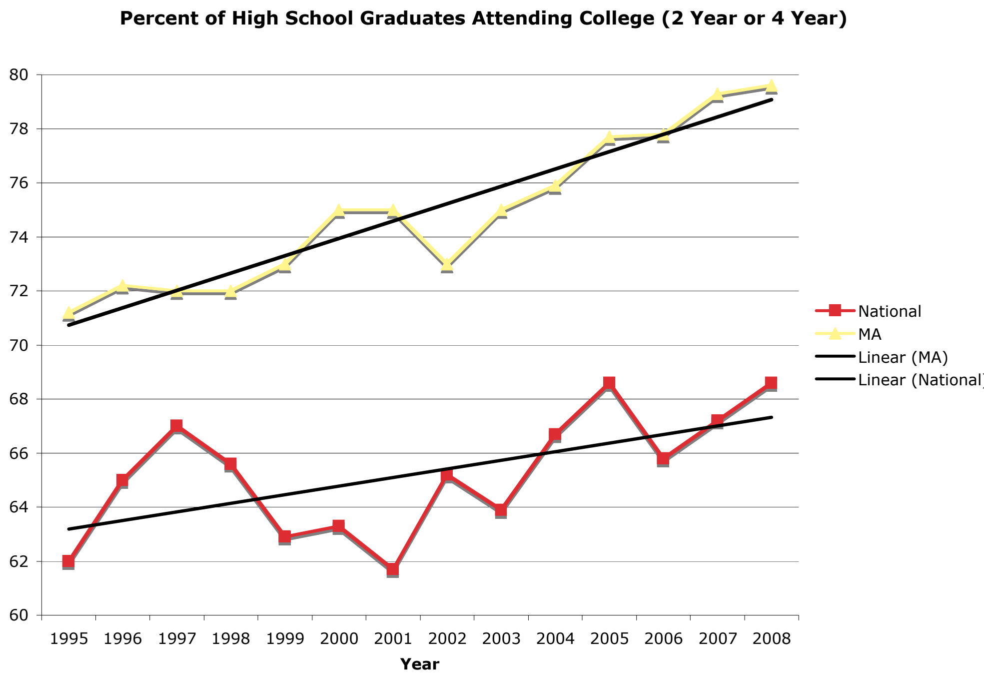
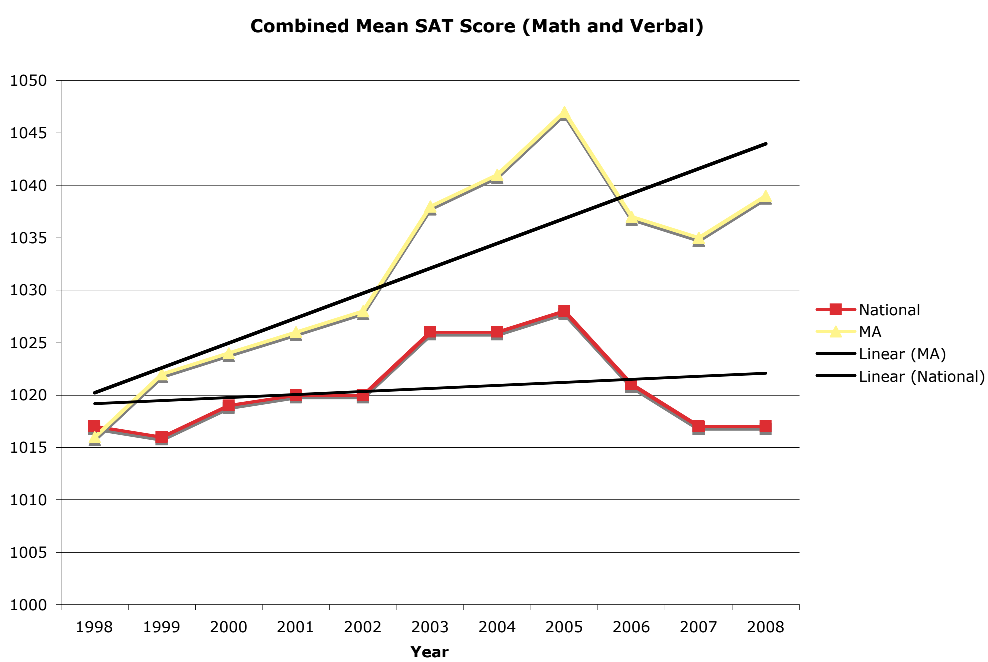


Figure 3.5: Mean SAT Scores Over Time



Note: Tables 3.1-3.4 refer to groupings from the 1996 MEAP at the 25th and 75th percentiles.

Table 3.1: Long term model (1)

	(1)	(2)	(3)
VARIABLES	dropout_rate	sr_plans	sat
y1994lower	0.798		
	(0.452)		
y1995lower	0.319	3.215	1.362
	(0.449)	(2.204)	(11.274)
y1996lower	0.068	3.191	
	(0.445)	(2.185)	
y1997lower	-0.030	1.823	1.580
	(0.443)	(2.175)	(11.104)
y1998lower	0.218	0.057	
	(0.440)	(2.155)	
y1999lower	0.581	1.978	
	(0.436)	(2.140)	
y2000lower	0.373	2.219	-3.866
	(0.432)	(2.120)	(10.813)
y2001lower	0.395	0.176	0.067
	(0.428)	(2.102)	(10.717)
y2002lower	-0.302	-0.297	
	(0.428)	(2.101)	
y2003lower	-0.099	2.984	
	(0.425)	(2.088)	
y2004lower	0.413	3.528	
	(0.419)	(2.058)	
y2005lower	0.568	1.619	1.772
	(0.417)	(2.048)	(10.436)
y2006lower	0.359	1.288	-3.035
	(0.416)	(2.041)	(10.429)
y2007lower	0.396	1.166	
	(0.415)	(2.039)	
percent_low_income	0.094**	-0.488**	-2.535**
	(0.003)	(0.013)	(0.094)
percent_special_ed	-0.005	-0.534**	0.699
	(0.015)	(0.076)	(0.538)
enrollment	0.069**	0.274**	-1.891**
	(0.011)	(0.055)	(0.390)

ppexpend	-2.497**	23.451**	112.466**
	(0.270)	(1.379)	(9.791)
Observations	1837	1715	859
Adj R-squared	0.755	0.837	0.850
SER	1.636	8.000	40.75
Yearly variable term F test p-value	.0106*	<.0001**	.0001**
Interaction term F test p-value	.5990	.7408	.9977
Covariates F test p-value	<.0001**	<.0001**	<.0001**
Standard errors in parentheses ** p<0.01, * p<0.05			

Table 3.2: Fixed Effects (3)

	(1)	(2)	(3)
VARIABLES	dropout_rate	sr_plans	sat
y1994lower_m	0.001		
	(0.336)		
y1995lower_m	0.290	-5.016**	-8.267
	(0.331)	(1.525)	(6.044)
y1996lower_m	0.392	-5.111**	
	(0.329)	(1.512)	
y1997lower_m	0.532	-3.132*	-3.837
	(0.328)	(1.510)	(5.920)
y1998lower_m	0.255	-2.914	
	(0.328)	(1.505)	
y1999lower_m	0.125	-4.822**	
	(0.326)	(1.496)	
y2000lower_m	0.237	-6.261**	-2.137
	(0.325)	(1.493)	(5.772)
y2001lower_m	0.134	-2.974*	-5.277
	(0.324)	(1.487)	(5.738)
y2002lower_m	0.344	-2.169	
	(0.325)	(1.490)	
y2003lower_m	-0.151	-4.735**	
	(0.324)	(1.486)	
y2004lower_m	-0.533	-2.442	

	(0.319)	(1.462)	
y2005lower_m	-0.506	-2.255	2.800
	(0.318)	(1.455)	(5.545)
y2006lower_m	0.107	-3.128*	1.097
	(0.317)	(1.453)	(5.530)
y2007lower_m	-0.008	-1.145	
	(0.317)	(1.449)	
percent_low_income	0.032**	-0.162**	-0.788**
	(0.011)	(0.052)	(0.280)
percent_special_ed	-0.026	-0.150*	0.089
	(0.016)	(0.076)	(0.397)
enrollment	-0.330	-0.466	7.372
	(0.189)	(0.941)	(4.742)
ppexpend	-0.176		
	(0.491)		
Observations	1837	1715	859
Adj R-squared	-0.039	-0.009	0.178
SER	1.241	5.677	21.58
Yearly variable term F test p-value	.9792	.0215	<.0001**
Interaction term F test p-value	.0512	.0027**	.5818
Covariates F test p-value	.0152*	.0033**	.0649
Standard errors in parentheses ** p<0.01, * p<0.05			

Table 3.3: District Trends (5)

	(1)	(2)	(3)
VARIABLES	dropout_rate	sr_plans	sat
percent_low_income	0.040**	-0.139	-0.093
	(0.014)	(0.074)	(0.398)
percent_special_ed	-0.042*	0.048	0.653
	(0.018)	(0.091)	(0.461)
enrollment	0.310	1.232	5.644
	(0.385)	(2.046)	(12.243)
ppexpend	-0.152	5.257	7.645
	(0.641)	(3.282)	(18.184)
y1995lower_m	0.283		

	(0.281)		
y1996lower_m	0.392	-0.368	
	(0.273)	(1.341)	
y1997lower_m	0.553*	1.233	2.228
	(0.267)	(1.303)	(4.626)
y1998lower_m	0.260	1.244	
	(0.263)	(1.272)	
y1999lower_m	0.143	-1.006	
	(0.258)	(1.248)	
y2000lower_m	0.252	-2.673*	1.957
	(0.258)	(1.244)	(4.468)
y2001lower_m	0.150	0.083	-2.321
	(0.256)	(1.229)	(4.409)
y2002lower_m	0.362	0.332	
	(0.259)	(1.236)	
y2003lower_m	-0.117	-2.822*	
	(0.259)	(1.239)	
y2004lower_m	-0.499	-0.783	
	(0.261)	(1.242)	
y2005lower_m	-0.489	-0.914	3.831
	(0.266)	(1.265)	(4.591)
y2006lower_m	0.119	-2.137	1.014
	(0.274)	(1.302)	(4.757)
y2007lower_m	0.002	-0.547	
	(0.282)	(1.342)	
Observations	1837	1715	859
Adj R-squared	0.123	0.073	0.343
SER	1.140	5.443	19.28
Yearly variable term F test p-value	.9061	.2745	<.0001**
Interaction term F test p-value	.0858	.0948	.8615
Covariates F test p-value	.0062**	.1320	.6225
Standard errors in parentheses ** p<0.01, * p<0.05			

Table 3.4: Mean Model (2, 4, 6)

NOTE: The first three regressions are long term models; the fourth through sixth regressions are fixed effects models; and the final three are district trends models. Additionally, post_p75m *represents* LOWER • DUMPOST*, noted in the methodology.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	dropout_rate	sr_plans	sat	dropout_rate	sr_plans	sat	dropout_rate	sr_plans	sat
percent_low_income	0.094**	-0.487**	-2.534**	0.038**	-0.231**	-0.896**	0.041**	-0.146*	-0.091
	(0.003)	(0.013)	(0.093)	(0.010)	(0.049)	(0.262)	(0.014)	(0.074)	(0.394)
percent_special_ed	-0.004	-0.532**	0.694	-0.027	-0.131	0.124	-0.043*	0.052	0.654
	(0.015)	(0.076)	(0.536)	(0.016)	(0.076)	(0.396)	(0.018)	(0.091)	(0.459)
enrollment	0.070**	0.276**	-1.893**	-0.355	-0.054	7.833	0.315	1.510	4.942
	(0.011)	(0.054)	(0.388)	(0.188)	(0.939)	(4.698)	(0.383)	(2.033)	(12.093)
ppexpend	-2.517**	23.336**	112.483**	-0.260	-1.072	-1.089	-0.467	5.079	7.230
	(0.269)	(1.371)	(9.732)	(0.483)	(2.417)	(13.012)	(0.632)	(3.241)	(18.007)
post_p75m	-0.016	-1.380	-2.440	0.290*	-1.133	-4.725	0.076	0.812	1.838
	(0.192)	(1.050)	(6.838)	(0.141)	(0.710)	(3.593)	(0.193)	(0.938)	(5.339)
Observations	1837	1715	859	1837	1715	859	1837	1715	859
Adj R-squared	0.755	0.837	0.851	-0.043	-0.020	0.180	0.118	0.068	0.346
SER	1.636	7.990	40.64	1.243	5.708	21.55	1.143	5.456	19.25
Yearly variable term F test p-value	<.0001**	<.0001**	<.0001**	.0004**	.0027**	<.0001**	.0008**	.4439	<.0001**
Covariates F test p-value	<.0001**	<.0001**	<.0001**	.0011**	<.0001**	.0107*	.0032**	.1084	.6293
Standard errors in parentheses ** p<0.01, * p<0.05									

References

Ballou, D. & Springer, M.G. (2008). Achievement Trade-Offs and No Child Left Behind.

Bolon, C. (2001). Significance of Test-Based Ratings for Metropolitan Boston Schools. *Education Policy Analysis Archives, 2001*(9).

Carnoy, M. & Loeb, S. (2002). Does External Accountability Affect Student Outcomes? A Cross-State Analysis. *Educational Evaluation and Policy Analysis, 2002*(24).

Clarke, M., Haney, W. & Madaus, G. (2000). High Stakes Testing and High School Completion. *The National Board on Educational Testing and Public Policy, 2000* (1).

Cataldi, E. F., Laird, J. & KewalRamani, A. (2009). High School Dropout and Completion Rates in the United States: 2007. *National Center for Education Statistics, 2009*.

College-Bound Seniors. *CollegeBoard*, *1998-2008*.

College Enrollment and Work Activity of High School Graduates. *Bureau of Labor Statistics*, *1994-2008*.

Cullen, J. B. & Reback, R. (2006). Tinkering Toward Accolades: School Gaming Under a Performance Accountability System. *2006*.

Downes, T., Zabel, J. & Ansel, D. (2009). Incomplete Grade: Massachusetts Education Reform at 15. *MassINC*, *2009*.

Dropout Rates. *Massachusetts Department of Elementary and Secondary Education. 1995-2008*.

Haney, W. Lake Woebeguaranteed: Misuse of Test Scores in Massachusetts, Part I. *Education Policy Analysis Archives, 2002*(10).

Harris, D. N. & Taylor, L. L. (2008). The Resource Costs of Standards, Assessments, and Accountability. *National Research Council*, *2008*.

Jacob, B. A. (2001). Getting Tough? The Impact of High School Graduation Exams. *Educational Evaluation and Policy Analysis, 2001*(23).

Lee, B., Shaefer, S., & Messner-Zidell, S. (2007). School or the Streets: Crime and California's Dropout Crisis. *Fight Crime: Invest in Kids, 2007*.

Loveless, T., Farkas, S. & Duffett, A. (2008). High-Achieving Students in the Era of NCLB. *Thomas B. Fordham Institute*, *2008*.

Massachusetts School and District Profiles. *Department of Elementary and Secondary Education*, *1994-2008*.

Meier, D. (2002). *In Schools We Trust* (Boston: Beacon Press).

NAEP State Comparisons. (2010). *National Center for Education Statistics*, *2010*.

Neal, D. & Schanzenbach, D.W. (2008). Left Behind by Design: Proficiency Counts and Test-Based Accountability.

Papay, J.P., Murnane, R.J. & Willet, J.B. (2008). The Consequences of High School Exit Examinations for Low-Performing Urban Students: Evidence from Massachusetts. *National Bureau of Economic Research*.

Plans of High School Graduates. *Massachusetts Department of Elementary and Secondary Education, 1995-2008*.

Reback, R. (2008). Teaching to the Rating: School Accountability and the Distribution of Student Achievement. *Journal of Public Economics, 2008*(92).

SAT Report. *Massachusetts Department of Elementary and Secondary Education*. *1995, 1997, 2000, 2001, 2005, 2006, 2007.*

Springer, M. G. (2008). Accountability Incentives: Do Schools Practice Educational Triage? *Education Next*.

Warren, J. R., Jenkins, K. N. & Kulick, R.B. (2006). High School Exit Examinations and State-Level Completion and GED Rates, 1975 through 2002. *Educational Evaluation and Policy Analysis, 2006*(28).