

An Evaluation of Student Achievement in Edison Schools Opened in 1995 and 1996

Gary Miron and Brooks Applegate

The Evaluation Center
Western Michigan University
Kalamazoo, MI 49008
www.wmich.edu/evalctr

December 2000

This evaluation was funded by the National Education Association (NEA), which sought an independent evaluation of schools operated by Edison Schools, Inc. In order to guarantee the evaluators' autonomy, the NEA granted the authors the right to publish and disseminate the report's findings, following delivery of the report to NEA and Edison Schools, Inc. and after allowing "a reasonable time to reply." Both the NEA and Edison Schools, Inc. were invited to comment and provide feedback on drafts of the report throughout the course of the study, although the authors retained responsibility for its content. These groups were also provided copies of this final report at least 10 working days prior to our release of the report. The findings, as well as the views and opinions expressed in this report, are those of the authors and do not necessarily represent the views of the NEA or Edison Schools, Inc.

The Evaluation Center
Western Michigan University
Kalamazoo, MI 49008

2000

Executive Summary

This evaluation’s principal aim was to determine whether or not the Edison model worked—over approximately a 3-year period—to improve student performance on norm- or criterion-referenced standardized tests in the 10 Edison schools that were opened in 1995 and 1996. This report is intended for a wide audience including policymakers, researchers, district personnel, and parents interested in judging the relative merits—and student achievement outcomes—of the first 10 schools opened/operated by Edison Schools Inc.

The intent of our analyses was to incorporate data from a variety of sources and apply a variety of analytical methods to determine the extent of achievement gain made by students enrolled in the schools involved. As an evaluation, we include our own summative judgments by case and across cases. We additionally provide readers with sufficient detail regarding our data types and analytical methods employed so they can make their own judgments regarding the extent of the achievement gains of students in the examined Edison schools. Providing this detail, however, resulted in a rather lengthy and technical report. The Executive Summary, therefore, serves to provide a concise overview of the methods we employed and the results organized by case and across all the cases.

It would seem that addressing student achievement is a rather tangible and noncontestable area to study. This is not the case, however. This study addresses issues that are much on the minds of the public, professional educators, and policymakers. We expect the results to be scrutinized closely. We have stated clearly in the report that we examined student achievement in terms of gains made relative to comparison groups, as opposed to Edison’s preference to evaluate gains made by schools relative to themselves. We don’t contest that the Edison students in this study made gains. The students obviously made gains from year to year, in part due to formal schooling and in part due to nonformal and informal learning activities. On criterion-referenced tests nearly all schools are making gains each year as they adjust their curriculum to state standards. Important questions are whether students in the Edison schools are making gains larger than expected or larger than the district or other relative comparison groups.

Another often heard question specific to Edison—and other for-profit EMOs—is whether the company can “do more for less.” There are two parts to this question: Does Edison do more? And do they do it for less? By focusing on student achievement in schools operated by Edison, this report addresses only the first question. Economists, districts that negotiate contracts with Edison, and even stockholders in the company are in a better position to determine whether Edison operates schools more efficiently than traditional schools. We do not presume that we can answer the financial question. In each of the 10 cases, however, we have included some financial figures based on state and district data. We are aware, however, that these data are limited and that many of the figures are contested by both districts and by Edison.

In terms of inputs, Edison (1999) reports that it is doing more in terms of a longer school day, a longer school year, implementing a rigorous and research-based curriculum, extensive use of technology, etc. Edison also reports that it is doing more in terms of gains made by its students on standardized achievement tests. While this study does not examine what Edison is doing in its schools, and the extent to which its model is implemented, it does examine how students in its first 10 schools did in terms of student achievement.

Recent years have seen a rapid growth in the number of for-profit educational management organizations (EMOs) starting and operating charter schools and competing for contracts to take over operation of poor performing district schools (Miron, 2000). The results from this study contribute to the still limited body of research on the impact and performance of for-profit EMOs.

Description of Edison Schools Inc.¹

Edison identifies three major goals for its schools: (1) move students forward, (2) operate schools for less money, and (3) provide more services. Edison's school model reportedly includes the following: (i) a longer school day and longer school year; (ii) organization of schools based on academies, houses, and teams; (iii) a rich liberal arts curriculum with research-based instruction closely aligned with assessment; (iv) four to six weeks of training for staff prior to the opening of a new school; (v) daily professional development and planning time for teachers; (vi) career ladders for teachers; (vii) a high level of parent and community involvement; (viii) computers and training provided to teachers, students, and families; and (ix) a national system of educational, operational, and financial supports.

Established in 1992 as the Edison Project, the company did not open its first schools until the 1995-96 school year. Edison soon became the largest private educational management organization operating public schools in the U.S. Approximately half of the schools Edison operates are charter schools, while the remaining are schools run under direct contract, or partnership, with local school districts. As of the 2000-01 school year, Edison reports that it is operating 113 schools in 21 states and the District of Columbia, with a combined enrollment of approximately 57,000 students. Edison considers itself to be the first "national system of public schools." Plans for the future assume more expansive growth, including the announcement of new contracts in 2000-01 that include a large commitment in Dallas, Texas, and Inkster, Michigan. All of Edison's initial contracts from 1995 have been renewed except for the one with Sherman Independent School District in Texas where Edison had operated two schools.

Organization of Edison schools. Edison schools are organized by grouping 2 or 3 grade levels into academies. Within the academies, the students are organized into multigrade houses of 100-180 students. The students in each house are largely taught by the same team of teachers throughout the time they are in that academy.

¹ Documentation for this section is derived from Edison's own Web site and from their annual reports (1999 and 2000).

Curriculum design and instructional methods. Edison Schools Inc. has a curriculum that includes reading, math, history/social studies, science, writing, and world language as the core subjects, with classes in character and ethics, physical fitness and health, music, dance, visual art, drama, and practical arts and skills offered at various levels. Four methodological approaches to instruction are reportedly used in the classrooms: project-based learning, direct instruction, cooperative learning, and differentiated learning.

Assessment and accountability. One of the 10 fundamentals of the Edison design recognizes that assessment must provide the foundation for accountability. Edison has developed four main evaluation/assessment areas: (i) state and district tests; (ii) benchmark assessments; (iii) structured portfolios; and (iv) quarterly learning contracts.

Past Studies and Reports

This report summarizes an effort to determine whether or not the Edison model worked to improve student performance on standardized tests. Some attempts have already been made to evaluate overall student achievement at Edison schools. Edison itself has prepared three annual reports on student achievement at its schools (Edison, 1997, 1999, 2000) and has presented general information regarding the gains made at its schools at conferences, workshops, and in the media. The findings contained in Edison's annual reports and the message it spreads in conference presentations and in the media indicate that the company is quite successful and that students enrolled in its schools are making large and substantial achievement gains.

A few evaluations have focused on one or two schools within a given district (Miami Dade, 1998, 1999, 2000; Minneapolis Public Schools, 2000; Wichita State University, 1996) or within a given subject area (see the case studies conducted by Dr. Robert Mislevy between 1996 and 1999). The results from these studies generally have been mixed. In 1998, Dr. Howard Nelson, from the American Federation of Teachers (AFT) prepared a more comprehensive evaluation across most of the then operating Edison schools (AFT, 1998). This study relied more heavily on available state test data and it provided more information on school context than did the Edison reports. The AFT study, while acknowledging that some of the Edison schools were succeeding in raising student test scores relative with comparison groups, found that overall, Edison was not demonstrating gains across its schools. Following the release of the AFT study, Edison employed a Harvard scholar, Dr. Paul Peterson, to reexamine and compare the analyses and results in Edison's own annual reports with the AFT report. While not addressing many of AFT's more critical questions about Edison assessment and reporting activities, Peterson (1998) developed an analytical framework to compare the study designs and found that the designs employed by Edison in its 1997 annual report were stronger than those behind many of the negative or flat trends reported in the 1998 AFT report. The main rationale for this was that more of the trends in the Edison report were based on tracing gains made by individual students, while the trends presented by the AFT relied on consecutive class cohorts (i.e., different groups of students taking the same grade level test from year to year).

In October 2000, the AFT released a second report that provided an overview of the performance of schools operated by Edison. This report relied largely on data from state-mandated criterion-

referenced tests. This report has less narrative to allow for a greater focus on the numbers/data. While this overview still lacked the individual norm-referenced data that Edison also reports on in its annual reports, it did include relevant comparison groups, and the source of the data was clearly marked so others could check the figures for themselves. The overall finding from this study was that “Edison schools mostly do as well or worse than comparable schools; occasionally they do better” (AFT, 2000, p. 6). This finding can be contrasted with the summary of the findings from Edison’s third annual report that was released in September 2000: “For the 1999-2000 school year, the average gain of Edison students, in the core areas of reading, language arts, spelling, writing, and mathematics was 5 percentiles on nationally normed tests and 7 percentage points on criterion-referenced tests, which also include science and social studies. These gains represent improvements of one point in each case over the gains reported for 1995-99, and are the highest gains reported by Edison to date” (Edison, 2000, p. 2).

There are clearly differences in the nature and vested interests of the organizations conducting past studies on Edison. Thus, it is no surprise that they have presented contrasting and contradictory results. Results questioning Edison’s findings have induced strong and critical rebuttals that have, in some cases, been met with equally strong and critical rejoinders.

Aims and Objectives of the Evaluation

This study examined 10 schools operated by Edison Schools Inc. for at least 4 years. We intended to include all 11 schools that opened during the first 2 years of operation (1995-96 and 1996-97). However, we were unable to secure any independently verifiable student achievement data for Dillingham Intermediate School in Sherman, TX. We believe the schools operated by Edison for 4 or 5 years, rather than those open for 3 years or less, provide a more convincing picture of the impact the Edison model can have on student achievement. While some of these schools may have had more difficult start-ups than others, they have been in operation beyond what many consider the start-up phase.

The overriding aim of this study was to examine the impact of Edison schools on student learning as measured by norm- or criterion-referenced tests. In order to achieve this objective, a number of specific tasks were identified:

1. Review and critically assess existing research and evaluations on the impact of Edison schools.
2. Describe the evaluation measures used by Edison Schools Inc.
3. Describe the nature and quality of the standardized test results available.
4. Compare Edison schools in terms of student achievement over time.
5. Compare Edison schools with state and national norms on standardized tests.
6. Compare Edison schools with local school district and state performance levels and—where possible—other similar comparison schools.

7. Develop cases for each of the 10 Edison schools that include (i) a description of the school based upon available literature and documentation, (ii) findings from analysis of norm-referenced and criterion-referenced test results, and (iii) a summary of the diverse results from the analyses of test results.
8. Based upon available literature and documentation, develop a framework for analyzing the 10 cases.
9. Analyze the case studies according to the framework and summarize the results of this analysis.

Sample of Schools and Sources of Information

There is a possibility of selection bias related to the schools selected for the evaluation; therefore, it might be argued that the schools we studied were either performing more poorly or superior to non-selected Edison schools. We examined this possibility by comparing Edison's own school ratings, published in its 2000 annual report, for the 10 schools in this evaluation relative to the remaining 32 schools not considered in this report. Edison rates each school on a 5-point scale, from Strongly Positive to Strongly Negative. There is no indication that the 10 schools we included in this study are rated by Edison any differently from the 32 schools that opened during or after 1997 and for which it reported trend data in its 2000 annual report (Wilcoxon rank-sum test, $p = .7709$). Thus, we believe that while there may be some selection bias in our sample, there is no strong indication that the schools we evaluated are different than the schools for which Edison currently has trend data. Nevertheless, this study is a population study of 10 schools during a specific time period; we do not represent it as an inferential study aimed at all Edison schools and over a more extended time period.

Student achievement data were obtained from a variety of sources. Table 2:1 in Chapter 2 lists the standardized tests in which each participating school took part, and Appendix A describes these tests. Data sets containing individual student results on the norm-referenced tests (ITBS, MAT-7, and SAT-9) were made available to us by Edison Schools Inc. We received 7 such data sets covering 6 of the 10 schools in our study. At the time of our request for data, 3 of these data sets did not contain the results for all possible years, which limited some of these longitudinal analyses. The data sets contained anonymous indicators so that we could trace individual student results over time. From districts and state education agencies we were able to obtain results on the criterion-referenced tests for all 10 schools, the local district, the state, as well as the comparison schools/groups, where applicable.

Description of Our Approach and Unique Attributes of Our Study

Given this context, we are aware that any evaluation of Edison Schools Inc. is going to be carefully scrutinized. This evaluation speaks to both public and private interests in the work of Edison Schools, Inc. and regardless of whether the results are positive or negative, the findings are almost certain to be contested by one group or another. Anticipating this, we employed a number of strategies that we believe have strengthened our analyses as well as our role as an independent evaluator. To

insure our independence from the agency funding the study, the National Education Association (NEA) and the organization being evaluated (Edison Schools Inc.), we have done the following:

- Refused to pursue a larger study including data collection in Edison schools that would require a confidentiality agreement with Edison, which would then control access and release of the results depending on whether or not Edison approved of the findings
- Insisted that our contract with NEA allows us to publish the findings, regardless of the outcomes

In order to strengthen our analyses and the overall credibility of our work, we took the following steps:

- We provided the NEA, Edison Schools Inc., several of the school districts that have Edison schools within their boundaries, and researchers in the field with copies of the draft report and welcomed comments and corrections, which we considered in finalizing the report.
- We sought the advice of experts in the field regarding the strategies and methods we chose to work with state achievement data.
- We sought a variety of data sources and made comparisons with multiple groups whenever possible.
- We clearly describe the assumptions behind our analyses.
- We established and explain the criteria we used to distinguish whether trends are positive, mixed, or negative. We also developed and applied a common method to calculate a mean score across the trends in order to label each school on a 5-point scale—as Edison does—ranging from Strongly Negative to Strongly Positive.
- We attempted to report all data and present all charts and figures clearly and consistently.
- We clearly described the limitations of the study.
- We relied on the program evaluation standards (Joint Committee on Standards for Educational Evaluation, 1994) to guide our work.

To the extent possible, we tried to replicate the presentation of findings that Edison includes in its annual report. This includes a presentation of trends on various norm-referenced or criterion-referenced tests and then a rating given to the school that reflects a summary of all the trends. When we could identify Edison's own methods and techniques for calculating trends and rating schools, we applied them. When we had to develop our own methods, we explained them as clearly as possible. The following lists some important differences between Edison's presentation of school results in its annual reports and our own.

- Edison includes only a brief summary of each school, with a table of general data and a page of charts outlining the achievement trends. By contrast, we developed in-depth cases ranging from 10-20 pages in length.
- Edison's definition of a trend is different than our own. It appears that Edison's approach to defining and counting trends results in a larger number of trends than our approach. We rate

overall average annual change in a trend, while Edison counts any difference between a two-year block of time as a trend (we explain this in detail in Chapter 2).

- Edison states that it does not intend to compare change in its schools with others (in its annual report there are a number of exceptions to this), while our analysis of gains is dependent on comparisons (see Section 2.3 “To Compare or Not to Compare” for more details).
- Edison reports reviewed contained no information on the number of students (N) reflected in the trends it presents and the source of data is often unclear. By contrast, we made a conscious effort to report the N for all trends and to identify the source of data so that others can check these sources and replicate our work if they wish.
- Edison’s annual reports appeared to us to be intended for a lay audience and provide little insight on the methodology behind the findings they present. Our report is a technical report that provides extensive detail on the methodology, but with an executive summary that presents the results for a lay audience.
- While Edison attempts to present the findings across all its schools, we focused only on those schools that Edison opened in 1995 and 1996 since they have the most data available and because Edison has had time to implement its program more fully in these schools.

Many of the differences between Edison’s reporting and our own are dealt with and explained more thoroughly in other parts of this report.

The conduct of the study was complicated by a number of factors (for example, late receipt of data files from Edison and then only for a limited number of schools, also changes in the members of our evaluation team). The study is marked by a number of other limitations (see Section 2.6). Nevertheless, we are satisfied with the overall results and think this evaluation makes a number of important contributions to the debate over student achievement in schools operated by Edison since it both adds to the record and provides an in-depth discussion and summary of related research on Edison Schools Inc. Also, because of the many years of test results available for our study and because of the use of multiple sources of student achievement data and multiple comparison groups, we believe this is the most in-depth and extensive assessment of student achievement in Edison schools to date.

Methodology

The study combines a variety of statistical methods to compare growth in student achievement in Edison schools with growth in control schools and districts, statewide achievement data, and with national norms, when available. The use of odds ratio analysis to examine criterion-referenced achievement data, given the extensive availability of statewide assessment data and the limitations in the current traditional analyses, may prove to be a significant contribution (Chapter 2 contains more details about odds ratio analysis).

The focus of the methodology employed in this evaluation was to identify and assess the academic achievement gains of students enrolled in Edison schools; therefore, a comparison was essential.

There are a multitude of possible comparisons that could be made, and it is beyond the scope of this report to include them all. Rather, this evaluation focused on two primary types: (1) within-subject gains evidenced by a group of students' year-to-year scores on norm-referenced achievement tests (NRT), and (2) cohort comparisons evidenced by changes across years in the criterion-referenced test (CRT) scores of successive groups of students at given grade levels. These analyses focus on gains over time (i.e., value added) and are based on the assumption that the more exposure to the Edison effect (i.e., the more time a student is enrolled at an Edison school), the better the students will perform compared to relevant comparison groups.

Analysis of Norm-Referenced Achievement Test Data

A variety of NRT data (see Table 2.1 and Appendix A) was provided to us by Edison for 6 of 10 schools. Unfortunately, these data files did not contain data on a comparison sample and tended to be rather incomplete. Consequently, the NRT analyses only examine within-subject change compared with national norms. Following a list-wise deletion strategy, repeated measures ANOVA was used to test for longitudinal trends over the available years. Often Edison provided us with a variety of scores (grade equivalent, standard or scaled score, percentile rank or national percentile rank, and/or the normal curve equivalent). We provide parallel analyses for all types of scores, although gave preference to the normal curve equivalents when summarizing the trends at the end of each case.

Analysis of Criterion-Referenced Achievement Test Data

CRT data were culled from a variety of sources but primarily from the Web (see Table 2:1). The CRT data we examined were based on the aggregate performance of all students taking the test by grade and subject. Thus, unlike the NRT data, where we could follow individual student data over a period of years, the CRT data represented the performance of consecutive cohorts of students. Since these data are available to the public, we were able to define and construct comparison groups (detailed in each case study) for these analyses. The ability to define a comparison group allowed our analyses to test if the relative proportion of students in an Edison school scoring among the various levels coincided with either district or state scoring proportions via a chi-square analysis. These analyses were examined for different years by grade level and subtest category of the state test.

The second analysis strategy we applied to the CRT data examined student learning outcomes as a prospective cohort study by collapsing the distribution of scores on the various state tests into pass/fail categories, reflecting the percent of students meeting or exceeding state standards vs. not meeting state standards. We constructed the 2x2 tables for these analyses to represent the relative odds for a student to fail a component of the state test. The odds ratio (OR) represents the proportion of students who fail the test in the Edison school relative to the proportion of students who fail the test in the comparison school/group.

In order to accomplish the OR analyses, we had to collapse the various CRT scoring categories into pass/fail categories. Although there are several possible ways to do this, we opted to define passing and failing as specified by each state. It should be noted that this reclassification could mask some

important gains evidenced by Edison schools. We acknowledge this possibility and therefore present both the uncollapsed (chi-square) and collapsed (OR) findings so that the reader can determine if there has been a shift in scoring categories that is masked in the collapsed analyses.

We believe this evaluation makes a valuable methodological contribution in its application of a statistical procedure not commonly utilized in educational research or evaluation. With the CRT analyses the use of the Breslow-Day statistic to test for longitudinal trends in consecutive cohorts makes an important methodological contribution that improves evaluators' ability to determine the impact and merit of educational reform over a long period of time. The conventional evaluation approach to CRT data typically involves charting passing/failing trends over time, visual inspection, and quantifying (charting) change scores. The methodological improvement employed in this evaluation was to test for statistically significant differences in passing/failing rates over time. The conventional chi-square often used with this type of data cannot do this, it is limited to examining only one cohort at a time. Thus, in an absolute sense the chi-square tests can identify if the target school is superior to the reference school/group in any given year/grade on any given test; however, it cannot identify any longitudinal gains over several years.

Standards of Comparison

We believe that judgments about the overall performance of the involved schools need to be made on a case-by-case basis. In order to limit potential bias and to establish a common method of making judgments, we thought it was important to establish criteria to distinguish whether or not there had been change over time and whether any identified change was positive or negative. Edison's criteria served as our starting point following which we modified several criteria and added one new criterion. In evaluating trends, we used the following criteria to distinguish when there was meaningful change and when this change was positive or negative:

- *Effect sizes or differences in effect sizes of .20 or greater.* The effect size calculated for the NRT data is the omega squared (ω^2) for a one way repeated measures ANOVA and only provides the reader with an overall effect for time. It does not adequately convey the direction of change nor if the change occurs all in one year or is reflective of a gradual cumulative gain.
- *Differences in national percentile scores of 5 percentage points or more per year*
- *Differences in percentage proficient scores of 5 percentage points or more per year*
- *Differences in grade equivalents of 2 months or more and annual gains in grade equivalents of 14 months or more per year*
- *Differences in DALT gain scores of 2 points or more per year*
- *Differences that are statistically significant (at the .05 level) when tests of significance are available.* The p-value criteria are only applied to the CRT data and not to the NRT data. Utilization of a p-value criteria in the longitudinal NRT analyses does not adequately convey the direction of a statistically significant change. That is, in a longitudinal analysis there can be a statistically significant change in both directions and therefore the p-value is ambiguous relative to the direction of change, only the presence of change. However, in the chi-square and OR

analysis, the p-value conveys a meaningful difference due to the configuration of the contingency tables, in that for these analyses the referent is to the comparison group.

○ *Differences in normal curve equivalents of 3.5 or more per year*

Part of the technical complexity of this report is a function of the variety and large number of analyses conducted. We have constructed summary tables that help guide and focus the reader in distilling the overall impact of Edison in a given school. We treated each analysis category (NRT, CRT) separately. Within each analysis category we rated a finding as negative (-1), mixed (0), or positive (+1) based upon the guidelines presented above.

We also based our rating on a prioritized hierarchy of data. We consider a trend in NRT data to reflect the findings on a group of students as they progress through the life of the analysis by subject and grade level. A trend in CRT data reflects the consecutive cohort findings for a specific grade and subject test over the life of the analysis.² Although we calculated outcomes relative to various comparison groups (e.g., national, state, district, or other), we only counted one trend in the combined table. For NRT data we prioritized the analyses as follows: we considered the normal curve equivalent trend first if available, followed by the percentile rank, then grade equivalent, and lastly the standard score. For CRT data we counted each grade and subject test separately based on the outcomes of the OR Breslow-Day findings relative to the district data.

In its 2000 annual report, Edison defined the 5-point scale it used to rate the overall trends in its schools (Edison assigns one to five stars for each of the categories, from Strongly Negative to Strongly Positive, respectively). Its cut points are as follows: Strongly Negative when 0-19 percent of the trends are positive; Negative when 20-39 percent of the trends are positive; Mixed when 40-59 percent of the trends are positive; Positive when 60-79 percent of the trends are positive; and Strongly Positive when 80-100 percent of the trends are positive.

Since we wanted to consider all the trends and not focus on the positive trends alone, we calculated a mean across the trends where a negative trend is equal to -1, a mixed trend is equal to 0, and a positive trend is equal to +1. We then applied the following 5-point rating scale to the mean trend:

- -1.00 to -0.60 corresponds with “Strongly Negative”
- -0.59 to -0.20 corresponds with “Negative”
- -0.19 to +0.19 corresponds with “Mixed”
- +0.20 to +0.59 corresponds with “Positive”
- +0.60 to +1.00 corresponds with “Strongly Positive”

² By contrast, Edison counts trends in one-year change segments, so a trend of data for a cohort of students over 4 years would be counted by Edison as 3 different trends, while we would count this as one trend and base our rating on the change over the life of the trend.

Methodological Limitations

Several inherent limitations in this evaluation need to be examined in order to provide a balanced interpretation of the findings we have reported and conclusion we have drawn. The limitations to this study can be grouped into three areas—methodology, data quality, and conceptual limitations. They include the following: (i) lack of a comparison group in the longitudinal analyses; (ii) limited and incomplete individual student achievement data supplied by Edison; (iii) validity of comparison groups used in the chi-square and OR analyses; (iv) validity and completeness of Web-based reporting of the district- and state-mandated testing results; (v) evaluation of schools based on student performance alone; and (vi) limited resources. These limitations, which are explained in Chapter 2, should temper all conclusions derived from this evaluation.

Summary of Cases

Roosevelt-Edison Charter School, Colorado Springs, Colorado

Roosevelt-Edison Charter School is a district charter school that was established in 1996 to serve grades K-5. Enrollments in this school totaled 674 during the 1999-00 school year. The school's population was more diverse ethnically than the district and state populations and had a higher proportion of low-income students than the district and state.

Edison provided us a data set that included individual student results on the Iowa Test of Basic Skills. However, this only covered two years (1997-98 and 1998-99). In terms of criterion-referenced tests, we utilized the results from the Colorado Student Assessment Program (CSAP), which is a state-mandated test, based on state standards. We were able to secure test results for the Roosevelt School, the local district, and the state from the Colorado Department of Education. Because the CSAP is a relatively new state assessment program and because it is still not fully implemented in terms of grades and subject tests, we were not able to collect data for all years that the Edison charter school was in operation. We were able to obtain grade 3 reading results for 2 years, grade 4 reading and writing results for 3 years, and grade 5 math results for one year.

Our analyses produced results on nine different trends, six of these based on norm-referenced test results, which examined a longitudinal change in individual student data. Three of the trends were based on the state-mandated assessment test, and we used odds ratio analysis to distinguish if the odds of meeting state standards increased or decreased over time.

Roosevelt-Edison	Positive (+1)	Mixed (0)	Negative (-1)
Norm-Referenced	2 of 6	4 of 6	0 of 6
Criterion-Referenced	0 of 3	2 of 3	1 of 3
TOTALS	2 of 9	6 of 9	1 of 9

This school had a mean trend rating of 0.11. According to our criteria for the 5-point rating scale, the trends in this school are categorized as Mixed. In its 1999 annual report, Edison rated this school as Strongly Positive; and in its 2000 annual report, it rated the achievement gains since opening as Positive and the achievement gains in 1999-00 as Negative.

Our findings are similar to district findings on the District Achievement Level Tests (DALT). Results on the DALT indicate that the school is consistently below the district levels, but the gains made by Roosevelt-Edison on the DALT largely parallel the gains made by the district (i.e., the charter school is not gaining more than the district over time). The results to date indicate that this Edison school does not differ substantially from other district schools.

Henry E.S. Reeves Elementary School, Miami-Dade County, Florida

Henry E.S. Reeves Elementary School is a district contract school serving grades K-5 that Edison began operating in 1996. The total enrollment has remained relatively stable, similar to both the district and the state, with a range from 1,081 students in 1996 to 1,193 students in 1999. Reeves has less diversity ethnically than the state or the district with a much higher proportion of African-American students and lower proportions of white and Hispanic students.

Edison did not provide any NRT data to us for this case, and thus we based our NRT summaries on a previously published evaluation conducted by Miami-Dade County Public Schools (MDCPS) and a dissertation by Sally Shay (2000). Additionally, we were able to extract three years of data on the FCAT and Florida Writes assessments.

MDCPS's ongoing evaluation of Henry E. S. Reeves Elementary School as well as the Shay (2000) study are very likely the most thorough evaluations of any Edison school to date. Because we could not obtain individual results to analyze for this school, we based our NRT trend ratings on results presented by Shay (2000). Her study presented partial effect size estimates (expressed in percentages) from repeated measures analyses. The percentage estimates reflected effect size contributions from both the main effect for group membership and the group by time interaction. Both the district evaluation and the Shay study found that the performance of the Edison students is comparable to but not better than the control groups.

Student results from the three state-mandated tests (FCAT reading and math and Florida Writes) indicate some gains for the Edison students; but in absolute scores, Edison students are still far behind the averages for the district and state. More importantly, the gains made by Reeves on the CRT are similar to those made by the district and state groups.

Henry E. S. Reeves	Positive	Mixed	Negative
Norm-Referenced	0 of 4	4 of 4	0 of 4
Criterion-Referenced	0 of 3	3 of 3	0 of 3
TOTALS	0 of 7	7 of 7	0 of 7

Based on our findings and the discussion of the results, we rated the trends in this school as Mixed with an overall mean trend rating of 0.0. In fact, all seven trends that we considered when summing up this case were Mixed. In its second annual report, Edison (1999) rated the trends on student performance as Positive. In its third annual report, Edison rated the trends for 1999-00 as Strongly Positive and the overall trends since opening as Strongly Positive. Extensive results from the SAT-7 are available for this school. Edison did not consider the findings from the district evaluation, nor from the Shay (2000) study, when presenting the results for this school in its 2000 annual report. In fact, while Edison reported same cohort SAT results for 1996-97 and 1997-98 in its 1999 annual report, it included only limited consecutive cohort SAT results in its 2000 annual report.

Dodge-Edison Elementary School, Wichita, Kansas

Dodge-Edison Elementary School is a district contract school established in 1995 to serve grades K-5. The school's enrollment has ranged from a low of 334 students in 1995 to a high of 657 students during the 1997/1998 school year. Data on the ethnic composition of the students enrolled in the school indicate a majority of white students (65 percent) with little fluctuation over the years. Approximately 58 percent of the students qualify for free or reduced lunches, which is slightly higher than district levels.

Edison provided us with three years of MAT-7 mathematics and reading subtest NRT data from which we were able to extract two cohorts to follow for the years 1995-98 (Cohort A) and 1996-99 (Cohort B). Additionally, we were able to extract a limited amount of data on the Kansas Reading, Math and Writing Assessments, yielding three CRT trends.

Dodge-Edison	Positive	Mixed	Negative
Norm-Referenced	2 of 4	2 of 4	0 of 4
Criterion-Referenced	1 of 3	2 of 3	0 of 3
TOTALS	3 of 7	4 of 7	0 of 7

The average annual gains on the MAT-7 in terms of normal curve equivalents were 5.8 in Math and 4.8 in reading for Cohort A, which were statistically significant. The average annual gains made by Cohort B (2.9 in math and 2.0 in reading), though respectable, were not statistically significant.

The change score for Dodge-Edison students on the Kansas Reading Assessment was more negative than for district students, while the Dodge-Edison students showed a slightly larger gain than the district on Kansas Math Assessment. Differences between Dodge-Edison and the district were small in both of these cases. On the Kansas Writing Assessment, however, the gain made by the Dodge-Edison students was quite substantial and represented 6 percent of the total range of the scale. District students made a very small gain on the Writing Assessment. Therefore, we rated one of three CRT trends as Positive and two as Mixed.

Given the total ratings for the seven trends, which are highlighted above, we rated this school as Positive with a mean trend rating of 0.43. In its 1999 annual report, Edison rated the trends in this school as Positive. In its 2000 annual report Edison rated the achievement gains since opening as Strongly Positive and the achievement gains in 1999-00 as Strongly Positive.

Dodge-Edison has been lauded by Edison as one of its success stories. While the overall label we use to categorize the trends in this school is clearly Positive, they are not as positive as Edison suggests in its third annual report. In any case, the gains made by students enrolled in Dodge-Edison are both substantial and consistent. The achievement gains in this school are clearly the most positive of the 10 cases we examined.

Jardine-Edison Junior Academy, Wichita, Kansas

Jardine-Edison Junior Academy is a district contract school established in 1996 to serve grades 6-8. The school enrolled 822 students during the 1999-2000 school year. Ethnic composition has fluctuated somewhat at this school with the percentage of white students declining from a high of 49 percent during the 1996-97 school year to 32 percent during the 1999-2000 school year. The proportion of students qualifying for free or reduced lunches has fluctuated between 61 and 74 percent, which is consistently higher than district averages, which are around 50 percent.

Edison did not provide us with individual student data for this school; however, we were able to make comparisons between Jardine-Edison and the district on overall performance on the Metropolitan Achievement Test (MAT-7) from data made available by Wichita Public Schools. Additionally, we were able to collect three years of data on the Kansas Reading, Math and Writing Assessments, yielding three CRT trends.

The design behind the trends in the norm-referenced results is based on tracing consecutive cohorts of students at three grade levels over three years. These trends were traced for performance on MAT-7 reading and math. We gave positive ratings to three trends where the Jardine-Edison students made large average annual gains in terms of their national percentile rank as well as large gains compared with districts gains, which were also large, particularly in math.

On the criterion-referenced test results, Jardine had a positive trend relative to the district in reading, but a negative trend relative to the district in writing. Gains made in math by both Jardine-Edison and the district were small and similar in size.

Jardine-Edison	Positive	Mixed	Negative
Norm-Referenced	3 of 6	3 of 6	0 of 6
Criterion-Referenced	1 of 3	1 of 3	1 of 3
TOTALS	4 of 9	4 of 9	1 of 9

Given the total ratings for the 9 trends that are highlighted above, we rated this school as Positive with a mean trend rating of 0.33. In its 1999 annual report, Edison rated the trends in this school as Positive. In its 2000 annual report, Edison rated the trends for the 1999-00 school year alone as Negative, yet at the same time labeled the trends since opening as Strongly Positive. This begs the question or explanation of how the Positive rating of overall trends in 1999 plus the Negative trends for 1999-00 could equal the Strongly Positive rating of overall trends cited in the 2000 annual report (Edison, 2000, p. 78). Although the label we used to categorize the overall trends in this school is Positive, the trends are clearly not Strongly Positive as Edison suggests in its 2000 school report.

Boston Renaissance Charter School, Boston, Massachusetts

The Boston Renaissance Charter School is an independent Charter School established in 1995 to serve grades K-5 with grades 6-8 added in 1996. During the 1998/99 school year there were 1,063 students enrolled in the school. Data on ethnic compositions indicates that the school is predominately nonwhite (86.2 percent), similar to the composition reported for the Boston Public School District; however, Boston Public Schools had more Hispanic and Asian students and fewer African-American students than Boston Renaissance. Both school districts had considerably larger proportions of minority students when compared with the overall state percentage of 77.1 percent white and 32.9 percent nonwhite. This Edison school has a lower proportion of students receiving special education services and a lower proportion of students qualifying for free or reduced lunches than the averages for Boston Public Schools.

Edison provided us with Stanford Achievement Test, Version 9 (SAT9) test data for four years (1996-1999), and this school is also required to take part in the state CRT assessment program. We were able to secure MCAS test results for two years, 1998 and 1999.

Boston Renaissance	Positive	Mixed	Negative
Norm-Referenced	0 of 12	12 of 12	0 of 12
Criterion-Referenced	0 of 6	3 of 6	3 of 6
TOTALS	0 of 18	15 of 18	3 of 18

We examined an abundance of data for this case. We derived 18 overall trends: 12 from NRT data and 6 from CRT data.

We rated this school as Mixed with a mean rating of -0.167. In its 1999 annual report, Edison rated this school as Mixed. In its 2000 annual report, it rated the 1999-2000 school year as Strongly Positive and the achievement gains since opening as Positive.

Seven Hills Charter School, Worcester, Massachusetts

Seven Hills Charter School is an independent charter school established in 1996 to serve grades K-8. The school opened with K-6, and in 1997 grades 7 and 8 were added. During the 1998/99 school

year, 662 students were enrolled in the school. Data on ethnic composition indicate a relatively equal distribution between white and nonwhite students and similar distribution to that found in the Worcester school district. This school has a slightly lower proportion of students with special educational needs than the local school district and has a slightly smaller proportion of students qualifying for free or reduced lunch than the district.

Edison provided us two data sets: Stanford Achievement Test (SAT-9) data for two years (1998-1999) and three years of student achievement data on the Metropolitan Achievement Test (MAT-7) (1997-1999). Similar to all Massachusetts public schools, this school is also required to take part in the Massachusetts Comprehensive Assessment System (MCAS) program. We were able to secure test result data for 1997-98 and 1998-99 only. This matches the CRT results reported by Edison in its 2000 annual school performance report.

Our analyses produced 18 different trends, 12 based on longitudinal norm-referenced tests data and 6 based on consecutive cohort odds ratio analysis of criterion-referenced test data.

Seven Hills	Positive	Mixed	Negative
Norm-Referenced	3 of 12	9 of 12	0 of 12
Criterion-Referenced	0 of 6	1 of 6	5 of 6
TOTALS	3 of 18	10 of 18	5 of 18

This school had a mean trend rating of -0.11 and, according to our 5-point rating scale, this mean score represents a school with Mixed results. In its 1999 annual report, Edison rated this school as Strongly Positive. In its 2000 annual report, it rated the 1999-2000 school year as Mixed and the achievement gains since opening as Mixed also.

It appears that Seven Hills' students evidenced mild achievement growth over the three academic years on the norm-referenced tests, but did not make any noteworthy gains in comparison with the national norm. The analysis of individual student results, as measured by the MAT-7, indicated largely no gains or losses relative to national norms over the three years we traced the students. Two exceptions to this are when the larger Cohort A showed a statistically significant decrease relative to the national norm for normal curve equivalent (NCE) over two years in language, and the smaller Cohort B showed a statistically significant gain on the NCE over one year in math.

The odds ratio analysis of the MCAS results indicated that the odds of failing (i.e., scoring in the "needs improvement" or "failure" categories) the state assessment test are higher at Seven Hills Charter School than in the district and in the state as a whole. What is most disconcerting is that the odds of Seven Hills' students failing increases over the two years considered in all subject areas but one in both grades 4 and 8.

In its 1999 annual report, Edison rated this school as Strongly Positive; and in its 2000 annual report it rated the 1999-2000 school year as Mixed and the achievement gains since opening as Mixed also. Our findings, based upon the analysis of individual student results on the MAT-7, SAT-9, and the odds ratio analysis of consecutive cohorts on the MCAS, indicate that the performance of this school—in terms of student achievement—can best be characterized as Mixed with a mean rating of -0.11. This is due to the fact that nearly all the trends were mixed although 3 NRT trends were positive and 5 of the CRT trends were moving in the wrong direction. This is the only school in this study where the rating given by Edison and The Evaluation Center are the same.

Dr. Martin Luther King Jr. Academy, Mt. Clemens, Michigan

Dr. Martin Luther King Jr. Academy (MLK) is a district contract school established in 1995 to serve grades K-5. The school had 877 students enrolled in 1998/99. Data on ethnic composition (for 1997/98) indicates relatively equal distribution between white (53.3 percent) and African-American students (44.3 percent), but scant inclusion of other ethnic groups. The school has around 6 percent of its students receiving special educational services and close to 30 percent qualifying for free or reduced lunches. The proportion of students in the district qualifying for free or reduced lunches has risen in recent years, while the proportion in this and other Edison schools in this district declined.

Edison provided us with a data set containing the Iowa Test of Basic Skills results for three years (1997, 1998, and 1999). Like all Michigan public schools, this school is required to take part in the state CRT assessment program (MEAP). We secured CRT test results for the years 1995-2000 from the Michigan Department of Education for the Edison school, for the only other elementary school in the district with students at grades 4 and 5 (i.e., George Washington Elementary), and for the state.

M.L. King Academy	Positive	Mixed	Negative
Norm-Referenced	0 of 3	3 of 3	0 of 3
Criterion-Referenced	0 of 4	2 of 4	2 of 4
TOTALS	0 of 7	5 of 7	2 of 7

Our analyses produced seven different trends: three based on longitudinal norm-referenced test data and four based on consecutive cohort odds ratio analysis of criterion-referenced test data. Although this represents a moderate number of trends, it is notable that the NRT data represent three-year longitudinal trends and the CRT data represent trends over four or more years.

Three trends were based on the individual student gains on the ITBS, and 4 trends were based on MEAP results. Five of the seven total trends were mixed, and two were negative. Based on these figures, the school had a mean trend rating of -0.29, which indicates an overall rating of Negative. Both the results from the NRT data and the CRT data indicate that students at this school are below national and district norms. This is essentially how the school performed before Edison took over operation. In its 1999 annual report, Edison rated this school as Strongly Positive. In its 2000

annual report, it also rated the 1999-2000 school year as Strongly Positive and the achievement gains since opening as Strongly Positive.

While some gains can be seen in the results from the Iowa Test of Basic Skills, they are not sustained over two years. There were gains in terms of normal curve equivalents (NCE) in the first year and then minor decreases in the NCE in the second year of the longitudinal analysis. One exception to this was in language, where the cohort of students made gradual gains over two years, with the NCE ranking in 1998-99 being statistically significant from the NCE in the 1996-97 school year.

Results from the state MEAP assessment were not promising, however. The odds of not meeting state standards on the subject tests at MLK are generally large. The odds of failing or not meeting state standards for the most part remained the same over time in comparison with the state total and with George Washington Elementary, the only other school in the district with students enrolled in grades 4-5. Both the ITBS and the MEAP results indicated that students at MLK are below national and district norms. This is essentially how the school performed before Edison took over operation.

Mt. Clemens Secondary Academies, Mt. Clemens, Michigan

Mt. Clemens Secondary Academies is a district contract school established in 1996. Both the Edison middle school and the Edison high school share a common building with their district's only other middle and high school. Edison reported that the total enrollment in the secondary academies totaled 481 for the 1999-00 school year. Data on ethnic composition (for 1997-98) indicates slightly more white (62.9 percent) than nonwhite students. According to data presented in Edison's annual reports, the proportion of students receiving special education has dropped in recent years (11.4 percent in 1997-98 and 7.9 percent in 1999-00). The proportion of students qualifying for free or reduced lunches has dropped even more sharply (27 percent in 1997-98 down to 12.5 percent in 1999-00).

Tests administered at the Mt. Clemens academies include the Michigan Educational Assessment Program (MEAP) as well as the Iowa Test of Basic Skills (ITBS). The results from the ITBS were not reported in Edison's second or third annual reports. We were unable to secure individual student achievement data on the ITBS from Edison for this school, although we did receive copies of the summaries of school results for 1996/97 that were supplied to the school by the test company. It was unclear whether or not the school summaries included both the Edison school and the district middle school that share the building.

Given the wide range of grades, this school should have produced a lot of data from the state criterion-referenced test (MEAP). Unfortunately, data were often unusable due to the fact that this school was a school-in-a-school. Thus, we could only secure disaggregated data from a limited number of years. We were able to include MEAP for grades 7 (1997/98 - 1999/00) and 8 (1999/98 and 1999/00), but for grade 11 we were able to examine only the 1998/99 data because instruction at grade 11 began in 1998.

Mt. Clemens Sec. Acad.	Positive	Mixed	Negative
Norm-Referenced	—	—	—
Criterion-Referenced	2 of 8	6 of 8	0 of 8
TOTALS	2 of 8	5 of 8	0 of 8

Because of the limited data available, the school could be grouped with the other schools that Edison opened later and which they categorized as “baseline” rather than attempting to place a label on the progress of the school. If we were to make a judgment based on the limited data available, the Secondary Academies would be rated as Positive with a mean rating of 0.25. In its 1999 annual report, Edison rated this school as Strongly Positive. In its 2000 annual report, it rated the 1999-00 school year as Strongly Positive and the achievement gains since opening as Strongly Positive also.

Mid-Michigan Public School Academy, Lansing, Michigan

Mid-Michigan Public School Academy is a charter school established in 1996 to serve grades K-5, with grades 6-8 added in 1997. This is the largest charter school in Michigan and enrolls more than 1,000 students. Data on ethnic composition indicate a much larger percentage of nonwhite students (79 percent) than is found in Lansing Public Schools. The proportion of students qualifying for free or reduced lunches is approximately 50 percent, which is similar to the district level.

Edison Schools Inc. provided individual student test results on the Metropolitan Achievement Test for one cohort of students between 1997-1999. Like all Michigan public schools, this school is also required to take part in the state CRT assessment program, and we secured test results for the years 1997-1999.

Our analyses produced 11 trends: 3 based on longitudinal norm-referenced tests data and 8 based on consecutive cohort odds ratio analysis of criterion-referenced test data.

Mid-Michigan PSA	Positive	Mixed	Negative
Norm-Referenced	0 of 3	3 of 3	0 of 3
Criterion-Referenced	0 of 8	0 of 8	8 of 8
TOTALS	0 of 11	3 of 11	8 of 11

This school had a mean trend rating of -0.73 and fell into the Strongly Negative category on our 5-point rating scale. Our assessment of this school is similar to the assessment made by the school itself in its 1999 annual report (Mid-Michigan PSA, 1999). Edison’s central office, however, reported a different picture. In its 1999 annual report, Edison rated this school as Mixed. In its 2000 annual report, it rated the 1999-2000 school year as Positive and the achievement gains since opening as Mixed, clearly inconsistent with our findings.

Washington Elementary School, Sherman, Texas

Washington Elementary School is a district contract school that Edison began to operate in 1995 to serve grades K-4. The school had 445 students enrolled during the 1998/99 school year. Data on student characteristics indicate that this school has more minorities and more students qualifying for free or reduced lunches, but fewer students receiving special educational services than other district schools.

Edison did not provide us with individual norm-referenced achievement test data for students in this school, although the ITBS is administered at this school, at least in the earlier grades. However, the Texas Assessment of Academic Skills (TAAS) in reading, math, and writing for Grades 3 and 4, the state-mandated CRT, is administered; and we secured test results for the 1995-96, 1996-97, 1997-98 and 1998-99 school years. The TAAS reporting differs from many state CRTs in that each school is compared with a control group of 40 demographically similar schools, based on the growth in the students' Texas Learning Index³ scores on the TAAS reading and mathematics tests given from one year to the next. The comparable improvement is based only on students whose TAAS results can be matched from the current to the prior year. Since only grades 3 and 4 are tested at Washington, the comparable improvement on the TAAS is based on the performance of fourth graders compared with how they performed on the TAAS in third grade (Texas Learning Index scores are used for comparing individual students from one year to the next).

Based on student performance on the TAAS, we characterize the overall trends in this school as Negative. Except for one successful year, the school is largely unable to even match the gains made in the comparison groups when we consider matched students. In terms of consecutive class cohorts, the school has only been able to come close to matching the control groups on the 4th grade writing test. The results dropped considerably in Edison's first year of operation compared with the previous year when the school was run by the district.

Washington Elementary	Positive	Mixed	Negative
Norm-Referenced	—	—	—
Criterion-Referenced	1 of 5	1 of 5	3 of 5
TOTALS	1 of 5	1 of 5	3 of 5

The proportion of students exempted from the state criterion-referenced test climbed from 4 percent in 1995-96 to around 25 percent in 1998-99, even while the school's performance on the state test remained very poor.

Given the total ratings for the trends, we rate this school as Negative with a mean rating of -0.40. In its 1999 annual report, Edison rated this school as Strongly Positive. In its 2000 annual report, Edison rated the 1999-2000 school year as Positive and the achievement gains since opening as Positive.

³ The Texas Learning Index (TLI) is a score that describes a student's performance on the TAAS reading or mathematics test. It can be used to tell how far a student is above or below the passing standard.

Summary of the Findings

There has been a paucity of evaluations of Edison schools, and those conducted have been typically limited to one or two schools and/or limited in time (Gomez & Shay, 1998, 1999, 2000; Minneapolis Public Schools, 2000; Wichita State University, 1996; AFT, 1998, 2000; Edison, 1997, 1999, 2000). Of particular interest in this evaluation is the relationship of our results to the results previously published in Edison annual reports on student achievement. Thus, we will present a brief overview of findings published in Edison's annual reports (Edison, 1997, 1999, 2000). Edison's First Annual Report on School Performance (Edison, 1997) was structured around five key performance areas: student achievement, customer satisfaction, the implementation of the school design, financial management, and systems growth.

Edison reported that it used the following rules to establish achievement trends: (i) whenever possible to measure trends using the same individual students; (ii) when the same individual students are not available, to measure the same group of students; and (iii) when the same individual students and the same group of students are not available, to measure the same grade level (Edison, 1999). Edison calculated the net gain for every trend from the first observation to the final observation of a trend. Some of its reported trends were three years long, others two years, and some only one.

Edison's second annual report (1999) provided the following summary statements about the achievement levels of its students/schools:

- The overwhelming majority of achievement trends—numbering nearly 200—were positive.
- On average, students were gaining more than 5 percentiles per year against state and national standards.
- Students in nearly every Edison school are achieving more today than when the school opened; in no Edison school are students achieving less.
- Fourteen of 17 Edison schools that have now established achievement trends have moved student achievement forward.
- In total the trends in 10 of the 17 schools were rated Strongly Positive, 4 were rated Positive, 3 were Mixed, and no schools were rated as Negative or Strongly Negative.

The principal aim of this evaluation was to determine whether or not the Edison model worked to improve student performance on norm- or criterion-referenced standardized tests. To that end, this evaluation provides a comparison of *achievement gain*, focusing on two primary types, (1) within-subject gains evidenced by longitudinal panel norm-referenced achievement test (NRT) data and (2) cohort comparisons evidenced by longitudinal cohort criterion-referenced test (CRT) data.

Norm-Referenced Achievement Test Data

Although a variety of NRT data was provided to us by Edison (ITBS, MAT-7 & SAT-9), it was limited in the number of schools and number of years of data collection and did not contain data on

a comparison sample. Consequently, all NRT analyses examined only within-subject change and did not provide an external comparison except for that imbedded in the score reporting scale, e.g., NCE scale. A list-wise deletion strategy, followed by repeated measures ANOVA to test for longitudinal trends over the available years, resulted in cases with complete data over the life of the analysis. However, this procedure likely raised attrition rates. Specific attrition rates can be calculated for each case by cohort group, subject test, and reporting scale and are elaborated in each case and in Chapter 13. Our ratings (positive, mixed, or negative) of NRT trends were based on a system developed and elaborated in Section 2.5.

Overall Results on Norm-Referenced Trends

Case	Positive	Mixed	Negative
Roosevelt-Edison	2 of 6	4 of 6	0 of 6
Henry E. S. Reeves	0 of 4	4 of 4	0 of 4
Dodge-Edison	2 of 4	2 of 4	0 of 4
Jardine-Edison	3 of 6	3 of 6	0 of 6
Boston Renaissance	0 of 12	12 of 12	0 of 12
Seven Hills Charter School	3 of 12	9 of 12	0 of 12
Dr. Martin Luther King Jr.	0 of 3	3 of 3	0 of 3
Mt. Clemens Sec. Academies	—	—	—
Mid-Michigan PSA	0 of 3	3 of 3	0 of 3
Washington Elementary	—	—	—
TOTALS	10 of 50	40 of 50	0 of 50

The main findings from the norm-referenced analyzes can be summarized as follows:

- We charted 50 norm-referenced trends.
- Overall, the norm-referenced trends were either mixed or positive; none were negative.
- Students in Edison schools are generally showing academic achievement gains consistent with grade level advancement on norm-referenced tests.
- Achievement gains do not consistently exceed grade level expectations on norm-referenced tests.

Criterion-Referenced Achievement Test Data

CRT data were culled primarily from the Web. These data typically were reported in aggregate for all students taking the test, broken down by grade and subject test. No individual performance data were obtainable. Thus, the CRT data represented the performance of consecutive cohorts of

students. The criterion-referenced tests were different for each state since they were usually part of a mandated state assessment program. Our ratings (positive, mixed, or negative) were based on a rating system developed and elaborated in Section 2.5. We analyzed the CRT data a number of different ways; however, we based our rating on the consecutive cohort odds ratio (OR) analyses.

A negative rating would be an OR greater than 1.0 with a confidence interval (CI) whose lower bound was greater than 1.0. A mixed rating would be reflective of even odds in an OR analysis, where the CI eclipses 1.0. A positive rating was a protective OR, signified by an OR less than 1.0 with a CI whose upper bound was less than 1.0. With this general scoring system we tried to determine if a trend was present when there were two years of data present by examining the Breslow-Day statistic. In some cases we could not implement these criteria as noted in the table.

Overall Results on Criterion-Referenced Trends

Case	Positive	Mixed	Negative
Roosevelt-Edison	0 of 3	2 of 3	1 of 3
Henry E. S. Reeves	0 of 3	3 of 3	0 of 3
Dodge-Edison	1 of 3	2 of 3	0 of 3
Jardine-Edison	1 of 3	1 of 3	1 of 3
Boston Renaissance	0 of 6	3 of 6	3 of 6
Seven Hills Charter School	0 of 6	1 of 6	5 of 6
Dr. Martin Luther King Jr.	0 of 4	2 of 4	2 of 4
Mt. Clemens Sec. Academies	2 of 8	6 of 8	0 of 8
Mid-Michigan PSA	0 of 8	0 of 8	8 of 8
Washington Elementary	1 of 5	1 of 5	3 of 5
TOTALS	5 of 49	21 of 49	23 of 49

The main findings from the criterion-referenced analyses can be summarized as follows:

- We charted 49 criterion-referenced trends, which is nearly equal to the number of norm-referenced trends.
- Student performance on criterion-referenced tests often lags behind district performance and almost always behind state performance levels.
- In nearly half the trends, we found that students enrolled in Edison schools were making smaller gains on the criterion-referenced tests than comparison groups (i.e., 23 out of 49 trends).
- In 21 out of 49 trends, Edison students showed gains or changes in test results that were similar to the local districts and other comparison groups, and in only 5 of the 49 trends did we find Edison students making larger gains than comparison groups

Overall Findings

Given the varied designs and available data used in this and past evaluations, one can realize that we are far from having a true scientific experiment that can render a conclusive decision on whether or not students succeed in schools operated by Edison Schools Inc. Our results do not differ greatly from past evaluations by Miami-Dade or Minneapolis or those studies conducted by Howard Nelson at AFT (AFT, 1998, 2000). Statewide evaluations of charter schools that included schools operated by Edison also indicate that these schools are not better than other schools (Colorado Department of Education, 2000; Horn & Miron, 1999, 2000; and Renewal Inspection Report, 1999). While there are differences in the quality, scope, and rigor of these studies, it is important to note that the most rigorous studies (Gomez & Shay, 2000; Shay, 2000; and Minneapolis Public Schools, 2000) although limited in scope, all indicated that Edison students were not doing better than comparison groups of students. We are not alone in finding that students in schools operated by Edison—while they often start at levels below national norms and districts averages—progress at rates comparable to students in other district schools. This conclusion indicates that the expectations of district and charter school boards that contract with Edison as well as the expectations of parents who enroll their children in an Edison school are not being met. These groups believe that an apparent goal behind Edison’s school model is to have achievement gains that exceed the gains at comparable schools. Edison advertises this goal in promotional presentations, and the data in its annual reports suggest that it is fulfilling this goal.

Examining all 99 trends in one table, we see an equal proportion of trends from norm- and criterion-referenced analyses. It is important to reflect on our trend definitions. A positive trend reflects students in the Edison school evidencing statistically significant and meaningful annual gains on norm-referenced tests or outperforming the comparison groups on state criterion-referenced tests. A mixed trend is indicative of Edison students gaining at grade-level expectation on norm-referenced tests or performing at the same levels as comparison group students on criterion-referenced tests. A negative trend is evidenced by lower than grade level gains on norm-referenced tests or a lower passing percentage on criterion-referenced tests than the comparison group.

Combined Overall Trends Across All 10 Edison Schools in Our Study

	Positive	Mixed	Negative
Norm Referenced	10 of 50	40 of 50	0 of 50
Criterion Referenced	5 of 49	21 of 49	23 of 49
TOTALS	15 of 99	61 of 99	23 of 99

Our trends for both the norm-referenced tests and the criterion-referenced tests are based on gains in performance and not absolute performance levels. To summarize the main findings in this table, we include the following points:

- The majority of the trends, both norm- and criterion-referenced, were mixed, indicating that students in Edison schools are achieving at levels similar to students in the comparison groups.

- Edison students tend to show larger gains on norm-referenced tests than on state or district criterion-referenced tests. This is likely due to a combination of different factors, two of which are noted: (i) Edison's curriculum might not be adequately oriented to state standards in the various states in which it works; and (ii) the norm-referenced tests are administered to fewer students, and selectivity may come into play.

Comparison Between Edison's Findings and The Evaluation Center's Findings

Our findings suggest that Edison students do not perform as well as Edison claims in its annual reports on student performance. The next table presents a summary of our numerical ratings (5-point scale from strongly positive to strongly negative) for the 10 cases and the ratings Edison gave to these cases in its third annual report. Although this table is based on a number of broad generalizations, it provides a tentative overview of how our findings compare to the findings reported in Edison's annual performance reports.

Comparison of Results by Edison Schools and The Evaluation Center

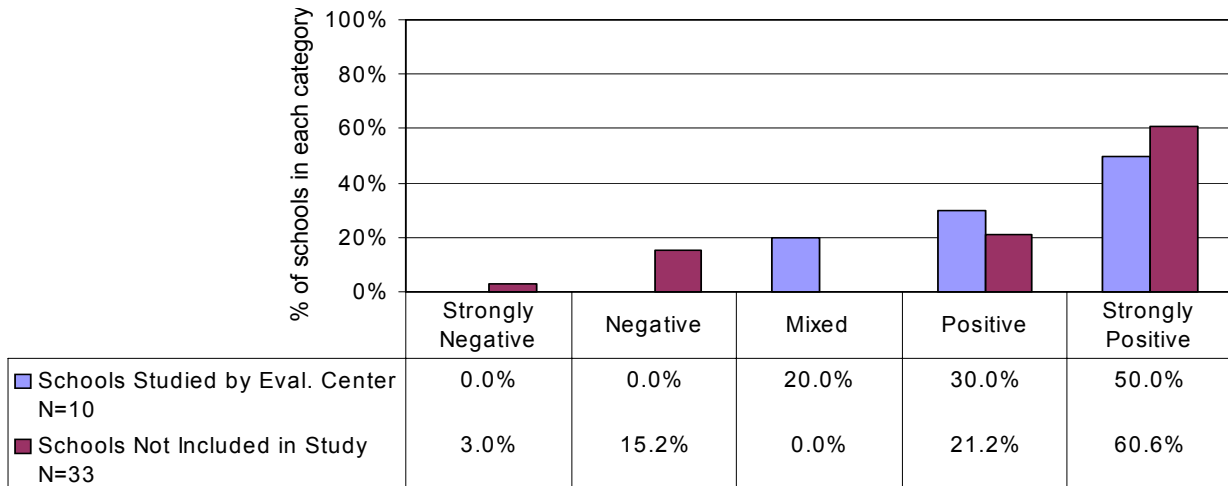
School	Edison' Schools Inc.				Evaluation Center	
	Edison Rating from 1999 Annual Report	Edison Rating in 2000 for 1999-00 Alone	Edison Rating in 2000 for All Years Since Opening	Numerical Rating	Rating for All Years Since Opening	Numerical Rating
Roosevelt-Edison Charter School	Strongly Positive	Negative	Positive	4	Mixed	3
Henry E.S. Reeves Elementary School	Positive	Strongly Positive	Strongly Positive	5	Mixed	3
Dodge-Edison Elementary	Positive	Strongly Positive	Strongly Positive	5	Positive	4
Jardine-Edison Junior Academy	Positive	Negative	Strongly Positive	5	Positive	4
Boston Renaissance Charter School	Mixed	Strongly Positive	Positive	4	Mixed	3
Seven Hills Charter School	Strongly Positive	Mixed	Mixed	3	Mixed	3
Dr. Martin Luther King Jr. Academy	Strongly Positive	Strongly Positive	Strongly Positive	5	Negative	2
Mt. Clemens Secondary Academies	Strongly Positive	Strongly Positive	Strongly Positive	5	Positive*	4
Mid-Michigan Public School Academy	Mixed	Positive	Mixed	3	Strongly Negative	1
Washington Elementary School	Strongly Positive	Positive	Positive	4	Negative	2

* Indicates very weak basis for rating the school

Average performance rating **4.30**

2.90

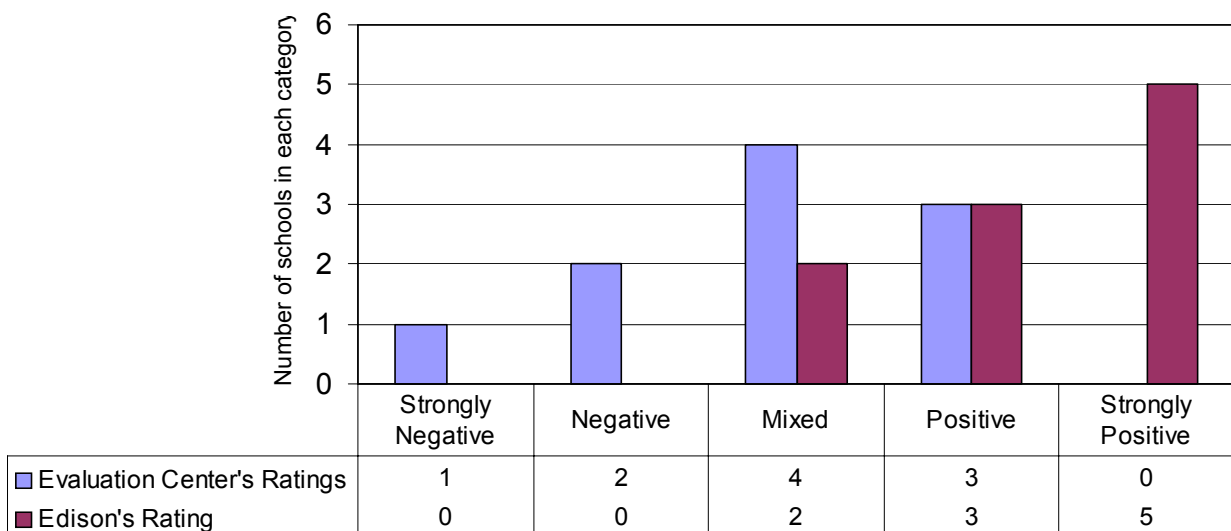
Figure A. Edison's Own Ratings of the 10 Sampled vs. Nonsampled Schools



In terms of “value-added” performance, over time we found that students at only three of the ten Edison schools were performing better than the comparison groups we examined (overall positive rating). On the other hand, there were three schools whose gains scores were less than those in the comparison groups (overall negative rating), and the remaining four schools showed mixed results.

There is a possibility of selection bias related to the schools included in this evaluation. However, when we analyzed the ratings Edison gave to the 10 schools in this evaluation relative to the 33 nonsampled schools with start-up dates after 1997, we found no significant difference in the ratings Edison published in its third annual report, $p = .7709$; see Figure A. There was, however, a statistically significant difference in the mean ratings we gave the 10 schools based on the findings in this study as compared with Edison's rating of these same 10 schools, $p < .001$; see Figure B.

Figure B. Ratings of the 10 Sampled Schools by Edison and The Evaluation Center



The differences in Edison’s and our own ratings for the 10 schools included in this study raises an important question. Since disparities exist between Edison’s and our own results for these 10 schools, might they also exist if we were to conduct an analysis of student achievement gains in its other schools?

Recommendations

Several different recommendations are described in detail in Chapter 13. We grouped them into two categories. The first group are general recommendations aimed at improving the general nature of reporting student achievement data in evaluations such as this one. The second group of recommendations is more directly aimed at individuals and groups that are considering contracting out the operation of its schools or already have existing contracts with companies such as Edison Schools Inc.

General Recommendations

There is obviously a large amount of data that a company such as Edison could share with participating schools, districts, and communities as well as with the research community. Listed below are some general recommendations to report student achievement more effectively.

- Prepare more comprehensive and complete annual reports.
- Define trends more clearly and how schools are rated, whether positive or negative.
- Follow the same trends from annual report to annual report.
- Report data across all years with consistent trends.
- Prepare academic reports as well as “lay” reports.
- Report all sample size information with analyses.
- Cite sources for data on schools, particularly when these differ from district and state data.

Specific Recommendations

It is especially important for district and charter school boards to ask questions and make requests in order to benefit from the collaboration with an outside company. Below, we have listed some relative suggestions to consider when contracting out schools. These recommendations are based on the findings presented in the report or reflect lessons learned during the course of the study.

- Require that all student achievement data will be available to the public/researchers.
- Require both external and internal evaluations.
- Obtain and analyze both norm-referenced and criterion-referenced tests
- Ensure that all students are included in the test results and require evaluators to report the number of students included in test results.

- Ensure that all evaluation findings are made available to right-to-know audiences.

Additional recommendations for district or charter school boards when contracting out educational services to EMOs can be found in Lin & Hassel (1999) and Miron (2000).

When seeking new contracts, Edison promises districts and charter school groups that its model is a successful one. In this report, we examined the question of whether Edison did more in terms of student performance on standardized achievement tests. We selected schools that have a long record to trace, and we looked at a variety of test results with its first 10 schools. While our findings do not suggest that Edison did less, they do not suggest that the company did more with these schools in terms of gains on standardized tests.

The differences in Edison's and our own ratings for the 10 schools included in this study raise an important question. Since disparities exist between Edison's and our own results for these 10 schools, might they also exist if we were to analyze student achievement gains in its other schools?

Policymakers and investors want clear and unquestionable results whereas social scientists have a tendency to weigh their findings with cautious interpretations. We attempted to present our findings in a clear and cohesive manner and have tried to alert readers about the many limitations that need to be considered. Nevertheless, we believe that because of the many years of test results examined in our study and because of the multiple sources of student achievement data and multiple comparison groups, this is the most in-depth and extensive evaluation of student achievement in Edison schools to date. We hope that the extensive documentation of our methodology contained in this report will serve those individuals and groups who are interested in looking beyond summary tables and charts and who wish to pursue systematic evaluation of alternative schools.

Table of Contents

Executive Summary	i
Table of Contents	xxix
Lists of Tables, Exhibits, and Appendices	xxxii
List of Acronyms	xxxvii
Acknowledgments	xxxviii
 Chapter 1 Introduction and Background to the Study	 1
1.1 Introduction	1
1.2 Description of Edison Schools Inc.	4
1.3 Description of Edison’s Evaluation and Assessment Activities	8
1.4 Overview of the Report	9
 Chapter 2 Methodology and Related Issues	 10
2.1 Aims and Objective of the Evaluation	10
2.2 Sources of Information	11
2.3 To Compare or Not to Compare	13
2.4 Statistical Methods Utilized	16
2.5 Criteria for Evaluating Trends and Summarizing Results for Each Case	22
2.6 Limitations of the Study	25
 Chapter 3 Roosevelt-Edison Charter School, Colorado Springs, CO	 29
3.1 Descriptive Summary of School	29
3.2 Past Studies and Evaluations	30
3.3 Longitudinal Analysis of Individual Student Data	31
3.4 Chi-Square Analysis of CSAP Data	35
3.5 Odds Ratio Analysis of the CSAP Data	37
3.6 Summary	44
 Chapter 4 Henry E. S. Reeves Elementary, Miami-Dade County, FL	 46
4.1 Descriptive Summary of School	46
4.2 Past Studies and Evaluations	49
4.3 Summary of Findings by the Miami-Dade County Public School District	49
4.4 Longitudinal Student Outcomes: The Shay Study	52
4.5 Performance on the FCAT and Florida Writes State Assessments	53
4.6 Chi-Square Analysis of FCAT Data	54
4.7 Odds Ratio Analysis of the FCAT Data	58
4.8 Summary	61

Chapter 5	Dodge-Edison Elementary, Wichita, KS	63
5.1	Descriptive Summary of School	63
5.2	Past Studies and Evaluations and Data Available for Analysis	64
5.3	Longitudinal Analysis of Individual Student Data	66
5.4	Comparison with District on the MAT-7	69
5.5	Comparison with District and State on Kansas Assessments	73
5.6	Summary	76
Chapter 6	Jardine-Edison Junior Academy, Wichita, KS	79
6.1	Descriptive Summary of School	79
6.2	Past Studies and Evaluations and Data Available for Analysis	80
6.3	Comparison with District on the MAT-7	80
6.4	Comparison with District and State on Kansas Assessments	84
6.5	Summary	87
Chapter 7	Boston Renaissance Charter School, Boston, MA	90
7.1	Descriptive Summary of School	90
7.2	Past Studies and Evaluations and Data Available for Analysis	91
7.3	Longitudinal Analysis of Individual Student Data	94
7.4	Chi-Square Analysis of MCAS Data	105
7.5	Odds Ratio Analysis of the MCAS Data	107
7.6	Summary	118
Chapter 8	Seven Hills Charter School, Worcester, MA	121
8.1	Descriptive Summary of School	121
8.2	Past Studies and Evaluations and Data Available for Analysis	122
8.3	Longitudinal Analysis of Individual Student Data	123
8.4	Chi-Square Analysis of MCAS Data	129
8.5	Odds Ratio Analysis of the MCAS Data	132
8.6	Summary	143
Chapter 9	Dr. Martin Luther King Jr. Academy, Mount Clemens, MI	147
9.1	Descriptive Summary of School	147
9.2	Past Studies and Evaluation and Data Available for Analysis	148
9.3	Longitudinal Analysis of Individual Student Data	149
9.4	Chi-Square Analysis of MEAP Data	153
9.5	Odds Ratio Analysis of the MEAP Data	156
9.6	Overall Performance on the MEAP Assessments	161
9.7	Summary	167

Chapter 10 Mt. Clemens Secondary Academies, Mt. Clemens, MI	170
10.1 Descriptive Summary of School	170
10.2 Data Available for Our Analyses	172
10.3 Chi-Square Analysis of MEAP Data	173
10.4 Odds Ratio Analysis of the MEAP Data	175
10.5 Overall Performance on the MEAP Assessments	179
10.6 Summary	186
Chapter 11 Mid-Michigan Public School Academy, Lansing, MI	188
11.1 Descriptive Summary of School	188
11.2 Past Studies and Evaluations and Currently Available Data	189
11.3 Longitudinal Analysis of Individual Student Data	190
11.4 Chi-Square Analysis of MEAP Data	194
11.5 Odds Ratio Analysis of the MEAP Data	197
11.6 Overall Performance on the MEAP Assessments	204
11.7 Summary	214
Chapter 12 Washington Elementary School, Sherman, TX	216
12.1 Descriptive Summary of School	216
12.2 Past Studies and Evaluations and Available Test Data	220
12.3 Analysis of Texas Assessment of Academic Skills Results	221
12.4 Comparable Growth on the TAAS Based on Consecutive Class Cohorts	227
12.5 Comparable Improvement Results Based on Individual Students' Gains on the Texas Learning Index	231
12.6 Chi-Square Analysis of TAAS Data	234
12.7 Odds Ratio Analysis of the TAAS Data	235
12.8 Summary	243
Chapter 13 Summary of the Findings	246
13.1 Edison's Findings	246
13.2 Evaluation Center Findings	247
13.3 Comparison Between Edison's Findings and The Evaluation Center's Findings ..	253
13.4 Recommendations Based on the Findings	255
References	257
Appendices	261

List of Tables, Exhibits, and Appendices

List of Tables

Table 2:1	Standardized Test Results from Edison Schools Included in the Study	12
Table 2:2	Construction of 2 x 2 Tables for Odds Ratio Analysis	20
Table 2:3	Types of Analyses Possible for Each School Included in the Study	22
Table 3:1	Table of Sample Sizes for Individual Student Data by Grade and Year	32
Table 3:2	Summary of Chi-Square Findings for Roosevelt-Edison Charter School	36
Table 3:3	Summary of Grade 3 CSAP Odds Ratio Findings for Roosevelt-Edison	38
Table 3:4	Summary of Grade 4 CSAP Odds Ratio Findings for Roosevelt-Edison	38
Table 3:5	Summary of Grade 5 CSAP Odds Ratio Findings for Roosevelt-Edison	43
Table 3:6	Summary of Results for Norm-Referenced Student Achievement Test	44
Table 3:7	Summary of Results on Criterion-Referenced Student Achievement Tests	45
Table 3:8	Combined Overall Trends for Roosevelt-Edison	45
Table 4:1	Distribution of Students by Ethnic Grouping (1997-98)	46
Table 4:2	School, Student and Teacher Background Information for Henry E.S. Reeves Elementary, Miami-Dade, and the State of Florida	48
Table 4:3	Descriptive Statistics for Reeves and Control Groups by Year	53
Table 4:4	Summary of Chi-Square Findings for Henry E. S. Reeves, 1999	58
Table 4:5	Summary of FCAT Odds Ratio Findings for Reeves Elementary School	59
Table 4:6	Summary of Florida Writes Odds Ratio Findings for Reeves Elementary School	59
Table 4:7	Summary Results on Norm-Referenced Tests	61
Table 4:8	Summary Results on Criterion-Referenced Tests	62
Table 4:9	Combined Overall Trends for Henry E.S. Reeves Elementary School	62
Table 5:1	Sample Sizes for Individual Student Data on the MAT-7 by Grade and Year	67
Table 5:2	National Percentile Ranks on MAT-7 Compared with the District	69
Table 5:3	Comparison of Gains in Terms of NPR Between Dodge-Edison and District	70
Table 5:4	Performance on Kansas Assessment Tests Compared with the District and State	73
Table 5:5	Summary of Results on Norm-Referenced Student Achievement Tests	76
Table 5:6	Summary Results on Criterion-Referenced Tests: Comparison of Gains on Kansas Assessments Between Dodge-Edison and Wichita Public Schools	77
Table 5:7	Combined Overall Trends for Dodge-Edison	77
Table 6:1	National Percentile Ranks on MAT-7 Compared with the District	81
Table 6:2	Performance on Kansas Assessment Tests Compared with the District, and State	85
Table 6:3	Comparison of Gains in Terms of NPR Between Jardine-Edison District	87
Table 6:4	Summary Results on Criterion-Referenced Tests: Comparison of Gains on Kansas Assessments Between Jardine-Edison and Wichita Public Schools	88
Table 6:5	Combined Overall Trends for Jardine-Edison	88
Table 7:1	Description of Longitudinal Cohorts	94
Table 7:2	Sample Sizes for Individual Student Data by Grade and Year on the SAT-9	95
Table 7:3	Summary of Chi-Square Findings for Boston Renaissance, Grade 4	106
Table 7:4	Summary of Chi-Square Findings for Boston Renaissance, Grade 8	107
Table 7:5	Summary of Results on Norm-Referenced Student Achievement Tests	118
Table 7:6	Summary Results on Criterion-Referenced Tests	119

Table 7:7	Combined Overall Trends for Boston Renaissance Charter School	120
Table 8:1	Table of Sample Sizes for Individual Student Data on the MAT-7 by Grade and Year ..	124
Table 8:2	Dependent t-Test Summary Findings for Cohort C and D	129
Table 8:3	Summary of Chi-Square Findings for Seven Hills Charter School, Grade 4	131
Table 8:4	Summary of Chi-Square Findings for Seven Hills Charter School, Grade 8	131
Table 8:5	Summary of Odds Ratio Findings for Seven Hills Charter School, Grade 4	133
Table 8:6	Summary of Odds Ratio Findings for Seven Hills Charter School, Grade 8	134
Table 8:6	Summary of Results on the Metropolitan Achievement Test (MAT-7)	144
Table 8:7	Summary of Results on the Stanford Achievement Test (SAT-9)	144
Table 8:8	Summary Results on MCAS Criterion-Referenced Tests	145
Table 8:9	Combined Overall Trends for Seven Hills Charter School	145
Table 9:1	Table of Sample Sizes for Individual Student Data by Grade and Year on the ITBS	150
Table 9:2	Summary of Chi-Square Findings for MLK, Grade 4	155
Table 9:3	Summaries of the Chi-Square Findings for MLK, Grade 5	156
Table 9:4	Summary of Results on Norm-Referenced Student Achievement Tests	167
Table 9:5	Summary Results on Criterion-Referenced Tests	168
Table 9:6	Combined Overall Trends for Martin Luther King Jr. Academy	169
Table 10:1	Descriptive Information for Mt. Clemens Secondary Academies and District Middle and High Schools	171
Table 10:2	Summary of Chi-Square Findings for Mt. Clemens Academies, Grade 7	174
Table 10:3	Summary of Chi-Square Findings for Mt. Clemens Academies, Grade 8	174
Table 10:4	Summary of Chi-Square Findings for Mt. Clemens Academies, HST	175
Table 10:5	Summary of Odds Ratio Findings for Mt. Clemens Academies, Grade 7	176
Table 10:6	Summary of Odds Ratio Findings for Mt. Clemens Academies, Grade 8	176
Table 10:7	Summary of Odds Ratio Findings for Mt. Clemens Academies, HST	179
Table 10:8	Summary Results on Criterion-Referenced Tests	186
Table 10:9	Overall Summary of Trends	187
Table 11:1	Sample Sizes for Individual Student Data by Grade and Year on the MAT-7	191
Table 11:2	Summary of Chi-Square Findings for Mid-Michigan Academy, Grade 4	195
Table 11:3	Summary of Chi-Square Findings for Mid-Michigan Academy, Grade 5	196
Table 11:4	Summary of Chi-Square Findings for Mid-Michigan Academy, Grade 7	196
Table 11:5	Summary of Chi-Square Findings for Mid-Michigan Academy, Grade 8	197
Table 11:6	Summary of Results on Norm-Referenced Student Achievement Tests	214
Table 11:7	Summary of Criterion-Referenced Test Results for Mid-Michigan	215
Table 11:8	Combined Overall Trends For Mid Michigan PSA	215
Table 12:1	Descriptive Information for Washington Elementary School (1996/97 - 1998/99)	217
Table 12:2	Trends in School Background Characteristics, Washington Elementary School	219
Table 12:3	Washington Elementary Performance on TAAS by Grade and Subject, 1994-1999	222
Table 12:4	Percent Hispanic Students Passing TAAS Compared with Exemptions Because of <i>LEP</i> ..	228
Table 12:5	Performance on TAAS for All Students Relative to Local District and State 1995-1999 ..	229
Table 12:6	Annual Gains/Loses on the Texas Learning Index Based on Matched Students	232
Table 12:7	Summary of Odds Ratio Findings for Washington Elementary School, Grade 3	236
Table 12:8	Summary of Odds Ratio Findings for Washington Elementary School, Grade 4	236
Table 12:9	Summary of Trends on the State Criterion-Referenced Test, TAAS	244
Table 12:10	Overall Achievement Trends, Washington Elementary School	244
Table 13:1	Approximate Attrition Rates in the Norm-Referenced Test Analysis	249
Table 13:2	Overall Norm-Referenced Trends	250
Table 13:3	Overall Criterion-Referenced Trends	251

Table 13:4 Combined Overall Trends Across All Ten Edison Schools in Our Study	252
Table 13:5 Comparison of Results by Edison Schools and The Evaluation Center	253

List of Exhibits

Exhibit 1:1 Growth of Edison Schools in Terms of Schools and Enrollments	5
Exhibit 3:1 Cohort A at Roosevelt-Edison Charter School: ITBS Results	33
Exhibit 3:2 Cohort B at Roosevelt-Edison Charter School: ITBS Results	34
Exhibit 3:3 Odds Ratio Results for Roosevelt-Edison Charter School	39
Exhibit 3:4 Performance on Grade 3 Reading for Roosevelt-Edison, District, and State	40
Exhibit 3:5 Performance on Grade 4 Reading for Roosevelt-Edison, District, and State	41
Exhibit 3:6 Performance on Grade 4 Writing for Roosevelt-Edison, District, and State	42
Exhibit 3:7 Performance on Grade 5 Math for Roosevelt-Edison, District, and State	43
Exhibit 4:1 Henry Reeves Elementary School, FCAT and Florida Writes Results 1997-1999 Performance on Grade 4 and 5 Tests	55
Exhibit 4:2 Henry E.S. Reeves Elementary School, FCAT and Florida Writes Results 1998-1999 Performance on Grade 4 and 5 Tests	56
Exhibit 4:3 Results of the Odds Ratio Analysis for Henry Reeves Elementary School	60
Exhibit 5:1 Dodge-Edison Results from the Analysis of Individual Students Results	68
Exhibit 5:2 Results on the MAT-7 Reading Between Dodge-Edison and District	71
Exhibit 5:3 Results on the MAT-7 Math Between Dodge-Edison and District	72
Exhibit 5:4 Results on the Kansas Assessments Between Dodge-Edison and District	75
Exhibit 6:1 Results on the MAT-7 Reading Between Jardine-Edison and District	82
Exhibit 6:2 Results on the MAT-7 Math Between Jardine-Edison and District	83
Exhibit 6:3 Results on the Kansas Assessments Between Jardine-Edison and District	86
Exhibit 7:1 Boston Renaissance Cohort A: Individual Student Results on the SAT-9	99
Exhibit 7:2 Boston Renaissance Cohort B: Individual Student Results on the SAT-9	100
Exhibit 7:3 Boston Renaissance Cohort C: Individual Student Results on the SAT-9	101
Exhibit 7:4 Boston Renaissance Cohort D: Individual Student Results on the SAT-9	102
Exhibit 7:5 Boston Renaissance Cohort E: Individual Student Results on the SAT-9	103
Exhibit 7:6 Boston Renaissance Cohort F: Individual Student Results on the SAT-9	104
Exhibit 7:7 Results of the Odds Ratio Analysis for Boston Renaissance (Grade 4)	109
Exhibit 7:8 Results of the Odds Ratio Analysis for Boston Renaissance (Grade 8)	110
Exhibit 7:9 Performance on Grade 4 English Language for Edison, District, and State	112
Exhibit 7:10 Performance on Grade 4 Math for Edison, District, and State	113
Exhibit 7:11 Performance on Grade 4 Science & Technology for Edison, District, and State	114
Exhibit 7:12 Performance on Grade 8 English Language for Edison, District, and State	115
Exhibit 7:13 Performance on Grade 8 Math for Edison, District, and State	116
Exhibit 7:14 Performance on Grade 8 Science & Tech for Edison, District, and State	117
Exhibit 7:15 Performance on Grade 8 History & Social Studies for Edison, District, and State	118
Exhibit 8:1 Seven Hills Charter School: Cohort A: Individual Student Results on the MAT-7	126
Exhibit 8:2 Seven Hills Charter School: Cohort B: Individual Student Results on the MAT-7	127
Exhibit 8:3 Results of Odds Ratio Analysis for Seven Hill Odds, Grade 4	135
Exhibit 8:4 Results of Odds Ratio Analysis for Seven Hill Odds, Grade 8	136
Exhibit 8:5 Performance on Grade 4 English Language for Edison, District, and State	137

Exhibit 8:6	Performance on Grade 4 Math for Edison, District, and State	138
Exhibit 8:7	Performance on Grade 4 Science and Technology for Edison, District, and State	139
Exhibit 8:8	Performance on Grade 8 English Language for Edison, District, and State	140
Exhibit 8:9	Performance on Grade 8 Math for Edison, District, and State	141
Exhibit 8:10	Performance on Grade 8 Science and Technology for Edison, District, and State	142
Exhibit 8:11	Performance on Grade 8 History and Social Studies for Edison, District, and State ...	143
Exhibit 9:1	MLK: Individual Student Results on the ITBS	152
Exhibit 9:2	Grade 4 Odds Ratio Results When Compared to the District and State	158
Exhibit 9:3	Grade 5 Odds Ratio Results When Compared to the District and State	159
Exhibit 9:4	Performance on Grade 4 Math for Edison, District and State	162
Exhibit 9:5	Performance on Grade 4 Reading for Edison, District and State	163
Exhibit 9:6	Performance on Grade 5 Science for Edison, District and State	164
Exhibit 9:7	Performance on Grade 5 Writing for Edison, District and State	165
Exhibit 9:8	Performance on Grade 5 Social Studies for Edison, District and State	166
Exhibit10:1	Results of the Grade 7 Odds Ratio Analysis for Mt. Clemens Secondary Academies ..	177
Exhibit10:2	Results of the Grade 8 Odds Ratio Analysis for Mt. Clemens Secondary Academies ..	178
Exhibit 10:3	Performance on Grade 7 Math for Edison, District and State	180
Exhibit 10:4	Performance on Grade 7 Reading for Edison, District and State	181
Exhibit 10:5	Performance on Grade 8 Science, Writing and Social Studies, 1998-99 and 1999-00 ..	182
Exhibit 10:6	Performance on Grade 8 Science, Writing and Social Studies, 1998-99	183
Exhibit 10:7	Performance on Grade 11 Math and Reading for Edison, District, and State	184
Exhibit 10:8	Performance on Grade 11 Science and Writing for Edison, District, and State	185
Exhibit 11:1	Mid Michigan PSA: Analysis of Individual Student Results on the MAT-7	193
Exhibit 11:2	Results of the Odds Ratio Analysis Grade 4 Reading and Math	198
Exhibit 11:3	Results of the Odds Ratio Analysis Grade 7 Reading and Math	199
Exhibit 11:4	Results of the Odds Ratio Analysis Grade 5 Science and Writing	200
Exhibit 11:5	Results of the Odds Ratio Analysis Grade 8 Science and Writing	201
Exhibit 11:6	Grade 4 Reading MEAP Results for Mid Michigan, District, and State	205
Exhibit 11:7	Grade 7 Reading MEAP Results for Mid Michigan, District, and State	206
Exhibit 11:8	Grade 4 Math MEAP Results for Mid Michigan, District, and State	207
Exhibit 11:9	Grade 7 Math MEAP Results for Mid Michigan, District, and State	208
Exhibit 11:10	Grade 5 Science MEAP Results for Mid Michigan, District, and State	209
Exhibit 11:11	Grade 8 Science MEAP Results for Mid Michigan, District, and State	210
Exhibit 11:12	Grade 5 Writing MEAP Results for Mid Michigan, District, and State	211
Exhibit 11:13	Grade 8 Writing MEAP Results for Mid Michigan, District, and State	212
Exhibit 11:14	Grades 5 & 8 Social Studies: MEAP Results for Mid Michigan, District, and State ...	213
Exhibit 12:1	Washington Elementary School TAAS Results 1994-1999 Grade 3 Comparisons Reading and Math	224
Exhibit 12:2	Washington Elementary School TAAS Results 1994-1999 Grade 4 Comparisons Reading and Math	225
Exhibit 12:3	Washington Elementary School TAAS Results 1994-1999 Grade 4 Comparisons Writing and All Tests	226
Exhibit 12:4	Average Gains/Losses on the Texas Learning Index, Washington Elementary School	233
Exhibit 12:5	Results of the Odds Ratio Analysis for Washington Elementary Grade 3 vs. Campus ..	237
Exhibit 12:6	Results of the Odds Ratio Analysis for Washington Elementary Grade 3 vs. District ..	238
Exhibit 12:7	Results of the Odds Ratio Analysis for Washington Elementary Grade 3 vs. State	239
Exhibit 12:8	Results of the Odds Ratio Analysis for Washington Elementary Grade 4 vs. Campus ..	240
Exhibit 12:9	Results of the Odds Ratio Analysis for Washington Elementary Grade 4 vs. District ..	241

Exhibit 12:10 Results of the Odds Ratio Analysis for Washington Elementary Grade 4 vs. State	242
Exhibit 13:1 Edison's Own Ratings of the Sampled Schools vs. Nonsampled Schools	254
Exhibit 13:2 Edison's Own Ratings As Compared to The Evaluation Center Ratings For the Ten Schools Considered in This Study	255

List of Appendices

Appendix A Description of Norm- and Criterion-Referenced Tests Considered in the Study . . .	261
Appendix B Normal Curve Equivalents Based on National Percentile Ranks	267
Appendix C Criteria for Rating Student Achievement Trends Presented by Edison and AFT . .	268
Appendix D Results Provided by Edison for Dillingham Intermediate School, Sherman, TX . .	270
Appendix E Chi-Square Results for All Schools	271

List of Acronyms

AEIS	Academic Excellence Indicator System	MAT-7	Metropolitan Achievement Test, 7 th Edition
AFT	American Federation of Teachers	MCAS	Massachusetts Comprehensive Assessment System
ANOVA	Analysis of variance	MDCPS	Miami-Dade County Public Schools
BRCS	Boston Renaissance Charter School	MDE	Michigan Department of Education
B-D	Breslow-Day statistic	MDOE	Massachusetts Department of Education
CDE	Colorado Department of Education	MEAP	Michigan Educational Assessment Program
CI	Confidence interval	MLK	Dr. Martin Luther King Jr.
CRT	Criterion-Referenced Test	NEA	National Education Association
CSAP	Colorado Student Assessment Program	NCE	Normal curve equivalent
CTBS	Comprehensive Test of Basic Skills	NPR	National Percentile Rank
DALT	District Achievement Level Tests	NRT	Norm-Referenced Test
DEPS	Dodge-Edison Partnership School	OR	Odds ratio
EMO	Educational management organization	PR	Percentile rank
EC	The Evaluation Center, WMU	PSA	Public School Academy
ES	Effect Size	SAT-9	Stanford Achievement Test, Version 9
ESL	English as a second language	SES	Socioeconomic status
FCAT	Florida Comprehensive Assessment Test	SFA	Success for All
FDE	Florida Department of Education	SPSS	Statistical Package for the Social Sciences
GE	Grade equivalent	SS	Standard score
HLM	Hierarchical linear model	Stn	stanine scores
HST	High School Test	TAAS	Texas Assessment of Academic Skills
ISD	Intermediate School District	TEA	Texas Education Agency
ITBS	Iowa Test of Basic Skills	TLI	Texas Learning Index
KDE	Kansas State Department of Education	WMU	Western Michigan University
LEP	Limited English proficiency		

Acknowledgments

This was a complex study, and it could not have been completed without the cooperation and commitment of a number of individuals. First and foremost, we want to thank the National Education Association for providing some of the funds for this important study. At the NEA, Heidi Steffens and Phil McLaurin, from the Office for Public Education Advocacy, provided oversight and guidance in the conduct of the study. Anthony Rolle, formerly a statistician at NEA, also contributed ideas and comments on the statistical methods employed. We also wish to thank Edison Schools Inc. for sharing with us individual data sets for 6 of the 10 schools in our study. John Chubb, Chief Education Officer at Edison, exchanged thoughts and ideas with us during various phases of the work; and he and Christopher Cerf, the Chief Operating Officer at Edison, provided us with general comments on draft chapters of the report.

We would also like to thank the various school districts that provided documentation and data and in many cases also reviewed and commented on draft cases. Likewise, the study benefited from a number of statisticians and researchers who provided suggestions on methods and design as well as feedback on some of the chapters.

We were fortunate to have a talented evaluation team at our side and a number of research assistants who helped us with a wide variety of tasks. Carla Howe, Deborah Lehmann, and Kim Reynolds were instrumental in helping us organize and summarize data from numerous annual school reports, Web sites, and a variety of district and state documents. They also assisted by reviewing drafts of many of the chapters. In addition, student assistants Santos Kalaskar and Archana Prakash provided data entry support. We were also fortunate to have expert assistance from Christopher Nelson, who reviewed and provided comments on some of the chapters, and Sally Veeder, who helped with editing. Finally, we express our appreciation to Daniel Stufflebeam, the director of The Evaluation Center, and Arlen Gullickson, Chief of Staff, who provided guidance in the overall organization and design of the evaluation.

From this long list of names and contributions, one can clearly see that this was a team effort. We are confident that the evaluation has benefited from the expertise and diverse perspectives of the various contributors and evaluation team members. While we recognize and express our appreciation for the contributions made by these many persons, we are mindful that we are responsible for the content of the report, including errata.

Gary Miron and Brooks Applegate
December 20, 2000

Chapter One

Introduction and Background to the Study

1.1 Introduction

This report summarizes an effort to determine whether or not the Edison model is working in terms of improving student performance on standardized tests. Some attempts have already been made to evaluate overall student achievement at Edison schools. Edison itself has prepared three annual reports on student achievement at its schools (Edison, 1997, 1999, 2000) and has presented general information at conferences, workshops, and in the media regarding the gains made at its schools. The findings contained in Edison's annual reports and the message it spreads in conference presentations and in the media indicate the company is quite successful and the students enrolled in its schools are making large and substantial achievement gains.

A few evaluations have focused on one or two schools within a given district (Miami Dade, 1998, 1999, 2000; Minneapolis Public Schools, 2000; Wichita State University, 1996) or within a given subject area (see the case studies conducted by Dr. Robert Mislavy between 1996 and 1999). The results from these studies generally have been mixed. In 1998, Dr. Howard Nelson, from the American Federation of Teachers (AFT), prepared a more comprehensive evaluation across most of the then operating Edison schools (AFT, 1998). This study relied more heavily on available state test data, and it provided more information on school context than do the Edison reports. The AFT study, while acknowledging that some of the Edison schools were succeeding in raising student test scores in comparison with control groups, found that *overall* Edison was not demonstrating gains across its schools. Following the release of the AFT study, Edison employed a Harvard scholar, Dr. Paul Peterson, to reexamine and compare the analyses and results in Edison's own annual reports with the AFT report. While not addressing many of AFT's more critical questions about Edison assessment and reporting activities, Peterson (1998) developed an analytical framework to compare the study designs and found that the designs employed by Edison in its 1997 annual report were stronger than those behind many of the negative or flat trends reported in the 1998 AFT report.

In October 2000, the AFT released a second report that provided an overview of the performance of schools operated by Edison. This report relied largely on data from state-mandated criterion-referenced tests. This report had less narrative and allowed the numbers to talk. While this overview still lacked the individual norm-referenced data that Edison also reports in its annual reports, it did include relevant comparison groups, and the source of the data was clearly marked so others could check the figures for themselves. The overall finding from this study was that "Edison schools mostly do as well or worse than comparable schools; occasionally they do better" (AFT, 2000, p. 6).

This finding can be contrasted with the summary of the findings from Edison's third annual report that was released in September 2000: "For the 1999-2000 school year, the average gain of Edison students, in the core areas of reading, language arts, spelling, writing, and mathematics was 5 percentiles on nationally normed tests and 7 percentage points on criterion-referenced tests, which also include science and social studies. These gains represent improvements of one point in each case over the gains reported for 1995-99, and are the highest gains reported by Edison to date" (Edison, 2000, p. 2).

There are clearly differences in the nature and vested interests of the organizations conducting past studies on Edison. Thus, it is no surprise that they have presented contrasting and contradictory results. Results questioning Edison's findings have induced strong and critical rebuttals that have—in some cases—been met with equally strong and critical rejoinders. Given this context, we are aware that any evaluation of Edison Schools Inc. is going to be carefully scrutinized. This is, in many respects, a "high stakes" evaluation and regardless of whether the results are positive or negative, the findings are almost certain to be contested by one group or another. Anticipating this, we employed a number of strategies that we believe have strengthened our analyses as well as our role as an independent evaluator. To insure our independence from the agency funding the study, the National Education Association (NEA)⁴, and the organization being evaluated (Edison Schools Inc.), we have done the following:

- Refused to pursue a larger study including data collection in Edison schools that would require a confidentiality agreement with Edison, which would then control access and release of the results
- Insisted that our contract with NEA allows us to publish the findings, regardless of the outcome

In order to strengthen our analyses and the overall credibility of our work, we took the following steps:

- We provided the NEA, Edison Schools Inc., several of the school districts that have Edison schools within its boundaries, and researchers in the field with copies of the draft report and welcomed comments and corrections, which we considered in finalizing the report..
- We sought the advice of experts in the field regarding the strategies and methods we chose to work with state achievement data.
- We sought a variety of data sources and made comparisons with multiple groups whenever possible.
- We clearly describe the assumptions behind our analyses.

⁴ NEA officials and researchers at The Evaluation Center began discussing the possibility of conducting an evaluation of Edison schools in the spring of 1998. While NEA currently has a policy to not work with for-profit operators, a number of NEA local affiliates have decided to work with Edison. NEA's interest in an evaluation was to inform decisions made within its national organization. In all our planning meetings with representatives from NEA, they have made it clear that they wish to have a rigorous and completely independent evaluation.

- We established and explain the criteria we used to distinguish whether trends were positive, mixed, or negative. We also developed and applied a common method to calculate a mean score across the trends in order to label each school on a 5-point scale—as Edison does—ranging from Strongly Negative to Strongly Positive.
- We attempted to report all data and present all charts and figures clearly and consistently.
- We clearly described the limitations of the study.
- We relied on the program evaluation standards (Joint Committee on Standards for Educational Evaluation, 1994) to guide our work.

The pursuit/conduct of the study was complicated by a number of factors (for example, late receipt of data files from Edison for a limited number of schools and changes in evaluation team members). The study is also marked by a number of limitations (see Section 2.6). Nevertheless, we are satisfied with the overall results and think this evaluation makes a number of important contributions to the debate over student achievement in schools operated by Edison since it provides an in-depth discussion and summary of related research on Edison Schools Inc. Also, because of the many years of test results available for our study and because of the use of multiple sources of student achievement data and multiple control groups, we believe this is the most in-depth and extensive assessment of student achievement in Edison schools to date.

To the extent possible, we have tried to replicate the presentation of findings that Edison includes in its annual report. This includes a presentation of trends on various norm-referenced or criterion-referenced tests and then a rating given to the school that reflects a summary of all the trends. When we could identify Edison’s own methods and techniques for calculating trends and rating schools, we applied them. When we had to develop our own methods, we explained them as clearly as possible. There are some important differences between Edison’s presentation of school results in its annual report and our own.

- Edison includes only a brief summary of each school, with a table of general data and a page of charts outlining the achievement trends. By contrast, we developed in-depth cases ranging from 10-20 pages in length.
- Edison’s definition of a trend is different than our own. It appears that Edison’s approach to defining and counting trends results in a larger number of trends than our approach. We rate overall average annual change in a trend, while Edison counts any difference between a two-year block of time as a trend (we explain this in detail in Chapter 2).
- Edison states that it does not intend to compare change in its schools with others (in its annual report there are a number of exceptions to this), while our analysis of gains is dependent on comparisons (see Section 2.3 “To Compare or Not to Compare” for more details).
- Edison reports reviewed contained no information on the number of students (N) reflected in the trends it presents, and the source of data is often unclear. By contrast, we made a conscious effort to report the N for all trends and to identify the source of data so that others can check these sources and replicate our work if they wish.

- Edison’s annual reports appeared to us to be intended for a lay audience and provide little insight on the methodology behind the findings they present. Our report is a technical report that provides extensive detail on the methodology, but with an executive summary that presents the results for a lay audience.
- While Edison attempts to present the findings across all its schools, we focused only on those schools that Edison opened in 1995 and 1996 since they have the most data available and because Edison has had time to implement its program more fully in these schools.

Many of the differences between Edison’s reporting and our own are dealt with and explained more thoroughly in other parts of this report.

In terms of methodology, the study combines a variety of statistical methods to compare growth in student achievement in Edison schools with growth in control schools, districts, statewide achievement data, and with national norms, when available. The use of odds ratio analysis to examine criterion-referenced achievement data is clearly a new contribution in terms of methodology; and given the extensive availability of statewide assessment data and the current limitations in the analyses, this may prove to be a significant contribution (Chapter 2 contains more details about odds ratio analysis).

Recent years have seen a rapid growth in the number of for-profit educational management organizations (EMOs) starting and operating charter schools and competing for contracts to take over operation of poor performing district schools (Miron, 2000). The results from this study contribute to the still limited body of research on the impact and performance of for-profit EMOs. Specific to Edison, many people are asking such key questions as “Can this company be profitable?” and “Has Edison been able to implement its model in the schools it operates?” Questions such as these, while important, are not addressed in this study. Rather, this study focuses on the performance of Edison students on standardized tests.

1.2 Description of Edison Schools Inc.

Edison identifies three major goals for its schools: (1) move students forward, (2) operate schools for less money, and (3) provide more services. In many respects, Edison is supporting a national school reform model that emphasizes a “comprehensive and fully integrated school reform.” Edison’s school model includes the following:

- A longer school day and longer school year
- Organization of schools based on academies, houses, and teams
- A rich liberal arts curriculum
- Research-based instruction closely aligned with assessment
- Four to six weeks of training for staff prior to the opening of a new school
- Daily professional development and planning time for teachers

- Career ladders for teachers
- A high level of parent and community involvement
- Computers and training provided to teachers, students, and families
- A national system of educational, operational, and financial supports

Growth of Edison Schools, 1995-2000

Although the Edison Project was established in 1992, its first 4 schools didn't open until the 1995/96 school year. Edison quickly became the largest private educational management organization operating public schools in the U.S. The company started with 2,249 students in 4 schools in the 1995-96 school year. As of the 2000-01 school year, Edison reported that it operates 113 schools in 21 states and the District of Columbia, with a combined student enrollment of approximately 57,000 students. Exhibit 1:1 illustrates the growth of the company in terms of schools and enrolled students. Edison considers itself to be the first "national system of public schools" (Edison, 1999, p. 2).

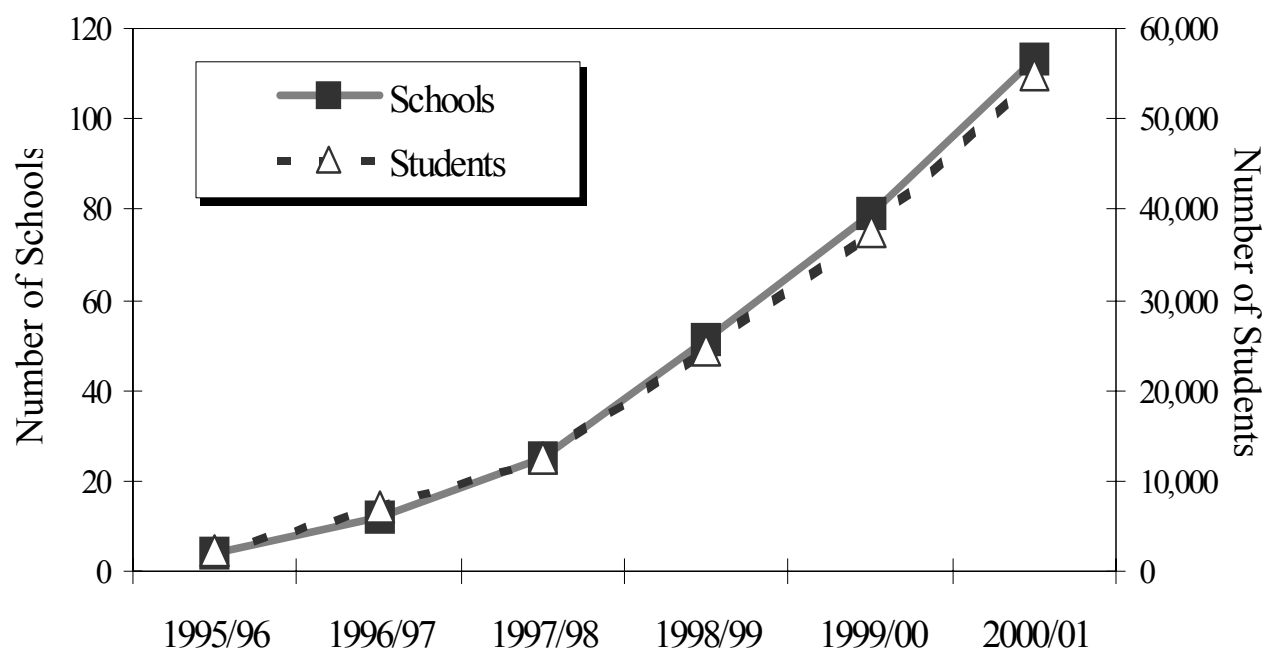


Exhibit 1:1 Growth of Edison Schools in Terms of Schools and Enrollments

Plans for the future assume more expansive growth. For the 2000-01 school year, Edison announced new contracts that include a very large commitment in Dallas, Texas, and Inkster, Michigan. Also, it has renewed its first contracts in Wichita, Kansas; Duluth, Minnesota; Boston, Massachusetts; and Mt. Clemens, Michigan. At the same time, Edison's contract with the Sherman Independent School District (Texas) to operate two schools was not renewed.

Approximately half of the schools Edison operates are charter schools, while the remaining are schools run under direct contract, or partnership, with local school districts. The charter schools have its own school boards and a charter granted by an authorizing agency. Some of the charter schools that Edison operates are new start-up schools; others are conversions of regular public schools. The contract schools, on the other hand, are operated under management contracts with public school districts. The contract schools are not new start-ups, since Edison takes over an existing building and school. While Edison can select teachers, it is assumed that the school will at least cater to the same catchment area so many of the students remain the same. Most of Edison's contracts with public schools are for five years.

Organization of Edison Schools

The names and terms used in describing the organization of Edison schools are unique and can be confusing to outsiders. Even the meaning of "school" becomes complex, since Edison defines a school differently than many states or districts. While Edison reports operating 79 schools in 1999-2000, the actual number of schools appears to be 60 according to how most states define a school. Edison sometimes considers new academies at the same school as new schools (for example when adding an junior academy which serves Grades 6-8 to a school that caters to elementary students), even though the district and state would still consider this as one school entity. In state and district records, only one school and building code are typically assigned for all the academies found in the same building. Each charter school has a common charter and a governing board, even though several academies may be contained in the school, or may occupy two or more separate buildings. In some cases, Edison considers what the state or district defines as one charter school as two or more schools, each containing one or more academies. In its annual reports, Edison sometimes reports data for all the schools at the same site, thus further complicating what it defines as a single school.

In our study, we included 10 of the 11 schools opened during Edison's first 2 years of school operation (i.e., 1995/96 and 1996/97 school years). In one of the Edison brochures, which contains a list of its schools, these 10 schools were referred to as 14 distinct schools. The 10 schools included in our study encompass a total of 23 academies.

Even though the actual number of schools that Edison operates can be disputed, the size of the schools' enrollments is very clear. Edison's schools are very large compared with traditional public schools. Many contain 2 or 3 academies, and the average size of an academy is around 475 pupils. The average Edison academy is nearly as large as the average public elementary school, which contained 481 pupils in 1996 (Digest of Education Statistics, 1997, Table 97).

Edison schools are organized by grouping grade levels into academies. Each academy may be considered a “school within a school” with its own administrative head, who reports to the principal of the school. A K-5 school, for example, would consist of a Primary Academy for grades K-2 and an Elementary Academy for grades 3-5. As additional grades are added to a school, they are organized into a Junior Academy for grades 6-8, a Senior Academy for grades 9-10, and a Collegiate Academy for grades 11-12. Edison’s contracts for charter schools are typically for K–12 schools, although they usually open as K–5 schools consisting of a Primary and an Elementary Academy, then expand by adding one additional grade or sometimes an additional academy each year.

Within the academies, the students are organized into multigrade houses of 100-180 students. The students in each house are taught by the same team of teachers throughout the time they are in that academy.

Edison’s national system of schools is incorporated in Delaware, but headquartered in New York City.

Curriculum design and instructional methods

Edison Schools Inc. has a curriculum that includes reading, math, history/social studies, science, writing, and world language as the core subjects. In addition, classes in character and ethics, physical fitness and health, music, dance, visual art, drama, and practical arts and skills are offered at various levels. Four methodological approaches to instruction are reportedly used in the classrooms: *project-based learning*, *direct instruction*, *cooperative learning*, and *differentiated learning*.

Project-based learning refers to short-term or long-term projects carried out by individuals, small groups, or a whole class. These projects are intended to increase motivation and provide for authentic learning and concrete problem solving.

Direct instruction is when explicit teaching is necessary and provides a structured approach to learning.

Cooperative learning is used extensively at Edison schools as part of its adapted Success for All reading program. Edison reports that all of its teachers are trained in the cooperative learning/instructional model developed by Dr. Robert Slavin at Johns Hopkins University.

Differentiated learning encompasses special education, English as a second language (ESL), gifted/talented programs, and remedial programs for students having academic difficulty. Special education services are provided primarily in the regular education classroom with special education consulting services provided indirectly to the classroom teacher. Although direct special education teacher services to students are available through both pull-asides and self-contained programs, the emphasis is on inclusion with indirect services. English as a second language (ESL) services are provided to students with limited English proficiency during times when most students are in Spanish instruction. Information in Edison’s brochures and web sites did not indicate whether specially trained teachers were hired to provide this instruction. No specialized programs are provided for gifted/talented students unless required by state or local regulations. Edison says that its program design provides many opportunities for student growth at all levels of learning abilities

and emphasizes that students may move on to the next academy as soon as they reach mastery at their current level.

This study does not examine the extent to which the Edison model has been implemented, nor Edison's level of commitment in terms of financial and human resources. Two district studies considered the extent that Edison has implemented its model in the schools it operates: Gomez and Shay (1998,1999) and Wichita State University (1996).

1.3 Description of Edison's Evaluation and Assessment Activities

One of the ten fundamentals of the Edison design recognizes that assessment must provide the foundation for accountability. In its effort to develop accountability, Edison has developed four main evaluation/assessment areas: (i) state and district tests; (ii) benchmark assessments; (iii) structured portfolios; and (iv) quarterly learning contracts. According to Edison the current record of student achievement is as follows:

Edison claims that student achievement in its schools has been heading steadily upward. In its second annual report (Edison, 1999) the company lists the following summary statements about the achievement levels of its students/schools:

- "The overwhelming majority of achievement trends—now numbering nearly 200—are positive.
- On average, students are gaining more than 5 percentiles per year against state and national standards.
- Students in nearly every Edison school are achieving more today than when the school opened; in no Edison school are students achieving less.
- Fourteen of seventeen Edison schools that have now established achievement trends have moved student achievement forward."

Structured portfolios, maintained electronically, act as the base of the Edison assessment system. These portfolios serve as a collection of student work over time and are intended to be a reliable and valid tool that exemplifies student progress. There are two main parts to this portfolio: (i) Common Performance Assessments, and (ii) Elective Elements. The Common Performance Assessments vary in time and length spanning from one day to one week and are scored by teachers using "Edison's Standards," common scoring guides and rubrics. Each quarter students and teachers work together to select one item from each main subject area to put in their portfolio. Edison reportedly provides suggestions for items to place in the portfolio, but teachers may use their discretion in the decision making. Any tangible that is chosen, a piece of art work or printed stories, is scanned into the portfolio.

Finally, the Quarterly Learning Contract involves a teacher adviser, the student, and student's family. It is described as a narrative report card that reports the level of work being done by the student and the quality of the student's performance. The adviser, student, and family members work to create

one goal for the student for the next quarter and identify how the goal will be attained. The Learning Contract is also maintained electronically. Using technology to maintain student assessment records allows teachers to modify and maintain the records with ease.

Assessing school performance

Edison has established performance standards in specific areas expected of schools during each year of implementation. Schools are expected to advance one level in each standard every year. Approximately 40 sets of performance standards are used to assess the implementation progress.

The Edison design includes four stages of schools implementation: Beginning–1st year; Developing–2nd year; Proficient–3rd year; and Exemplary–4th year. According to Edison’s second annual report (1999), most of its schools have progressed as expected, although three schools were noted as not reaching expected levels on performance standards: Henry Reeves in Miami during Year 1; Washington Elementary in Sherman, Texas, during years 1 and 2; and Boston Renaissance in years 2 and 3. It is interesting to note that in Edison’s annual report, Reeves received a “positive” rating in terms of trends for academic performance of students on standardized tests, Washington received a “strongly positive” rating, and Boston Renaissance received a “mixed” rating. All three of these schools are considered in this study.

In the summer of 2000, Edison announced that it would have RAND Corporation conduct an external evaluation of its schools, including an examination of student progress on standardized tests. Further information about Edison Schools Inc. can be obtained from its headquarters in New York or from its web site <<http://www.edisonschools.com>>.

1.4 Overview of the Report

This report is organized into three distinct parts. The first part, including chapters 1 and 2, provides a framework for the study. Chapter 2 describes the objectives of the study, sources of data and information, and a lengthy description of the methodology. This chapter also contains a discussion of why comparisons need to be made in order to derive meaning from the gains made in schools operated by Edison. Finally, Chapter 2 also contains a list and description of the limitations to the study.

Chapters 3 to 12 contain in-depth case studies of each of the schools included in our study. The order of the cases is arranged alphabetically by state and then by school name. Therefore, schools from the same state that have similar data will be located next to each other. Each case is organized to include a descriptive summary of the school, a review of previous research on that school, a presentation of findings from both norm-referenced test data and criterion-referenced test data, and a summary of the trends.

Chapter 13 summarizes and describes the findings across all the cases and compares them with Edison’s findings. This closing chapter also includes recommendations and suggestions for further research.

Chapter Two

Methodology and Related Issues

2.1 Aims and Objectives of the Evaluation

This study examined 10 schools that have been operated by Edison Schools Inc. for at least 4 years. It was our intention to include all 11 schools that opened during the first two years that Edison was operating schools (1995-96 and 1996-97). However, one school was left out of our analyses, Dillingham Intermediate School in Sherman, TX, because we were unable to secure any independently verifiable student achievement data.⁵ The rationale for selecting schools that have been operated by Edison for 4 or 5 years is that we believe these schools, rather than those open for 3 years or less, provide a more convincing picture of the impact the Edison model can have on student achievement. While some of these schools may have had more difficult start-ups than others, they have been in operation beyond what many consider the start-up phase.

The overriding aim of this study was to examine the impact of Edison schools on student learning as measured by norm- or criterion-referenced tests. In order to achieve this objective, a number of specific tasks were identified:

1. Review and critically assess existing research and evaluations on the impact of Edison schools.
2. Describe the evaluation measures used by Edison Schools Inc.
3. Describe the nature and quality of the standardized test results available.
4. Compare Edison schools in terms of student achievement over time.
5. Compare Edison schools with state and national norms on standardized tests.

⁵ More specifically, this was a school within a school, and the available data from the state and district were not disaggregated so that we could separate the results for the Edison half of the school. Edison did provide us some of this disaggregated data, but the district claimed it did not have this data and could not confirm its validity. These data were also limited in that they did not cover all the years when data should have been available and did not include information on the number of test takers in each group. Appendix D includes a summary of the data that Edison provided to us.

6. Compare Edison schools with local school district and state performance levels and—where possible—other similar comparison schools.
7. Develop cases for each of the 10 Edison schools that include (i) a description of the school based upon available literature and documentation, (ii) findings from analysis of norm-referenced and criterion-referenced test results, and (iii) a summary of the diverse results from the analyses of test results.
8. Based upon available literature and documentation, develop a framework for analyzing the 10 cases.
9. Analyze the case studies according to the framework, and summarize the results of this analysis.

2.2 Sources of Information

Student achievement data were obtained from a variety of sources. Table 2:1 lists the standardized tests in which each participating school took part. The column with district and state mandated tests include all the criterion-referenced tests (CRT), and the last three columns in the table include the three norm-referenced tests that are being used in the ten schools included in this study.

From districts and state education agencies we were able to obtain results on the criterion-referenced tests for all ten schools, as well as for the local district, the state, and control schools, where applicable. We secured data for all years these tests were administered, with the exception of Mt. Clemens Secondary Academies. Because the Mt. Clemens Junior and Senior Academies share their building with the only other middle school in the district, their test results were reported as one building/school. Data provided by Edison helped us to disaggregate the data between the two schools. The analysis conducted on the CRT took place in the spring of 2000. New data that became available after this point in time were also reported in the case studies, but in only a few cases did we rerun the chi-square and odds ratio analyses to include the new test results.

Data sets containing individual student results on the norm-referenced tests were made available to us by Edison Schools Inc. We received seven such data sets. Three of these data sets did not contain results for all possible years, which limited the length of some of the longitudinal analyses. We did not receive any norm-referenced test data for four schools. For one school, we were provided with two data sets from different norm-referenced tests. While Edison didn't inform us as to the reasons for its restricted release of data, it is possible that some districts, or the company, did not wish to share such data. The data sets contained individual student data with anonymous indicators, which enabled the tracing of individual student results over time.

Table 2:1 Standardized Test Results from Edison Schools Included in the Study

SCHOOL NAME	State- or District-Mandated Criterion-Referenced Tests	Stanford Achievement Test (SAT-9)	Metropolitan Assessment Test (MAT-7)	Iowa Test of Basic Skills (ITBS)
Roosevelt-Edison Charter School Colorado Springs, CO (1996)	DALT (district test) 1996/97, 97/98, 98/99 CSAP (state test) 1996/97, 97/98, 98/99, 1999/00			1996/97 1997/98 1998/99
Henry E.S. Reeves Elementary Miami, FL (1996)	Florida Writing Assessment 1996/97, 1997/98, 1998/99, 1999/00 Florida Comprehensive Assessment Test 1997/98, 1998/99, 1999/00	1996/97 1997/98 1998/99		
Dodge-Edison Elementary Wichita, KS (1995)	Kansas Reading, Math, and Writing Assessments 1995/96, 1996/97, 1997/98, 1998/99, 1999/00		1995/96 1996/97 1997/98 1998/99	
Jardine-Edison Junior Academy Wichita, KS (1996)	Kansas Reading, Math, and Writing Assessments 1996/97, 1997/98, 1998/99, 1999/00		1996/97 1997/98 1998/99	
Boston Renaissance Charter School Boston, MA (1995)	Massachusetts Comprehensive Assessment System (MCAS) 1997/98, 1998/99	1996/97 1997/98 1998/99	1995/96 1996/97	
Seven Hills Charter School Worcester, MA (1996)	Massachusetts Comprehensive Assessment System (MCAS) 1997/98, 1998/99	1996/97 1997/98 1998/99	1996/97 1997/98 1998/99	1996/97 1997/98 1998/99
Dr. Martin Luther King Jr. Academy Mt.Clemens, MI (1995)	Michigan Educational Assessment Program (MEAP) 1995/96, 1996/97, 1997/98, 1998/99, 1999/00		1995/96 1996/97 1997/98 1998/99	1996/97 1997/98 1998/99
Mt. Clemens Secondary Academies Mt. Clemens, MI (1996)	Michigan Educational Assessment Program (MEAP) 1996/97, 1997/98, 1998/99, 1999/00			1997/98
Mid-Michigan Public School Academy Lansing, MI (1996)	Michigan Educational Assessment Program (MEAP) 1996/97, 1997/98, 1998/99, 1999/00		1996/97 1997/98 1998/99	
Washington-Edison Elementary School Sherman, TX (1995)	Texas Assessment of Academic Skills (TAAS) 1995/96, 1996/97, 1997/98, 1998/99			1996/97 1997/98 1998/99

Note: For the three norm-referenced tests, the table indicates all years of available data when conducting our analysis. The years marked in blue indicate data made available to us by Edison Schools. Where we are uncertain if test data actually exists, we have used *italics*.

Other sources of data and information that were reviewed throughout the course of the evaluation included relevant documents, research reports, and literature.

The ten schools that were included in the evaluation are listed below (the year when the school began operation is indicated in parentheses):

1. Roosevelt-Edison Charter School, Colorado Springs, CO (1996)
2. Henry E.S. Reeves Elementary, Dade County, FL (1996)
3. Dodge-Edison Elementary, Wichita, KS (1995)
4. Jardine-Edison Junior Academy, Wichita, KS (1996)
5. Boston Renaissance Charter School, Boston, MA (1995)
6. Seven Hills Charter School, Worcester, MA (1996)
7. Dr. Martin Luther King Jr. Academy, Mt. Clemens, MI (1995)
8. Mt. Clemens Secondary Academies, Mt. Clemens, MI (1996)
9. Mid-Michigan Public School Academy, Lansing, MI (1996)
10. Washington-Edison Elementary School, Sherman, TX (1995)

2.3 To Compare or Not to Compare

Since children grow and develop over time, we expect gains and learning to take place, regardless of the school program and even regardless of whether or not they attend school at all. The only way to separate the impact of the Edison model on students' formal learning from their learning in nonformal and informal settings is to compare students enrolled in Edison schools with students not enrolled in Edison schools. There are many ways to make such comparisons. The first method we utilized in this study was to compare an individual's achievement performance relative to the norms on nationally normed student achievement tests. By examining the relative ranking of Edison students in terms of national percentiles or normal curve equivalents, we can see whether or not they are gaining ground or losing ground compared with other students across the nation. In this study we made such comparisons using the MAT-7, SAT-9, and the ITBS.

A second way to make comparisons is to monitor the gains made by Edison students as compared with students in a local school with similar characteristics, with the district average, or with the state average. In some states, a comparison group of schools have students with similar characteristics and provide educational services at the same level (i.e., elementary, intermediate, or secondary levels). When we used comparisons in this area, we tried not to limit ourselves to one comparison group; rather, we compared against several groups. In nearly all cases the positive, negative, or

similar results are the same no matter which group is used. In this study, we use these comparisons with the state- and district-mandated tests.

In its first two annual reports on student progress, Edison strayed away from comparing its students with control groups. A noteworthy exception to this was the reports prepared by Dr. Robert Mislevy. John Chubb, the Chief Education Officer at Edison, informed us that the most important comparison was to follow Edison students/schools over time. He pointed out that gains made in the public schools surrounding the Edison schools should, in part, be credited to Edison. This may be partially true, since competition is bound to incite the local schools to improve the quality of their services. Nevertheless, the Edison schools are equally as free or more free to compete with the local schools and advance their students.

Edison claims its program leads to higher levels of achievement, and since—in a few cases—it refers to the fact that its schools are outperforming local district schools, it seems both fair and reasonable to make comparisons with control groups across all its schools. Also, Edison’s partners are expecting the schools they contract out or charter to improve student learning. It is obvious that they also expect the gains in these schools to exceed the gains made in the other local public schools. If they thought their own schools could match or exceed the gains of Edison’s schools, they probably would not have contracted out to Edison.

Why Edison does not wish to compare

Edison’s second annual report (Edison, 1999) makes three major arguments against comparing Edison schools with other public schools.⁶

1. “It is often impossible to find achievement trends for students and schools closely matched to Edison students and schools” (Edison, 1999, p. 12).
2. “Because Edison schools are launched by partnership communities to raise achievement not only in the Edison school but, through healthy competition and the diffusion of innovations, in all schools in a community, it is not a straightforward matter to estimate the relative success of an Edison school. In a successful partnership the achievement gains in other community schools might not match those in the Edison school, but they should be substantial as well. A successful Edison school, then, might not build an ever-widening advantage over other local schools; all schools might progress together with the Edison school leading the way” (Edison, 1999, p.12).
3. “A statistician would not compare the achievement of the Edison school and other local schools as if each school were performing independently. The achievement of all of the schools would be modeled as ‘endogenous’ variables, the achievement of the Edison school influencing the achievement of the other local schools, and vice versa” (Edison, 1999, p.12).

⁶ These arguments sum up the reasons we have heard from John Chubb, the Chief Education Officer at Edison, for why one should not compare Edison schools with others

In answer to the first point, we have identified extensive achievement data, although of varying quality, with which to make comparisons. While the control groups we have identified are not always equivalent in all areas, the expected advantages are almost always in Edison's favor. The Edison advantages are outlined below:

- Since half of Edison's schools are charter schools, and three of the ten Edison schools in our study are charter schools, these schools have the possibility of attracting families from throughout their districts who are more involved in their children's education and willing to search out and arrange an alternative to the school to which they were assigned. In the Edison contract or partnership schools, Edison negotiates that students who move out of the local neighborhood but remain within the district shall have the right to remain enrolled in their school. Therefore, parents who are actively involved will seek out, and if they move, arrange to have their children enrolled in the Edison school. Thus, selection bias should be in favor of Edison schools, unless they are perceived to be worse schools and parents act to get their children out. We think this should serve as an advantage to Edison. However, if communities and parents view the Edison program negatively, the selection bias could work against them. In our descriptive analysis of the schools, we traced changes in background demographics in the schools. From these analyses, it would appear that parents who are more likely to exercise choice are doing so in favor of Edison. Given that the differences in background characteristics of the control groups largely change over time in Edison's favor, one cannot claim that using control groups will be to Edison's disadvantage.
- A second advantage is that among the schools where we could identify student demographic data over time, there was often a slight tendency for the Edison schools to exhibit a decrease in numbers of students qualifying for free and reduced lunches compared with the control groups.
- A third advantage is that over time the Edison schools appear to exclude more students from tests than the control groups.
- The last advantage is Edison's own school model. The control groups are not equal in terms of expenditures or in terms of time at school. Edison advertises that it invests heavily in its schools (an average of \$1.5 million per school) and generally spends more money per pupil than the control groups. The Edison schools have a longer school day and longer school year than the students in the control group.

In response to the point that comparisons with local schools fail to capture the impact that Edison has had on these schools due to "healthy competition and the diffusion of innovations," we should point out that we used state-level comparisons in addition to the local district and/or control school comparison to control for this. Where possible, we also used national percentiles or normal curve equivalents (NCEs) to compare the levels of achievement in the Edison schools with national norms for students.

The argument that is made by Edison in the third point refers to implicit nesting of schools within districts, regions, or intermediate school districts. The question we ask is whether this should inhibit comparison or can it be taken into consideration in the design of the study? Important factors resulting from the nesting that would/could affect student achievement include budget/financial

allocations and local building level autonomy (site-based management of schools). Also, at the elementary level, the nesting impact is less, due to the “local” nature of each school. The nesting effect probably increases as you move from elementary into intermediate and upper-secondary schools. Because elementary schools usually include students from the immediate neighborhood, they are more heavily influenced by the local neighborhood since the schools at the intermediate and upper secondary levels are fed by many smaller neighborhood schools.

The argument that Edison’s schools contribute to gains made in local schools is not a hindrance to comparing, but rather a testable hypothesis. While the available data and the scope of this study did not permit us to fully test this hypothesis, we took it into consideration in the comparisons we made and in the discussion of findings. Since we selected control groups from the state level statistics and also made comparisons with national and state norms where available, we think we controlled for this factor in our evaluation. For example, if the performance in an Edison school and in the district goes up while the performance level of the state remains stable or declines, one might suggest that an Edison school helped raise performance levels in the district in which it resides. However, if the trends for the state and district are similar and the trend at the Edison school does not match the state and district, one might suggest that the Edison school is not having an impact on raising performance levels in the district in which it resides. As one will see in our case studies, the latter pattern is far more prevalent than the former pattern.

In summary, we agree in part with Edison’s third argument that the schools in the communities are interlinked. In half of Edison’s schools (the contract schools) there is one governance structure: the local district school board that contracted with Edison to operate one or more of its schools and which also governs the local public schools. However, in terms of policy and day-to-day decision making, there is a clear separation between the Edison schools and the local public schools in all but a few cases where the Edison school has a district employee serving as the principal administrator (this occurs in the schools within schools where the principal is a district employee but two separate vice principals are assigned to the two school entities that share the building). While the leadership of the Edison schools is based in New York City, the local public schools are governed locally. Edison also has a separate budget from the local public schools. Because the Edison schools have largely separate governance, educational programs, and budgets from the local district schools, we think that comparison with the local district is valid. In handouts prepared by Edison and in a few instances in its annual reports on student progress (Edison, 1999, 2000), Edison does state that its students are gaining more than students in local district schools. For these reasons, we believe that comparisons can and should be made between the Edison schools and local district schools.

2.4 Statistical Methods Utilized

We employed several different statistical methods and analyses which, taken together, provide a composite picture of student performance on standardized tests at ten Edison schools. These methods are described and discussed in the sections that follow.

Assumptions guiding our analyses

It is important to recognize some of the more general assumptions we made in order to make the various statistical comparisons and how these affected our conclusions regarding the impact of the Edison model on student achievement. First, we utilized a wide variety of student achievement data, e.g., nationally administered achievement tests and various state- and district-mandated tests. These tests represent only a subset of possible indices of student academic performance, and in many situations it can be argued that nationally and state normed tests do not adequately describe a student's achievement level. Consequently, our first assumption was that national and state normed tests do provide a common and valid assessment of student achievement that allows for meaningful comparisons. Appendix A contains a description of each standardized test considered in the ten cases.

Secondly, we assumed that attrition rates are low and stable over time in the longitudinal cohort⁷ analyses (panel analyses) and at a level similar to the comparison samples in the consecutive cohort analyses. Edison (1999) noted that its rate of mobility is very low (7 percent as compared with a national average of 17 percent) and indicated that this is a form of market accountability. In most of the 10 schools included in this study, we found a higher rate of mobility; however, the rates of mobility are similar to, and seldom exceed, the mobility rates of the local districts in which the Edison schools reside.

A third assumption, particularly important in the consecutive cohort analyses, was that the later cohorts would have more exposure to the Edison effect. That is, once a student enters an Edison school, he or she is assumed to be matriculating through the consecutive grades. For example, the Martin Luther King Jr. Academy (MLK) in Mt. Clemens, Michigan, became an Edison school in 1995. First graders entering MLK in that year had nearly four years of exposure to the Edison effect when the 4th grade state assessment test (MEAP) was administered in the spring of 1999. Fourth grade students taking the state test in spring 1998 had three years of exposure to the Edison program, fourth grade test takers in 1997 had only two years of exposure, and fourth grade test takers in 1996 had just completed their first year in an Edison school. Given the expectation that students enrolled in an Edison school will improve achievement levels faster than students in a traditional public school, we should see a gradual rise in academic performance on a test like the MEAP from 1996 to 1999.

We also conducted the analysis under the assumption that students enrolled in Edison schools would improve more quickly than students not enrolled in its schools. We are aware that many of the schools that Edison operates have average performance levels that are lower than those in the local schools. In fact, many of the schools that districts contract out to Edison are the lowest performing schools in their districts. Because of this, we are more interested in the value added (i.e., gain scores) over time, rather than on absolute performance levels. Among the factors that underlie our

⁷ In the text of this report we use the term “cohort” to reference multiyear longitudinal trends in individual student achievement data (what is often referred to as a “panel”). We use the term “consecutive cohort” to describe groups of students who consecutively pass through a particular grade.

assumption that Edison students should demonstrate larger gains than comparable groups of students, are the following:

- Edison has a longer school day and longer school year than traditional public schools.
- Edison reportedly invests an additional \$3,000 per student, above regular per-pupil funding, when it starts each new school (Edison, 1999). In 1995, Edison reported capital investments of \$5,114,000. This figure grew to \$70,233,000 in 1999 (Edison, 1999).
- Edison has a program and curriculum that incorporate a number of research-based practices.
- Perhaps the strongest factor supporting this third assumption is that Edison claims that its schools will gain more. At conferences and meetings, Edison personnel report that its schools are making large gains, and in its second annual report (Edison, 1999) it was reported that students in schools operated by Edison are making average annual percentile point gains of 5 percent on norm-referenced tests and 6 percent average annual percentage point gains on criterion-referenced tests. Edison's third annual report suggests that the performance during the 1999-2000 school year was even better than during earlier years (Edison, 2000).

Decisions by district or charter school boards to contract with Edison are based on this assumption as are decisions by parents who chose to enroll their children in a school operated by Edison.

Description of statistical analyses utilized

We utilized three principal statistical analyses to gauge the effect of Edison schools on student learning. However, we first present and discuss descriptive summary data for each school, identifying important school-, teacher-, and student-related factors that may have an impact on student learning. Unfortunately, we are not in the position to relate these known moderators of student learning to achievement outcomes due to the limited available building-level data and also because the nature and type of indicators vary from case to case. We include the descriptive summary for each school so that readers will have a greater understanding of the context in which the schools operate and so that readers can judge for themselves the relevance and validity of the comparison groups and the differences between the comparison groups and the Edison schools.

Longitudinal trend analysis. The first analytical strategy we utilized on six cases was a longitudinal trend analysis on individual norm-referenced student achievement data provided to us by Edison. Identifying variables were coded to retain student confidentiality. The outcome variables (results on standardized achievement tests) differ by school and within schools by grade and number of connective years due to the nature of the data provided to us by Edison. A detailed description of the tests with information on the grades and subjects they cover is presented in Appendix A as well as in each school case study. A repeated measures ANOVA (list-wise deletion) was examined for longitudinal trends over the available years. Parallel analyses are reported for all types of scores reported, e.g., grade equivalent (GE), standard (or scaled) score (SS), percentile rank (or national percentile rank) (PR), and normal curve equivalent NCE) score. In all models, the assumption of sphericity was evaluated and if found to be violated, the Huynh-Feldt adjusted p-values are reported.

If the overall effect for time was found to be statistically significant, unadjusted (alpha) pair-wise comparisons were examined to identify where a difference in the means might be located.

We received individual student results on norm-referenced tests for six of the ten schools included in the study. Most of the data sets contained results on four separate scales: GE, SS, PR, and NCE. GE and SS scores should show increases over time for all cohorts for obvious reasons: the students are maturing and learning. Likewise, grade equivalents rise as grade level placements advance each year. The important question here is whether or not the students are gaining the equivalent of one year's knowledge between two separate test administrations, which roughly occur at the same time each year. In some cases, we found that students were progressing less than one grade equivalent in a given year, suggesting that while learning is occurring, the students are not learning at a rate suggested by national norms for the tests. In other cases, we found that students' grade level equivalents were increasing more than one grade for each year, suggesting that students were advancing more quickly than the national norm. In many, but not all the schools in our study, the students started out with grade level equivalents lower than their current grade level placement.

The last two scales that we report on regarding the longitudinal data are the national percentile rank (PR) and the normal curve equivalent (NCE). The percentile rank indicates the relative rank of a student or school in comparison with the national norm. A PR of 70 percent indicates that only 30 percent of the students in the national sample scored higher and 70 percent scored at the same level or lower. Given the nature of the data we worked with, we believe that the NCE is a better indicator of a student's or school's relative status.⁸ The NCE is a normalized standard score with a mean of 50 and a standard deviation of 21.06. Percentile ranks and NCEs have a direct fixed relationship as shown in Appendix B. NCE scores are a preferred method for measuring and comparing gains made by a school over time. Percentile ranks in a normal distribution clearly do not represent the same score scale distance between equal differences in PR values. For example, the difference between PRs of 20 and 30, or between 45 and 55, means the same in terms of percent of the normal distribution, e.g, 10 percent. But in terms of raw score distance they are not equivalent. Indeed, the z-score distance between a PR of 20 and 30 is .32 (-.84 – -.52), whereas the z-score distance between a PR of 45 and 55 is .26 (-.13 – +.13). This is even more exaggerated when one compares the same difference in PRs at the extreme end of the distribution versus those in the middle.

The relationship between PRs and NCEs is linear between PR 5 and PR 95, but at the extremes it goes curvilinear.

$$\text{NorCurPR} = 1.21 + \text{NCE} - 10.6$$

⁸ Since percentile ranks are so popular, it was decided that a better scale would be one that looked like the PR scale but where the difference in raw score distance was the same for the same difference in distance on the new scale, the NCE scale. The rationale behind that was that pre- to postcomparison on an NCE score scale would be a more meaningful comparison, since a difference of 10 NCE points for one student or school corresponds with the 10 NCE points for another student or school, even if they started at different points on the pretest.

Values in normative tables for converting PRs to NCEs are obtained by determining the normalized z-score associated with the PR of interest and making a transformation of the form.

$$\text{NCE} = 50 + 21.06(z)$$

Some of the concerns that arose regarding the use of NCE include the following:

- The NCE is very close in scale and meaning with the big T-score.
- The NCE and PR are too easily confused, especially by lay people.
- Using NCE and finding PR equivalents with a formula (like what you see in published tables), assumes that the distribution of NCE scores is normal. This should be empirically verified.

A number of other strategies for analyzing the individual student data might have been utilized but were not due to the limited amount of available individual student achievement data. For example, a hierarchical linear model would be a superior method for identifying the longitudinal growth of student achievement. However, the data necessary to accomplish this type of analysis were not available. Moreover, the longitudinal analyses presented in this evaluation do not incorporate a comparison sample against which to gauge student learning, which represents a major limitation of this analysis. Nevertheless, national norms represent a point of comparison; but because we cannot control for the characteristics of the students considered in the national norms, this is somewhat limited.

Chi-square analysis. The second analysis strategy focused on student learning outcomes as measured by district- and state-mandated tests. The next section of this report describes these largely criterion-referenced tests. We accessed composite outcomes by grade level for schools within our sample that fall under state testing guidelines. The state tests are scored along various ordinal scales (detailed in each case study). Since these data are open to the public, we were able to construct comparison groups (detailed in each case study) for these analyses. Chi-square analyses were examined to determine if the relative proportion of students falling in the various performance levels on the state-mandated criterion-referenced tests (CRT) differed between the Edison school and the comparison group. These analyses were examined separately by year, grade level, and subtest category of the state test.

Table 2:2 Construction of 2 x 2 Tables for Odds Ratio Analysis

	Fail	Pass
Edison School	a	b
Comparison School or Comparison Group	c	d

Note: “Fail: corresponds with not meeting state standards and “pass” corresponds with meeting or exceeding expected state standards.

Although there are several possible ways to define passing, we opted to define passing and failing as specified by each state. For example, if the CRT is scored along a 4-point scale (Level 1 [lowest]

to Level 4 [highest]) and the state criteria for passing is a score of 3 or 4 then we collapsed level 4 into level 3 and level 1 into level 2 to define passing and failing respectively. It should be noted that this reclassification could mask some important gains evidenced by the students in either the Edison school or the comparison group. Appendix E contains the complete results from the chi-square analyses.

Odds ratio analysis. The third analysis strategy examined student learning outcomes within a prospective cohort study by analyzing the collapsed ordinal responses (pass/fail categories) on the state tests. A cohort study is when subjects are selected before they are exposed to possible determinants of interest (i.e., being in an Edison school), and their exposure to possible determinants of interest (i.e., “the Edison effect”) are then recorded along with the outcome (i.e., passing or failing the state test or in other words, meeting or exceeding state standards vs. not meeting state standards). The critical design factor in a cohort study is the comparability (similarity) of the two groups at the beginning of the time period under study. If the two groups are similar, then an observed association between being in an Edison school and passing (or failing) a component of the state test can be reasonably defended. However, if the two groups are not similar, then any observed association between being in an Edison school and passing (or failing) a component of the state test may or may not be truly a function of attending an Edison school. We constructed the 2x2 tables for these analyses in such a way to represent the relative odds for a student to fail a component of the state test (see Table 2:2).

The odds ratio (OR) (McNeil, 1996) is defined as $OR = ad/bc$ and represents the proportion of students who fail the test in the Edison school relative to the proportion of students failing the test in the comparison school. An odds ratio can take values from zero to positive infinity. Interpretation of an OR is straightforward. An OR value of 1.00 represents equal odds for failing (or passing) relative to the comparison group. Values from 0.00 to 1.00 are representative of a “protective” effect; that is, the odds of failing are lower in the Edison school. Values greater than 1.00 would represent increasing odds for failing the test if enrolled in the Edison school. As with any point estimate, a $(1-\alpha)$ confidence interval (CI) needs to be constructed for accurate interpretation. Thus, if the CI around the OR includes 1.00, the conventional interpretation would be that there is no statistically significant difference in the relative failing rate between the two schools (i.e., if the CI included 1.00, there is no statistically significant difference). However, if the CI does not include 1.00, the OR is generally interpreted as statistically significant, either representing a statistically significant protective effect or a statistically significant increase in the odds for failing the test. Due to the truncated nature of the sampling distribution of the OR, the standard error of the OR is calculated based on the natural logarithm of the OR, similar to converting a correlation to a Fisher’s Z before constructing a $(1-\alpha)$ confidence interval around a correlation.

The standard error of the natural log of the OR is

$$SE(\ln OR) = \sqrt{\frac{1}{a} + \frac{1}{n_1 - a} + \frac{1}{b} + \frac{1}{n_2 - b}}$$

Table 2:3 Contains a summary of the various analyses used for each of the ten cases. Each case also discusses other research and evaluation studies that have been conducted in the past. In addition to these analyses, we have provided a summary and discussion of findings for both the norm-referenced and criterion-referenced test results for each case.

Table 2:3 Types of Analyses Possible for Each School Included in the Study

School	Descriptive Analysis	Longitudinal Study	Chi-Square	Cohort Study Odds Ratio
Roosevelt-Edison Charter School	Yes	Yes	Yes	Yes
Henry E. S. Reeves Elementary	Yes	Yes*	Yes	Yes
Dodge-Edison Elementary	Yes	Yes		
Jardine-Edison Junior Academy	Yes			
Boston Renaissance Charter School	Yes	Yes	Yes	Yes
Seven Hills Charter School	Yes	Yes	Yes	Yes
Dr. Martin Luther King Jr. Academy	Yes	Yes	Yes	Yes
Mt. Clemens Secondary Academies	Yes		Yes	Yes
Mid-Michigan Public School Academy	Yes	Yes	Yes	Yes
Washington Elementary School	Yes		Yes	Yes**

* We included the findings from Shay (2000), a dissertation that was based on the analysis of longitudinal results by Reeves' students and a control group.

** Breslow-Day statistic and confidence intervals could not be calculated because we lacked information on the number of test takers.

2.5 Criteria for Evaluating Trends and Summarizing Results for Each Case

When analyzing and comparing our findings, we are conscious that one cannot simply count the number of positive and negative findings to come up with a conclusion. The approach we took for this evaluation focused on developing individual school cases (see Chapters 3-12), with each case developed according to the available data and subsequent analyses. At the end of each case, a short description and discussion of the relative quality of the various analyses can be found along with the overall findings from that case.

Our discussions of the relative strength of the analyses will consider the quality of the study design, the sample size, quality of instruments used, etc. In terms of judging and rating research designs, we considered the analytical framework developed by Peterson (1998). Peterson was asked by The Edison Project to compare the results presented by AFT and Edison. In his analysis, he outlined nine categories of designs, with randomized experiments rated as the "gold medal" design.

The overall availability of student achievement data varies extensively among the schools included in our study. So, too, does the quality of the available data along with the quality of the designs behind the trends, as described above. For example, some of the data we analyzed were provided to us by Edison Schools, Inc., other data were obtained from internet Web sites, and still other data came from various state and local reports. Moreover, given the varied mechanisms we utilized to garner our data, we often had to rely on the reporting agencies for data accuracy and could not verify the accuracy of the data. There are even instances where our data are incomplete due to incomplete information received from the reporting agency.

Given the range of possibilities before us, that we perceive designs that involve tracing individual students to be superior to designs that only measure change in consecutive class cohorts. In some cases we were able to conduct analyses on both longitudinal analysis of individual student achievement data and consecutive class cohorts. In these cases it is particularly salient when the two different design strategies converge in their findings and interpretation.

In our summaries of trends for each case, described in detail in the next section, we rated the various trends as negative (-1), mixed (0), or positive (+1). We included only one trend for each subject grade level test. For example, for Dodge-Edison we were able to establish trends for the Metropolitan Achievement Test. We had individual student data and so were able to conduct a longitudinal panel analysis and compare the gains made by Edison students with the national norm. With this same data, we were able to make comparisons with the district based on national percentile ranks for each subject test by grade. Because the individual student data provided a stronger design, we included the trends based on individual student data in the summary for this case, but did not include the trends based on consecutive cohorts of students in the summary (see Chapter 5 for more details). Also, when various comparison groups are available for the criterion-referenced test data, we include only one trend in our summary, based on the comparison group that is most relevant.

We believe that judgments about the overall performance of a school need to be made on a case-by-case basis. In order to limit potential bias, and in order to establish a common method of making judgments, we thought it was important to establish criteria to distinguish, first, whether or not there had been change over time, and second, if any change was positive or negative. The criteria we chose are based upon Edison's own criteria, included in its first annual report (Edison, 1997, p. 6) to distinguish when changes in achievement levels are positive, mixed, or negative. It is not clear whether these criteria were used by Edison in its second and third annual reports.

Edison's criteria served as our starting point; however, we modified several criteria and added one new criterion. In evaluating trends we use the following criteria to distinguish meaningful change and when this change is positive or negative:

- *Effect sizes (ES) or differences in effect sizes of .20 or greater.* The effect size calculated for the NRT data is the omega squared (ω^2) (Kepple, 1991) for a one way repeated measures ANOVA and only provides the reader with an overall effect for time. It does not adequately convey the direction of change nor if the change occurs all in one year or is reflective of a gradual cumulative gain.

- *Differences in national percentile scores of 5 percentage points or more per year*
- *Differences in percentage proficient scores of 5 percentage points or more per year*
- *Differences in grade equivalents of 2 months or more, and annual gains in grade equivalents of 14 months or more per year*
- *Differences in DALT gain scores of 2 points or more per year*
- *Differences that are statistically significant (at the .05 level) when tests of significance are available.* The p-value criterion is only applied to the CRT data and not to the NRT data. Utilization of a p-value criterion in the longitudinal NRT analyses does not adequately convey the direction of a statistically significant change. That is, in a longitudinal analysis there can be a statistically significant change in both directions; therefore, the p-value is ambiguous relative to the direction of change, only the presence of change. However, in the chi-square and OR analysis, the p-value conveys a meaningful difference due to the configuration of the contingency tables, in that for these analyses the reference is to the comparison group.
- *Differences in normal curve equivalents of 3.5 or more per year*

Part of the technical complexity of this report is a function of the variety and large number of analyses conducted. We have constructed summary tables for the reader's benefit that help guide and focus the reader in distilling the overall impact of Edison in a given school. First we treated each analysis category (NRT, CRT) separately. Within each analysis category, we have rated a finding as negative (-1), mixed (0), or positive (+1) based upon the guidelines presented above. A negative finding would be an effect that meets one of the above criteria but in the opposite direction and should be relatively unusual. For example, in an NRT analysis on NCE, a negative finding would be a reflective of a 3.5 NCE drop per year over the number of years covered by the analysis. In a CRT OR analysis, a negative finding would be a statistically significant (p-value criteria) risk of failing the test relative to the comparison sample. A mixed finding would be reflective of grade-level improvement in an NRT, or even odds in an OR analysis. A positive effect could be illustrated by an average annual gain in NCEs of 3.5 points per year or more over the life of the analysis, or an OR that is statistically significant and protective. To this general scoring system we tried to determine if a trend was present when there was more than one score scale present, e.g., NRT data, or more than two years of data were present, e.g., OR analysis.

We also based our rating on a prioritized hierarchy of data. We consider a trend in NRT data to reflect the findings of a longitudinal panel of students as they progress through the life of the analysis by subject and grade. A trend in CRT data reflects the consecutive cohort findings for a specific grade and subject test over the life of the analysis.⁹ Although we have calculated outcomes relative to various comparison groups (e.g., national, state, district, or other), we count only one trend in the combined table. For NRT data we prioritized the analyzes as follows: we considered the NCE trend

⁹ By contrast, Edison counts trends in one-year change segments so a trend of data for a cohort of students over four years would be counted by Edison as three different trends, while we would count this as one trend and base our rating on the change over the life of the trend.

first if available, followed by the PR or NPr, then GE, and lastly SS. For CRT data we counted each grade and subject test separately based on the outcomes of the OR Breslow-Day findings relative to the district data.

Each case was then summarized by combining the NRT and CRT ratings into one table to derive an overall school rating. In its 2000 annual report, Edison defined the 5-point scale they used to rate the overall trends in its schools (Edison assigns one to five stars for each of the categories, from Strongly Negative to Strongly Positive, respectively). Its cut points are as follows:

- Strongly Negative when 0-19 percent of the trends are positive
- Negative when 20-39 percent of the trends are positive
- Mixed when 40-59 percent of the trends are positive
- Positive when 60-79 percent of the trends are positive
- Strongly Positive when 80-100 percent of the trends are positive

Since we considered all the trends and did not focus on the positive trends alone, we calculated a mean across the trends where a negative trend is equal to -1, a mixed trend is equal to 0, and a positive trend is equal to +1. We then applied the following 5-point rating scale to the mean trend:

- -1.00 to -0.60 corresponds with “Strongly Negative”
- -0.59 to -0.20 corresponds with “Negative”
- -0.19 to +0.19 corresponds with “Mixed”
- +0.20 to +0.59 corresponds with “Positive”
- +0.60 to +1.00 corresponds with “Strongly Positive”

For example, in Dodge-Edison (see Chapter 5, Table 5:7) we report a total of seven trends, four NRT and three CRT. Based on the criteria listed above, of the four NRT trends, two are positive (+2) and two are mixed (+0). Of the three CRT trends, 1 is positive (+1) and two are mixed (+0). Averaged together, we rate Dodge-Edison as a Positive school with a mean rating of 0.43.

2.6 Limitations of the Study

Several inherent limitations in this evaluation needed to be examined in order to provide a balanced interpretation of the findings we reported and the conclusions we have drawn. The limitations to this study can be grouped into three areas: methodology, data quality, and conceptual limitations. In this section we highlight what we think are the major limitations of the study that should temper all conclusions derived from this evaluation.

Evaluation of schools based on student performance alone

Ideally, evaluations of schools should not be based on student performance data alone. Nor should such evaluations be based on measures of market accountability, such as head counts. We believe that evaluations of schools, such as those included in our study, should be based on measures of market accountability, performance accountability, and regulatory accountability. Clearly, this evaluation is focused only on performance accountability. As suggested in the title of this report, we are evaluating the performance of students enrolled in Edison schools and not the Edison schools themselves.

Selection of Edison schools included in this evaluation

There is a possibility of selection bias related to the schools selected for the evaluation, in that it might be argued that the schools we studied were either performing more poorly or superior to nonselected Edison schools. We examined this possibility by conducting a chi-square analysis on the school ratings published in Edison's 2000 annual report for the 10 schools in this evaluation relative to the remaining 32 schools included in the report. Edison rates each school on a 5-point scale, from Strongly Positive to Strongly Negative. Our analysis indicates that there is no indication that the 10 schools we included in this study are rated by Edison any differently from the 32 schools that opened during or after 1997 and for which it reported trend data in its 2000 annual report. Thus while there may be some selection bias in our sample, there is no strong indication that the schools we evaluated are not an accurate representation of the schools for which Edison currently has trend data.

Lack of a comparison group in the longitudinal analyses

One of the principal advantages of this evaluation also suffers from one of the major limitations. Edison Schools Inc. provided us with seven different test data sets that included individual student achievement data from a variety of different standardized achievement tests, i.e., SAT-9, MAT-7, and ITBS. These data sets covered six schools and varied in quality and quantity. Nevertheless, they allowed us to examine longitudinal trends based on individual student data in six of the ten schools in our study. Unfortunately, Edison was not in a position to include data on similar schools for comparisons. Although these longitudinal analyses provided the most complete picture of student achievement in an Edison school, there is no comparison group to gauge gains and losses against. Thus, we were left with comparisons against national norms and interpretations focused on grade equivalent scores and NCEs.

Composition of comparison groups used in the chi-square and odds ratio analyses

The primary purpose of this evaluation was to develop a composite understanding of the effect of attending an Edison school on students' achievement. In order to do this, we needed to make performance comparisons on state-mandated testing programs relative to some comparison group; that is, we needed to construct suitable comparison groups. Much of the state-mandated data was

extracted from Internet web sites, and there was considerable variation in the quantity and quality of data from state to state. Moreover, we decided to try to define our comparison groups in such a way that the data sources would represent the same quality of information and thus have similar meaning from state to state. This reasoning, unfortunately, resulted in comparison groups that were at the least sophisticated end of the spectrum, e.g., district and state. Obviously, superior comparison groups are at least theoretically possible to identify. However, even if identified—for example, equated on gender, grade, SES, mobility, teacher experience, teacher mobility, etc.—identifying the necessary data to construct the contingency tables may or may not be readily available. For example, different states report different resolutions of performance data. Thus, it was extremely difficult to identify a common, minimum set of variables to begin developing more sophisticated comparison groups. Consequently, the validity of our comparison groups can be questioned. For example, if student mobility in the Edison school is dramatically higher or lower than in the district or state, then the longitudinal comparisons may not reflect the true magnitude of the “Edison effect.”

Variability and completeness of Web-based reporting of the district- and state-mandated testing results

Much of the data we used in the chi-square and odds ratio analyses was extracted from the Internet. Much of these data, perhaps all, had undergone significant data filtering and cleaning by the various state agencies reporting the data. We cannot be sure that there were no data posting errors at these web sites. The data we extracted and analyzed cannot be checked for accuracy beyond what is posted on the Internet. If there were data reporting errors (we know there are rounding errors), we cannot identify or examine the data for possible bias resulting from this possibility. All we can do is assume that posting errors on the Internet, if any, are randomly distributed across the various Internet sites we used. In the event that there are posting errors, any impact would be to increase the background noise in our analyses, thus making it harder to detect differences among the groups.

Marginal cooperation of Edison in supplying individual student achievement data, regardless of type

As stated above, Edison provided some individual student achievement data. These data, however, were very inconsistent and often incomplete. In many cases we were given only two or three years of data; yet it is documented (see Table 2:1) that much more data are available. Even without a comparison sample, the longitudinal analyses would be stronger had all available data been utilized. While promised access to the individual student data in July 1999, we did not receive any data files from Edison until the end of November 1999. During that time we expended considerable time and money attempting to build data sets with individual student data for the MEAP or to secure data sets containing individual student results from other sources. None of these efforts resulted in any usable data sets containing individual student data.

Limited resources

The budget for the study was limited in size. NEA contracted with The Evaluation Center, but the Center ended up covering 15 percent of the expenditures for the study. During the latter part of the study, the authors contributed their own time since allocated funds for the study had been expended. Conducting evaluations of this nature with limited budgets leaves little room for errors in planning or exploration of alternative sources of data or methods of analysis. Additional resources for the study would have allowed us to conduct a meta-analysis and would have helped us include additional outside experts. Nevertheless, the evaluators are grateful for the interest and advice of a number of persons outside the Center who contributed generously with their time and expertise.

Controversial and polarized nature of Edison Schools Inc.

Evaluations of this sort are made more complicated by the controversial and polarized nature of the reform (i.e., private, for-profit operators of public schools) and the strong vested interest of many of the stakeholders.

Chapter Three

Roosevelt-Edison Charter School

Colorado Springs, Colorado

3.1 Descriptive Summary of School

Roosevelt-Edison Charter School is a district charter school that was established in 1996 to serve grades K-5. In 1997, the same charter was used to open Emerson-Edison Partnership School.¹⁰ The mission of the Roosevelt-Edison Charter School is to prepare its diverse population of students for success in their lives by providing a “world class education” at prevailing public school costs. Enrollments at Roosevelt-Edison have been quite consistent, with 681 students during the 1997-98 school year, 684 in the 1998-99, and 674 in 1999-00. While enrollments have remained steady at Roosevelt, they dropped in the district as a whole. For example, the district enrolled a total of 32,815 students in the 1997-98 school year, but by the 1998-99 school year the enrollments had dropped to 31,586.

The student-to-teacher ratio has also remained relatively stable at Roosevelt with 13.7 students per teacher, increasing only from 12.2 students per teacher during the 1997/1998 school year when there were 56 instructional staff members. In comparison, the Colorado state average for student-to-teacher ratio decreased from 18.5 in the 1996/1997 school year to 18.2 during the 1997/1998 school year. The district reported an 18.3 student-to-teacher ratio during the 1998/1999 school year. Roosevelt-Edison has provided a consistently lower student-to-teacher ratio than both the state and the district. Roosevelt-Edison’s average attendance rate was 94.9 percent in the 1997/1998 school year and decreased slightly during the 1998/1999 school year to 92.9 percent (CDE, 1999, 2000).

In its third annual report, Edison (2000) reported that of the 674 students enrolled during the 1999/00 school year, 28.3 percent were African American, 2.5 percent were Asian/Pacific, 34 percent were Caucasian, and 20.9 percent were Hispanic, and 14.2 percent were categorized as “other” (students in this group are largely Native American Indians). According to the Colorado Charter Schools Evaluation, 51 percent of Roosevelt’s students were white in the 1999/2000 school year, about a 14

¹⁰ Even though Emerson and Roosevelt are operated under the same charter, we were asked by Edison officials to distinguish the two as separate schools, one at the elementary level and one at the middle school level. The two entities occupy separate facilities and serve students at different levels. In any case, there was only one year of data available for grade 7 students at Emerson, so we could not trace the trends in achievement.

percent increase from the 1997/1998 school year. The population of white students in the state was consistently in the lower 70 percent range between 1995 and 1999. The district school population was 71.2 percent white during the 1999/2000 school year, similar to the state. The state and district school populations have approximately 20 percent fewer minorities than at Roosevelt, which is about 50 percent minority. Also, approximately 7.1 percent of the Roosevelt students were labeled as bilingual or ESL (English as Second Language) during the 1999-00 school year.

The number of students qualifying for free and reduced lunches rose from 54 percent during the 1996/1997 school year (AFT, 1998) to 64.6 percent during the 1999/2000 school year (CDE, 2000). The size of the waiting list decreased from 375 in 1998 to 325 in 1999. As one can see, the list is still quite large. Approximately 25 percent of all families volunteer at the school and logged more than 7,000 hours of work (CDE, 2000).

Special education students represented 9.7 percent of the population during 1999-00 (CDE, 2000). There was a discrepancy in the findings for the percentage of the population receiving special education services during the 1998/1999 school year. According to the 1998 Colorado Charter School Evaluation, 2.3 percent of the student population needed special education services. In its 1997/1998 self-reported school profile, Edison stated that 8.5 percent of its students required special education services.

The average teacher salary at Roosevelt-Edison was \$38,876 during the 1998/1999 school year. The state average was up from \$35,364 during 1995/1996 to \$37,240 during 1997/1998. With an increase of about \$1,000 each school year, we can speculate that during the 1998/1999 school year the average teacher salary in Colorado was around \$38,240, still slightly lower than Roosevelt.

In terms of accountability, Roosevelt-Edison administers two norm-referenced tests: the Iowa Test of Basic Skills (ITBS) and the Comprehensive Test of Basic Skills (CTBS). The school is also required to take part in the district criterion-referenced test: District Achievement Levels Test (DALT). Finally, in terms of performance assessments, the school reports that it uses individual learning programs and student portfolios (CDE, 2000).

3.2 Past Studies and Evaluations

According to the two most recent state charter school evaluations (CDE, 1999, 2000) Roosevelt School provided data for both the 1997/98 and the 1998/99 school years that “generally indicated that they were meeting expectations defined for their performance.” During the 1998/99 school year, 33 percent of the charter schools exceeded expectations, and the Edison school was joined by another 25 charter schools (51 percent of all charter schools in the state) in “generally meeting expectations.” Another 16 percent of the charter schools did not provide sufficient information to indicate if they were meeting expectations (CDE, 2000, p. xvi-xvii).

Mislevy (1998) researched the area of the *Success for All* reading program that Edison implements in its schools. Roosevelt was compared with a control school from the local district in grades K-2.

Students were matched by grade, gender, and participation in free- or reduced-lunch programs, to see how *Success for All* affected reading achievement. Both groups took pretests and in every grade, the control school outscored Roosevelt in the second year. The analyses of 1997/1998 were done with checks on attrition, comparison of prettest and 1997/1998 posttest scores, comparison of effect sizes across years, comparison of cohorts across years, and change over years. In turn, the “posttest differences cannot be interpreted as estimates of program effects,” although a matched control was attempted. Unfortunately, both groups began from starkly different beginnings in reading achievement which weaken the argument that differences in results could be attributed to the program.

Although the differences in the pretests “precluded interpreting posttest differences as program effects, a pattern did emerge”(Mislevy 1998). There is the possible indication that the program is having positive effects when we see that, although the pretests revealed vast differences among the groups, there was a pattern of decreasing posttest differences among the younger cohorts for the 1997/1998 school year. However, even with adjustments to compensate for the initial differences, there was still a small difference favoring the control school for second graders; there were mixed differences for first graders, meaning neither school was favored; and there was a small difference that favored Roosevelt kindergartners. But even after statistical adjustments were made, there was little difference among the groups (Mislevy 1998).

The Colorado Springs District 11 administers a set of tests to measure student performance and growth in basic skills. The District Achievement Levels Tests (DALT) are series of tests in reading, language, and math that are constructed to align with the district’s curriculum. These tests are administered at the beginning and end of each school year. The DALT’s validity is enhanced by the use of Rasch scaling, which is a statistical method that employs test items representing narrow bands of increasing difficulty. Progress is monitored through the use of growth scores referred to as Rasch units or, more simply, RIT scores. The district reported results for Roosevelt-Edison in order to compare gains made at grades 3, 4, and 5 with district gains. Generally, Roosevelt-Edison is consistently below the district levels, but the gains made by Roosevelt-Edison on the DALT largely parallel the gains made by the district (i.e., the charter school is not gaining more than the district over time).

3.3 Longitudinal Analysis of Individual Student Data

Edison Schools Inc. provided individual student test results only for the Iowa Test of Basic Skills (ITBS). ITBS scores were provided in three separate scales: stanine scores (Stn), national percentile rank scores (PR), and normal curve equivalent scores (NCE). Unlike the other data sets we received, there were no data for grade equivalent scores. Parallel analyses are reported for each of the three scales for which we received data.

Data indicated that Roosevelt-Edison had 498 students during the 1997-98 and 1998-99 school years. The ITBS tests were administered during the spring of each year. We were able to trace 2 cohorts representing a small portion of the students: between 64 and 68 students in Cohort A, depending on

the subject test, and 71 students in Cohort B for all 3 subject tests. For example, ITBS language stanine scores from 1998 show that there were 106 possible third graders to follow; by 1999, 63 percent remained (71/106 pair-wise analysis). This represents a 37 percent drop in students in the Cohort A that progressed from grade 3 to grade 4. In Cohort B 109 students began in grade 4 and 71 were available in grade 5, or about 65 percent. Table 3:1 presents sample size information for each ITBS score scale for each subtest by year and cohort. Also depicted in this table are the sample sizes for the one-year gain analysis.

Table 3:1 Table of Sample Sizes for Individual Student Data on the ITBS by Grade and Year

		Grade 3		Grade 4		Grade 5	
		1998	1999	1998	1999	1998	1999
Stanine Score							
	Language	106	105	109	115	85	112
	Math	105	106	106	112	87	111
	Reading	108	105	108	113	84	111
National Percentile Rank (PR)							
	Language	106	105	109	115	85	112
	Math	105	106	106	112	87	111
	Reading	108	105	108	113	84	111
Normal Curve Equivalent (NCE)							
	Language	106	105	109	115	85	112
	Math	105	106	106	112	87	111
	Reading	108	105	108	113	84	111

	Cohort A (3rd to 4th grade)			Cohort B (4th to 5th grade)		
	Stanine	PR	NCE	Stanine	PR	NCE
Language	67	67	67	71	71	71
Math	64	64	64	71	71	71
Reading	68	68	68	71	71	71

Longitudinal analysis findings

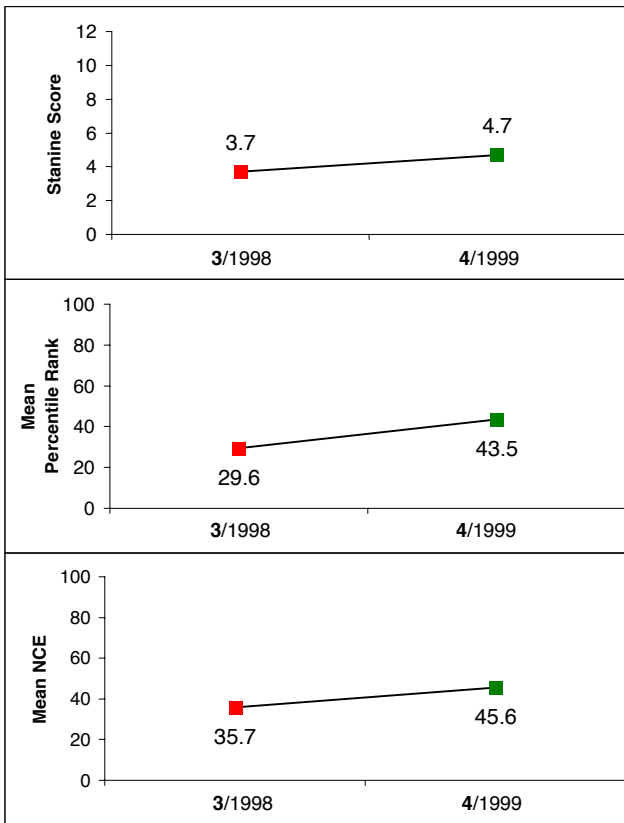
Dependent t-tests were examined for detecting a one-year change (gain) in the two cohorts between 1998 and 1999 in the Roosevelt-Edison charter school. Outcome data were individual student data on the Iowa Test of Basic Skills language, mathematics, and reading subtests. Parallel analyses are reported for all types of scores reported by Edison, e.g., stanine score, percentile rank, and the normal curve equivalent (NCE) score. The tables and charts in Exhibits 3:1 and 3:2 illustrate the results in terms of stanine scores, percentile ranks, and normal curve equivalents (NCEs). The results are grouped by subject area tests: language, math, and reading.

Exhibit 3:1 Cohort A at Roosevelt-Edison Charter School: Results From the Iowa Test of Basic Skills

COHORT A

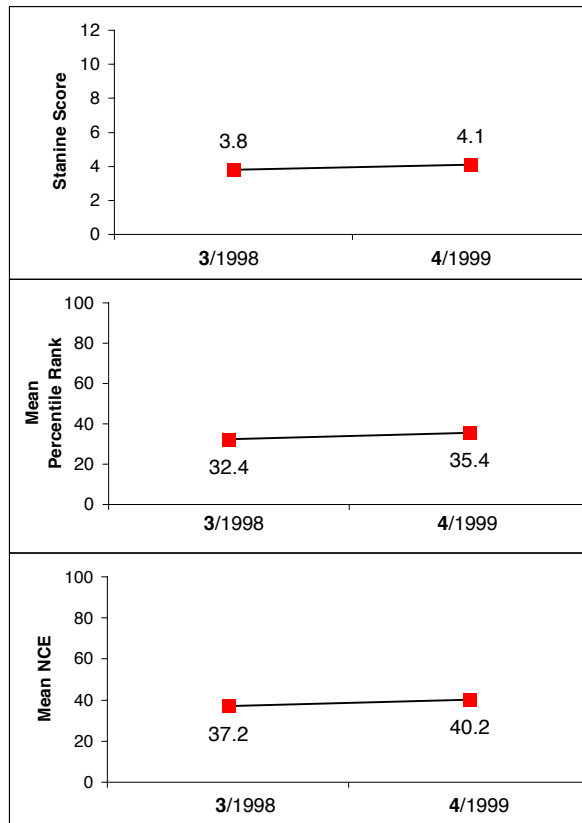
LANGUAGE

n=67	1998	1999	
Grade	3th	4th	(Gain)
Stn Language	3.7	4.7	0.985 t(66)=8.2539, p<.0001
NP Language	29.6	43.5	13.92 t(66)=7.5911, p<.0001
NCE Language	35.7	45.6	9.94 t(67)=7.8872, p<.0001



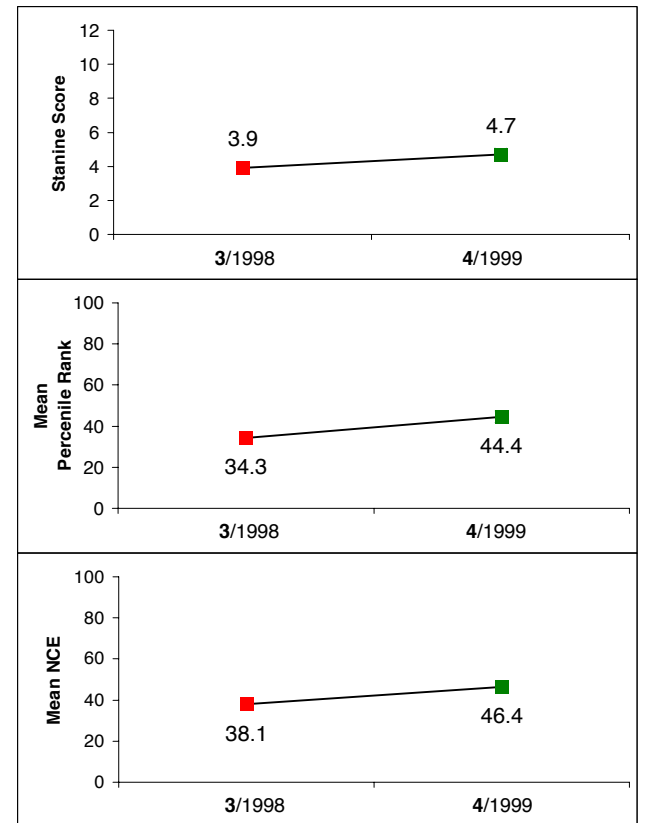
MATH

n=64	1998	1999	
Grade	3th	4th	(Gain)
Stn Math	3.8	4.1	0.234 t(63)=1.4921, p=.1407
NP Math	32.4	35.4	3.02 t(63)=1.4568, p=.1501
NCE Math	37.2	40.2	3.03 t(63)=1.8512, p=.0688



READING

n=68	1998	1999	
Grade	3th	4th	(Gain)
Stn Reading	3.9	4.7	0.794 t(67)=5.6153, p<.0001
NP Reading	34.3	44.4	10.07 t(67)=5.3388, p<.0001
NCE Reading	38.1	46.4	8.32 t(67)=5.5254, p<.0001



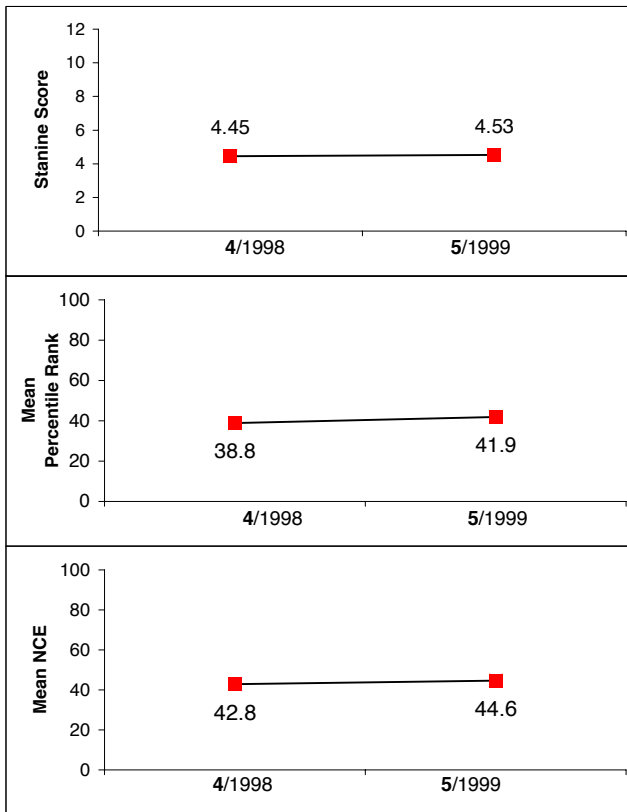
A color change represents a statistically significant change in the means. A trend with two color changes represents differences among all three means. A dual colored charting point (red and green) represents a statistically significant difference between one mean but not the other. The reader is encouraged to examine the individual table of means for these cases.

Exhibit 3:2 Cohort B at Roosevelt-Edison Charter School: Results From the Iowa Test of Basic Skills

COHORT B

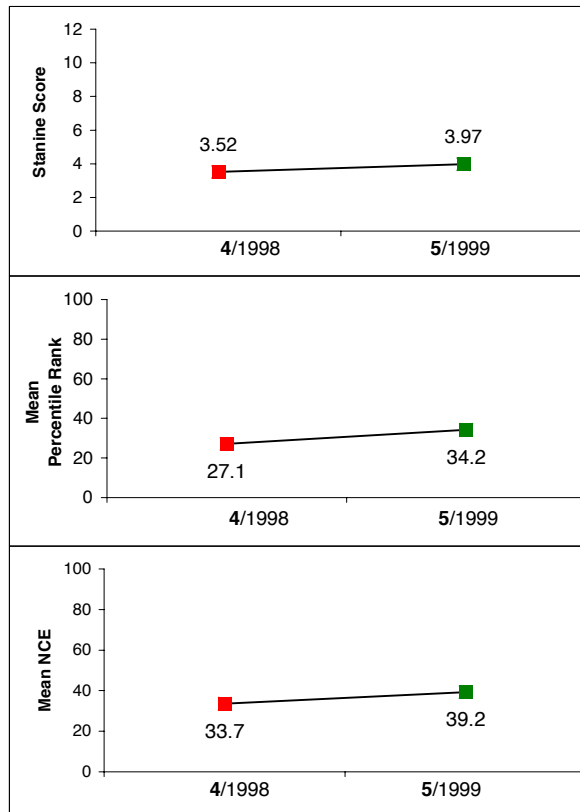
LANGUAGE

n=71	1998	1999	
Grade	4th	5th	(Gain)
Stn Language	4.45	4.53	0.084 t(70)=0.6298, p=.5309
NP Language	38.8	41.9	3.07 t(70)=1.6108, p=.1117
NCE Language	42.8	44.6	1.74 t(70)=1.2492, p=.2157



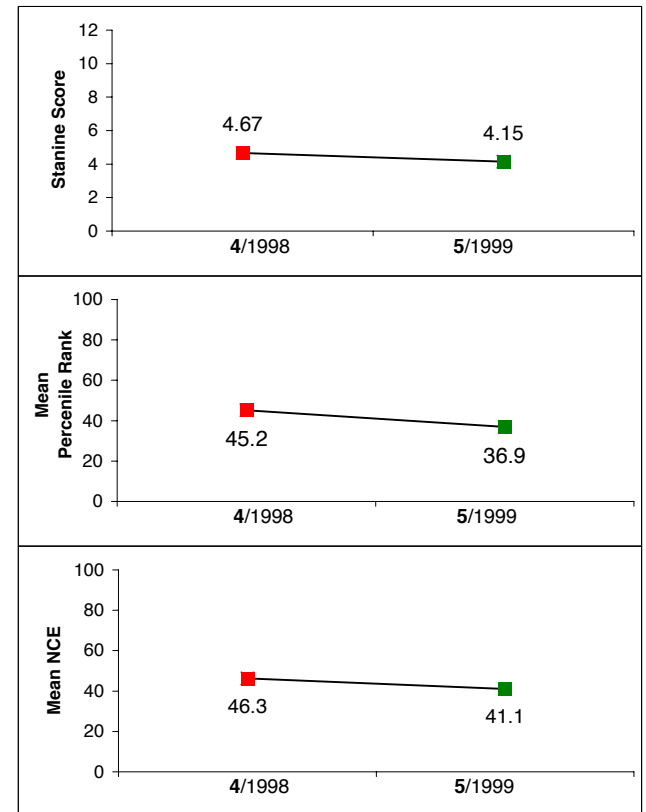
MATH

n=71	1998	1999	
Grade	4th	5th	(Gain)
Stn Math	3.52	3.97	0.450 t(70)=3.3952, p=.0011
NP Math	27.1	34.2	7.15 t(63)=3.5987, p=.0006
NCE Math	33.7	39.2	5.56 t(70)=4.0135, p<.0001



READING

n=71	1998	1999	
Grade	4th	5th	(Gain)
Stn Reading	4.67	4.15	-0.52 t(70)= -3.022, p=.0035
NP Reading	45.2	36.9	-8.33 t(70)= -3.653, p=.0005
NCE Reading	46.3	41.1	-5.27 t(70)= -2.929, p=.0046



A color change represents a statistically significant change in the means. A trend with two color changes represents differences among all three means. A dual colored charting point (red and green) represents a statistically significant difference between one mean but not the other. The reader is encouraged to examine the individual table of means for these cases.

In terms of stanine scores, national percentile ranks, and NCEs on the ITBS, Cohort A students showed statistically significant increases in language and reading but not in mathematics. Because these gains were statistically significant and the scales are relative to the ITBS national norms, one can be certain that these students are moving ahead and are in fact progressing faster than the national norms except in mathematics. Between the 1997-98 and 1998-99 school years, the students in Cohort A increased their performance on the language subtest by 10 NCEs, and on the reading subtest they increased by 8.3 NCEs. The students also increased by 3 NCEs on math, but this gain was not statistically significant.

The performance by students in Cohort B was nearly the opposite, with the students performing well in math, but equal to or less than the national norm in language and reading. The students moved from an NCE of 33.7 in the 1997-98 school year up to 39.2 NCEs in the 1998-99 school year. The students in this cohort gained 1.7 NCEs in language, which indicates no significant change over the two years. In reading, however, the students in Cohort B performed very poorly and dropped from an NCE of 46.3 in the 1997-98 school year to 41.1 NCEs in the 1998-99 school year. This decrease was statistically significant.

On the whole, we can see that the students at Roosevelt-Edison Charter School are below the national norm and in the areas where they had a higher ranking in 1997-98, they were more likely to make small gains or actually lose ground on the national norms. In the areas where they had a lower starting point in 1997-98, the students were more likely to make larger and statistically significant gains.

3.4 Chi-Square Analysis of CSAP Data

Available data

One of the main sources of student data was the Colorado Student Assessment Program (CSAP). We were able to secure test results for the Roosevelt School, the local district, and the state from the Colorado Department of Education. Because the CSAP is a relatively new state assessment program and because it is still not fully implemented in terms of grades and subject tests, we were not able to collect data for all years that the Edison charter school was in operation. We were able to obtain grade 3 reading results for two years, grade 4 reading and writing results for three years, and grade 5 math results for one year (1999-00).

A chi-square analysis was initiated on data made available by the Colorado Department of Education on the outcomes of the Colorado Student Assessment Program (CSAP), the state mandated criterion-referenced test. The CSAP is administered in grade 3 (reading), grade 4 (reading and writing), and grade 5 (mathematics, first results were available in March 2000). For all tests across the grades and subject areas, a similar four-category ordinal scale is used to report results: Unsatisfactory, Partially Proficient, Proficient, and Advanced.

Construction of the comparison groups

We constructed two different comparison groups for the Roosevelt-Edison chi-square analyses (see Table 3:2). Since we were interested in examining the number/proportion of students who meet state standards (“passing”) or, conversely, the number/proportion of students who do not meet state standards (“failing”) on the CSAP, we needed to define a suitable comparison group. Our first comparison was with the local district (Colorado Springs, District 11), and the second comparison group we selected was the state of Colorado. While the state demographics vary between Roosevelt-Edison and the district and state, we believe that comparisons with district and state performance levels, particularly since we focus on gains, can yield further information regarding the relative performance levels at the Edison school. Also, since Edison claims that advances in other district schools is—in part—due to its presence, we use the state as a more distant point of comparison that cannot be easily influenced by the presence of Edison’s schools.

Table 3:2 Summary of Chi-Square Findings for Roosevelt-Edison Charter School

	1997	1998	1999
3 rd Grade Reading			
Roosevelt vs. District	NA	sig/sig	sig/sig
Roosevelt vs. State	NA	sig/sig	sig/sig
4 th Grade Reading			
Roosevelt vs. District	NA	sig/sig	sig/sig
Roosevelt vs. State	sig/sig	sig/sig	sig/sig
4 th Grade Writing			
Roosevelt vs. District	NA	sig/sig	sig/ns
Roosevelt vs. State	sig/sig	sig/sig	sig/sig
5 th Grade Math			
Roosevelt vs. District	NA	NA	sig/sig
Roosevelt vs. State	NA	NA	sig/sig

Note: Each result cell in the matrix is divided with the results for the 2x4 analysis on the left-hand side, and the results for 2x2 analysis on the right-hand side (i.e., 2x4/2x2). Red color indicates a statistically significant difference that favors the comparison group; blue color indicates a significant difference that favors the Edison school.

General procedure

Utilizing published data from the Colorado Department of Education, we were able to make yearly comparisons (consecutive cohorts) for grade 3 from 1998 to 1999, in grade 4 for 1997 through 1999, and for grades 5 and 7 for only one year, 1999.

Percentage data (students in each scoring category) were converted to raw frequency data prior to chi-square analysis. To insure independence of the rows in the chi-square tables, the raw frequencies for each scoring category of the CSAP in the district and state comparisons were down-weighted by subtracting the number of students in that category from Roosevelt-Edison. Thus, the district and state numbers reflect all students exclusive of those in Roosevelt-Edison.

Four chi-square analyses were evaluated for each subtest nested within year and grade level. Two of these analyses were on uncollapsed data, that is, all scoring levels were represented in the contingency table (e.g., a 2x4) for the district and state comparisons. Two follow-up analyses were conducted on the data after collapsing the multilevel scoring into a dichotomy (pass, fail), thus producing 2x2 contingency tables. According to the Colorado Department of Education, a score in

the “proficient” and “advanced” categories constitutes “passing” or meeting the state standard for that particular grade and subject. On the other hand, the “unsatisfactory” and “partially proficient” categories refer to scores that have not met state standards and fall into the “fail” category in our 2x2 chi-square and odds-ratio analyses.

Chi-square findings

Results of the chi-square analyses for grade 3 covered the CSAP administration years for 1998 and 1999 for reading. Individual contingency tables are presented in Appendix E. Eight analyses, (four 2x4 for 1998 results) and four 2x2 for 1999 results) were examined, four each for the district and state comparisons.

The results of these chi-square analyses for grade 3 evidence a reversal between 1998 and 1999 (see Table 3:2). In 1998 Roosevelt-Edison students showed a higher proportion (statistically significant) of students scoring in the lower categories relative to students in the district or state. However, in 1999 this pattern flipped; there were a higher proportion of students from Roosevelt-Edison were at the proficient and advanced scoring levels.

In grade 4 two CSAP tests are administered: reading and writing. We were able to make district and state comparisons for 1998 and 1999 and state comparisons for 1997. The results were consistent across test, comparison group, and type of analyses (see Table 3:2). All chi-square analyses except one were statistically significant, with lower proportions of Roosevelt students achieving at the proficient and advanced levels. The exception to this general pattern was in the 2x2 district comparison for 1999. Roosevelt students performed at comparable levels relative to district students on the grade 4 writing test. Fifth grade CSAP mathematics data were only available for 1999. Higher proportions of Roosevelt students scored in the lower two categories than either the district or the state.

3.5 Odds Ratio Analysis of the CSAP Data

One of the many possible statistics that can be derived from a 2x2 contingency table is the odds ratio statistic and corresponding 1- α confidence interval. As presented in Section 2.4 of this report, the 2x2 tables analyzed in the previous section can be thought of as representing consecutive class cohorts in a prospective design. From a classical epidemiological perspective, the students in the Edison school can be thought of as the “exposed” group, that is, exposed to the “Edison-effect,” and students in the comparison group as the unexposed group. From this perspective each yearly comparison is a “new” cohort; and measured over a period of years, there are consecutive class cohorts. There is a minimal possibility for cohort contamination if a number of students in one group are not promoted to the next grade level. However, we think this represents a very small number of possible cases and therefore has minimal impact on the validity of these analyses.

Section 2.4 details the OR statistic and corresponding 1- α confidence interval. We have calculated and charted OR for each of the 2x2 tables constructed from the chi-square analyses presented above. Exhibit 3:3 presents these findings.

Odds ratio findings

Consistent with state interpretations of passing score on CSAP, we grouped the scoring categories “proficient” and “advanced” as representing “passing” and scoring categories “partially proficient” and “unsatisfactory” as representing “fail” for our odds ratio analyses. We calculated and charted OR for each of the 2x2 tables constructed from the chi-square analyses presented above. Exhibit 3:3 and Tables 3:3 and 3:4 present these findings. In grade 5 only one year of data could be obtained for these comparisons; thus, a trend could not be estimated. For this reason we have not graphed the odds ratio analysis for grade 5.

As seen in Table 3:3, grade 3 students in Roosevelt-Edison significantly improved relative to the district, showing a statistically significant decrease in OR from 1998 to 1999 (note that the Breslow-Day statistic was significant). The OR in 1998 dropped from 2.344 to 0.591 in 1999. This is a dramatic improvement for Roosevelt-Edison students relative to students in the rest of the district. However, there was only a marginal improvement in OR relative to students in the state comparison that did not reach statistical significance. The Breslow-Day statistic was not significant, indicating that a common OR for the two-year period can be calculated (see Table 3:3). Students from Roosevelt-Edison were more than two times as likely to fail the 3rd grade CSAP reading test as were students in the state.

Exhibits 3:4 to 3:7 illustrate performance levels for Roosevelt, the district, and the state. These exhibits also include the 1999-00 data that were not considered in the odds ratio analysis.

Table 3:3 Summary of Grade 3 CSAP Odds Ratio Findings for Roosevelt-Edison

	OR	95% CI
Common OR		
<i>Odds of not meeting standard compared with district</i>		
1998	2.344	1.617 - 3.397
1999	0.591	0.402 - 0.870
<i>Odds of not meeting standard compared with state</i>		
1998 - 1999	2.344	1.180 - 3.035

Table 3:4 Summary of Grade 4 CSAP Odds Ratio Findings for Roosevelt-Edison

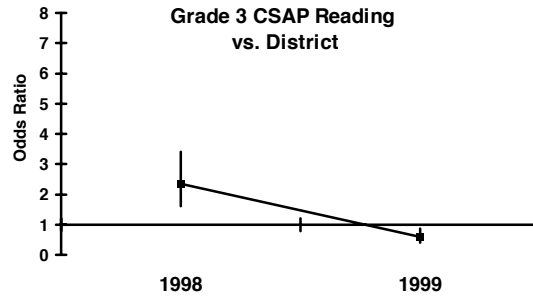
	OR	95% CI
Common OR		
<i>Odds of not meeting standard compared with district</i>		
1998-99 Reading	3.014	2.305 - 3.942
1998 Writing	2.641	1.657 - 4.209
1999 Writing	0.972	0.587 - 1.612
<i>Odds of not meeting standard compared with state</i>		
1997 - 1999 Reading	2.414	1.940 - 3.003
1997 - 1999 Writing	2.044	1.546 - 2.701

Exhibit 3:3 Odds Ratio Results for Roosevelt-Edison Charter School (Two and Three Year Trends)

Grade 3 CSAP Reading Comparison to District

Year	U CI	L CI	OR
1998	3.397	1.617	2.344
1999	0.870	0.402	0.591

Breslow-Day for Homogeneity of Odds Ratio
Chi-Sq (1, N=5,080) = 25.381, p < .001

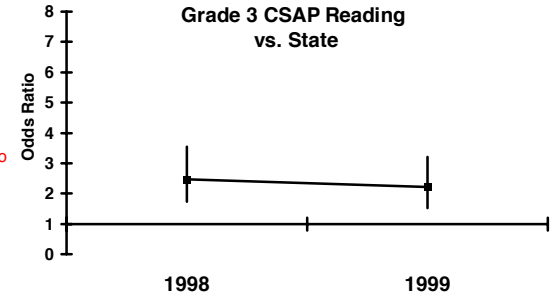


Grade 3 CSAP Reading Comparison to State

Year	U CI	L CI	OR
1998	3.543	1.725	2.472
1999	3.211	1.528	2.215

Breslow-Day for Homogeneity of Odds Ratio
Chi-Sq (1, N=103,713) = 0.163, p = .686

OR = 2.344
LB = 1.810
UB = 3.035

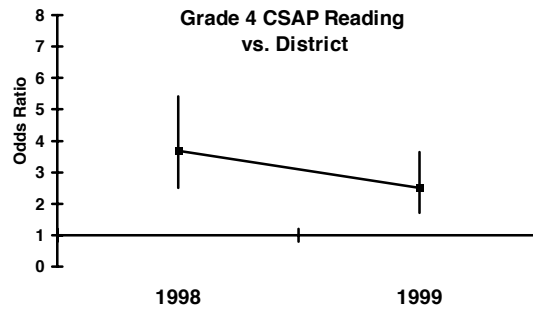


Grade 4 CSAP Reading Comparison to District

Year	U CI	L CI	OR
1998	5.412	2.503	3.680
1999	3.639	1.715	2.498

Breslow-Day for Homogeneity of Odds Ratio
Chi-Sq (1, N=5,090) = 1.819, p = .177

OR = 3.014
LB = 2.305
UB = 3.942

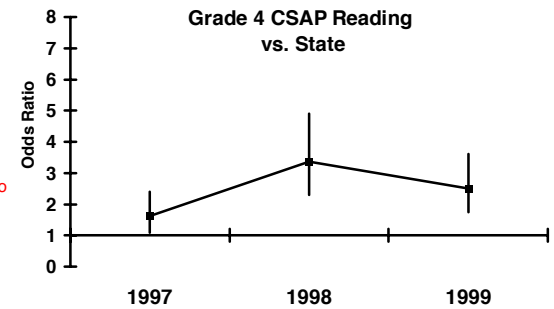


Grade 4 CSAP Reading Comparison to State

Year	U CI	L CI	OR
1997	2.408	1.079	1.612
1998	4.907	2.303	3.362
1999	3.610	1.735	2.503

Breslow-Day for Homogeneity of Odds Ratio
Chi-Sq (2, N=148,638) = 6.539, p = .038

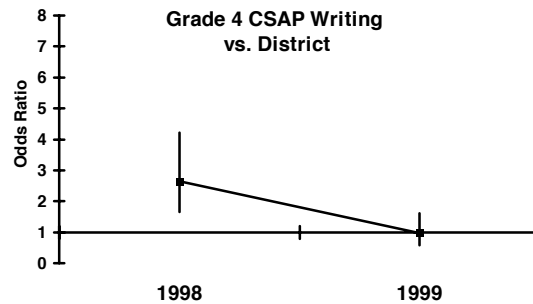
OR = 2.414
LB = 1.940
UB = 3.003



Grade 4 CSAP Writing Comparison to District

Year	U CI	L CI	OR
1998	4.209	1.657	2.641
1999	1.612	0.587	0.972

Breslow-Day for Homogeneity of Odds Ratio
Chi-Sq (1, N=4,608) = 8.097, p = .004



Grade 4 CSAP Writing Comparison to State

Year	U CI	L CI	OR
1997	3.296	1.201	1.989
1998	3.905	1.556	2.465
1999	2.735	1.027	1.676

Breslow-Day for Homogeneity of Odds Ratio
Chi-Sq (2, N=148,864) = 1.227, p = .541

OR = 2.044
LB = 1.546
UB = 2.701

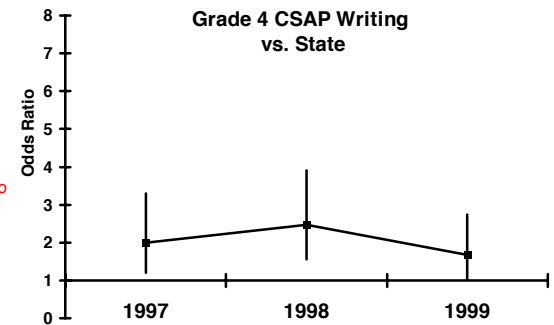


Exhibit 3:4 Performance on Grade 3 Reading for Roosevelt-Edison, District, and State

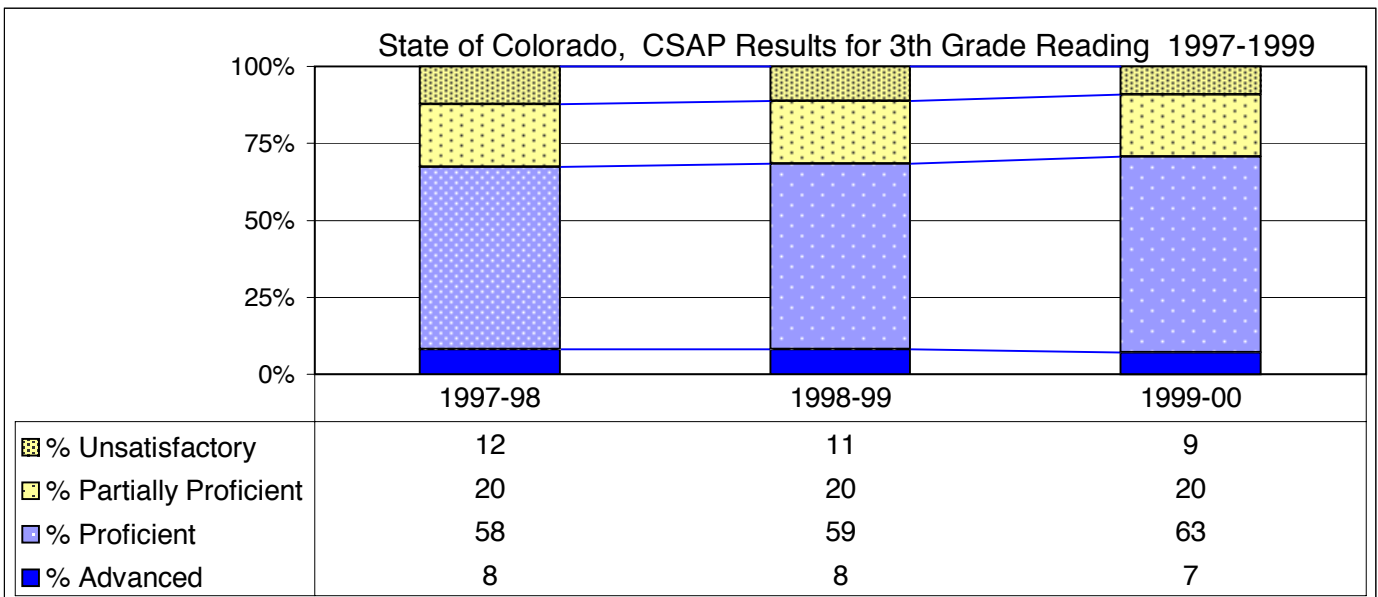
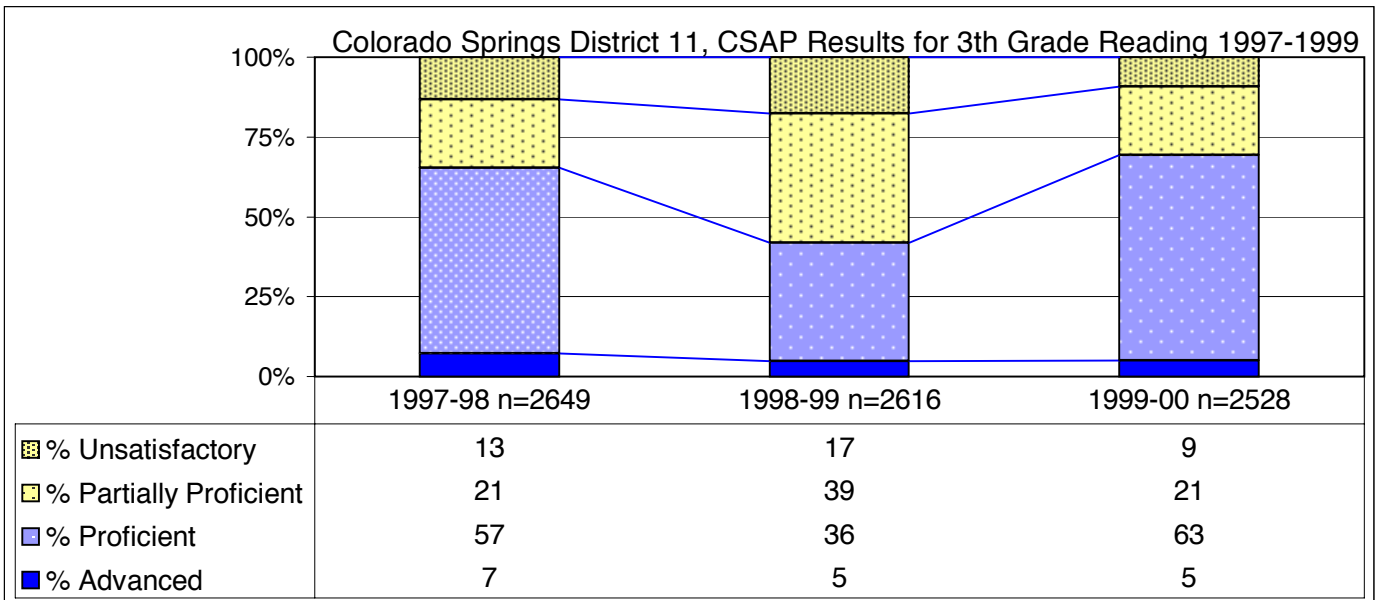
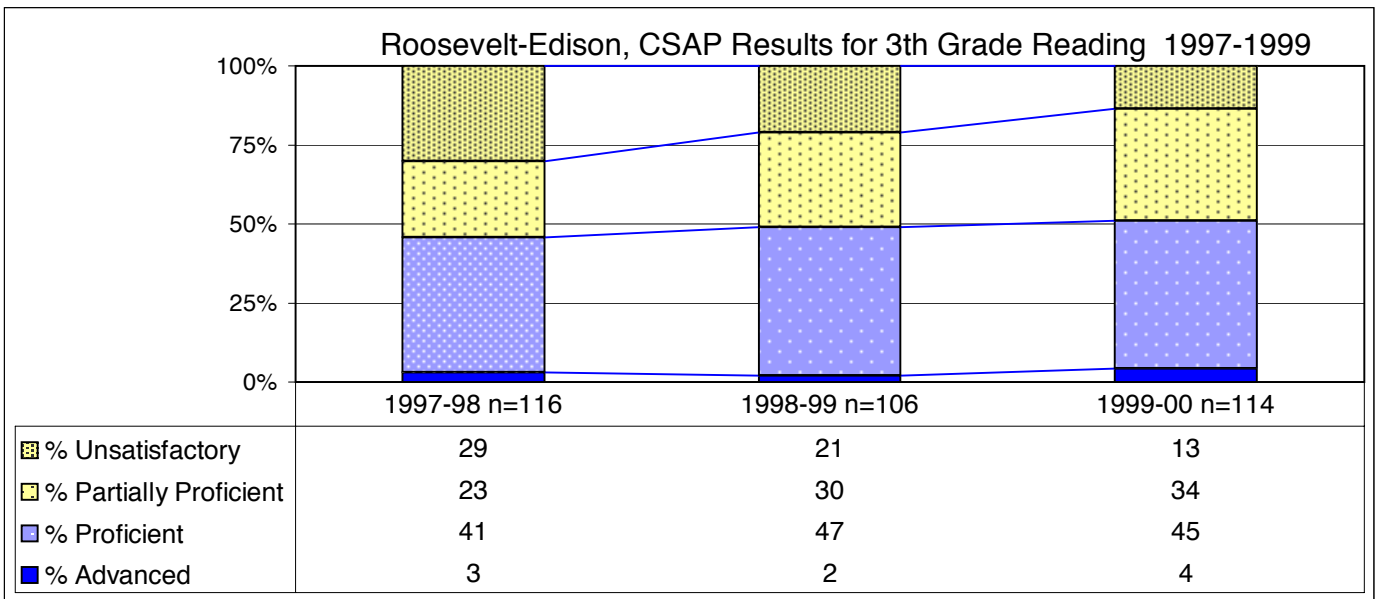


Exhibit 3:5 Performance on Grade 4 Reading for Roosevelt-Edison, District, and State

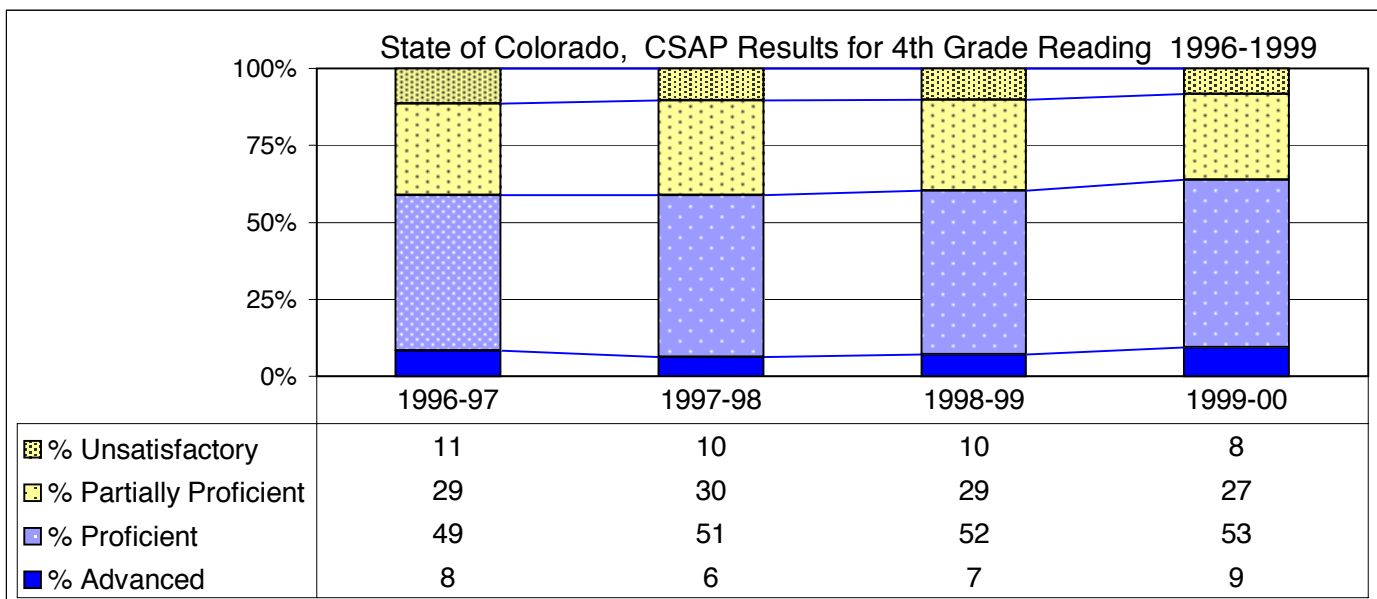
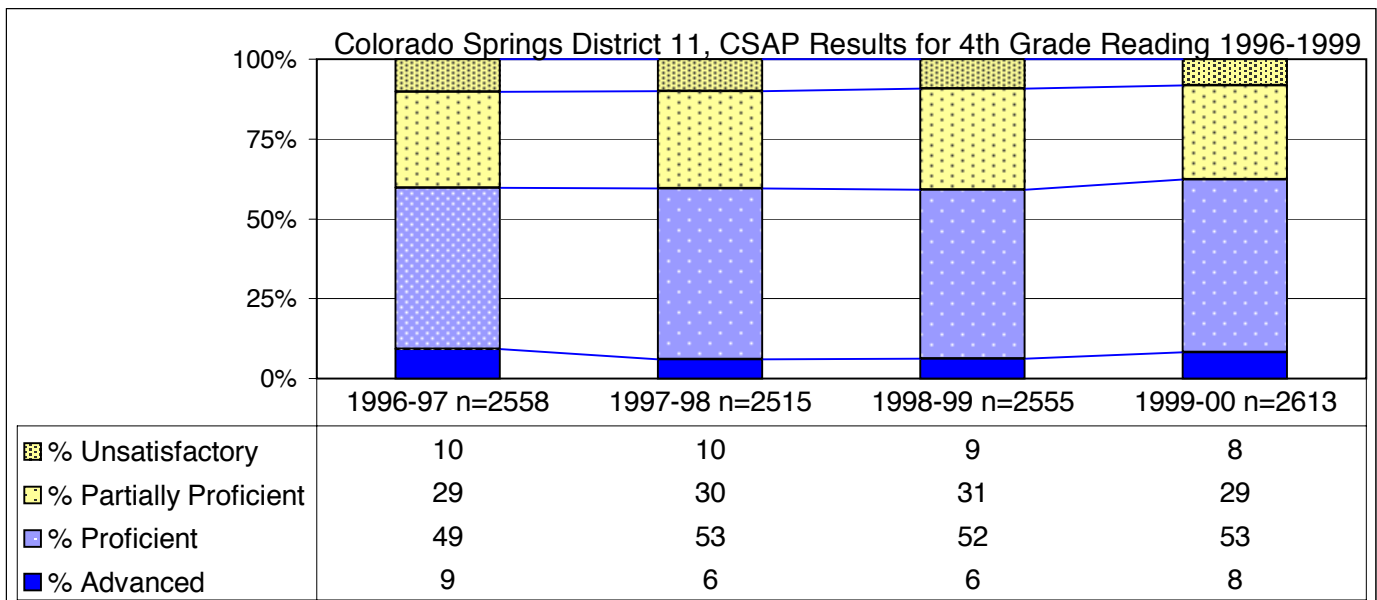
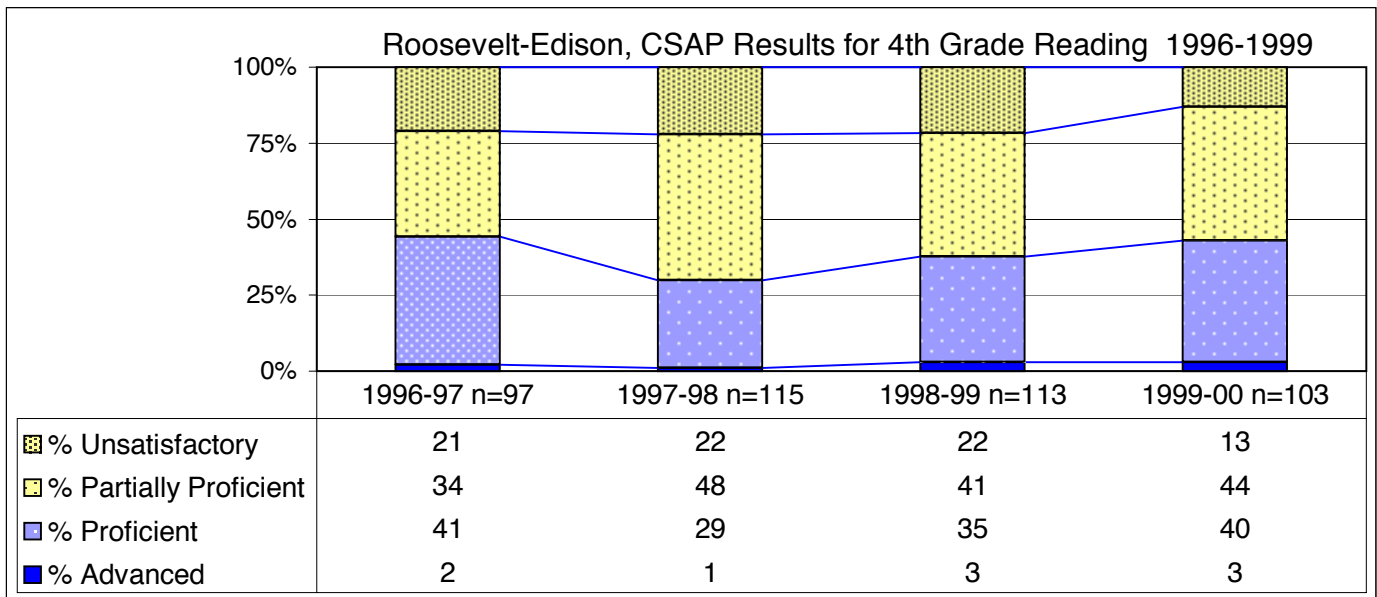


Exhibit 3:6 Performance on Grade 4 Writing for Roosevelt-Edison, District, and State

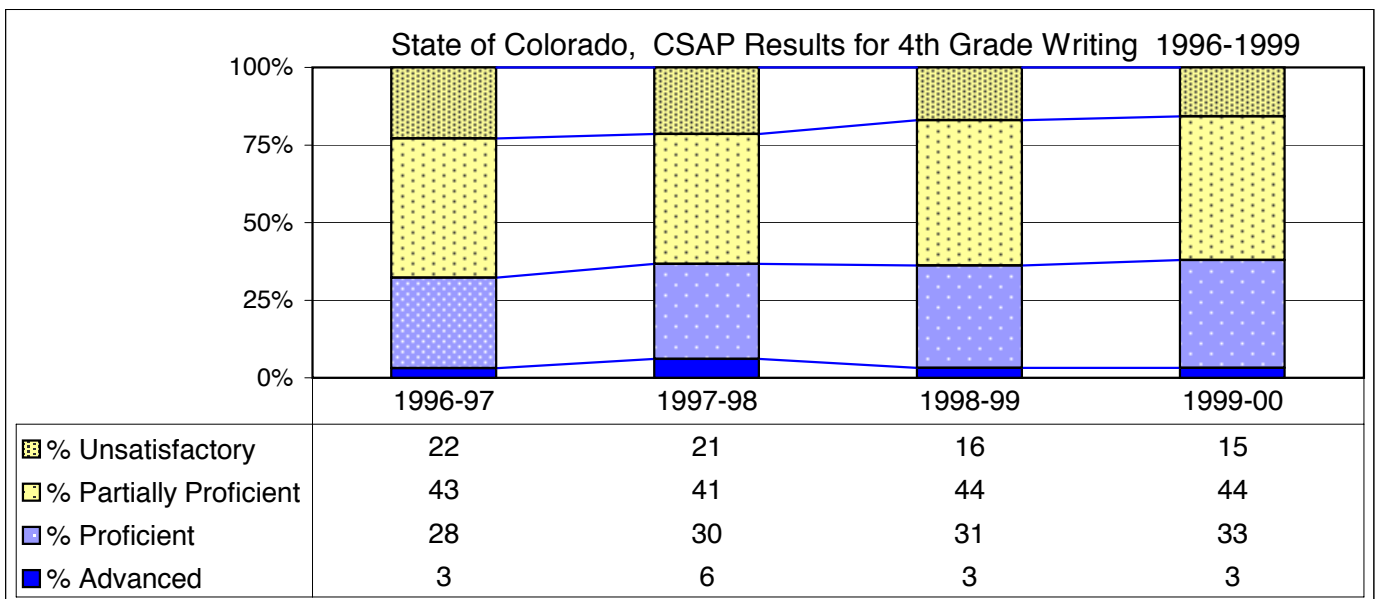
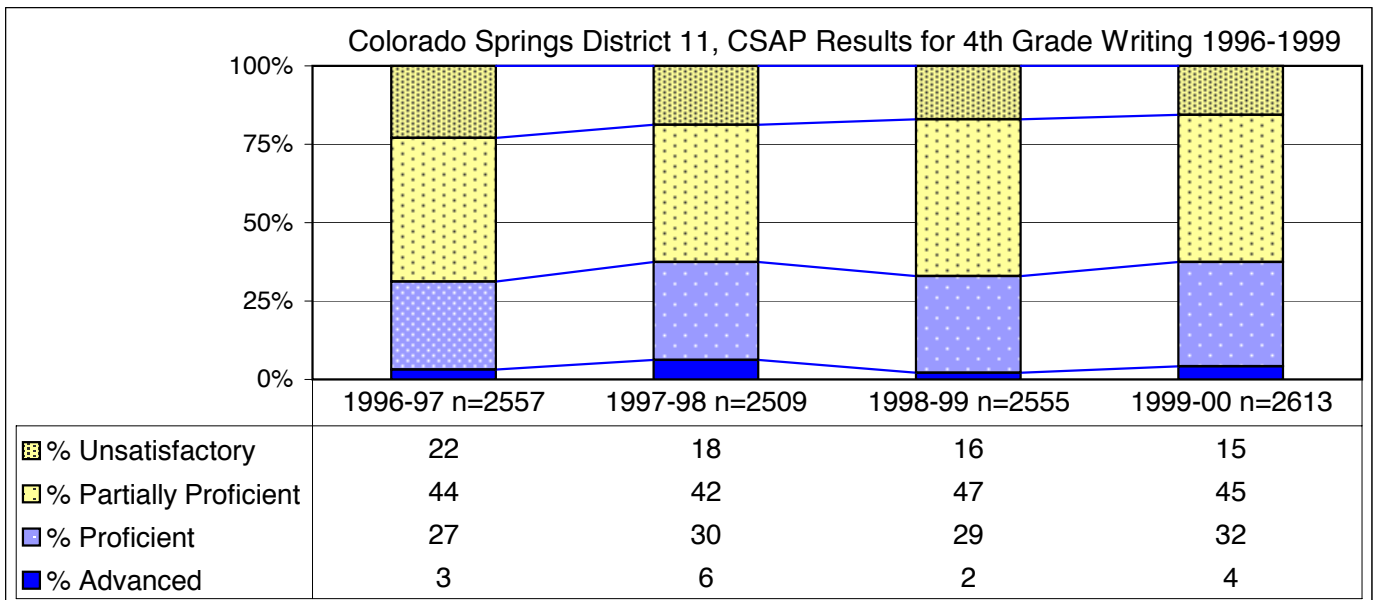
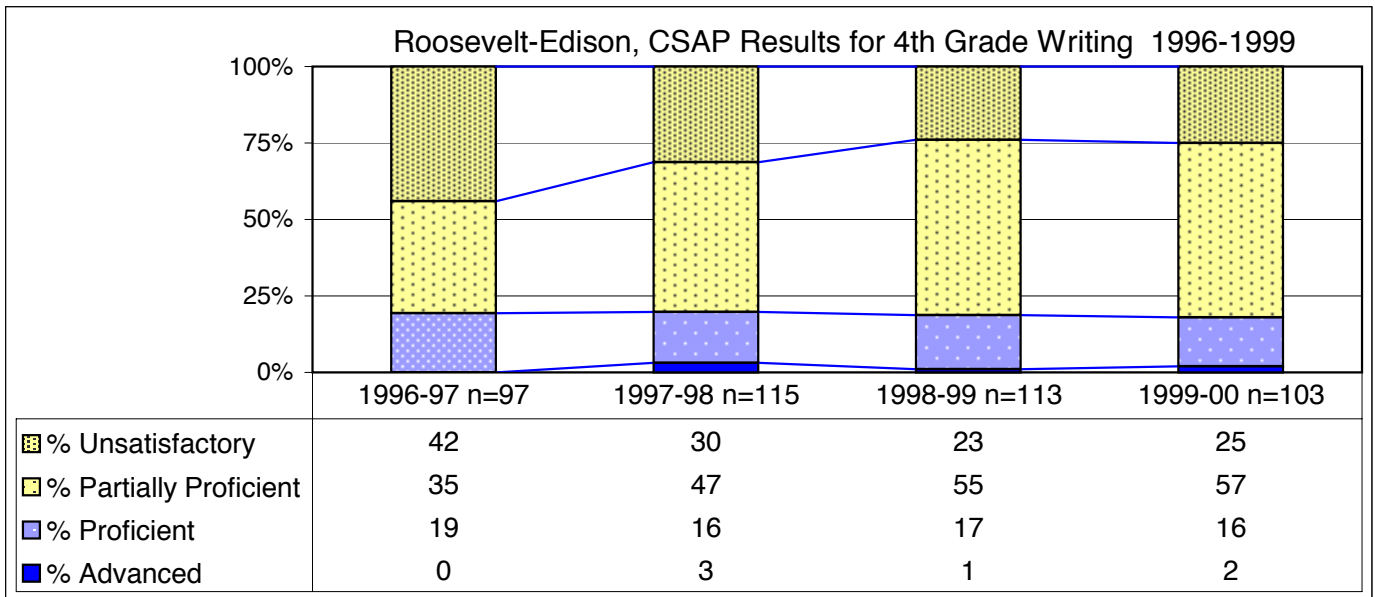


Exhibit 3:7 Performance on Grade 5 Math for Roosevelt-Edison, District, and State

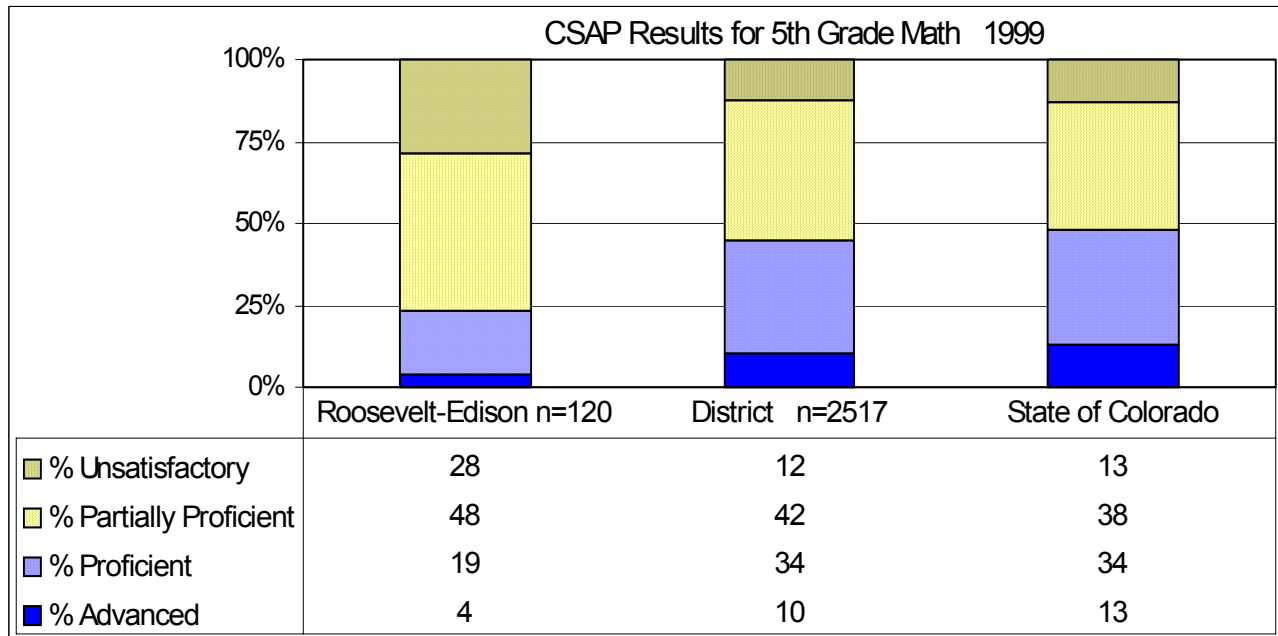


Table 3:4 presents the OR summary data from the fourth grade analyses. Breslow-Day analysis indicated that a common OR for the reading test was meaningful. Students at Roosevelt-Edison were about three times more likely to fail this test relative to students in the district and about twice as likely to fail relative to students in the state. The fourth grade CSAP also includes a writing test. Roosevelt-Edison students showed a statistically significant improvement in performance on this test over the two-year period relative to district students. As can be seen from Table 3:4, the OR dropped from 2.641 to 0.972. Roosevelt students improved from being about 2.5 times more likely to fail to even odds relative to district students. Performance relative to the state did not show this dramatic improvement. A common OR still indicated that Roosevelt-Edison students were about twice as likely to fail the CSAP writing test in the fourth grade.

Only one year of data was available for the fifth and seventh grade CSAP test. Table 3:5 presents the summary findings for the OR analyses. In both analyses students at Roosevelt-Edison performed at levels significantly below students in the district or the state. Individual ORs ranged from a low of about 2.7 to 3.0, suggesting that students at Roosevelt-Edison were much more likely to fail the CSAP mathematics test in the fifth grade than were fifth graders in the district and state.

Table 3:5 Summary of Grade 5 CSAP Odds Ratio Findings for Roosevelt-Edison (1999)

	OR	95% CI
<i>Odds of not meeting standard compared with district</i>		
5 th Grade Mathematics	2.769	1.768 - 4.058
<i>Odds of not meeting standard compared with state</i>		
5 th Grade Mathematics	3.029	2.025 - 4.531

3.6 Summary

While the students, on the whole, made gains on the norm-referenced test (i.e., the Iowa Test of Basic Skills), they did not gain as much, on the whole, as the district and state on the criterion-referenced test mandated by the state (i.e., Colorado Student Assessment Program).

Norm-referenced test findings

A summary score of -1 indicates an unfavorable result for the Edison school, a score of 0 indicates a neutral finding, and a score of 1 indicates a favorable result according to the criteria specified in Section 2.5. The effect size (ES) is the omega squared (ω^2) for a one way repeated measures ANOVA (Kepple, 1991). Table 3:6 contains our trend ratings and the basis for these ratings for the norm-referenced achievement tests.

Table 3:6 Summary of Results on Norm-Referenced Student Achievement Tests

Cohort A ITBS Grades 3, 4 (1998-99)	Stanine		NPR			NCE			Trend
	p-value	ES	p-value	Δ	ES	p-value	Δ	ES	
Language	<.0001	.332	<.0001	13.9	.297	<.0001	9.9 (+1)	.314	positive (+1)
Math	.1407	.010	.1501	3.0	.009	.0688	3.0 (0)	.019	mixed (0)
Reading	<.0001	.183	<.0001	10.1	.168	<.0001	8.3 (+1)	.178	positive(+1)
Cohort B ITBS Grades 4, 5 (1998-99)	Stanine		NPR			NCE			Trend
	p-value	ES	p-value	Δ	ES	p-value	Δ	ES	
Language	.5309	-.004	.1117	3.1	.011	.2157	1.7 (0)	.004	mixed (0)
Math	.0011	.069	.0006	7.2	.078	<.0001	5.6 (0)	.096	mixed (0)
Reading	.0035	.054	.0005	-8.3	.080	.0046	-5.3 (0)	.051	mixed (0)

Criterion-referenced test findings

Decisions regarding the OR were based on whether or not the $(1-\alpha)$ C.I. included 1.0. If the $(1-\alpha)$ C.I. fell completely below 1.0, this was interpreted as a protective odds ratio (1), thus favoring the Edison school. If the $(1-\alpha)$ C.I. included 1.0 (0), this was interpreted as an equal odds situation. If the $(1-\alpha)$ C.I. fell completely above 1.0 (-1), this was interpreted as an increase in odds for failing the state CRT relative to the comparison sample.

Interpretation of the OR tables: If the Breslow-Day statistic (B-D) is nonsignificant, one overall OR and $(1-\alpha)$ C.I. can be used to represent the odds for failing the CRT relative to the comparison group. Thus, there are no trends reported for each specific year in the tables, only a rating in the B-D

column. If the B-D statistic is found to be statistically significant, then an overall common OR cannot be meaningfully interpreted; that is, there is a statistically significant change in the OR over years and yearly ORs are necessary. Thus, our summary ratings appear for each year of data and not in the B-D column. Table 3:7 highlights the trend ratings we have given to the grade level tests.

Table 3:7 Summary of Results on Criterion-Referenced Student Achievement Tests

Edison vs. District	1998	1999	B-D	Trend
Grade 3 Reading	-1	0		mixed (0)
Grade 4 Reading			-1	negative (-1)
Grade 4 Writing	-1	0		mixed (0)

Note: All comparisons with state as comparison group were negative

Combined ratings

Given the total ratings for the trends that are highlighted in Table 3:8, this school has a mean rating of 0.11 which corresponds to an overall Mixed rating. In its 1999 annual report, Edison rated this school as Strongly Positive; and in its 2000 annual report, it rated the achievement gains since opening as Positive and the achievement gains in 1999-00 as Negative.

Table 3:8 Combined Overall Trends for Roosevelt-Edison

	Positive	Mixed	Negative
Norm Referenced	2 of 6	4 of 6	0 of 6
Criterion Referenced	0 of 3	2 of 3	1 of 3
TOTALS	3 of 9	4 of 9	2 of 9

The design behind the trends in the norm-referenced results are based on tracing individual students over 2 years. While some would argue that this is a better design than was used with the criterion-referenced results (tracing consecutive cohorts of students), one also has to consider the sample sizes. The sample size for the norm-referenced test was very small. While we received a database from Edison with nearly 500 records to perform this analysis, we were able to trace only 64 students in 1 cohort and 71 students in the other cohort. On the other hand, the test administration procedures are likely to be more regulated for the state-mandated CSAP, and the results we have for these trends include a much larger proportion of the total enrollment at the school.

Our findings are similar to what the district found on the District Achievement Level Tests (DALT). Results on the DALT indicate that the school is consistently below the district levels, but the gains made by Roosevelt-Edison on the DALT largely parallel the gains made by the district (i.e., the charter school is not gaining more than the district over time). The results to date indicate that this Edison school does not differ substantially from other district schools.

Chapter Four

Henry E. S. Reeves Elementary School, Miami-Dade County, Florida

4.1 Descriptive Summary of School

Henry E. S. Reeves Elementary School is a district contract school that Edison began operating in 1996 to serve grades K-5. As reported in Table 4:2, the school has a total of nearly 90 staff members, and during the 1999-00 school year, the school enrolled 1,161 students, making this a very large elementary school. According to Edison Schools Inc. (Edison, 1999), the student mobility in 1997-98 was 2 percent, and the student/staff ratio was 15.6:1. Data provided by the state and district (see Table 4:2) conflict with the Edison figures. This may be due to a difference of definitions.

With just over 1 percent of its students being white, Reeves had a far lower percentage of white students than the state of Florida (55.8 percent) and the district (13 percent). The school also had a lower percentage of Hispanic students (17 percent) when compared with the district (52 percent) and the state (16.4 percent) in 1997-1998. Table 4:1 contains additional figures for the 1997-98 school year. During the 1999-00 school year, 87.9 percent of Reeves' students qualified for free/reduced lunches, while the district average was 70.1 percent and the state average was 53.2 percent. Table 4:2 contains further details about student characteristics.

The total enrollment for Henry E.S. Reeves Elementary School has remained relatively stable, only fluctuating by 20 students over the years between 1996 and 1999 (1,081 students to 1,193 students). Both the district and the state have similarly stable enrollment patterns, with no dramatic increases or decreases. The average class size has grown during the time the school has been operated by Edison, increasing from 25.7 students per class in 1996-97 to 29.8

Table 4:1 Distribution of Students by Ethnic Grouping (1997-98)

Ethnicity/Race of students	Reeves Elementary	Miami-Dade District	State of Florida
Black	81.0%	33.4%	25.3%
Hispanic	17.0%	52.0%	16.4%
White	1.0%	13.0%	55.8%

students per class in 1998-99, and then back to 27.6 in 1999-00. During the same period of time, the average class size remained consistent in the district at approximately 25.5 students per class and at the state level at approximately 23.5 students per class.

The percentage of students at Reeves with limited English proficiency has increased from 15 percent in 1996-1997 to 17.2 percent in 1998-1999, and then down to 9.2 percent in 1999-00. The district had a higher percentage of students with limited English proficiency (about 25 percent in 1999-00), while the state had a similar percentage of students (9.3 percent in 1999-00) with limited English proficiency. The percentage of Reeves' students who have disabilities has remained at about 3.7 percent, while both the district and the state have maintained much higher percentages (i.e., 9.8 percent and about 15.1 percent, respectively in 1999-00). In terms of gifted students, the Edison school had 4.9 percent of its students in the gifted program in 1999-00, as compared with 6.5 percent in the district and 3.9 percent in the state.

The promotion rate overall for Henry E. S. Reeves Elementary, the district, and the state has decreased slightly, but has remained above 95 percent. The district and state attendance rates reflect a steady 8 to 9 percent of students absent for 21 or more days, but the rate at Reeves has been higher with 20.7 percent in 1998-1999 and 9.2 percent in 1999-00.

The total number of staff has increased from 69 in 1996-1997 to 89 in 1999-00. The district and state both increased their total number of staff during these years also. Of the staff population at Henry E. S. Reeves Elementary in 1999-00, 70.8 percent are instructional staff, 2.2 percent administrative, and 27 percent are support staff. The district has similar staffing figures (close to 72 percent of the district staff are instructional, 3.3 percent are administrative, and 25.1 are support staff). The instructional staff at Henry E. S. Reeves is very different in terms of formal training and years of experience. Reeves teachers have substantially fewer years of experience, and a much lower proportion of their teachers have advanced degrees. Table 4:2 contains specific details about the differences between Reeves, the district, and the state. A close examination of the figures in Table 4:2 suggest that there is a high rate of attrition among teachers at Reeves since the average years of experience are decreasing each year rather than increasing, as one would expect. The average years of experience of teachers at Henry E. S. Reeves has decreased from 3.4 years in 1996-1997 to 1.8 years in 1999-00. During the same period the average years of experience of district teachers increased from 11.6 in 1996-1997 to 12.1 years in 1999-00.

Expenditures per pupil for exceptional, regular, and at-risk students, as well as for the school operating costs per pupil at Henry E. S. Reeves, have all increased over the past 4 years. Table 4:2 includes specific figures broken down by student groups and years. The per-pupil expenditure at Reeves increased from \$3,370 in 1997-1998 to \$5,117 in 1999-00, while the district increased from \$4,925 to \$5,365 during the same period.

Table 4:2 School, Student and Teacher Background Information for Henry E.S. Reeves Elementary, Miami-Dade, and the State of Florida

	Henry E.S. Reeves				Miami-Dade District				State of Florida			
	1996-97	1997-98	1998-99	1999-00	1996-97	1997-98	1998-99	1999-00	1996-97	1997-98	1998-99	1999-00
Student Characteristics												
Number of Students (Fall)	1,081	1,076	1,193	1,161	176,674	174,393	175,161	176,705	1,127,315	1,138,979	1,150,251	1,164,051
Average Class Size	25.7	26.9	29.8	27.6	25.6	25.5	25.6	24.8	23.8	23.6	23.7	23.3
Free/Reduced-Price Lunch %	86.3%	90.9%	87.2%	87.2%	70.7%	70.1%	69.8%	70.1%	52.4%	52.2%	52.8%	53.2%
Limited English Proficient %	15.0%	12.6%	17.2%	9.2%	21.7%	21.5%	21.2%	25.3%	7.7%	7.9%	8.2%	9.3%
Students with Disabilities (%)	3.6%	3.7%	4.2%	3.7%	8.8%	9.2%	9.5%	9.8%	14.4%	14.8%	14.9%	15.1%
Gifted %	2.6%	2.7%	5.8%	4.9%	5.0%	0.1%	0.6%	6.5%	3.8%	3.9%	3.9%	3.9%
Mobility %	99.7%	169.8%	608.8%	142.3%	31.3%	31.4%	43.5%	34.5%	33.3%	32.1%	35.0%	33.3%
Promotion Rate %	99.6%	95.8%	98.7%		99.0%	98.6%	98.2%		97.8%	97.0%	96.3%	
Absent 21+ Days (%)	9.1%	9.4%	20.7%	9.2%	9.7%	8.5%	7.4%	6.7%	9.3%	8.7%	7.5%	6.2%
School Staff												
Total	69	82	82	89	13,977	14,019	14,283	14,653	104,662	107,331	110,007	113,338
Instructional (%)	85.5%	80.5%	78.0%	70.8%	72.0%	72.0%	71.7%	71.7%	63.1%	63.0%	63.1%	62.6%
Administrative (%)	1.4%	125.0%	1.2%	2.2%	3.3%	3.3%	3.2%	3.3%	2.6%	2.5%	2.6%	2.5%
Support (%)	13.0%	18.3%	20.7%	27.0%	24.8%	24.7%	25.1%	25.1%	34.3%	34.4%	34.3%	34.8%
Teachers												
Master's Degree or Higher %	36.2%	23.5%	20.9%	16.4%	42.3%	43.0%	42.9%	43.2%	31.8%	31.5%	32.0%	32.5%
Average Years of Experience	3.4	2.6	2.5	1.8%	11.6	11.5	12.2	12.1%	12.6	12.3	12.5	12.9%
Per Pupil Expenditures/Finance												
Exceptional (\$)		\$4,180	\$5,866	\$6,365		\$7,385	\$8,068	\$8,051		\$6,555	\$6,880	\$7,092
Regular (\$)		\$3,118	\$4,935	\$4,752		\$4,315	\$4,263	\$4,608		\$3,902	\$4,024	\$4,247
At-Risk (\$)		\$6,571	\$7,897	\$10,026		\$5,106	\$5,427	\$6,194		\$4,827	\$5,081	\$5,383
Vocational (\$)		\$0	\$0	\$0		\$4,243	\$4,438	\$4,949		\$4,422	\$4,714	\$4,879
School Operating Costs (per pupil)		\$3,370	\$5,155	\$5,117		\$4,925	\$5,091	\$5,365		\$4,507	\$4,692	\$4,874

Note: The data in this table were derived from various Florida School Indicators Reports (Florida Department of Education, 2000b).

4.2 Past Studies and Evaluations

Miami-Dade County Public Schools (MDCPS) contracted with Edison to take over the operation of the Henry E. S. Reeves Elementary School. The contract was for a period of five years, beginning with the 1996-97 school year, and stipulated yearly evaluations to be conducted by the public school district in conjunction with Edison. To date, the Office of Educational Evaluation, in conjunction with Edison, Inc., has conducted three annual Interim Reports that were prepared by Drs. Joseph Gomez and Sally Shay for the 1996-97, 1997-98, 1998-99 school years (Gomez & Shay, 1998, 1999, 2000). A fourth evaluation of longitudinal student achievement outcomes was conducted by Sally Shay and reported in her dissertation (Shay, 2000).

4.3 Summary of Findings by the Miami-Dade County Public School District

In this section, we provide a summary and comparison of the methods and findings as reported in the first, second, and third year interim reports dated February 1998, March 1999, and June 2000 respectively.

Objectives and evaluation questions

According to MDCPS evaluation reports (Gomez & Shay, 1998, 1999, 2000) four general areas were evaluated: (1) actual implementation of the Edison model in the school, (2) raising the academic achievement of all students to the highest level possible, (3) increasing parent involvement and satisfaction to levels consistent with educational excellence, and (4) improve school climate in the many ways necessary to foster greater learning.

Methodology

To address the four evaluative questions, the MDCPS evaluation team utilized a wide variety of research methods. The various methods used by the MDCPS evaluation team are described below.

Implementation of the Edison model. Unstructured interviews were conducted with the principal. The interviews primarily focused on the school's progress in implementing the basic elements of the Edison model. Teacher surveys attempted to (1) assess the teachers' perceptions of the Edison model, (2) rate the extent of implementation of the basic elements of the model, and (3) compare the Edison model with their previous experiences with other district schools. Classroom observations were also conducted. In the first interim report, classroom observations were conducted near the end of the fourth grading period (end of the year). For the second and third interim reports, classroom observations were obtained at the beginning of the third grading period and near the end of the fourth grading period. Classroom visitations, which were conducted by members of the evaluation team, were unannounced and randomly determined.

Student's academic achievement. The primary source of student achievement data were student scores on the Stanford Achievement Test 8th Edition (SAT-8). MDCPS routinely administers the SAT-8 in the spring to all students in all grade levels except kindergarten and grades 10 and 12. The analysis of student performance on the SAT-8 was examined by way of a nonequivalent control group design limited to students who were enrolled in Reeves in year 1. Details of this evaluation component can be found in the third interim report: 1998-99 (Gomez & Shay, 2000). In addition to the SAT-8 analyses, a comparative analysis of student performance on the Florida Writing Assessment was conducted. Finally, the evaluation team examined overall student progress in attaining the curriculum standards of the Edison model.

Involvement of parents. Various parent surveys were conducted. In the first year the School Climate Survey was used (annually administered by MDCPS). In the second year this instrument was supplemented with the parent satisfaction survey. A final source of information on parent involvement was from data obtained from project records that detailed parents' participation in school-related activities.

School climate. The school climate was evaluated by comparing survey responses from teachers at Reeves with responses from teachers in control schools.

As can be seen, MDCPS has conducted and made available its extensive evaluation of the progress Edison has made in Henry E. S. Reeves Elementary School relative to the four evaluation goals it set out to examine. For a fuller appreciation of these evaluations, please refer to the three interim reports (Gomez & Shay, 1998, 1999, 2000).

Summary of findings from MDCPS evaluations

Below we have summarized the main findings and, where appropriate, referred to the year of the evaluation. The findings are grouped according to specific questions that were addressed in the MDCPS evaluation.

Implementation of Edison model. The first interim report covering the academic year 1996-97 indicated that all elements of the model (21 in all) were either fully or partially implemented. There was general consensus that the model was fully implemented in year 2, an improvement from year 1. However, in the third interim report, the data did not support full implementation of the Edison model. Specifically, the teachers reported that at least one element of the model was not fully implemented. However, the third interim report indicated that the model as a whole was fully implemented.

Students' academic achievement. Data pertaining to this goal were treated separately depending on the specific type of student data analyzed: SAT-8, Florida Writes, or students' attainment of Edison's curriculum standards. In year one the performance of the control students was better at nearly every grade level in both reading and math than the performance of the Edison students, and the evaluation team concluded that "the analysis of the SAT test results have revealed that to date the project students have not performed as well academically as their counterparts in the regular MDCPS

program” (Gomez & Shay, 1998, p. 41). However, Henry E. S. Reeves’ students significantly outperformed control students on the Florida Writes assessment in year one. Finally, in its interim report for 1996-97, the MDCPS evaluation team summarized the findings from their evaluation: “To date, the analysis of the data has yielded no evidence that the model will eventually ‘raise the academic achievement of the [project] students to the highest level’” (Gomez & Shay, 1998, p. 44). The third data component for evaluating student achievement outcomes was based on subjective judgments of the percentage of students attaining the 100 curriculum standards specified by Edison. Results indicated a mixed picture of whether or not students actually met the curriculum standards.

In the year two report, Edison students generally compared favorably with control students in SAT reading, and the differences from year one mathematics were much narrower. Contrary to year one findings, on the Florida Writes assessment where the project students significantly outperformed control students, in year two the difference was lost. Edison and control students scored at equivalent levels on this test. In year two the MDCPS evaluators’ report was generally positive in terms of the students’ attainment of the curriculum standards set by Edison Schools Inc.

In the year three interim report, analysis of the SAT achievement indicated that the gains in year two were not capitalized on. In fact, in the year three analysis, some of the achievement gains in year 2 were actually lost. Gomez & Shay (2000) offered this interpretation: In year one the achievement of the project students was below that of the control students, but in year two these differences were significantly reduced. However, in year three much of the gains in year two were lost, especially in mathematics. Thus, by year three, the project students were able to overcome their poor performance in year one, but by year three only the reading performance had risen to a level comparable to the control students. Project students’ performance in mathematics had not yet reached a level comparable to the control students. Performance on the Florida Writes assessment in year three presented a similar picture, There was no evidence, based on the FWA, “that indicates the Edison model has produced an advantage in student’s writing proficiency” (Gomez & Shay, 2000, p. 50). In the analysis of the attainment of Edison curriculum standards, Gomez and Shay (2000) concluded that there was still insufficient evidence to report that the teachers had fully attained this goal. And in their summarizing sentence (2000, p. 52) they concluded “. . . the project has yet to attain the first and most important of its stated objectives: To raise the academic achievement of all students to the highest level possible. . . the project students’ academic achievement at its best can be deemed only comparable to that of their counterparts in the regular MDCPS program.”

Involvement of parents. The third goal was to increase parent involvement and satisfaction to levels consistent with educational excellence. In the first year, the MDCPS evaluators’ reported that the available data were sufficiently mixed and thus reported that there was inconclusive evidence to report that the project had met this goal. However, in year two they reported a direct improvement regarding this goal. Moreover, in year three, continued progress was made toward increasing “parent involvement and satisfaction to levels consistent with educational excellence.”

School climate. The final goal of Edison, Inc. is to “improve school climate in the many ways to foster greater learning.” In year one MDCPS evaluators concluded that during the initial year the project failed to make adequate progress in attaining this objective. In year two progress was made toward meeting this objective, but in year three the progress evidenced in year two had waned.

4.4 Longitudinal Student Outcomes: The Shay Study

Shay (2000) recently conducted a comprehensive longitudinal (growth curve) analysis of achievement outcomes of the students at Henry E. S. Reeves Elementary School. In her study, Shay examined the academic achievement (SAT-8, scaled scores) of students in Reeves relative to a control group from 1996-97 to 1998-99. This study builds upon the official MDCPS interim reports (Gomez & Shay, 1998, 1999, 2000) by conducting the first longitudinal analysis of student achievement. In this study two panels of students were followed. Panel A (n=114) progressed from second grade to fourth grade and Panel B (n=159) students progressed from third to fifth grade. Control groups for each panel were constructed by taking stratified random samples of students who attended other district schools. Stratification variables insured comparability among the groups in terms of demographics and pretest performance, despite attrition. Achievement results in reading and mathematics were examined for each panel. Two different analytical approaches were examined, a traditional mixed model repeated measures ANOVA and a three-level hierarchical linear model (HLM) employing individual growth curves. Both the more traditional repeated measures analysis and the HLM analysis indicated statistically significant levels of growth over the three year period in both subject areas for both groups (Edison and control). In reading, no statistically significant differences could be attributed to group membership. However, the results for mathematics presented a more complex picture. The repeated measures analysis indicated a significant group by year interaction in both panel analyses, suggesting that the rate of growth was different in the two groups. Table 4:3 presents descriptive statistics for each group by year.

In parallel analyses, Shay used a 3-level HLM model with individual variation over time as the level-1 effect, individual variation with the same group as the level-2 effect, and between group variation as the level-3 effect to test for differences in longitudinal growth in achievement between the two groups. HLM results for reading paralleled the previous repeated measures analyses for both panels. Significant growth was observed as students progressed through the grades, while individual growth curves varied significantly, growth rates did not differ significantly, and there were no differences that could be attributed to group membership. As expected from the repeated measures analysis, HLM results for mathematics depicted a different picture. In panel A, not only was there significant growth within individuals, there was significant difference in growth among individuals within a group but not between groups. Panel B results presented a different picture. Once again there was significant individual growth but at comparable rates within a group. However, in Panel B there was also a statistically significant difference in the group growth rate.

Table 4:3 Descriptive Statistics for Reeves and Control Groups by Year

	Henry Reeves				Control			
	Mathematics		Reading		Mathematics		Reading	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Panel A								
Year 1	561.03	38.30	553.64	41.79	564.82	37.53	570.82	37.68
Year 2	595.10	37.08	595.11	53.56	599.79	35.59	594.74	39.26
Year 3	607.62	39.04	615.24	58.88	613.81	35.68	627.18	47.86
Panel B								
Year 1	576.29	32.64	563.51	42.54	579.39	33.28	573.93	45.51
Year 2	592.72	37.87	597.02	46.07	594.86	36.68	608.09	48.99
Year 3	617.65	35.47	627.97	42.77	619.16	35.93	628.13	44.69

Source: Adopted from Shay (2000, p. 49).

Fairly consistent results were obtained in Shay's study, which indicated that at the end of three years of enrollment in an Edison school there were no differences in achievement, as measured by the SAT reading and mathematics subtests, that could be accounted for by group membership. The main research question posed by Shay was whether the students who attended one elementary school operated by Edison Schools Inc. in Miami-Dade County made greater academic progress than comparable students who attended other district schools. This question is consistent with the focus and purpose of this evaluation. Shay's study is the most rigorous study of Edison achievement to date. In summary, after three years of exposure to the Edison model, students enrolled in the Henry E. S. Reeves Elementary School are no better off than comparable students in other district schools as measured by the SAT.

4.5 Performance on the FCAT and Florida Writes State Assessments

Individual student achievement data (longitudinal) were not provided to us by Edison or by the administration of Henry E. S. Reeves Elementary School, so we could not include this information in this report. The data we have utilized for our analyses came from the two different state-mandated testing programs: FCAT and Florida Writes assessments. Since these data only recently became available, we have only one year of data to report.

Henry E. S. Reeves Elementary School participates in two state-mandated assessments: the Florida Comprehensive Assessment Test (FCAT) and the Florida Writes test. Before we break down the results on these tests for the chi-square and odds ratio analyses, we shall examine the absolute scores and measure gains in the average total scores in comparison to the Miami-Dade County Public School District and the state.

The FCAT is administered in grade 4 (reading) and grade 5 (mathematics), and the Florida Writes is administered in grade 4. Scores on the FCAT can range from 100 to 500. Five student achievement levels categorize students based on the scale score ranges. Level 5 is the highest level and Level 1 is the lowest. The Florida Writes assessment is scored along a 6 point scale with 6 point being the highest and 1 the lowest.

As one can see from the results in Exhibit 4:1, this school performs substantially lower than the district and state on these tests. However, in terms of gains in average scores, the size of the gains for two years on the reading and math tests and for three years on the writing test showed that Henry E. S. Reeves gained more than the district and state. The largest gains were on the reading test where the Edison students gained 13 points, while the district lost 2 points and the state lost 3 points. Because the total scores run up to 500, and because the level of performance at the Edison school is considerably lower, it is hard to determine if this gain is substantial or not. In the next sections, we will examine the data more closely and measure whether or not the differences between the Edison school and the district and state are significant, as well as whether the odds of not meeting state standards have increased or decreased at the Edison school.

Exhibit 4:2 illustrates the performance of the school in terms of the established performance levels for the FCAT and Florida Writes test. The charts in Exhibit 4:2 include results for the 1998-99 school year only since we were not able to obtain data broken down by levels for the previous year.

4.6 Chi-Square Analysis of FCAT Data

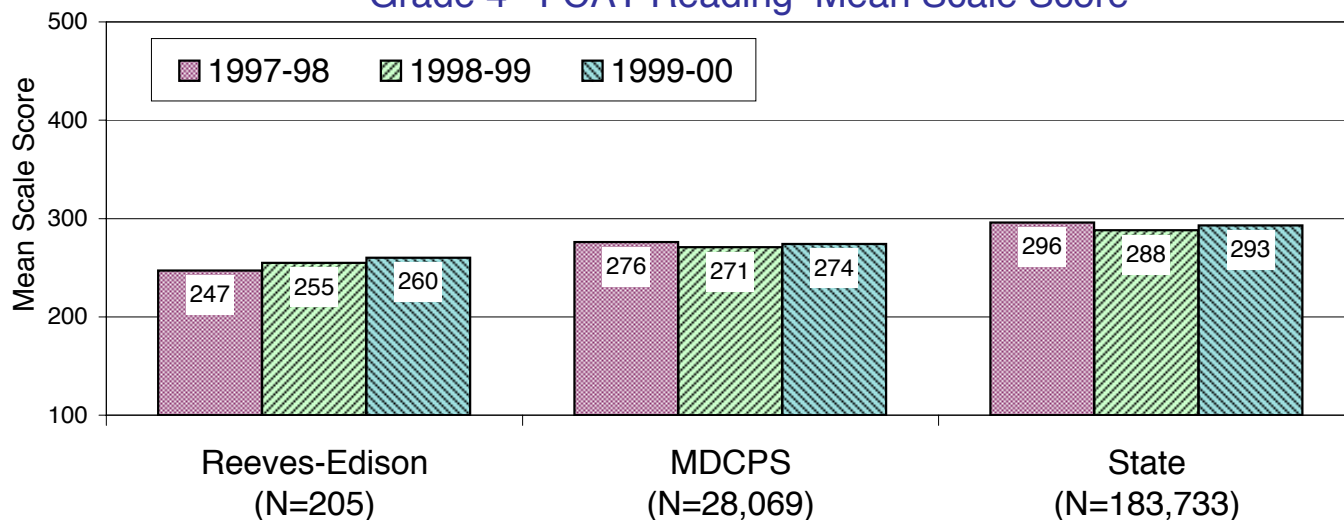
Available data

A chi-square analysis was initiated on data made available from the state of Florida on the outcomes of the Florida Comprehensive Assessment Test (FCAT) and the Florida Writes test, the state-mandated criterion-referenced tests. While the chi-square results can help us distinguish the size and strength of the differences between Henry E. S. Reeves Elementary School and the two control groups we compare it with, our main interest is in the change over time at this Edison school.

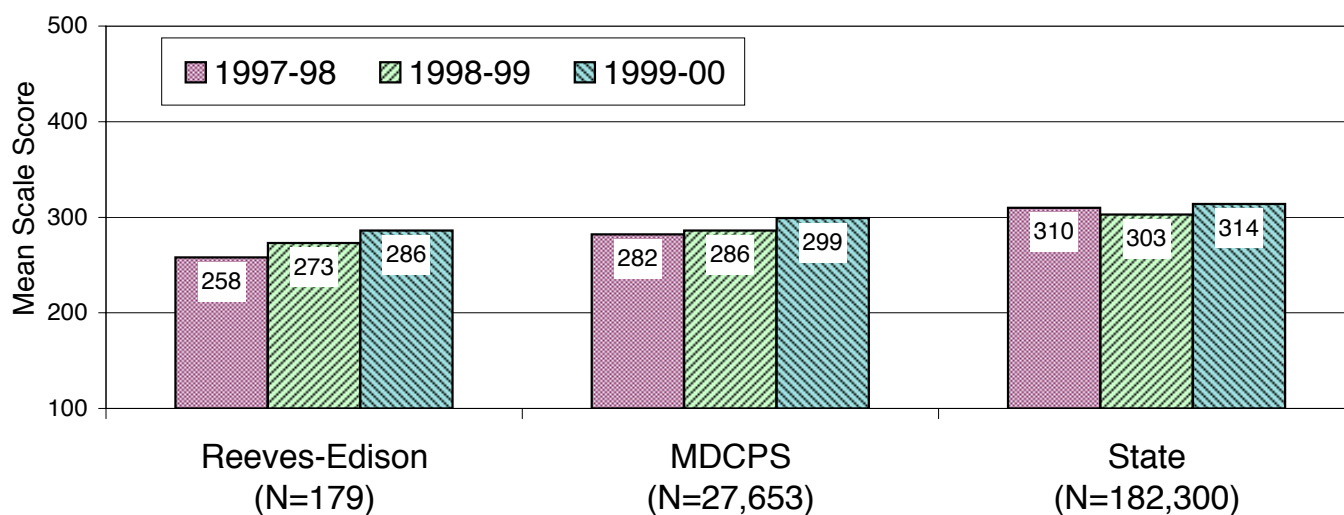
The FCAT was administered in January and February 1998 in order to establish baseline information on the achievement of Florida students and schools. The results for this first year were not broken down in the five scoring levels, so we could not use this data for chi-square and odds ratio analyses. The results from the February 2000 analysis were not available when we conducted our analysis for this case, so we were left with just one year of data for the chi-square and odds ratio analyses. However, we were able to obtain the mean standard score for three years on the Grade 4 and 5 FCAT results, which are described in Exhibit 4:1.

Exhibit 4:1 Henry E. S. Reeves Elementary School, FCAT and Florida Writes Results 1997-1999
Performance on Grade 4 and 5 Tests Compared with Local School District and the State

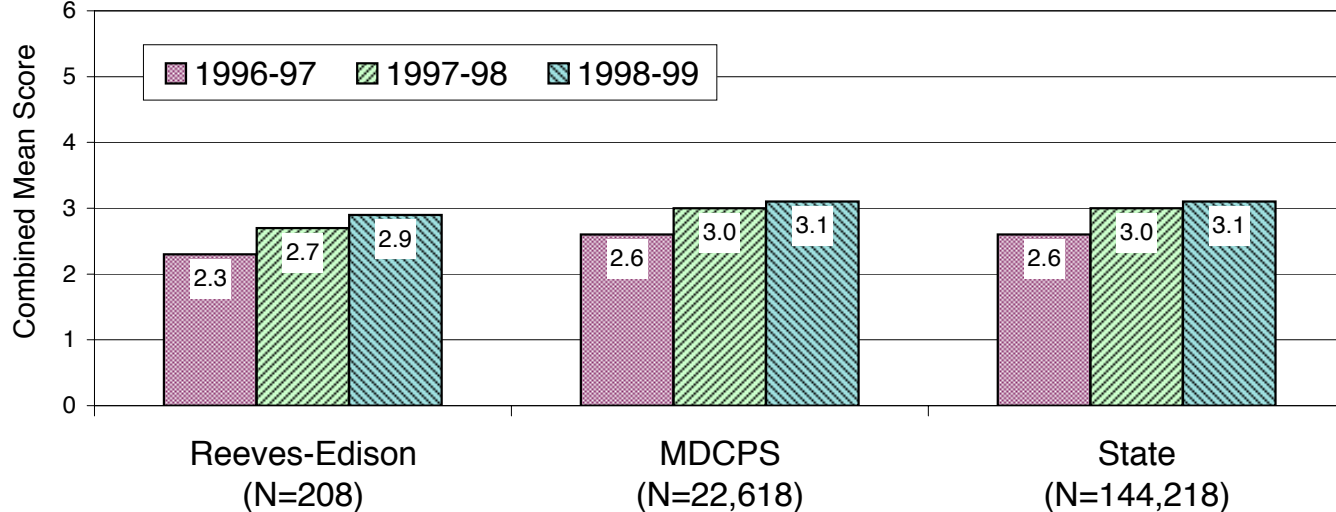
Grade 4 FCAT Reading Mean Scale Score



Grade 5 FCAT Math Mean Scale Score



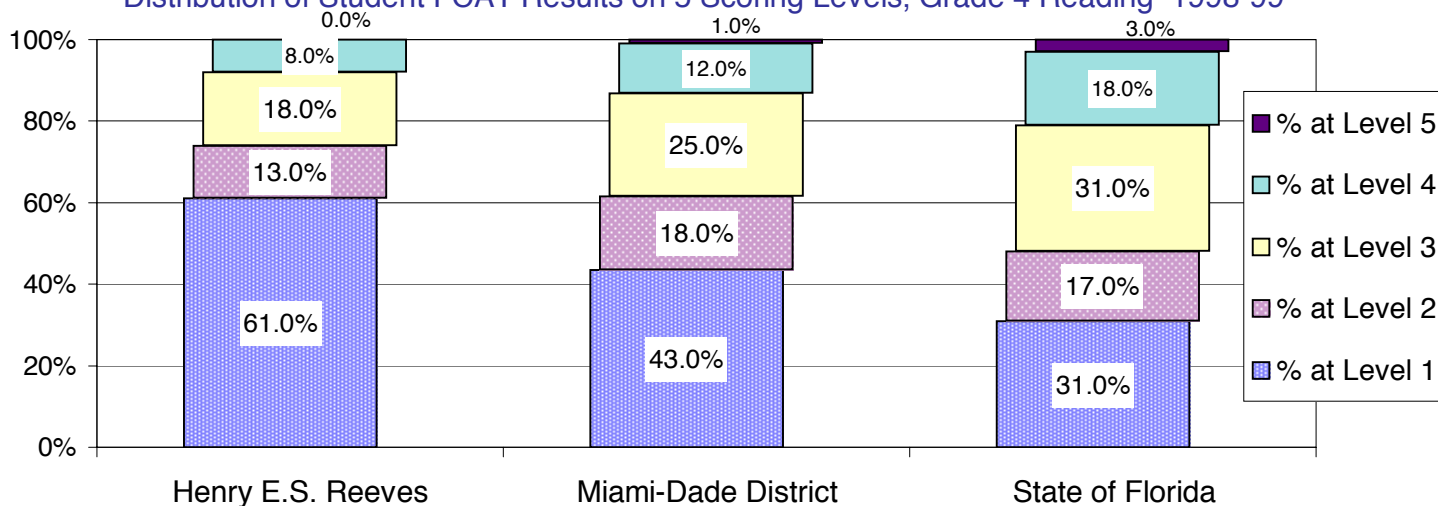
Grade 4 Florida Writes Combined Mean Score



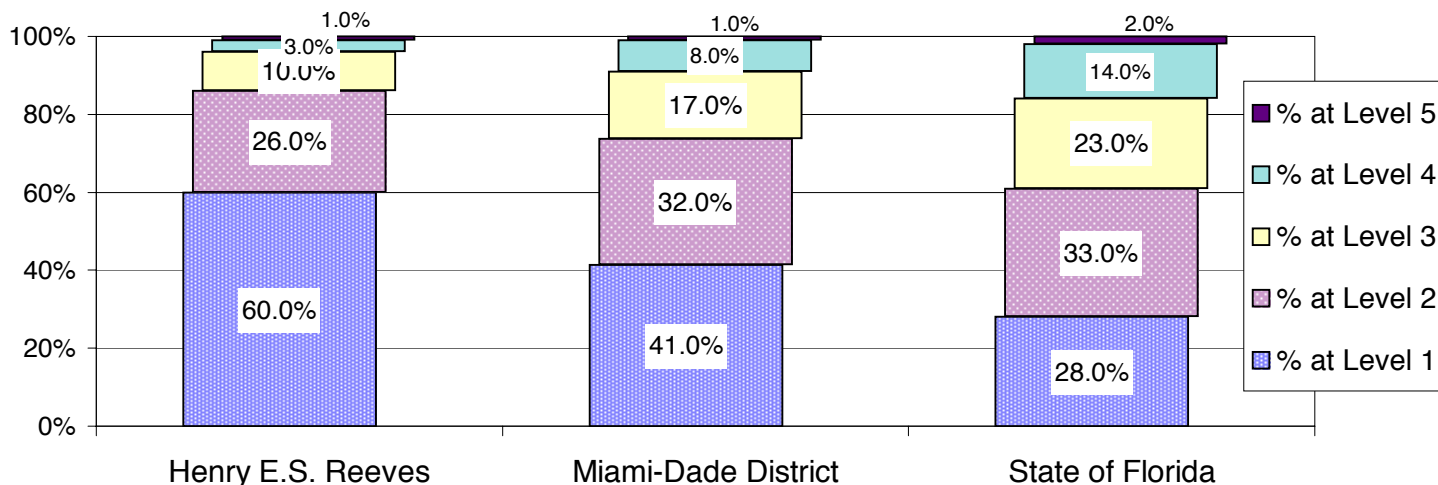
Note: The N represents the number of students taking the test in 1999-00.

Exhibit 4:2 Henry E.S. Reeves Elementary School, FCAT and Florida Writes Results 1998-1999
Performance on Grade 4 and 5 Tests Compared with Local School District and the State

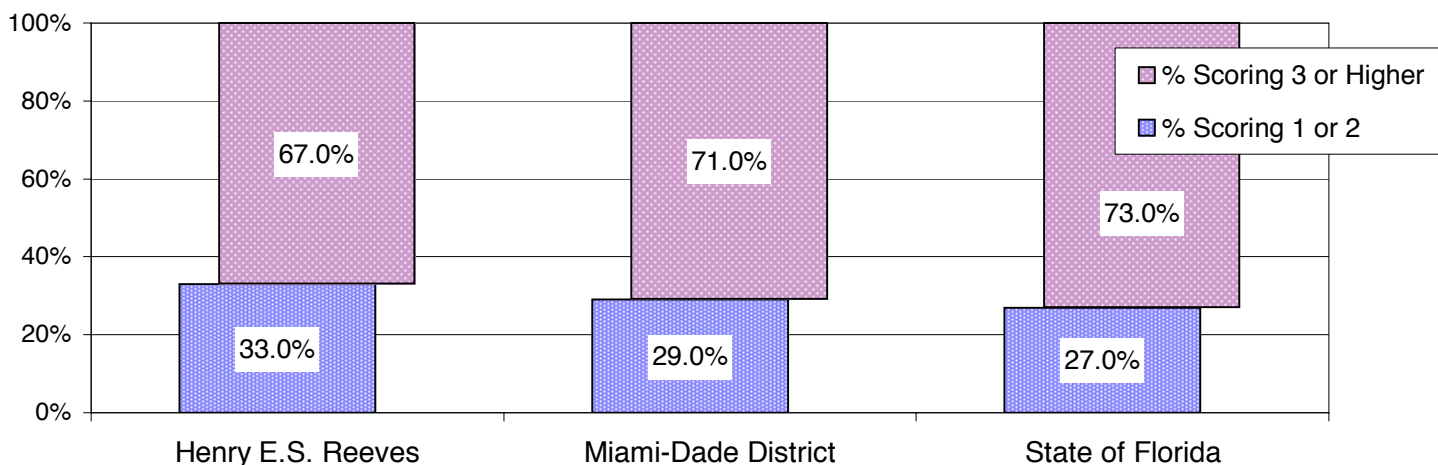
Distribution of Student FCAT Results on 5 Scoring Levels, Grade 4 Reading 1998-99



Distribution of Student FCAT Results on 5 Scoring Levels, Grade 5 Math 1998-99



Distribution of Students Results on the Florida Writes Test Grade 4 1998-99



Source: Results were derived from Florida Indicators Report, 1996-1999 (FDE, 2000).

Note: The FCAT results are reported according to 5 scoring levels with Level 5 being the highest and Level 1 the lowest.
The Florida Writes assessment is scored along a 6-point scale, with 6 being the highest and 1 the lowest.

Construction of the comparison groups

We constructed two different comparison groups for the chi-square analyses. Since we were interested in examining the number/proportion of students who met state standards (“passing”) or conversely the number/proportion of students who did not meet state standards (“failing”) on the FCAT, we needed to define a suitable comparison group. Our first comparison is with the district (i.e., Miami-Dade Public Schools). The second comparison group we selected was the state average passing/failing rates.

General procedure

Utilizing published data from the state of Florida, we made comparisons for the 1999 administrations for both grades 4 and 5. Percentage data (students in each scoring category) were converted to raw frequency data prior to chi-square analysis. To insure independence of the rows in the chi-square tables, the raw frequencies for each scoring category of the FCAT in the district and state comparisons were down-weighted by subtracting the number of students in that category from Henry E. S. Reeves. Thus, both the district and state numbers reflect all students in the district or state exclusive of those in Henry E. S. Reeves Elementary School.

Four chi-square analyses were evaluated for each grade level. Two of these analyses were on uncollapsed data, that is, all scoring levels were represented in the contingency table (e.g., a 2x5) for the district and state comparisons. Two follow-up analyses were conducted on the data after collapsing the multilevel scoring into a dichotomy (pass, fail), thus producing 2x2 contingency tables. For the purpose of our chi-square and odds ratio analyses, we defined Level 1 as “not proficient or fail” and levels 2 and higher as “proficient or pass.” We chose this grouping because between 40 and 60 percent of the students fall into Level 1, and the description of the five levels suggests that students who fall into Level 1 have not met the Sunshine State Standards. For the Florida Writes test, we rated scores 3 and above as passing, since this is both close to the mean score and many Florida districts as well as the Florida Department of Education report the proportion of students that score 3 or higher.

Chi-square findings

These chi-square analyses are testing the null hypothesis that the relative frequency (of students) in the four (or two) scoring categories are the same for Henry E. S. Reeves and the comparison group (either district or the state).

Results of the chi-square analyses for grade 4 covered the 1999 FCAT administration for reading, grade 5 mathematics, and the grade 4 Florida Writes assessment. Individual contingency tables are presented in Appendix E.

The first set of comparisons were made against district data as the comparison group. Four separate chi-square statistics were evaluated from 2x5 contingency tables and four from 2x2 contingency tables (see Appendix E) for the FCAT. Only 2x2 analyses were available for the Florida Writes assessment; however, we were able to secure data from 1998 and 1999.

District and state comparisons on the reading subtest administered in grade 4 closely parallel each other. Statistically significant differences were observed relative to the district and state proportions among the 5 scoring levels. Fewer Reeves students fell into the more advanced performance levels relative to the overall state and district. This pattern was also replicated in the collapsed analyses (2x2).

In grade 5, the mathematics subtest is administered. As seen in Table 4:4 the results of the grade 5 chi-square analyses parallel the grade 4 findings. That is, students at Reeves performed lower relative to students in the district or state in both analyses.

Exhibit 4:1 compares the Reeves' results on the FCAT and Florida Writes assessment with the district and state. The three charts in Exhibit 4:1 indicate that Reeves is behind the district and state but made noticeable progress in Grade 4 reading while the district and state lost some ground. The results for Grade 5 math and Grade 4 writing indicate that Reeves, the district, and the state all show improvements.

Exhibit 4:2 illustrates the breakdown of the FCAT and Florida Writes assessments that were administered in the spring of 1999 according to the five established scoring levels for the FCAT and according to the state standard (3 or higher) on the Florida Writes assessment.

4.7 Odds Ratio Analysis of the FCAT Data

Consistent with state interpretations of FCAT, we grouped all scoring levels above 1 as “proficient or pass” and kept Level 1 as “not proficient or fail” for our odds ratio analyses. As presented in Section 2.4 of this report, the 2x2 tables analyzed in the previous section can be thought of as representing a class cohort in a prospective design. From a classical epidemiological perspective, the students in the Edison school can be viewed as the “exposed” group, that is, exposed to the “Edison-effect,” and students in the comparison group as the unexposed group. From this

Table 4:4 Summary of Chi-Square Findings for Henry E. S. Reeves, 1999

	2x5	2x2
<i>Grade 4, Reading</i>		
Reeves vs. District	sig	sig
Reeves vs. State	sig	sig
<i>Grade 5, Mathematics</i>		
Reeves vs. District	sig	sig
Reeves vs. State	sig	sig
<i>Grade 4, Florida Writes</i>	1998	1999
Reeves vs. District	sig	n.s.
Reeves vs. State	sig	n.s.

Red color indicates a statistically significant difference that favors the comparison group, blue color indicates a significant difference that favors the Edison school.

perspective, each yearly comparison is a new cohort, measured over a period of years. There is a minimal possibility for cohort contamination if a number of students in one group are not promoted to the next grade level. However, we think this represents a very small number of possible cases and therefore has minimal impact on the validity of these analyses. Section 2.4 details the OR statistic and corresponding 1- α confidence interval. We calculated and charted OR for each of the 2x2 tables constructed from the chi-square analyses presented above. Exhibit 4:3 and Tables 4:5 and 4:6 present these findings. Only one year of data could be obtained for these comparisons; thus, a trend could not be estimated. However, Table 4:6 reveals that students in Reeves did not fare well relative to students in Miami-Dade County or the rest of the state. Generally, Reeves students were at least two times more likely to score in the Level 1 category than these other students in reading and more than 3.5 times more likely to score in the Level 1 category in mathematics.

Table 4:6 presents the odds ratio analyses for two years of data on the 4th grade Florida Writes examination. Here, students enrolled at Reeves performed only slightly below their counterparts in the district and state. In both district and state comparisons, the Breslow-Day statistics for testing the homogeneity of OR over the two years were not rejected (significant). Thus, a common OR can be meaningfully calculated. Table 4:6 presents these OR and corresponding 95 percent confidence intervals.

The results from the odds ratio analysis suggests that there is no significant change over time between Reeves Elementary and the district in terms of the odds of meeting or not meeting state standards.

Table 4:5 Summary of FCAT Odds Ratio Findings for Reeves Elementary School

1999	
<i>Odds of not meeting standard compared with district</i>	
4 th grade FCAT Reading	2.043 (1.508 – 2.768)
5 th grade FCAT Math	2.092 (1.546 – 2.832)
<i>Odds of not meeting standards compared with state</i>	
4 th grade FCAT Reading	3.479 (2.603 – 4.650)
5 th grade FCAT Math	3.788 (2.844 – 5.046)
<i>Odds of not meeting standards compared with state</i>	
Reading	2.341
Math	1.809

Red color indicates a statistically significant difference that favors the comparison group, blue color indicates a significant difference that favors the Edison school.

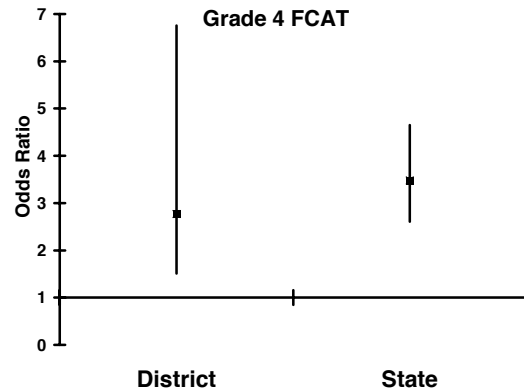
Table 4:6 Summary of Florida Writes Odds Ratio Findings for Reeves Elementary School

	Common OR	95% CI
<i>Odds of not meeting standard compared with district</i>	1.372	1.103 - 1.708
<i>Odds of not meeting standard compared with state</i>	1.472	1.184 – 1.830

Exhibit 4:3 Results of the Odds Ratio Analysis for Henry E. S. Reeves Elementary School (Grades 4 and 5)

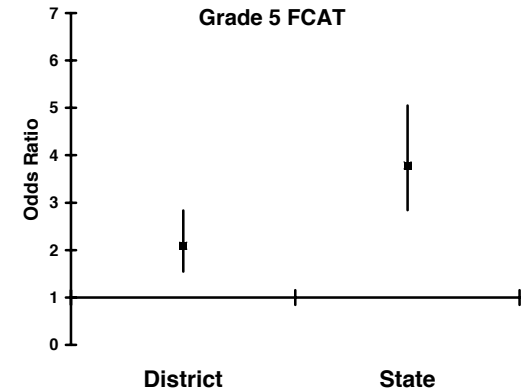
Grade 4 FCAT Reading

Year	U CI	L CI	OR
District	6.756	1.508	2.768
State	4.650	2.603	3.479



Grade 5 FCAT Math

Year	U CI	L CI	OR
District	2.832	1.546	2.092
State	5.046	2.844	3.788

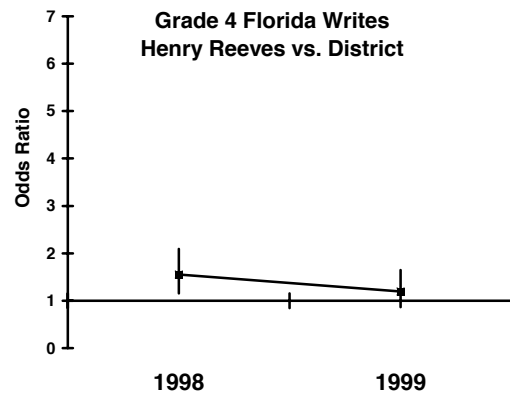


Grade 4 Florida Writes

Year	U CI	L CI	OR
1998	2.092	1.153	1.553
1999	1.648	0.864	1.193

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=45,926) = 1.376, p = .241

OR = 1.372
UB = 1.708
LB = 1.103

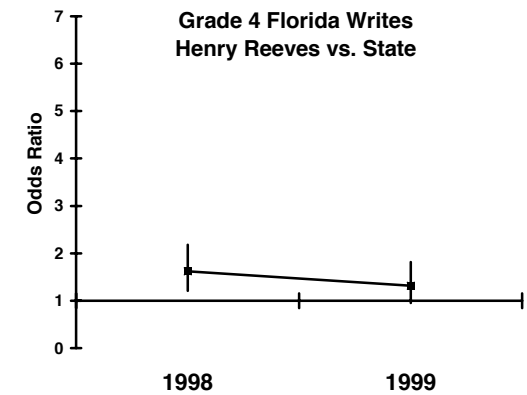


Grade 4 Florida Writes

Year	U CI	L CI	OR
1998	2.182	1.207	1.623
1999	1.815	0.955	1.316

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=304,828) = 0.873, p = .350

OR = 1.472
UB = 1.830
LB = 1.184



4.8 Summary

MDCPS' ongoing evaluation of Henry E. S. Reeves Elementary School is very thorough, likely the most thorough evaluation of any Edison school to date. It utilizes a Silver Medal design (see Appendix C), but falls short in that it only examines change from one year to the next, not longitudinally. For example, the 1998 interim report examined the 1997-98 school calendar while the 1999 report examined the 1998-99 year. Likewise, the 2000 report evaluated only the 1999-2000 year. It would be highly desirable to link these three years of data into one longitudinal evaluation with both cohort and panel samples available. This is exactly what the Shay (2000) study has done. On the positive side, MDCPS developed a plausible and defensible set of criteria to define comparison schools and utilized appropriate statistical techniques for the comparisons they made. This evaluation also looked at other factors, such as parent satisfaction and school climate.

Norm-referenced test findings

Limited data were available for us to gauge trends in NRT data. However, Shay (2000) presented partial effect size estimates (expressed in percentages) from the repeated measures analyses she conducted. Table 4:7 presents our NRT trend ratings based on these estimates. The percentage estimates in this table reflect effect size contributions from both the main effect for group membership and the group by time interaction.

Table 4:7 Summary Results on Norm-Referenced Tests

SAT-8	Partial Effect Size Estimate	Trend
Panel A - Mathematics	$\approx 3.2\%$	Mixed (0)
Panel A - Reading	$< 1\%$	Mixed (0)
Panel B - Mathematics	$\approx 2.1\%$	Mixed (0)
Panel B - Reading	$< 1\%$	Mixed (0)

The gains at Reeves during its second year under Edison are highlighted in Edison's second annual report (1999). But as the MDCPS (Gomez & Shay, 1999) evaluation report points out, the performance of the Edison students is comparable to the control groups, but not better after two years of operation as an Edison school. Based on the thorough analyses conducted by the evaluation team at MDCPS, the student achievement results suggest that improvements between the second and third year helped the Edison students make up for losses in test performance between the first and second years of operation in comparison with control groups. Unfortunately, these gains apparently were not maintained in the third year.

Criterion-Referenced Test Findings

Decisions regarding the OR were based on whether or not the $(1-\alpha)$ C.I. included 1.0. If the $(1-\alpha)$ C.I. fell completely below 1.0, this was interpreted as a protective odds ratio (1), favoring the Edison School. If the $(1-\alpha)$ C.I. included 1.0 (0), this was interpreted as an equal odds situation. If the $(1-\alpha)$ C.I. fell completely above 1.0 (-1), this was interpreted as an increase in odds for failing the state

CRT relative to the comparison sample (see Table 4:8). If the Breslow-Day statistic (B-D) is nonsignificant, one overall OR and (1- α) C.I. can be used to represent the odds for failing the CRT relative to the comparison group. Thus, there are no trends reported for each specific year in the tables, only a rating in the B-D column. If the B-D statistic is found to be statistically significant, then an overall common OR cannot be meaningfully interpreted; that is, there is a statistically significant change in the OR over years and yearly OR are necessary. Thus, our summary ratings appear for each year of data and not in the B-D column.

Table 4:8 Summary Results on Criterion-Referenced Tests

Reeves vs District	1998	1999	BD	Finding/Trend*
FCAT Grade 4 Reading		-1	N/A	mixed (0)
FCAT Grade 5 Math		-1	N/A	mixed (0)
FL Writes Grade 4			-1	mixed (0)

* We have adjusted our trends to reflect a mixed rating due to the recent availability of 2000 test results discussed in Section 4.5 and presented graphically in Exhibit 4:1.

Student results from the three state-mandated tests (FCAT reading and math and Florida Writes), indicate some gains for the Edison students, but absolute scores are still far behind the averages for the district and state). More importantly, the gains made by Reeves on the CRT are similar to those made by the district and state groups.

Combined ratings

Table 4:9 Combined Overall Trends for Henry E.S. Reeves Elementary School

	Positive	Mixed	Negative
Norm Referenced	0 of 4	4 of 4	0 of 4
Criterion Referenced	0 of 3	3 of 3	0 of 3
TOTALS	0 of 7	7 of 7	0 of 7

Based on the findings presented in this chapter and based on the discussion of the results, we rate the trends in this school as Mixed with an overall mean trend rating of 0.0. In fact, all seven trends that we considered when summing up this case were Mixed. In its second annual report, Edison (1999) rated the trends on student performance as Positive. In its third annual report, Edison rated the trends for 1999-00 as Strongly Positive and the overall trends since opening as Strongly Positive. There are extensive results from the SAT-7 available for this school. Unfortunately, Edison did not consider the findings from the district evaluation, nor from the Shay (2000) study, when presenting the results for this school in its 2000 annual report. In fact, while Edison reported same cohort SAT results for 1996-97 and 1997-98 in its 1999 annual report, it only included limited consecutive cohort SAT results in its 2000 annual report.

Chapter Five

Dodge-Edison Elementary School

Wichita, Kansas

5.1 Descriptive Summary of School

The Dodge-Edison Elementary School is a district contract school established in 1995 to serve grades K-5. In 1999-00, the school enrolled close to 630 students, and Edison reported that the school had 44 instructional staff (Edison, 2000). The total enrollment of Dodge-Edison has fluctuated from a low of 334 students, when Edison began operating the school, to a high of 657 students during the 1997/1998 school year and then back down to 626 students during the 1999/2000 school year. The enrollment in the district during the same time has shown a steady increase from 46,000 students to 47,637. Edison reported a low rate of student mobility at this school (i.e., 2.9 percent for 1997-1998) and a student/staff ratio of 17.2/1 (Edison, 1999).

Data on the ethnic composition of the students enrolled in the school vary according to the source. The Kansas School Building Report Card (Kansas State Department of Education [KSDE], 1999) indicated that for the 1998-99 school year, 65 percent of the students were white, 17 percent African American, 12 percent Hispanic, 4 percent Native American Indian, and 1 percent Asian. Edison's own data suggest that the school has slightly more minorities than the state data indicate. Over the years that Edison has operated the school, it appears that the ethnic composition of the school has fluctuated very little. The percentage of the district population that is white has steadily decreased from 62.5 percent during the 1994/1995 school year to 54.1 percent during the 1999/2000 school year. This decrease occurred while the total district population increased.

Edison reports that 6.2 percent of the students are classified for special education services and 58.8 percent qualify for free or reduced lunches. Around 11.6 percent of the district students receive special education services, and just over 50 percent of the district students qualified for free or reduced priced lunches. Average attendance rates for both Dodge-Edison and the state are relatively similar (both report attendance rates around 95 percent between the years 1995 and 1999).

The total number of certified teachers in Dodge-Edison Elementary has fluctuated between 29 and 27 during the years that the school has been operated by Edison. The total number of certified special education teachers has decreased sharply in Dodge-Edison Elementary, from 8.8 certified

special education teachers the year before Edison took over operation of the school (i.e., 1994-95) to 4 in 1995-96 and only 1 in 1999-00 (building level reports generated from the Kansas Department of Education Web site). During the same period of time, the number of certified special education teachers in the district increased from 439 in 1994-95 to 558 in 1999-00.

5.2 Past Studies and Evaluations and Data Available for Analysis

Mislevy reading study

Between 1995 and 1998 Dr. Robert Mislevy conducted a controlled comparison of reading levels at Dodge-Edison and a matched group of students from similar but non-Edison schools in the district. The first year results showed “Large significant effect in favor of Dodge-Edison at the Kindergarten level; mixed significant and non-significant effects favoring Dodge-Edison at first grade . . . ; no significant differences at Grade 2” (Mislevy, 1996; p. 8). The second- year results for grades 1-3 were comparable to the first-year findings. The third year results concluded that “Unadjusted effects favor the Dodge group in second grade and the Control group in the fourth, but when pretest differences are taken into account no adjusted effects differ significantly from zero” (Mislevy, 1998, p. 9).

When considering metric effect size, “the difference favoring the Dodge groups in the younger cohorts has *decreased*”; however, the metric grade equivalents favor Dodge and show in the youngest cohorts that there have been increases (Mislevy, 1998, pp. 11, 12). While it seems contradictory, the scores reflect differences in the Dodge group and the control group; both were tested when neither group read well, so the scores are low. However, students in the Dodge group were beginning to read earlier than students in the control group, and the results seemed to show a larger gap in achievement than what actually existed (Mislevy, 1998).

Wichita State University study

The Department of Administration, Counseling, Educational and School Psychology at Wichita State University was commissioned by the Wichita Unified School District #259 to conduct a three-year program evaluation of the Dodge-Edison Partnership School (DEPS). The purpose of the First Year Interim Report (Wichita State University, 1996) was to provide answers to initial research questions and establish a baseline for second year formative and third year summative evaluations. Areas covered in this report included governance, technology, curriculum, student assessment, personnel, professional development, student achievement, customer satisfaction, and program cost. Unfortunately, after the first-year report was submitted, the evaluation was discontinued.

Data collection methods for this study included interviews; surveys of staff, students and parents; focus groups; observation; document review; and analysis of student achievement data. The main findings from the study include the following:

- DEPS was implementing the governance model as described in the contract, and DEPS staff had a clear understanding of internal governance components (School Leadership Team, lead teachers, principal); however, there was a lack of understanding regarding role, function, and authority of community components (Board of Friends and Dodge Leadership Council), which diminished the effectiveness of these groups.
- DEPS was implementing the curricula as described in the contract, and students were being placed in flexible ability groups for reading and mathematics instruction. Both staff and students viewed the curriculum as positive for learning.
- DEPS was implementing the technology plan; classrooms had computers and teachers had laptops. However, not all homes had received computers (as of 2/29/96). For the homes that had computers, the training that was provided was not well coordinated with the delivery of the computers. Therefore, the development of skills was difficult. Aside from these difficulties, staff, students, and parents thought the technology was contributing to improved student achievement even though the staff thought more time was necessary to be able to do more with the students.
- DEPS hired more than half of the full-time certified, nonadministrative staff from outside USD 259, although the personnel were less ethnically diverse than USD 259. The certified staff earned less than the staff at comparison schools, while the principal received a larger salary package than any other USD 259 principal. There was some dissatisfaction among staff about compensation, and some were unclear as to how compensation increases and promotions were determined.
- DEPS provided time each day for professional development or for preparation time. In the first year, however, no personal development plans were implemented.
- Staff, students, and parents perceived the assessment program to be a valid assessment of student achievement, although student portfolios were neither fully nor uniformly implemented as outlined in the program design. The Quarterly Learning Contracts were implemented, but the implementation of on-demand assessments varied and were not uniformly understood by the staff.
- Parents, students, and staff reported an overall level of satisfaction with most aspects of the program. However, they were less satisfied with overall cleanliness of the facilities and with student social behaviors. The teachers were satisfied with administration and corporate support, more than other staff members, but felt stress from work loads, large classes, and the lack of time available for preparation, even with an extended day, with which some parents and students were less satisfied.
- The Edison Project promised that it could deliver quality education programs with no additional costs to parents, students, district, or state. However, the USD provided \$10,000 in repairs to the district-owned facility, which initially was provided for Edison use at no cost.

Available data for our analyses

The data available for this case were rather limited in both scope and detail. Edison provided us with a data set of individual student data on the MAT-7. This included normal curve equivalents results for only the first four years of operation (1995-96 to 1998-99), although there are usually several different scales available. Wichita School District (USD 259) provided us with the average national percentile rank results on the MAT-7 and the Kansas Assessment Tests for both Dodge-Edison and the district. Unlike the state data, the data provided by the district included information on the number of students considered in each test group. From the Kansas Department of Education we obtained supplemental information and student results on the Kansas Reading, Math, and Writing Assessments. Nevertheless, the nature of the data provided by the state and the district did not contain information on variability, nor did these results indicate the number or proportion of students who were meeting state standards. The latter was necessary in order to conduct the odds-ratio analysis.

5.3 Longitudinal Analysis of Individual Student Data

Edison Schools Inc. provided us with individual student MAT-7 test results for the 1995/96-1997/1998 school years for grades 3, 4, and 5. Although MAT-7 test results are normally reported in four separate scales—grade equivalent scores (GE), standard scores (SS), national percentile rank scores (PR), and normal curve equivalent scores (NCE)—we were provided only with NCE scores for math and reading. Thus, our longitudinal analyses were restricted to the NCE scale and these two subject tests. This norm-referenced test was always administered in the fall of each school year.

Table 5:1 presents the various sample size breakdowns by subject tested, grade, and group for 681 students covering the 1995/96 -1997/98 academic years. It was possible to trace 2 different cohorts. Cohort A included 33 students over 3 consecutive academic years (i.e., grade 3 in 1995/96, grade 4 in 1996/97, and grade 5 in 1997/98). Cohort B contained 46 students across the same 3 years, but differed in academic years (grade 3 in 1996/97, grade 4 in 1997/98, and grade 5 in 1998/99). Due to the low numbers of subjects in Cohort A, we think these results should be interpreted cautiously.

Similar to many of the other Edison schools in this study, there appears to be either a high rate of attrition or substantial inconsistencies in who is taking the tests at these schools. With regard to the larger Cohort B, there were 107 possible students to follow in 1995. By 1996 about 68 percent remained (73/107 pair-wise analysis), and by 1998 only about 43 percent of the students remained (longitudinal trend analysis). This represents a 57 percent drop in the cohort of students that progressed from grade 3 to grade 5 between the 1996/97 school year and the 1997/98 school year. The pattern for Cohort A was similar, with 82 students starting in third grade in 1995, dropping to 61 sixth graders (74 percent), and finally to 33 subjects in 1998 (40 percent) for a 60 percent drop in the three years.

Table 5:1 Sample Sizes for Individual Student Data on the MAT-7 by Grade and Year

	1995		1996		1997		1998	
	Math	Reading	Math	Reading	Math	Reading	Math	Reading
Grade 3	86	82	107	107	91	92	108	107
Grade 4	85	82	87	87	87	89	87	86
Grade 5	76	76	88	88	53	53	83	83

Sample sizes for 1 year gains

	3 to 4 (1995/96)	4 to 5 (1995/96)
Math	66	63
Reading	61	60
	3 to 4 (1996/97)	4 to 5 (1996/97)
Math	73	43
Reading	75	43
	3 to 4 (1997/98)	4 to 5 (1997/98)
Math	57	54
Reading	57	54

Cohort A	Cohort B
Sample Sizes for Longitudinal Analysis 1995-97(Grades 3 to 4 to 5)	Sample Sizes for Longitudinal Analysis 1996-98 (Grades 3 to 4 to 5)
33	46

Longitudinal analysis findings

Repeated measures ANOVAs were examined for longitudinal trends over a three-year period for Dodge-Edison. Outcome data were individual student data on the MAT-7 math and reading subtests. Only NCE scores were provided by Edison. In all models the assumption of sphericity was evaluated, and if found to be violated, the Huynh-Feldt adjusted p-values are reported. If the overall linear model was found to be statistically significant, unadjusted (alpha) pair-wise comparisons were examined to identify where a difference in the means might be located.

The tables and charts in Exhibit 5:1 illustrate the results of the NCE analyses for Cohorts A and B, respectively. On the whole the students in the two cohorts are performing at about the national median. What is of particular importance for our analysis is not where they are at one point in time, but rather the rate of learning or the relative size of the gains they are making each year. Cohort A evidenced a statistically significant gain in both mean math and reading NCE scores over the three-year period. However, this gain was not evidenced until the third year (1997), which surpassed both of the previous two years (p 's < .05). Cohort B's performance was not as positive, with students failing to show any significant change in mean math or reading NCE scores over the three years. Although there was some gradual improvement, it did not reach a level to be declared statistically significant.

Exhibit 5:1 Results from the Analysis of Individual Student Results on the Metropolitan Achievement Test, Dodge-Edison (Wichita, Kansas)

Cohort A

n=33

Grade

NCE Math

NCE Reading

1995 1996 1997

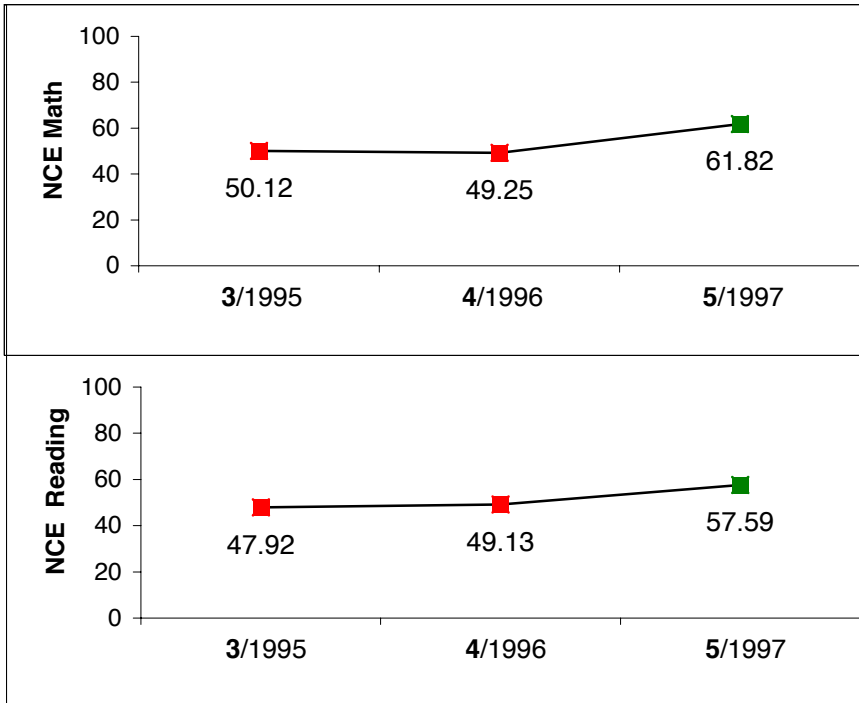
3 4 5

50.12 49.25 61.82

47.92 49.13 57.59

$F(2,64)=10.96, p<.0001$

$F(2,64)=9.25, p=.0003$



Cohort B

n=46

Grade

NCE Math

NCE Reading

1996 1997 1998

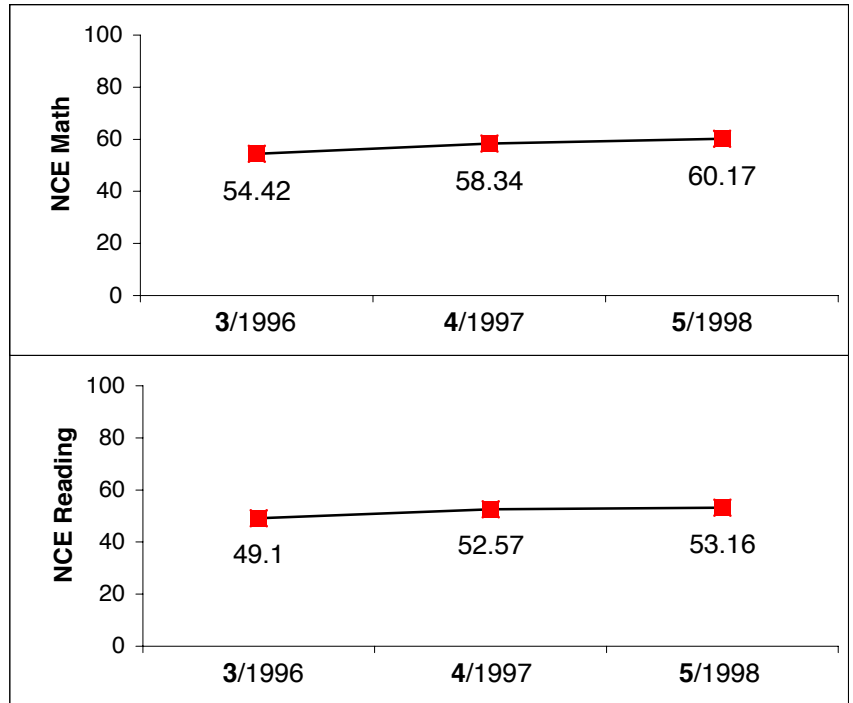
3 4 5

54.42 58.34 60.17

49.1 52.57 53.16

$F(2,90)=2.58, p=.0898$

$F(2,90)=1.26, p=.2881$



A color change represents a statistically significant change in the means. A trend with two color changes represents differences among all three means. A dual colored charting point (red and green) represents a statistically significant difference between one mean but not the other. The reader is encouraged to examine the individual table of means for these cases.

5.4 Comparison with District on the MAT-7

From data made available by Wichita Public Schools, we made comparisons between Dodge-Edison and the district on overall performance on the Metropolitan Achievement Test (MAT-7). This test is required by all schools in the district and is a part of the district's overall accountability system. While we compared gains made by Edison students relative to national norms in the previous section, this section compares Dodge-Edison and the district. The analysis in the previous section is based on individual student gains, while the analysis in this section is based on overall performance by consecutive groups of students at the same grade level. Table 5:2 includes the national percentile ranks for each of the five years that data are available. This table also includes information on the number of test takers.

Table 5:2 National Percentile Ranks on MAT-7 Compared with the District, Grades 3, 4, and 5

	Dodge-Edison			Wichita Public Schools		
	N	Reading	Math	USD 259	Reading	Math
Grade 3						
1995/96	(75)	39	39	(2962)	50	43
1996/97	(103)	47	58	(3322)	52	49
1997/98	(90)	67	70	(3564)	54	50
1998/99	(105)	52	63	(3458)	54	50
1999/00	(102)	61	71	(3375)	61	57
Change in Percentile Rank		22	32		11	14
Grade 4						
1995/96	(84)	47	37	(2921)	52	45
1996/97	(84)	49	52	(3135)	54	50
1997/98	(87)	52	67	(3351)	56	56
1998/99	(83)	56	65	(3191)	56	54
1999/00	(91)	59	70	(3264)	59	57
Change in Percentile Rank		12	33		7	12
Grade 5						
1995/96	(67)	54	49	(2932)	56	54
1996/97	(85)	60	65	(3034)	58	56
1997/98	(77)	60	69	(3030)	58	63
1998/99	(78)	54	62	(3096)	60	62
1999/00	(90)	61	77	(2954)	59	64
Change in Percentile Rank		7	28		3	10

The change in percentile ranks over time is highlighted in Table 5:3. Gains were being made by both the district and Dodge-Edison, although the gains made by Dodge-Edison are much larger. In reading, Dodge-Edison was substantially lower than the district in overall NPR in 1995-96, but by the 1999-00 school year they had approximately the same results. In math, Dodge-Edison was slightly lower than the district in 1995-96, but had a substantially higher NPR in 1999-00.

Table 5:3 Comparison of Gains in Terms of National Percentile Rank Between Dodge-Edison and Wichita Public Schools (MAT-7)

MAT-7 Results	Dodge-Edison Gain on NPR Between 1995-99	Wichita Public Schools Gains on NPR Between 1995-99	Difference in gains
Grade 3 Reading	22	11	11
Grade 3 Math	32	14	18
Grade 4 Reading	12	7	5
Grade 4 Math	33	12	21
Grade 5 Reading	7	3	4
Grade 5 Math	28	10	18

When we look at the overall gains in terms of national percentile ranks (NPR), we see that Dodge-Edison consistently had larger gains than the district. In fact, the average difference in gain scores (see the last column in Table 5:3), between the Dodge-Edison and the district was 12.8 NPR. The substantial improvement made by Edison students was more prominent in math than in reading.

Exhibits 5:2 and 5:3 illustrate the change over time in terms of national percentile ranks on the MAT-7 for Dodge-Edison and the district. While the gains in the district are incremental and rather consistent, the gains and—at times—decreases in results for Dodge-Edison tend to be large. This is largely due to the smaller number of test takers at Dodge-Edison compared with the district as a whole.

Improvements made by Edison students were more prominent in math than in reading. In math, the Dodge-Edison students increased an average of 31 NPR over the 5-year period, while in reading the average increase over the 5-year period across the 3 grade levels was 13.6 NPR. These gains when broken down over 4 periods reflect an average annual gain of 7.7 NPR in math and 3.4 NPR in reading.

Exhibits 5:2 and 5:3 illustrate the achievement gains made by Dodge-Edison students as compared with gains made by district students. Exhibit 5:2 contains the MAT-7 reading results, and Exhibit 5:3 contains the math results.

Exhibit 5:2 Results on the Metropolitan Achievement Test, Grades 3, 4, and 5 Reading Comparison Between Dodge-Edison and Wichita School District, 1995/96 -1999/00

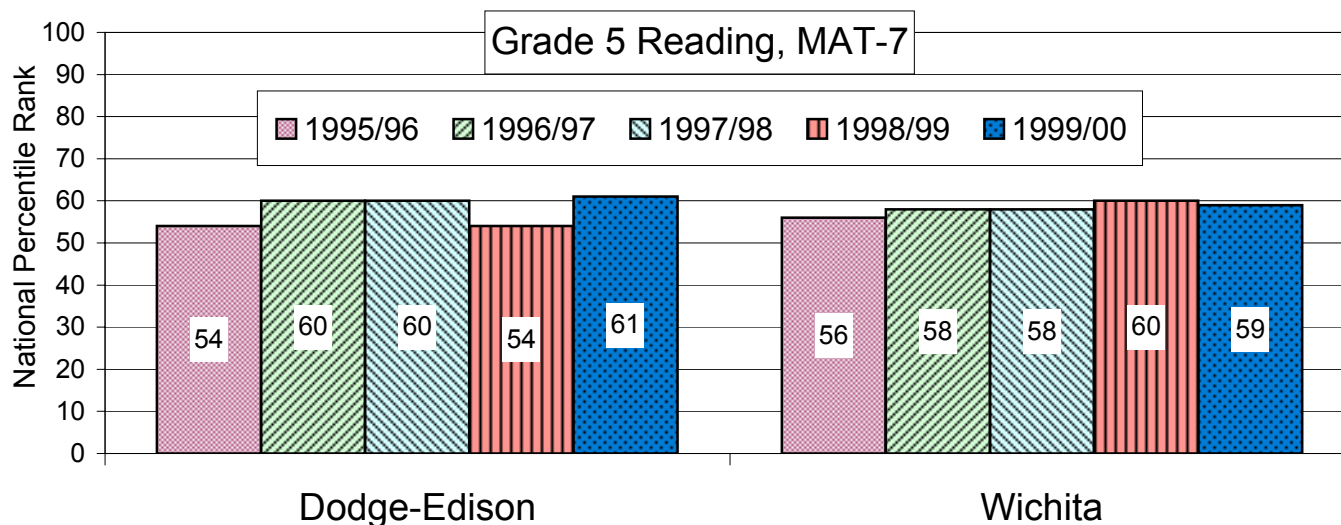
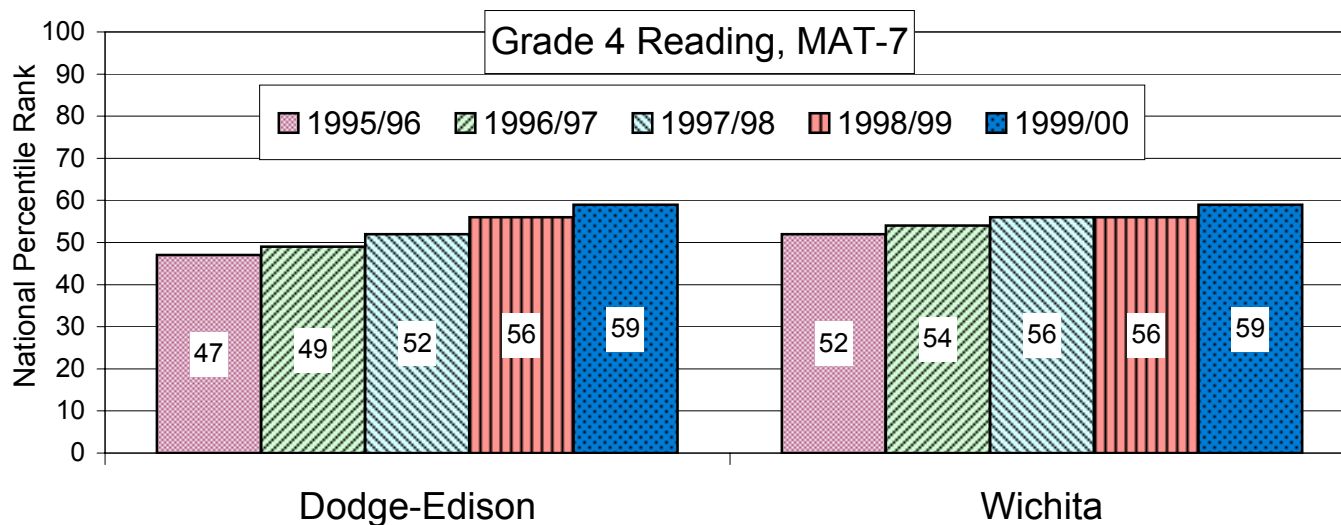
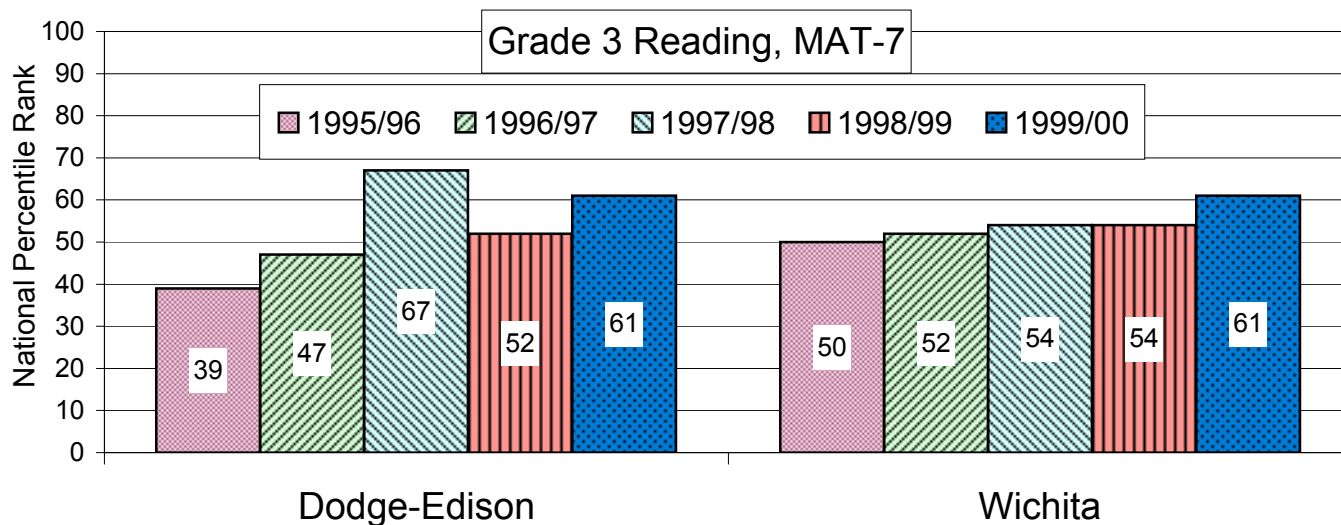
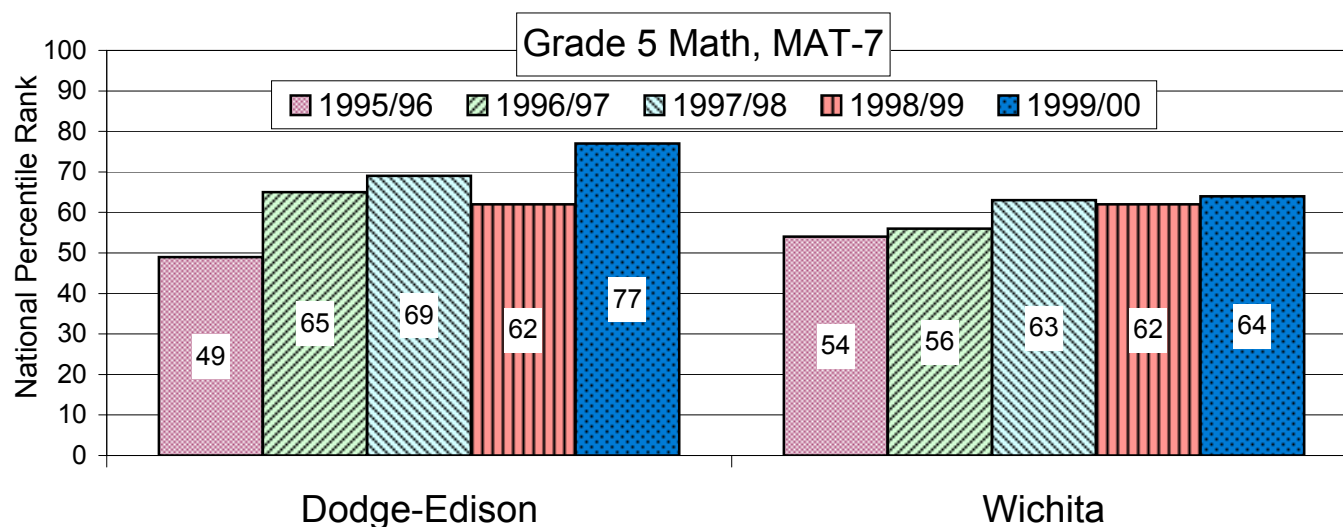
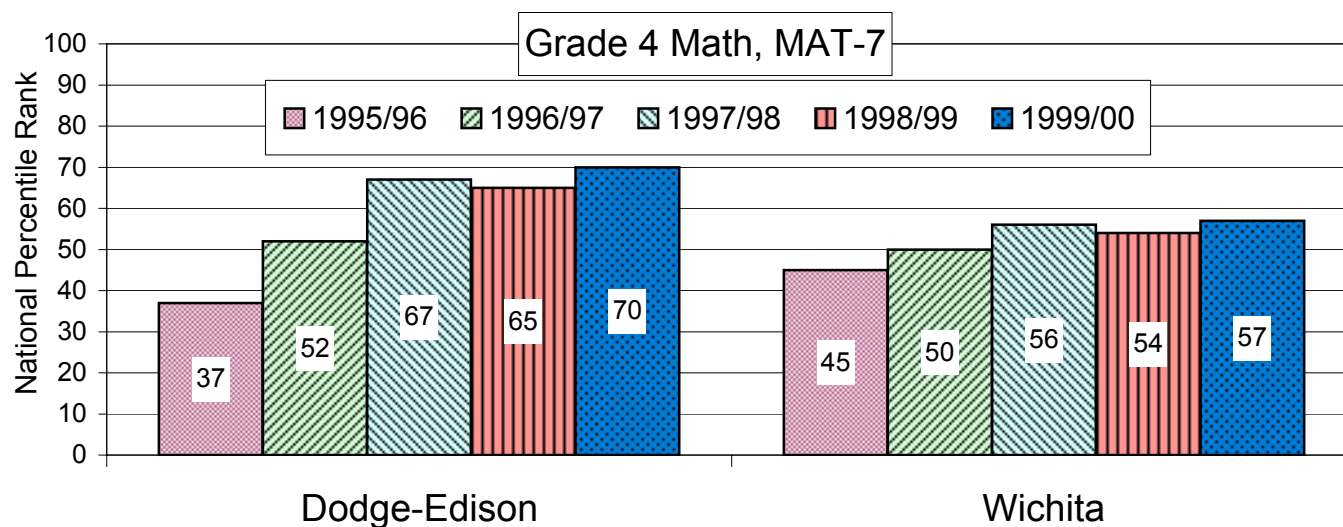
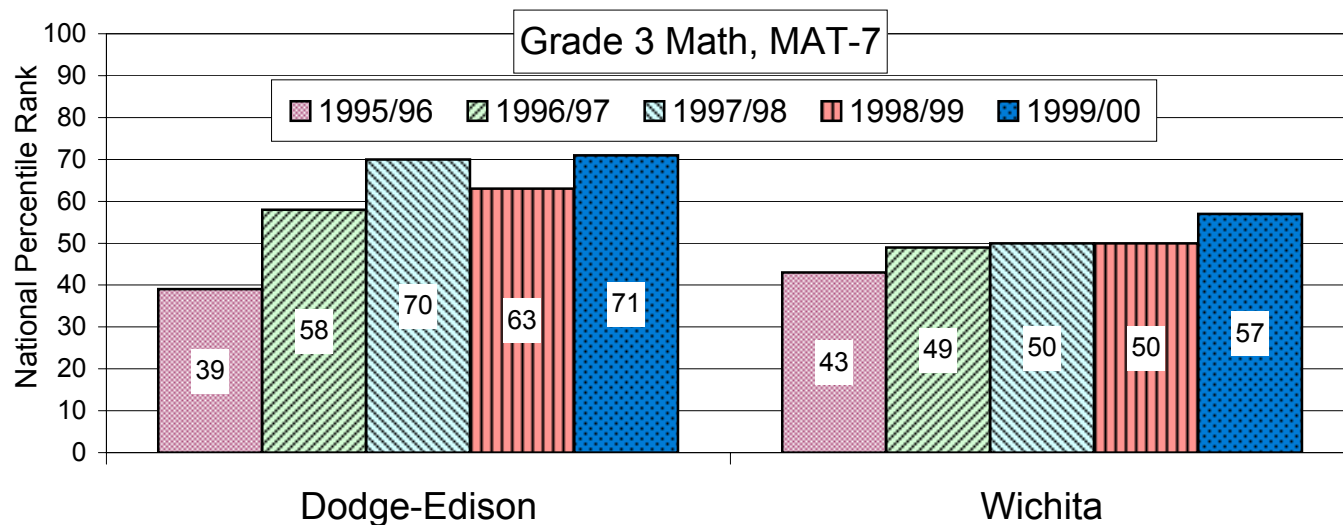


Exhibit 5:3 Results on the Metropolitan Achievement Test, Grades 3, 4, and 5 Math Comparison Between Dodge-Edison and Wichita School District, 1995/96 -1999/00



5.5 Comparison with District and State on Kansas Assessments

Because the results on the state assessment tests did not include information on the number or proportion of students meeting state standards, we could not conduct a chi-square analysis nor odds-ratio analysis. Also, because the results on the assessment tests did not include information on variability, we could not conduct other tests to determine whether or not the changes over time were statistically significant. Therefore, we are limited to providing a descriptive summary of changes in test results for consecutive groups at a particular grade level, just as we did in the previous section with the MAT-7 results.

Table 5:4 Performance on Kansas Assessment Tests Compared with the District and State

Kansas Reading Assessment (Index Score)	Dodge-Edison		Wichita		State
	N	Reading	N	Reading	Reading
Grade 3					
1996/97	(93)	54.86	(3258)	61.05	64.92
1997/98	(88)	56.84	(3240)	60.01	65.38
1998/99	(108)	52.54	(3534)	60.15	65.20
Change in Index Score		-2.32		-0.90	0.28
Kansas Math Assessment (Power Score)	Dodge-Edison		Wichita		State
	N	Math	N	Math	Math
Grade 4					
1996/97	(79)	46.39	(3076)	54.46	58.73
1997/98	(90)	48.17	(3202)	54.04	59.69
1998/99	(87)	49.26	(3157)	54.91	60.90
Change in Power Score		2.86		0.45	2.17
Kansas Writing Assessment Composite Score (0-5)	Dodge-Edison				State
		Writing			Writing
Grade 5					
1995/96		2.83			3.08
1997/98		2.92			3.03
1998/99		3.25			3.20
Change in Composite Score		0.42			0.12

Notes: Wichita Public Schools provided results for Dodge-Edison and the district on reading and math, while the Kansas Department of Education was the source of all state level results as well as all results on the Grade 5 writing assessment. All assessment tests were administered in the spring of each year, except for the writing assessment, which was not administered in the spring of 1997. These scores do not include special education.

The set of results we received from Wichita Public Schools contained the results for Dodge-Edison and the district as a whole. These data also included information on the number of test takers. There were some missing figures, which we filled in with data from the state. The state results were obtained from the web site of the Kansas Department of Education. Table 5:4 contains the results for Grades 3 reading, Grade 4 math, and Grade 5 writing. These results are also illustrated in Exhibit 5:4.

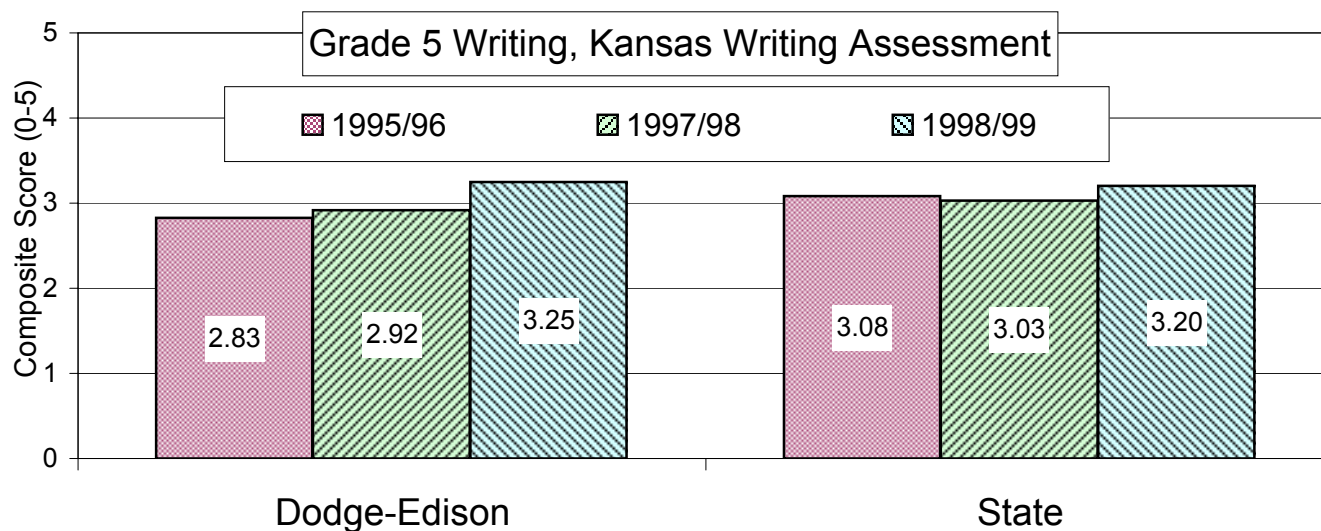
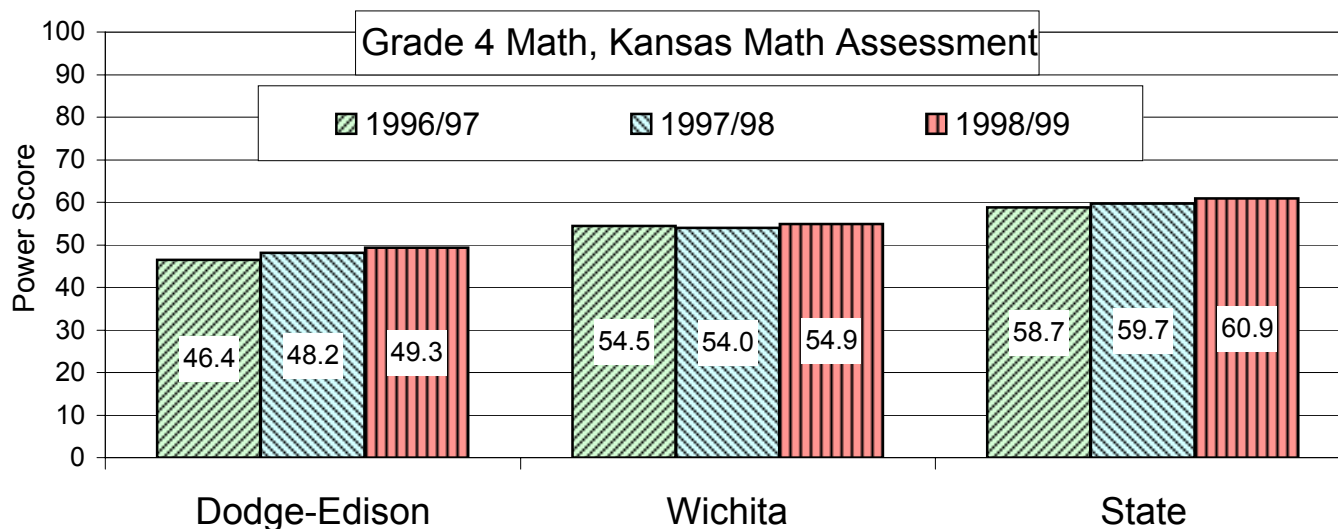
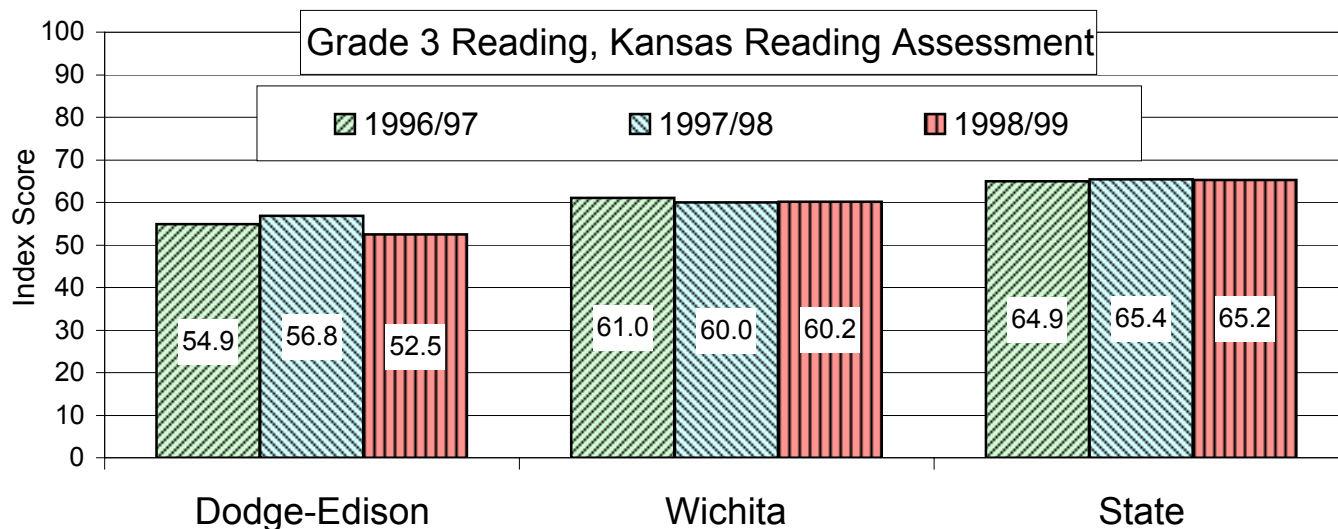
On the reading and math assessments, Dodge-Edison remained noticeably lower than the district and state in terms of absolute scores; however, on the writing assessment, Dodge-Edison started lower but surpassed the state average.

The reading results for grade 3 point out that Dodge-Edison students lost ground over the three years of reported data (i.e., -2.32 on the index score), while the district results went down by less than one point (-0.9). At the same time, overall state performance on the reading assessment went up slightly (0.28). The index score corresponds with the overall average percent correct on the expository and narrative components of this assessment test.

In math, Dodge-Edison had a larger gain than both the district and state. The gain on the power score for Dodge-Edison was 2.86 between 1997 and 1999, while the gains for the same period of time were 0.45 for the district and 2.17 for the state. The total power score is an equally weighted average of the scores based on percent correct on the three components of this assessment test: problem solving, communication, and reasoning.

Results for the Kansas Writing Assessment were available for Dodge-Edison and the state. Here Dodge-Edison gained substantially more than the state over the span of four years (note that the test was not administered in 1997, so the results are for 1996, 1998, and 1999). The Dodge-Edison score increased from 2.83 in the 1995/96 school year to 3.25 in the 1998/99 school year (i.e., an increase of 0.42). The state average increased from 3.08 in 1995/96 to 3.20 in 1998/99 (an increase of 0.12). The composite score is the average score on six tests or rating scales, each ranging from 0 to 5, with 5 being the highest and 0 the lowest.

Exhibit 5:4 Results on the Kansas Assessments, Grades 3, 4, and 5
Comparison Between Dodge-Edison and Wichita School District, 1995/96 -1999/00



5.6 Summary

A summary score of -1 indicates a result that is unfavorable for the sample school, a score of 0 indicates a neutral finding, and a score of 1 indicates a favorable result according to the criteria specified in Section 2.5. The effect size (ES) is the omega squared (ω^2) for a one-way repeated measures ANOVA.

Norm-referenced test findings

We analyzed gains on the Metropolitan Achievement Test (i.e., a norm-referenced test) based on individual student gains over three years. This analysis was based on normal curve equivalents (NCEs). We also used the MAT-7 results to examine consecutive cohorts of students at the same grade level over five years. The latter analysis was based on gains on average national percentile ranks for consecutive cohorts of students at the same grade level. Both analyses indicated that students at Dodge-Edison were making substantial gains when compared with the district and state.

With the longitudinal analysis we were able to determine whether or not the gains were significant or not. This was not possible with the analysis of consecutive cohorts because we were not able to obtain details on the variability of the data. Because it is our intention to avoid establishing more than one trend with the same data, we counted only the longitudinal analysis in this summary and not the analysis of the consecutive cohorts of students at a particular grade level. The design for the longitudinal analysis is a stronger design; however, it is limited by the small number of students that we could trace over three years. The average annual gains in NCE made by Cohort A (5.8 in math and 4.8 in reading) were significant, but the gains made by Cohort B (2.9 in math and 2.0 in reading) were not significant. Table 5:5 contains the summary results and trend ratings for the NRT.

Table 5:5 Summary of Results on Norm-Referenced Student Achievement Tests

Cohort A (1995-98)			NCE		Trend
MAT-7	Grades 3, 4, and 5	p-value	Δ	ES	
Math		<.0001	11.7 (+1)	.163	positive (+1)
Reading		<.0003	9.7 (+1)	.139	positive (+1)
Cohort B (1996-99)			NCE		Trend
MAT-7	Grades 3, 4, and 5	p-value	Δ	ES	
Math		<.0898	5.8 (0)	.022	mixed (0)
Reading		<.2881	4.1 (0)	.004	mixed (0)

Criterion-referenced test findings

Given the nature of the data and the lack of details on variability, we could not do much with this criterion-referenced data. The change in the index score and the power scores are clearly very small, especially given that these scores correspond to the average percent correct on this assessment test (i.e., 0 - 100 percent). The composite score on the Kansas Writing Assessment is based on a 0-5 scale, so the .42 gain made by the Dodge-Edison students is quite substantial. The difference in gain scores in favor of Dodge-Edison (0.3) represents 6 percent of the total range of the scale. Even though we do not have a cut-off level for the composite score to determine whether this trend is positive, mixed, or negative, we have rated this as a positive trend in favor of Dodge-Edison.

Table 5:6 Summary Results on Criterion-Referenced Tests: Comparison of Gains on Kansas Assessments Between Dodge-Edison and Wichita Public Schools

	Dodge-Edison Gains on Index or Power Scores Between 1996/97-98/99	Wichita Public Schools Gains on Index/Power Scores 1996/97 --1998/99	Difference	Trend
Grade 3 Reading Index Score (0-100)	-2.35	-0.9	-1.45	mixed (0)
Grade 4 Math (0-100) Power Score	2.86	0.45	2.41	mixed (0)
	Dodge-Edison Gain on Composite Score Between 1995/96-98/99	State of Kansas Gain on Composite Score Between 1995/96-98/99		
Grade 5 Writing Composite Score (0-5)	.42	.12	.30	positive (+1)

Combined ratings

Given the total ratings for the trends that are highlighted in Table 5:7, we rate this school as Positive with a mean trend rating of 0.43. In its 1999 annual report, Edison rated the trends in this school as Positive. In its 2000 annual report, Edison rated the achievement gains since opening as Strongly Positive and the achievement gains in 1999-00 as Strongly Positive.

Table 5:7 Combined Overall Trends for Dodge-Edison

	Positive	Mixed	Negative
Norm Referenced	2 of 4	2 of 4	0 of 4
Criterion Referenced	1 of 3	2 of 3	0 of 3
TOTALS	3 of 7	4 of 7	0 of 7

The designs behind the trends in the norm-referenced results are based on tracing two cohorts of individual students over three years. One limitation to this analysis was the small number of students that we could trace over three years (i.e., 33 students in Cohort A and 46 students in Cohort B). On the other hand, the test administration procedures are likely to be more regulated for the state assessment tests in reading, math, and writing. The latter data, however, did not allow us to determine the number or proportion of students who met state standards. Therefore, we could not conduct the odds ratio analysis. The results from the state assessment tests also lacked information on variability so it was impossible for us to determine whether the changes over time were significant or not.

Dodge-Edison has been lauded by Edison as one of its success stories. In presentations made by Edison officials, Dodge-Edison is the school that most often comes up as an example of the achievement gains being made by schools operated by Edison. While the overall label we use to categorize the trends in this school is clearly Positive, the trends are not as positive as Edison suggests in its third annual report. In any case, the gains made by students enrolled in Dodge-Edison are both substantial and consistent. The achievement gains in this school are clearly the most positive of the ten cases we examined.

Chapter Six

Jardine-Edison Junior Academy

Wichita, Kansas

6.1 Descriptive Summary of School

Jardine-Edison Junior Academy is a district contract school established in 1996 to serve grades 6-8. In 1999-00, the school enrolled 822 students and Edison reported that the school had 55 instructional staff (Edison, 2000). The student mobility in 1997-1998 was 3.7 percent, and the student/staff ratio was 15.1/1 (Edison, 1999).

Enrollment at Jardine increased when Edison began operating the school in 1996. During the 1995-96 school year, only 568 students were enrolled in the school, but this jumped to 900 students during the 1996-97 school year. Enrollment continued to increase the next school year (946 students), but then decreased to 890 and 822 students during the 1998-99 and 1999-00 school years, respectively.

As the population has fluctuated at this school, the total percentage of white students has steadily declined from a high of 49 percent during the 1996-97 school year to 32 percent during the 1999-00 school year. The percentage of the district population that is white also steadily decreased during the same period of time; currently, 54 percent of the district's students are white. Jardine-Edison is clearly a diverse school in terms of ethnicity: 36.6 percent of the students are African American, 17.5 percent Hispanic, 11.3 percent Asian or Pacific Islander, and 2 percent are American Indian or Alaskan Native (KSDE, 1999).

Average attendance for Jardine increased slightly from 92.3 percent in 1995 to 95 percent in 1999 with small fluctuations in 1996 (88.2 percent) and 1998 (96.2 percent). The district average attendance rate followed a similar trend, beginning at 92.25 percent in 1995 and reaching 95.04 percent in 1999. In comparison, the average attendance rate for the state was around 94 percent between 1995 and 1999. The dropout rate for Jardine increased from 0 percent in 1995-96 to 2.5 percent for the 1996-97 school year, the year Edison began operating the school, but then decreased to 1 percent in 1997. The state's dropout rate is typically around .20 percent while the district dropout rate remained at .01 percent from 1995 to 1997.

Jardine has a low rate of violent acts against both students (0.2 percent or less for 1997-1999) and teachers (0 percent; no reported cases). The total number of suspensions at Jardine increased when Edison assumed operation of the school: in 1995-96 there were 199 suspensions, and the following year this increased to 432. The number of suspensions has decreased since then, and in 1998-99 the number of suspensions was 329.

The percentage of students eligible for free and reduced lunch at Jardine has been consistently very high, with between 61 and 74 percent of the students qualifying for free or reduced lunches during the years that Edison has operated the school. The figure for the 1999-00 school year was the highest with 74 percent. In contrast, the proportion of students in the district qualifying for free or reduced lunches is just over 50 percent. In the state the figure is around 31 percent.

The proportion of special education students has also remained stable after an initial drop after Edison took over the operation of the school. In 1995-96, 14 percent of the students received special education services. The following year, this figure dropped to 9.4 percent and has fluctuated between 9 and 10 percent since then.

6.2 Past Studies and Evaluations and Data Available for Analysis

Jardine-Edison was not considered in the evaluation started by Wichita State University, since that study was terminated after delivering its first report in June 1996 and Edison's contract to operate Jardine did not commence until the beginning of the 1996-97 school year. Jardine was not included in the Mislevy studies because these reading studies focused only on lower elementary grades. We are not aware of any other studies or evaluations that included Jardine, aside from the reports prepared by the Wichita Public Schools, by AFT, and by Edison, which include summative results of student achievement data.

Jardine-Edison has administered the Metropolitan Achievement Test every year since Edison took over operations. Edison Schools Inc. did not provide us with a data set of individual student results for Jardine-Edison, so we could not conduct the longitudinal analyses as we did for several of the other schools. Wichita Unified School District (USD 259) provided us with the average national percentile rank results on the MAT-7 and the Kansas Assessment Tests for Jardine-Edison and the district. Unlike the available state data, the data provided by the district included information on the number of students considered in each test group. We obtained supplemental information and student results on the Kansas Reading, Math, and Writing Assessments from the Kansas Department of Education. Nevertheless, the nature of the data provided by the state and the district did not contain information on variability; nor did these results indicate the number or proportion of students who were meeting state standards. The latter was necessary in order to conduct the odds-ratio analysis. Kansas Assessment Tests were administered to seventh graders (reading and math) and eighth graders (writing). State science assessments and state social studies assessments will not be implemented in the schools until the 2000-01 school year.

6.3 Comparison with District on the MAT-7

From data made available by Wichita Public Schools, we were able to make comparisons between Jardine-Edison and the district on overall performance on the Metropolitan Achievement Test (MAT-7). This test is required by all schools in the district and is a part of the district's overall accountability system. Because Edison did not provide us with individual student data for this

school, we could not conduct a longitudinal analysis as we did for Dodge-Edison. In this section, the analysis is based on overall performance by consecutive groups of students at the same grade level. Table 6:1 includes the national percentile ranks for each of the five years that data are available. This table also includes information on the number of test takers.

The change in percentile ranks over time is highlighted in Table 6:1. As one can see, gains were made by Jardine-Edison in all subjects and grades except grade 6 reading. The district also made gains in all subjects and grades except for reading in grades 6 and 8. The gains made by Jardine-Edison were noticeably higher in both reading and math for grades 7 and 8, while the grade 6 results were similar to the district.

In terms of absolute results, Jardine started behind district averages and remains behind in terms of national percentile ranks in all subjects and grades except for Grade 7 reading. In most cases, Jardine is 10 NPR or more below the district average.

Table 6:1 National Percentile Ranks on MAT-7 Compared with the District, Grades 6, 7, and 8

	Jardine-Edison			Wichita School District		
	N	Reading	Math	N	Reading	Math
Grade 6						
1996/97	(315)	52	36	(3012)	58	47
1997/98	(274)	54	37	(2918)	56	52
1998/99	(252)	58	50	(3020)	60	52
1999/00	(244)	48	42	(2913)	55	54
Change in Percentile Rank		-4	6		-3	7
Grade 7						
1996/97	(258)	41	41	(2871)	52	56
1997/98	(295)	50	43	(2955)	56	58
1998/99	(235)	52	58	(2897)	54	60
1999/00	(211)	54	55	(2805)	55	64
Change in Percentile Rank		13	14		3	8
Grade 8						
1996/97	(181)	43	39	(2843)	60	52
1997/98	(239)	45	37	(2870)	58	54
1998/99	(263)	58	45	(2885)	62	58
1999/00	(221)	52	50	(2747)	59	59
Change in Percentile Rank		9	11		-1	7

Exhibits 6:1 and 6:2 illustrate the achievement gains made by Dodge-Edison students as compared with gains made by district students. Exhibit 6:1 contains the MAT-7 reading results, and Exhibit 6:2 contains the math results.

Exhibit 6:1 Results on the Metropolitan Achievement Test, Grades 6, 7, and 8 Reading Comparison Between Jardine-Edison and Wichita School District, 1996/97 -1999/00

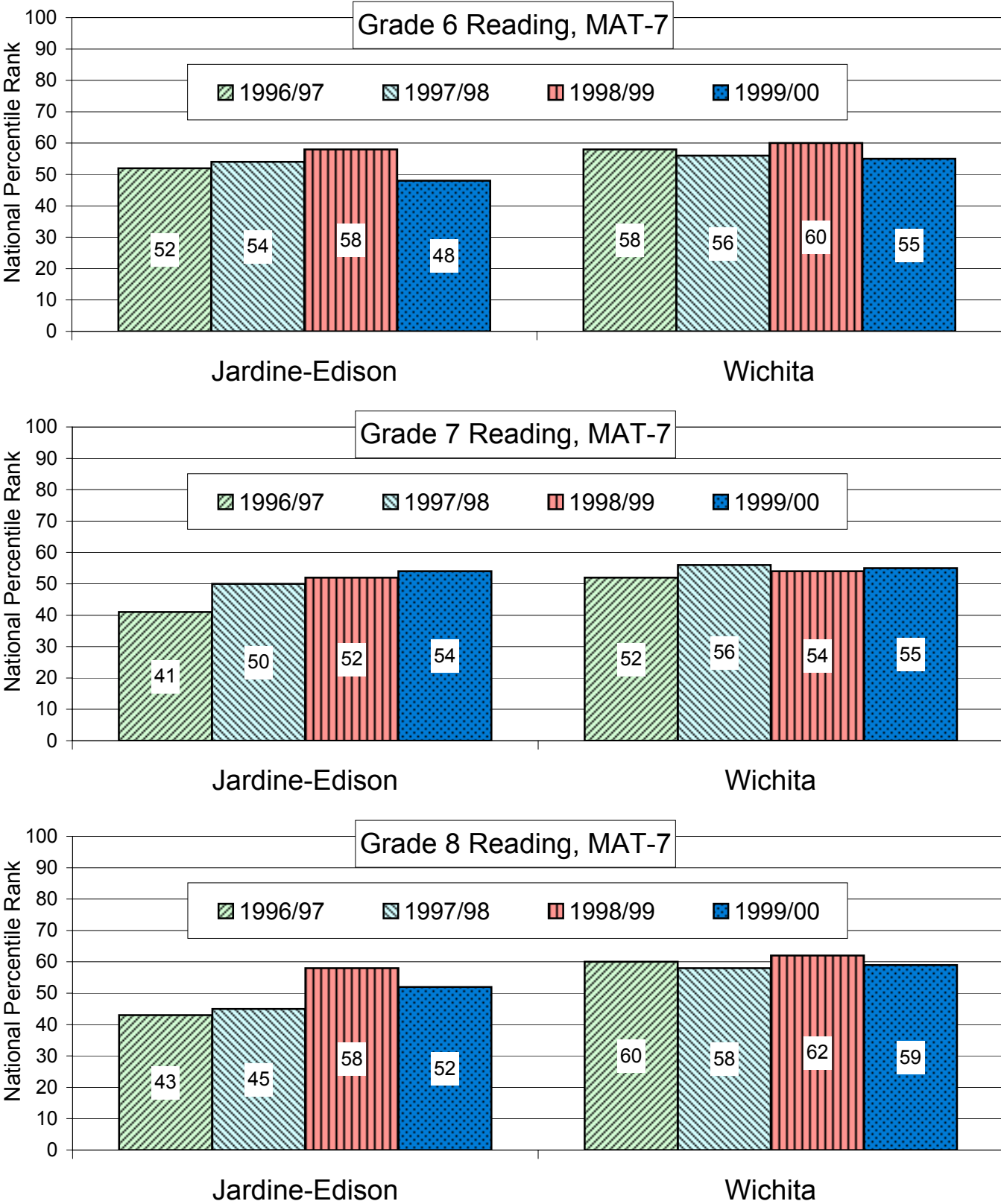
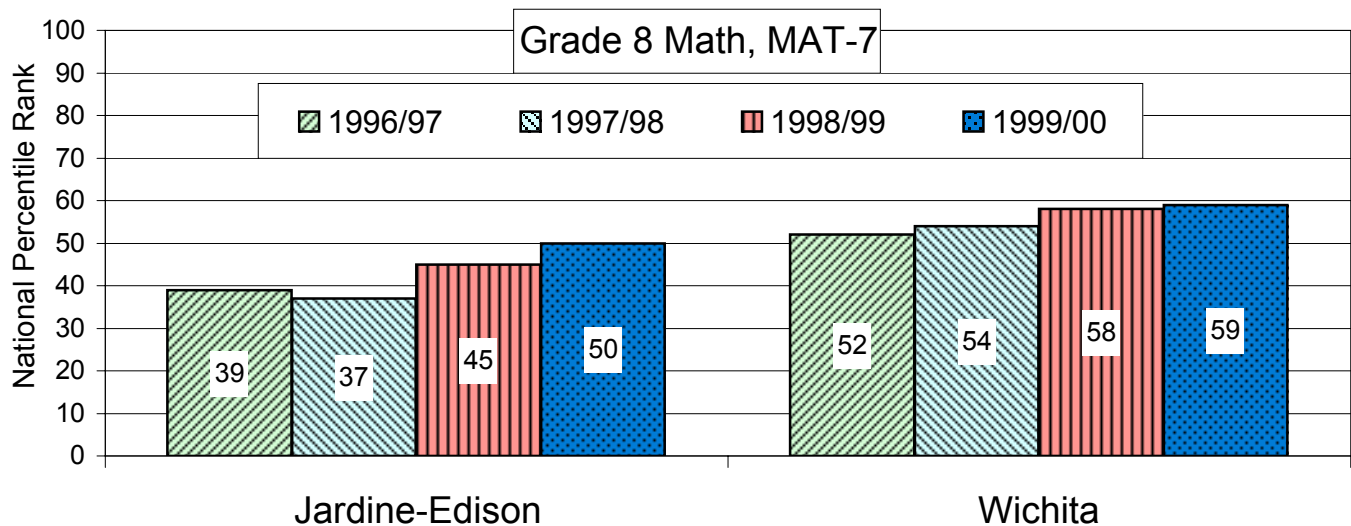
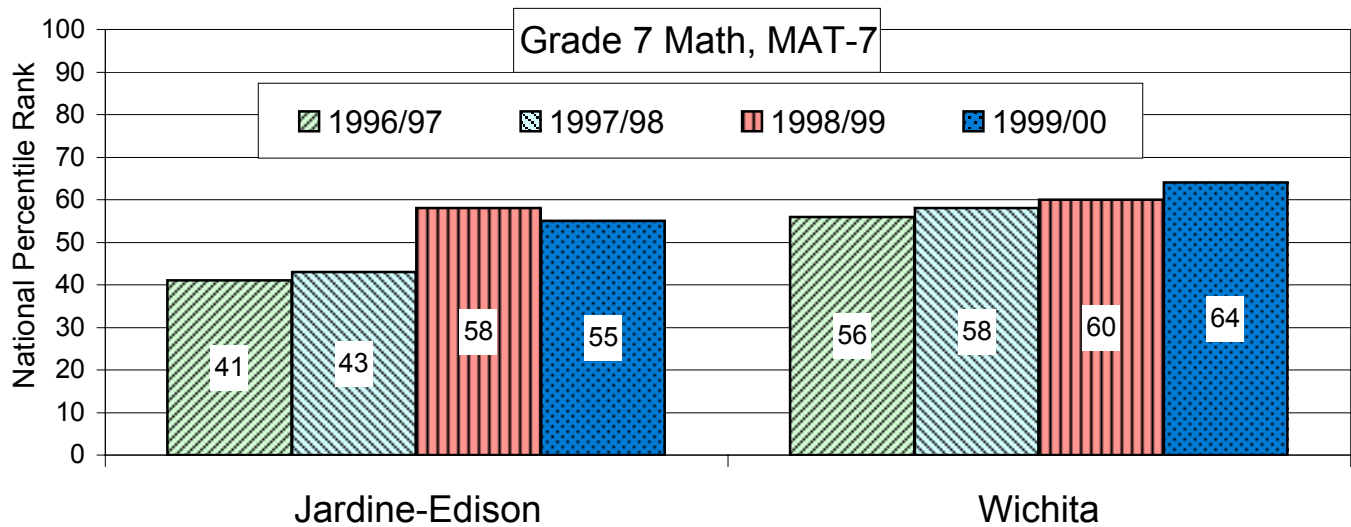
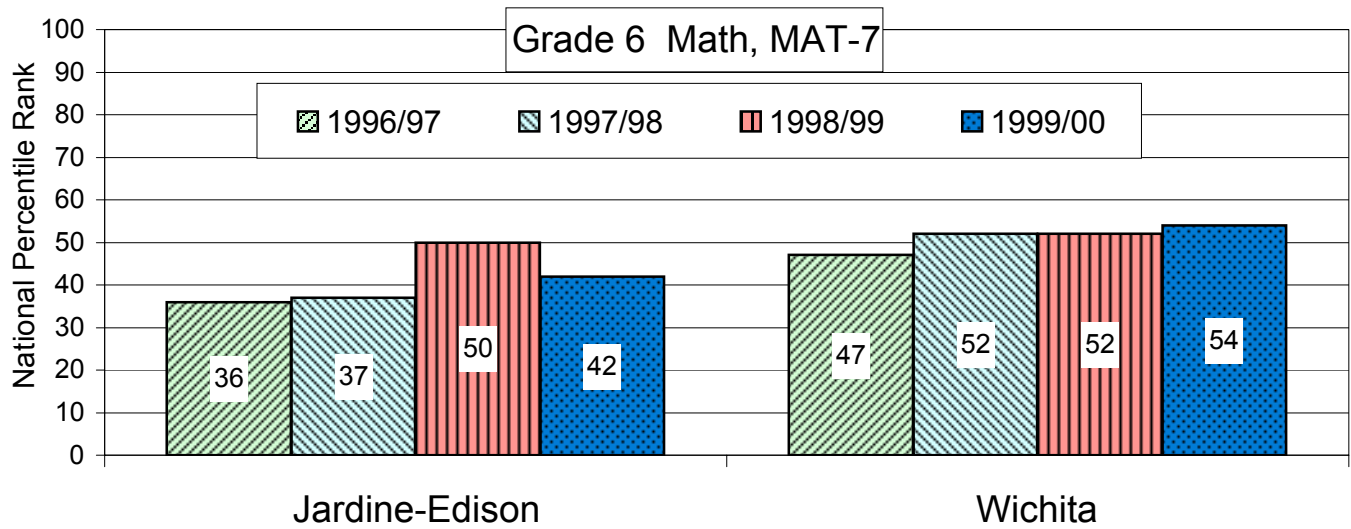


Exhibit 6:2 Results on the Metropolitan Achievement Test, Grades 6, 7, and 8 Math
Comparison Between Jardine-Edison and Wichita School District, 1996/97 -1999/00



Exhibits 6:1 and 6:2 illustrate the change over time in terms of national percentile ranks on the MAT-7 for Jardine-Edison and the district. It is worth noting that the national percentile ranks the year before Edison began operating the school were substantially higher in reading for grades 7 and 8 reading and grade 7 math. On the other hand, the pre-Edison results for grade 6 reading and math were substantially lower than the first-year results recorded by Edison.

The performance of Jardine students on these tests dropped noticeably during the 1999-00 school year, prompting even Edison to rate the 1999-00 trends as negative (Edison, 2000). However, when we look at the overall gains in terms of national percentile ranks (NPR), we see that Jardine-Edison had larger gains than the district, both in reading and math for grades 7 and 8, but the district change score was slightly better than Jardine-Edison's for grade 6. The average difference in change scores between the Jardine-Edison and the district was 4.6 NPR.

6.4 Comparison with District and State on Kansas Assessment Tests

Because the results of the state assessment tests did not include information on the number or proportion of students meeting state standards, we could not conduct a chi-square analysis nor odds-ratio analysis. Also, because the results on the assessment tests did not include information on variability, we could not conduct other tests to determine whether or not the changes over time were statistically significant. Therefore, we are limited to providing a descriptive summary of changes in test results for consecutive groups of students at each of three particular grade levels.

Wichita Public Schools provided a set of results that contained data for Jardine-Edison and the district as a whole. These data also included information on the number of test takers. There were some missing figures, which we filled in with data made available by the state. The state results were obtained from the Web site of the Kansas Department of Education. Table 6:2 contains the results for grade 7 reading and math, and grade 8 writing. These results are also illustrated in Exhibit 6:3.

On the math and writing assessments, Jardine started out and remained noticeably lower than the comparison groups, whether they are the district or the state. On the reading assessment, however, Jardine-Edison started much lower, but its 1989-99 results were very close to the district results.

The reading results for grade 7 show that Jardine-Edison gained substantial ground over the three years of reported data (i.e., a gain of 6.85 on the Index Score), while the district and state results remained largely the same over the same period. The Index Score corresponds with the overall average percent correct on the expository and narrative components of this assessment test.

Table 6:2 Performance on Kansas Assessment Tests Compared with the District and State

Kansas Reading Assessment		Jardine-Edison		Wichita		State
(Index Score)	N	Reading	N	Reading	Reading	
Grade 7						
1996/97	(256)	50.75	(3013)	57.44	64.68	
1997/98	(298)	51.91	(2877)	59.30	64.51	
1998/99	(227)	57.60	(2787)	58.43	64.78	
Change in Index Score		6.85		1.00	0.10	
Kansas Math Assessment		Jardine-Edison		Wichita		State
(Power Score)	N	Math	N	Math	Math	
Grade 7						
1996/97	(243)	35.99	(2855)	43.15	48.90	
1997/98	(294)	35.23	(2942)	43.64	49.81	
1998/99	(231)	37.95	(2761)	44.13	51.13	
Change in Power Score		1.96		0.98	2.23	
Kansas Writing Assessment		Jardine-Edison				State
Composite Score (0-5)		Writing				Writing
Grade 8						
1997/98		3.02				3.35
1998/99		2.91				3.40
Change in Composite Score		-0.11				0.05

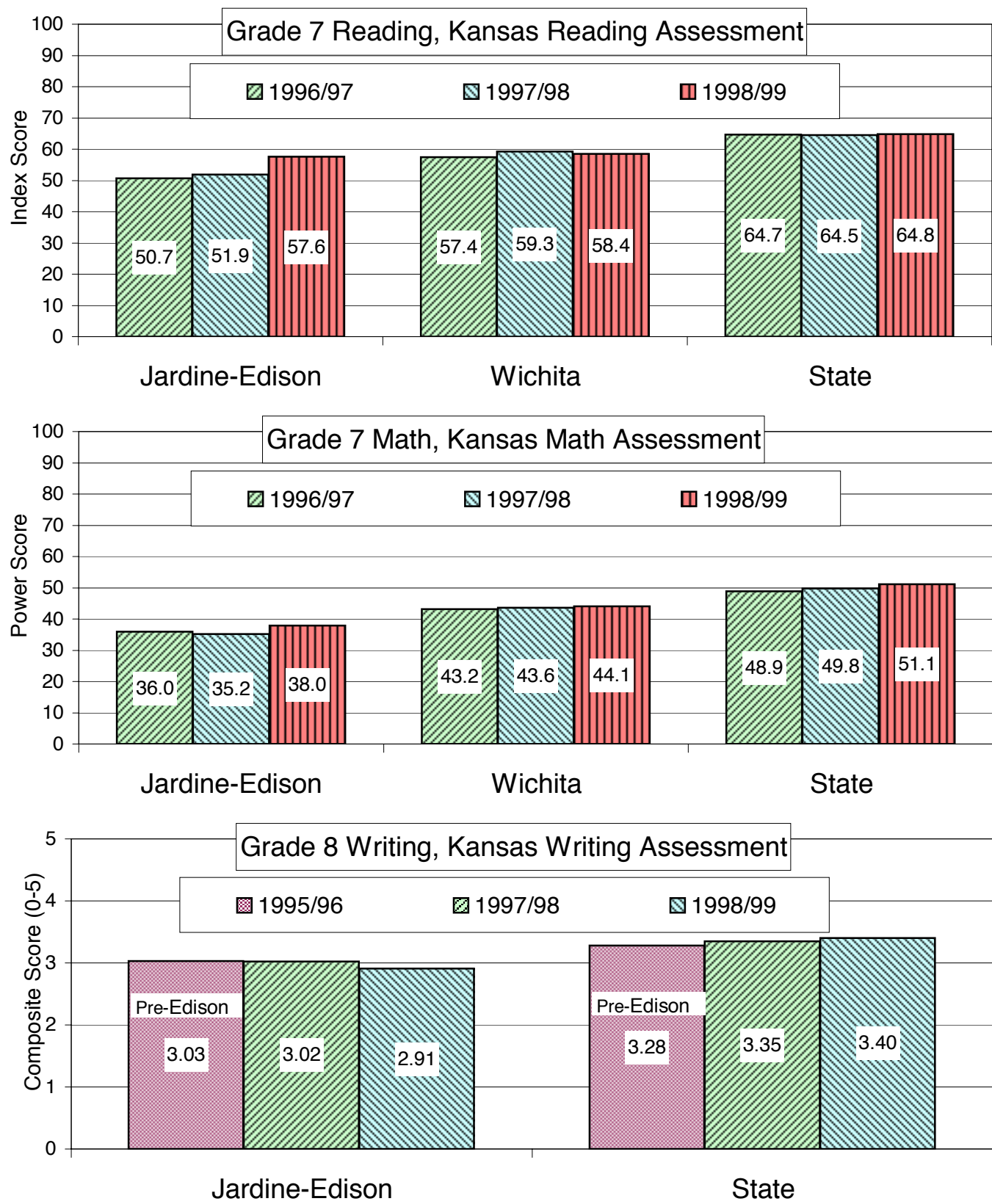
Note: Wichita Public Schools provided results for Jardine-Edison and the district on reading and math, while the Kansas Department of Education was the source of all state level results as well as all results on the grade 5 writing assessment. All assessment tests were administered in the spring of each year, except for the writing assessment, which was not administered in the spring of 1997. These scores do not include special education.

In math, Jardine-Edison had a slightly larger gain than the district, but the gain made by the state was still larger. The gain on the power score for Jardine-Edison was 1.96 between 1997 and 1999, while the gains for the same period of time were 0.98 for the district and 2.23 for the state. The total power score is an equally weighted average of the scores based on percent correct on the three components of this assessment test: problem solving, communication, and reasoning.

Results for the Kansas Writing Assessment were available for Jardine-Edison and the state. Here Jardine-Edison lost ground over the span of two years (notice that the test was not administered in the spring of 1997, so the results are for 1997/98 and 1998/99). The Jardine-Edison score decreased from 3.02 in 1997/98 to 2.91 in the 1998/99 school year (i.e., a decrease of 0.11).

The state average increased from 3.35 in 1997/98 to 3.40 in 1998/99 (an increase of 0.05). The composite score is the average score on six tests or rating scales, each ranging from 0 to 5, with 5 being the highest and 0 the lowest.

Exhibit 6:3 Results on the Kansas Assessments, Grade 7, Reading and Math; Grade 8 Writing
Comparison Between Jardine-Edison and Wichita School District, 1995/96 -1999/00



6.5 Summary

A summary score of -1 indicates a result that is unfavorable toward the sample school, a score of 0 indicates a neutral finding, and a score of 1 indicates a favorable result according to the criteria specified in Section 2.5.

Norm-referenced test findings

The design behind the trends in the norm-referenced results is based on tracing consecutive cohorts of students at three grade levels over three years. These trends were traced for performance on both reading and math. Due to limitations in the data, we could not distinguish the proportion of students who met state standards. Therefore, we could not conduct the odds ratio analysis. The results from the state assessment tests also lacked information on variability, so it was impossible for us to determine whether the changes over time were significant or not. Table 6:3 contains the summary results from the comparison between Jardine-Edison and the district on the Metropolitan Achievement Test. We have given positive ratings to three trends where the Jardine-Edison students made large average annual gains, in terms of their national percentile rank, as well as large gains compared to district gains, which were also large, particularly in math.

Table 6:3 Comparison of Gains in Terms of National Percentile Rank Between Jardine-Edison and Wichita Public Schools on the Metropolitan Achievement Test

	Jardine-Edison Gain on NPR Between 1996-99	Wichita Public Schools Gains on NPR Between 1996-99	Difference	Trend
Grade 6 Reading	-4	-3	-1	mixed (0)
Grade 6 Math	6	7	-1	mixed (0)
Grade 7 Reading	13	3	10	positive (+1)
Grade 7 Math	14	8	6	positive (+1)
Grade 8 Reading	9	-1	10	positive (+1)
Grade 8 Math	11	7	4	mixed (0)

Criterion-referenced test findings

Given the nature of the data and given the lack of details on variability, we could not do much with this criterion-referenced data. The change in the reading index score was still quite substantial and thus marked as a positive trend. The difference in the math power scores was quite small; this was

labeled as a mixed trend. The composite score on the Kansas Writing Assessment is based on a 0-5 scale, so the .5 gain made across the state is quite substantial even while Jardine was losing some ground. The difference in change scores (-.61) represents 12 percent of the total range of the scale. Even though we do not have a cut-off level for the composite score to determine whether this trend is positive, mixed, or negative, we have rated this as a negative trend because of the very large difference in change scores.

Table 6:4 Summary Results on Criterion-Referenced Tests: Comparison of Gains on Kansas Assessments Between Jardine-Edison and Wichita Public Schools

	Jardine-Edison Gains on Index or Power Scores Between 1996/97-98/99	Wichita Public Schools Gains on Index/Power Scores 1996/97 --1998/99	Difference	Trend
Grade 7 Reading Index Score (0-100)	6.85	1.0	5.85	positive (+1)
Grade 7 Math (0-100) Power Score	1.96	0.98	0.98	mixed (0)
	Jardine-Edison Gain on Composite Score Between 1995/96-98/99	State of Kansas Gain on Composite Score Between 1995/96-98/99		
Grade 8 Writing Composite Score (0-5)	-0.11	0.5	-0.61	negative (-1)

Combined ratings

Table 6:5 contains a summary of the trend ratings for both the norm-referenced and criterion-referenced test results.

Table 6:5 Combined Overall Trends for Jardine-Edison

	Positive	Mixed	Negative
Norm-Referenced	3 of 6	3 of 6	0 of 6
Criterion-Referenced	1 of 3	1 of 3	1 of 3
TOTALS	4 of 9	4 of 9	1 of 9

Given the total ratings for the trends that are highlighted in Table 6:5, we rate this school as Positive with a mean trend rating of 0.33. In its 1999 annual report, Edison rated the trends in this school as

Positive. In its 2000 annual report Edison rated the trends for the 1999-00 school year alone as Negative, yet at the same time labeled the trends since opening as Strongly Positive. This begs the question or explanation for how the Positive rating of overall trends in 1999, plus the Negative trends for the 1999-00 school, could equal the Strongly Positive rating of overall trends cited in the 2000 annual report (Edison, 2000, p. 78).

Although the label we used to categorize the overall trends in this school is positive, the trends are clearly not Strongly Positive as Edison suggests in its 2000 school report.

Chapter Seven

Boston Renaissance Charter School

Boston, Massachusetts

7.1 Descriptive Summary of School

The Boston Renaissance Charter School (BRCS) is an independent charter school established in 1995 to serve grades K-5 with grades 6-8 added in 1996. The mission of the Boston Renaissance Charter School is to prepare a diverse cross section of Boston children for success with a high quality education, all at prevailing public school costs. This mission statement is consistent for both Edison schools in Massachusetts.

During the 1998/99 school year there were 1,063 students enrolled in the school. The student body was 71.2 percent African American, 0.8 percent Asian, 13.5 percent Hispanic, 0.7 percent native American, and 13.8 percent white. The total proportion of minorities was similar to the Boston Public School District. However, Boston Public Schools had more Hispanic (26.1 percent) and Asian (8.9 percent) students and fewer African-American (49 percent) students. Both the Edison school and the district had considerably large proportions of minority students compared with an overall state percentage of 77.1 percent white and 32.9 percent minority.

According to Edison's second annual report of student achievement (Edison, 1999), there was an instructional staff of 75 with a ratio of 14.2 pupils per staff member for the 1997/98 school year. This teacher-to-student ratio differs from that reported in the Boston Renaissance Charter School District Profile, 1998-1999 (MDOE, 1999), which indicated the number of students per teacher was 16.8. This ratio is similar to the state average (18:1), but different from the Boston Public School District's ratio of 27 students per teacher. Student mobility at Boston Renaissance was 9.4 percent (Edison, 1999).

The number of students identified as having special educational needs at Boston Renaissance Charter School was constant for the 1997/98 and 1998/99 school years (10.9 percent), as reported in the state's school district profile (MDOE, 1999). This was down slightly from the 12.46 percent reported in the school's 1996-97 annual report (Boston Renaissance, 1997). These figures are consistent with those reported by Farber (1998) in an article that reported on questionable special education practices and policies at Boston Renaissance Charter School. The article reported that several teachers from the school suggested the school had an explicit policy against children with disabilities. One teacher even testified to the existence of such a policy before a committee of the Massachusetts legislature. The article also pointed out that there was an even greater discrepancy between students with severe

disabilities enrolled in Boston Public Schools (10 percent) and Boston Renaissance (1 percent). If these allegations are true, it is understandable why the Boston Renaissance percentage of students identified for special education services was relatively low when compared with 21.1 percent statewide and 16.6 percent for the Boston Public School District. Students identified as eligible for free or reduced priced lunch were also lower at Boston Renaissance (50.2 percent) compared with Boston Public School District (71.4 percent), but higher than the statewide average (25.8 percent).

In terms of students with limited English proficiency, the state's school district profile for Boston Renaissance (MDOE, 1999) indicated no students identified as limited English proficient, compared with 21.7 percent for Boston Public School District and 4.7 percent statewide. This varies only slightly from the 0.5 percent identified as ESL (English as second language) reported by Edison (Edison, 1999).

According to the state charter school report for 1997/98 (MDOE, 1998), Boston Renaissance had a total of 200 instructional days with a range of 1,151 to 1,225 hours of structured learning time. In addition, it was reported that 47 percent of the teachers were certified and there was a 9 percent turnover in staff. It also stated that 1,553 students were on a waiting list for enrollment.

The 1997/98 school year per pupil expenditure of \$7,666 (MDOE, 1998) at Boston Renaissance was greater than the reported state expenditure of \$6,361, but less than the reported \$8,118 for Boston Public School District. Teacher salary ranges were similar for Boston Renaissance (\$33,601-\$59,065) and Boston Public (\$33,281-\$59,847). Both were higher than the statewide average teacher salary range of \$27,387-\$52,037 (MDOE, 1999); however, this would be expected in an urban area.

What we can derive from the descriptive data we reviewed for this school is that the size and composition of the school have changed somewhat since it opened in 1995. Although only two additional grade levels were added, total enrollment nearly doubled. While the number of minority students increased from 78 percent to 86.2 percent, the number of low income students decreased from just over 60 percent to about 50 percent. In addition, the number of students receiving special education services was reduced by approximately 1.5 percent. In later sections of this chapter we will compare the performance of students in this school with the district and the state. Therefore, it is important to consider the differences in key background characteristics among the school, the district, and the state as a whole.

7.2 Past Studies and Evaluations and Data Available for Analysis

Mislevy reading study

The Mislevy Reading Achievement Analysis compared reading test scores gathered by using the Durrell Analysis of Reading Difficulty and the Woodcock Reading Mastery Tests. Progress in reading for Edison students was examined within the Edison school and with control groups. The

Mislevy studies conducted for the 1995-96 and 1996-97 school years could not report direct program effects; however, they suggested indirect program effects in favor of Edison.

In his 1998 report, Mislevy summarized the results of the *Success for All* reading/language arts program as mostly inconclusive. Mislevy also noted that the study design was undermined by such factors as attrition and the lack of a matched control group. Therefore, actual results with a specific positive link to this reading program were hard to determine without making adjustments in data. The results suggested that students who participated in the *Success for All* program the longest performed better on the tests.

Students who participated in the reading program for either two or three years have nearly the same average scores, showing that there may not be a significant factor that helps improve students' skills by using this program more than any other. Mislevy (1998) states that "main differences [in test results] are between sites rather than between programs," based on student background and the characteristics with which they entered Boston Renaissance. Nevertheless, because the students at Boston Renaissance were compared with an unmatched control group, it is difficult to draw clear conclusions.

School reports and renewal inspection report

A number of reports contain summative information for Boston Renaissance in terms of its academic performance. These include the school's own annual reports as well as statewide charter school reports. In 1999, the school was reviewed by an external inspection team that was reviewing the school's application to renew its charter. The findings from the review are listed below according to key questions the inspection team addressed (Renewal Inspection Report, 1999, pp. 4-14).

Is the academic program a success?

1. Student performance on external assessment did not achieve school expectations. Stanford 9 scores have generally declined between 1996-99. Grade 8 MCAS scores are below Boston Public School averages.
2. The schools's adaptation of Edison's internal assessment design does not function effectively and systematically at present: There is no Structured Portfolio System in place.
3. The Edison project has announced its intention to replace Annual Comprehensive Performance Assessments with monthly benchmark assessments, but no assessments were included in student folders and no summary information about these assessments was provided in the school's renewal application.
4. The volume of staff turnover and technological impediments has compromised student Quarterly Learning Contracts.
5. Scrutiny of student work and the observation of student classroom responses confirm that achievement is below standards set by the school and Edison.

Is the school a viable organization?

1. The board of trustees has demonstrated its strong commitment and has played an appropriate role in the school's history. It recognizes that its fundamental responsibility is at the policy and not at the management level.
2. The board of trustees, Edison staff, school administrators, and faculty do not share a common understanding of the Edison-Renaissance partnership and its lines of accountability. . . . Lines of accountability and reporting established in the original contract have become blurred.
3. The appointment of the new principal has been applauded consistently across every constituency. He has exhibited vision, thoughtful decision-making, and interest in the welfare of students and staff.
4. The school's environment is not safe and orderly. At the same time, the amount of emphasis and time spent on behavior control has encroached on the academic program.
5. The high turnover of faculty and shortage of veteran educators have compromised the mission of the school.
6. There is an acute need for professional development of all staff surrounding academic content, child development, the craft of teaching, and curriculum.

Has the school been faithful to the terms of its charter?

1. The academic program as presently implemented does not sufficiently challenge students to reach Edison-Renaissance standards and expectations.
2. The Edison aim to 'use technology as a second language' is not being met. The school has taken the first step to begin to address this issue by installing new hardware.
3. Significant progress has been made toward the goal of involving the Boston community in all aspects of school life.
4. A newly established parent center and the addition of a new position, Parent and Family Coordinator, are significant steps toward the school's commitment to increase family involvement.
5. Parental support of the school is strong and represents general satisfaction with the school and its program.

If the school's charter is renewed, what are its plans for the next five years of the charter?

1. The school, in its unusually candid and reflective response to this renewal question, identifies many of the same institutional issues remarked by the inspection team. Ambitious action planning, still to be prioritized and organized into shorter-term targets, clearly demonstrates the school's commitment to move forward towards its charter vision.

Data available for our analyses

Edison provided us with an SPSS data set containing the Stanford Achievement Test, Version 9 (SAT-9) test data for four years (1996-1999). Like all Massachusetts public schools, this school is also required to take part in the state assessment program (MCAS). We were able to secure MCAS test results for only two years, 1998 and 1999, from the Massachusetts Department of Education,

Tests administered at this school include the Massachusetts Comprehensive Assessment Systems (MCAS) in Spring 1998 and Spring 1999; the Stanford Achievement Test (SAT) in 1996, 1997, 1998, 1999, and 2000; the Metropolitan Assessment Test (MAT) in 1995-96, 1996-97,; and the Mislevy Reading Study beginning with the 1995-96 school year.

7.3 Longitudinal Analysis of Individual Student Data

Edison Schools Inc. provided individual student achievement data for the SAT-9 over a four-year administration period. SAT-9 scores were provided for two subtests—mathematics and reading—in four separate scales: grade equivalent scores (GE), standard scores (SS), national percentile rank scores (PR), and normal curve equivalent scores (NCE). Parallel analyses are reported for each scale.

Data records included 1,107 students covering the 1997-1999 academic years. We defined six different longitudinal cohorts of students over the four years. Table 7:1 details the various cohorts, grades, and years. For example, Cohort A progressed from third to fifth grade in the academic years 1997-1999.

Table 7:2 presents sample size information for each SAT-9 score scale for each subtest by year and grade level. Also depicted in this table are the sample sizes for the one-year gain analysis and for each cohort on the longitudinal analyses.

Table 7:1 Description of Longitudinal Cohorts

Cohort	Years	Grades
A	97 - 98 - 99	3 - 4 - 5
B	97 - 98 - 99	4 - 5 - 6
C	97 - 98 - 99	5 - 6 - 7
D	97 - 98 - 99	6 - 7 - 8
E	97 - 98 - 99	7 - 8 - 9
F	96 - 97 - 98 - 99	5 - 6 - 7 - 8

Care must be maintained when interpreting these analyses due to their nonindependence. Each cohort overlaps with another. Thus, findings related to one cohort will tend to parallel the adjacent cohort. For example, Cohorts A and B will tend to be more similar than Cohorts A and C.

Table 7:2 Sample Sizes for Individual Student Data by Grade and Year on the SAT-9

Grade 3								
	1996		1997		1998		1999	
	Reading	Mathematics	Reading	Mathematics	Reading	Mathematics	Reading	Mathematics
GE	110	110	110	110	103	103	111	111
SS	110	110	110	110	103	103	111	111
PR	105	108	100	99	95	100	108	110
NCE	105	108	100	99	95	100	108	110
Grade 4								
	1996		1997		1998		1999	
	Reading	Mathematics	Reading	Mathematics	Reading	Mathematics	Reading	Mathematics
GE			108	108	100	100		
SS			108	108	100	100		
PR			101	106	97	92		
NCE			101	106	97	92		
Grade 5								
	1996		1997		1998		1999	
	Reading	Mathematics	Reading	Mathematics	Reading	Mathematics	Reading	Mathematics
GE	103	103	108	108	112	112	109	109
SS	103	103	108	108	112	112	109	109
PR	102	103	103	108	105	112	107	108
NCE	102	103	103	108	105	112	107	108
Grade 6								
	1996		1997		1998		1999	
	Reading	Mathematics	Reading	Mathematics	Reading	Mathematics	Reading	Mathematics
GE			165	165	119	119	128	128
SS			165	165	119	119	128	128
PR			154	153	119	119	123	128
NCE			154	153	119	119	123	128
Grade 7								
	1996		1997		1998		1999	
	Reading	Mathematics	Reading	Mathematics	Reading	Mathematics	Reading	Mathematics
GE			102	102	146	146	100	100
SS			102	102	146	146	100	100
PR			100	96	141	130	99	100
NCE			100	96	141	130	99	100
Grade 8								
	1996		1997		1998		1999	
	Reading	Mathematics	Reading	Mathematics	Reading	Mathematics	Reading	Mathematics
GE			86	86	103	103	132	132
SS			86	86	103	103	132	132
PR			81	79	102	98	124	127
NCE			81	79	102	98	124	127

Grade 9											
1996		1997		1998		1999					
Reading		Mathematics		Reading		Mathematics		Reading		Mathematics	
GE								52		52	
SS								52		52	
PR								50		51	
NCE								50		51	
Gains in 1997											
3 to 4 grade		5 to 6 grade									
	Reading	Math	Reading	Math							
GE	84	87	84	85							
SS	84	89	84	85							
PR	84	87	84	85							
NCE	84	87	84	85							
Gains in 1998											
3 to 4 grade		4 to 5 grade		5 to 6 grade		6 to 7 grade		7 to 8 grade			
	Reading	Math	Reading	Math	Reading	Math	Reading	Math	Reading	Math	
GE	83	78	90	95	85	89	94	85	74	69	
SS	83	78	90	95	85	89	94	85	74	69	
PR	83	78	90	95	85	89	94	85	74	69	
NCE	83	78	90	95	85	89	94	85	74	69	
Gains in 1999											
4 to 5 grade		5 to 6 grade		6 to 7 grade		7 to 8 grade		8 to 9 grade			
	Reading	Math	Reading	Math	Reading	Math	Reading	Math	Reading	Math	
GE	83	78	93	103	80	81	109	106	44	43	
SS	83	78	93	103	80	81	109	106	44	43	
PR	83	78	93	103	80	81	109	106	44	43	
NCE	83	78	93	103	80	81	109	106	44	43	
97 to 98 to 99											
Cohort A		Cohort B		Cohort C		Cohort D		Cohort E			
SAT-9	Reading	Math	Reading	Math	Reading	Math	Reading	Math	Reading	Math	
GE	72	68	73	79	55	60	74	73	33	31	
SS	72	68	73	79	55	60	74	73	33	31	
PR	72	68	73	79	55	60	74	73	33	31	
NCE	72	68	73	79	55	60	74	73	33	31	
Cohort F (96 to 99)											
SAT-9	Reading	Math									
GE	41	43									
SS	41	43									
PR	41	43									
NCE	41	43									

Repeated measures ANOVAs were examined for the six separate cohorts for longitudinal trends over a three-year period (four-year period for Cohort F) for Boston Renaissance Charter School. Parallel analyses are reported for all types of scores reported by Edison, e.g., grade equivalent (GE), standard scores (SS), percentile rank (PR) (or national percentile rank), and normal curve equivalent (NCE) score for individual student data on the SAT-9 mathematics and reading subtests. In all models, the assumption of sphericity was evaluated and, if found to be violated, the Huynh-Feldt adjusted p-values were reported. If the overall linear model was found to be statistically significant, unadjusted (alpha) pair-wise comparisons were examined to identify where a difference in the means might be located.

Cohort A. In terms of all four reported scores, statistically significant increases in the means scores were noted in the SAT-9 reading subtest for GE and SS (see Exhibit 7:1). Alternatively, results showed statistically significant changes for the PR and NCE scores, such that there was an initial decrease in year two (grade 4), which was recovered in year three (grade 5). Generally, students in BRCS appear to be meeting grade level expectations in reading as measured by the SAT-9.

A similar picture emerges for the SAT-9 mathematics subtest. BRCS students show statistically significant increases on the GE and SS scales with an initial drop in year two that is recovered in year three on the PR and NCE scales. It should be noted that the Cohort A students, by the end of year three (grade 5), are achieving at a level considerably above grade level.

Cohort B. In terms of all four reported scores, statistically significant increases in the means scores were noted in each year in the SAT-9 reading subtest for GE and SS (see Exhibit 7:2). A similar increasing trend in the means was noted for the PR and NCE scores. In these two analyses, there was a statistically significant overall gain over the three-year period, but not a stepwise gain. Generally, students in Cohort B appear to be meeting grade level expectations in reading as measured by the SAT-9.

Paralleling the Cohort A finding in mathematics, BRCS students showed statistically significant increases each year on the GE and SS scales. However, on the PR and NCE scales, these same students evidenced an initial drop in year two, which was recovered in year three and even exceeded year one means. It should be noted that by the end of year three (grade 6), the Cohort B students were achieving at a level considerably above grade level.

Cohort C. Statistically significant increases in the means scores were noted each year for the SAT-9 reading subtest for the GE and SS scores (see Exhibit 7:3). However, there was no change noted for the PR and NCE score scales. Generally, students in Cohort C, although showing achievement gains, did not advance relative to the SAT-9 normative group (e.g., no statistically significant increase in PR or NCS scores). Moreover, this cohort appeared to be lagging behind in relation to grade level expectations in reading as measured by the SAT-9.

Cohort C findings in mathematics show a somewhat different picture. Statistically significant increases are noted only in the third year on the GE and SS score scales. Regarding the PR and NCE scales, these same students evidenced an initial drop in year two, which was recovered in year three.

By the end of year three (grade 7), Cohort C students were struggling to meet grade level expectations in mathematics.

Cohort D. Statistically significant increases in the means scores were noted for each year in the SAT-9 reading subtest for the GE and SS scores (see Exhibit 7:4). Alternatively, results showed statistically significant changes for the PR and NCE scores. There was a statistically significant decrease in year two (grade 7), which was recovered in year three (grade 8) in the NCE score scale and exceeded year one on the PR score scale. Students in Cohort D met, even exceeded, grade level expectations in reading as measured by the SAT-9.

Cohort D findings in mathematics closely parallel SAT-9 reading subtest findings. Statistically significant increases are noted each year on the GE and SS score scales. On the PR and NCE scales, these same students evidenced an initial drop (statistically significant) in year two, which was recovered and then exceeded year one in year three. by the end of year three (grade 8), Cohort D students were meeting grade level expectations in mathematics.

Cohort E. Statistically significant increases in the means scores were noted in each of the three years in the SAT-9 reading subtest for the GE and SS scores (see Exhibit 7:5). However, there were no statistically significant gains in this cohort until the third year for the PR and NCE scores. Students in Cohort E were below grade level expectations in reading as measured by the SAT-9.

Cohort E gains in mathematics were weaker, but tended to parallel the SAT-9 reading subtest findings. Statistically significant increases were noted only after the three-year period on the GE and SS score scales. However, there were no statistically significant gains or losses for the PR and NCE scores on the mathematics subtest. By the end of year three (grade 9), Cohort E students were well below grade level expectations in mathematics.

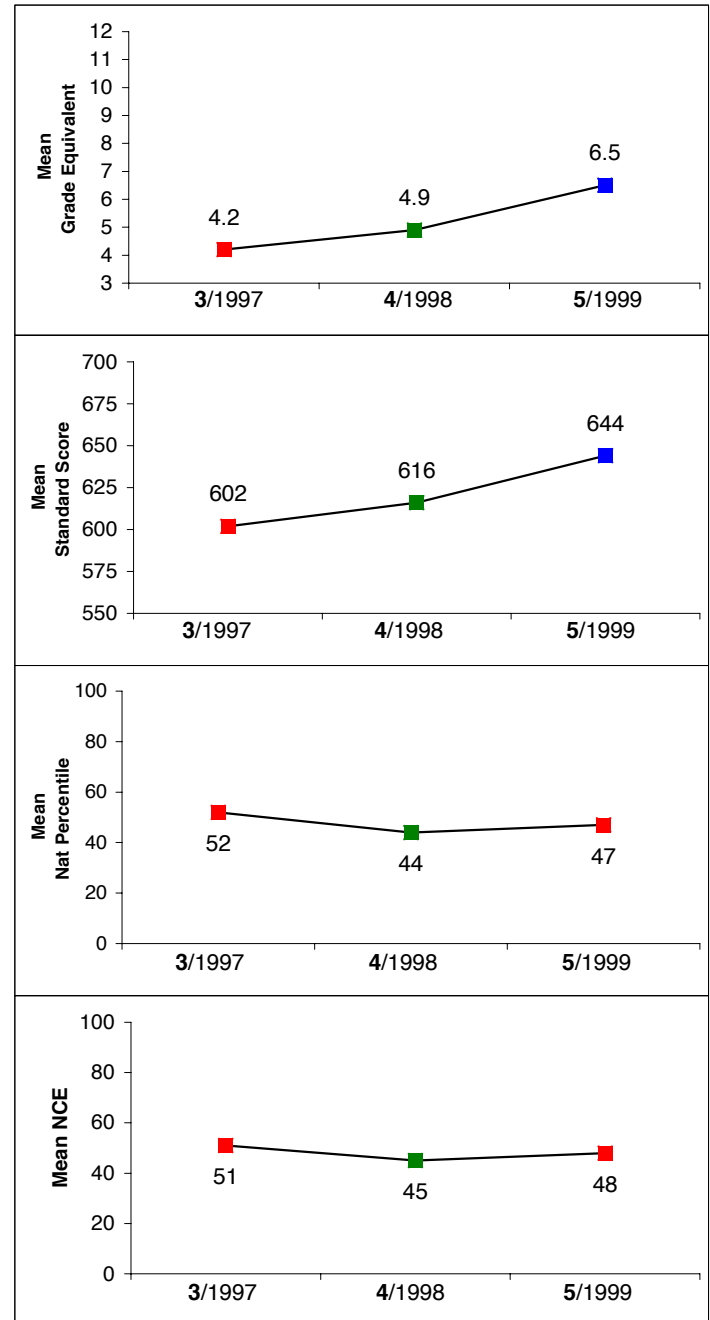
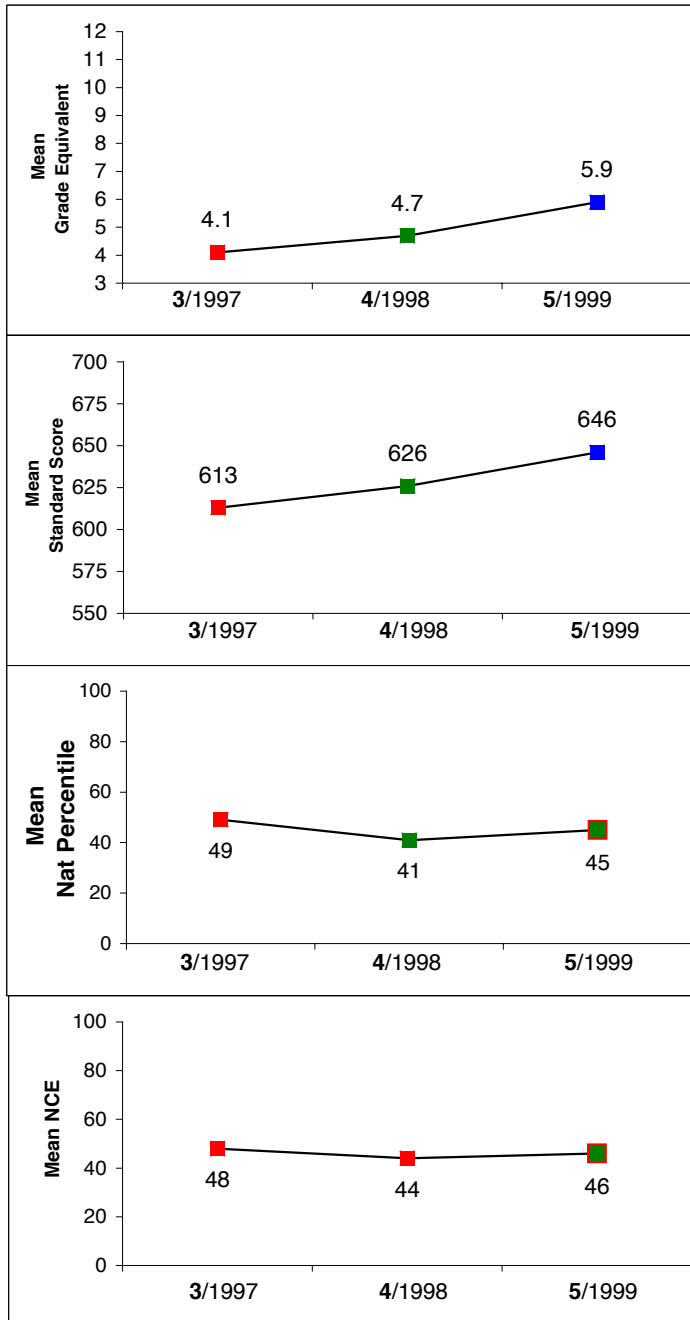
Cohort F. Cohort F represents a four-year look at SAT-9 achievement (see Exhibit 7:6). In the last two years statistically significant increases in the means scores were noted in the SAT-9 reading subtest for the GE and SS scores. However, there was a decreasing trend in PR and NCS score scales during the first three years that bottomed out by year three (grade 7) such that it represented a statistically significant decrease relative to year one (grade 5). Year 4 (grade 9) represented a turnaround year for this cohort in that there was a statistically significant gain over the year three means. Students in Cohort F appeared to be meeting grade level expectations in reading as measured by the SAT-9.

Cohort F gains in mathematics closely paralleled the SAT-9 reading subtest findings. However, by the end of year four (grade 9), Cohort F students were well below grade level expectations in mathematics.

Exhibit 7:1 Boston Renaissance Charter School, Cohort A: Findings From the Analysis of Individual Student Results on the Stanford Achievement Test

Cohort A	READING			
n=72	1997	1998	1999	
Grade	3rd	4th	5th	
GE Reading	4.1	4.7	5.9	$F(2,142)=64.93, p<.0001$
SS Reading	613	626	646	$F(2,142)=73.08, p<.0001$
PR Reading	49	41	45	$F(2,142)=6.62, p=.0018$
NCE Reading	48	44	46	$F(2,142)=4.94, p=.0084$

Cohort A	MATH			
n=68	1997	1998	1999	
Grade	3rd	4th	5th	
GE Math	4.2	4.9	6.5	$F(2,134)=60.99, p<.0001$
SS Math	602	616	644	$F(2,134)=98.96, p<.0001$
NP Math	52	44	47	$F(2,134)=5.67, p=.0063$
NCE Math	51	45	48	$F(2,132)=5.29, p=.0075$

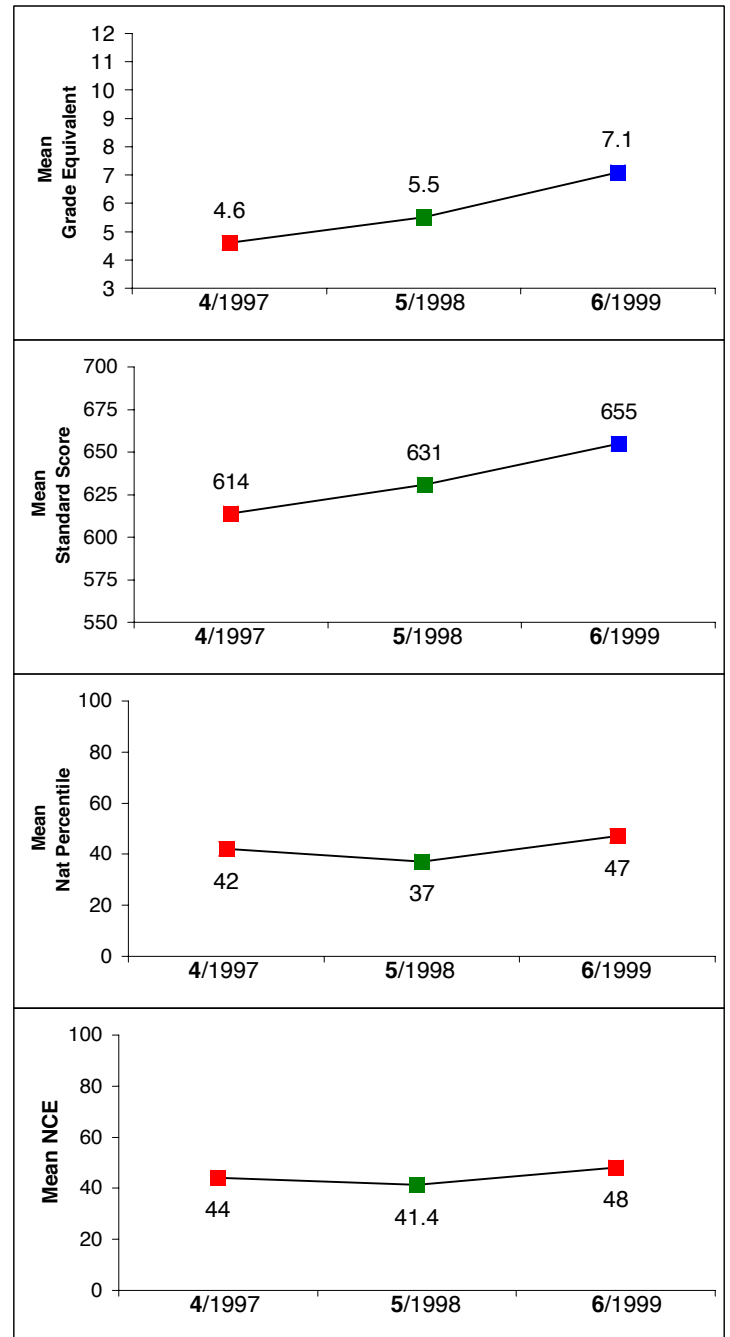
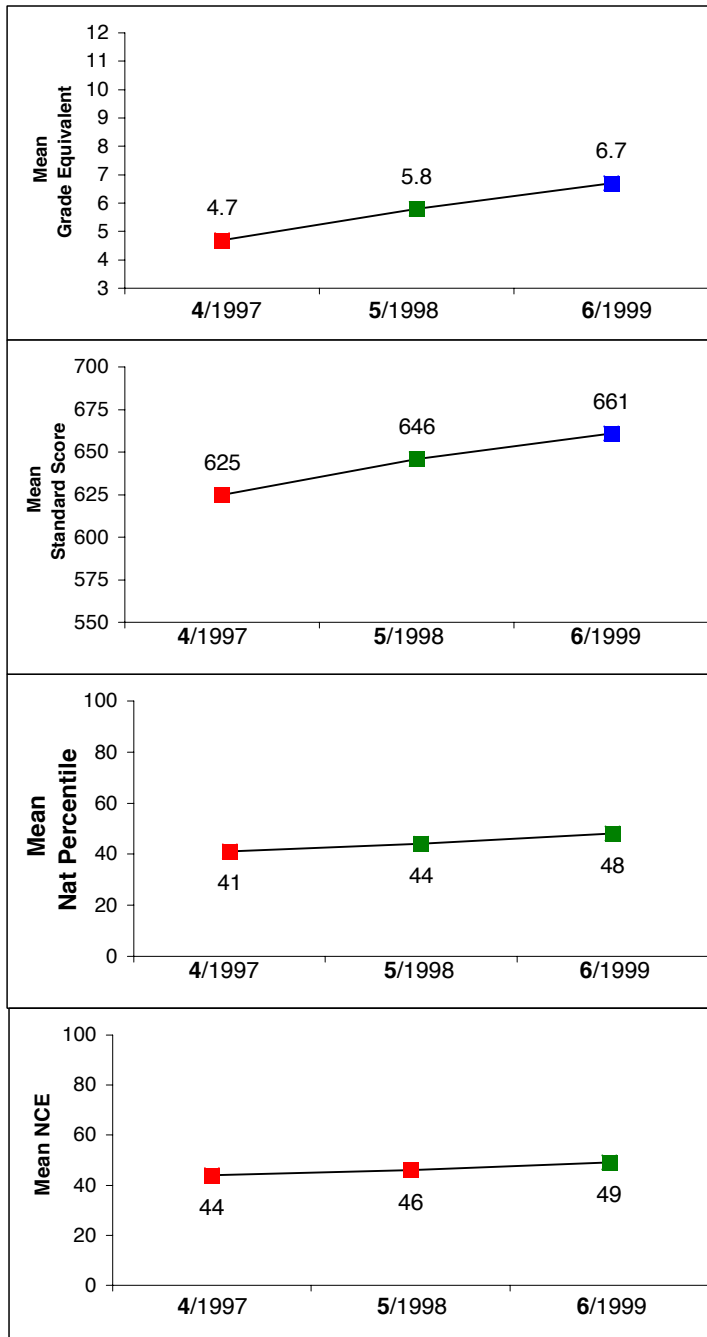


A color change represents a statistically significant change in the means. A trend with two color changes represents differences among all three means. A dual colored charting point (red and green) represents a statistically significant difference between one mean but not the other. The reader is encouraged to examine the individual table of means for these cases.

Exhibit 7:2 Boston Renaissance Charter School, Cohort B: Findings From the Analysis of Individual Student Results on the Stanford Achievement Test

Cohort B	READING			
n=73	1997	1998	1999	
Grade	4th	5th	6th	
GE Reading	4.7	5.8	6.7	$F(2,144)=76.63, p<.0001$
SS Reading	625	646	661	$F(2,144)=109.11, p<.0001$
PR Reading	41	44	48	$F(2,144)=7.25, p=.0010$
NCE Reading	44	46	49	$F(2,144)=8.77, p=.0003$

Cohort B	MATH			
n=79	1997	1998	1999	
Grade	4th	5th	6th	
GE Math	4.6	5.5	7.1	$F(2,156)=83.78, p<.0001$
SS Math	614	631	655	$F(2,156)=152.98, p<.0001$
NP Math	42	37	47	$F(2,156)=14.48, p<.0001$
NCE Math	44	41.4	48	$F(2,156)=16.69, p<.0001$

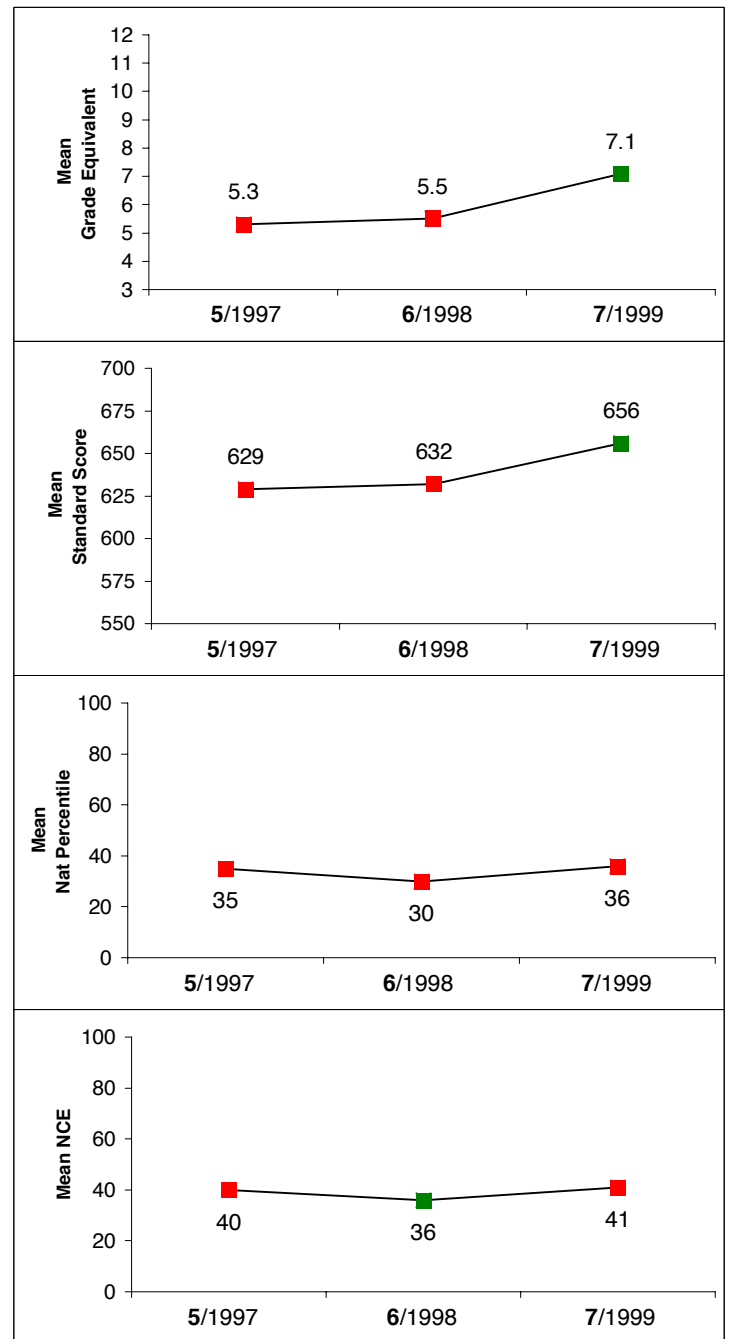
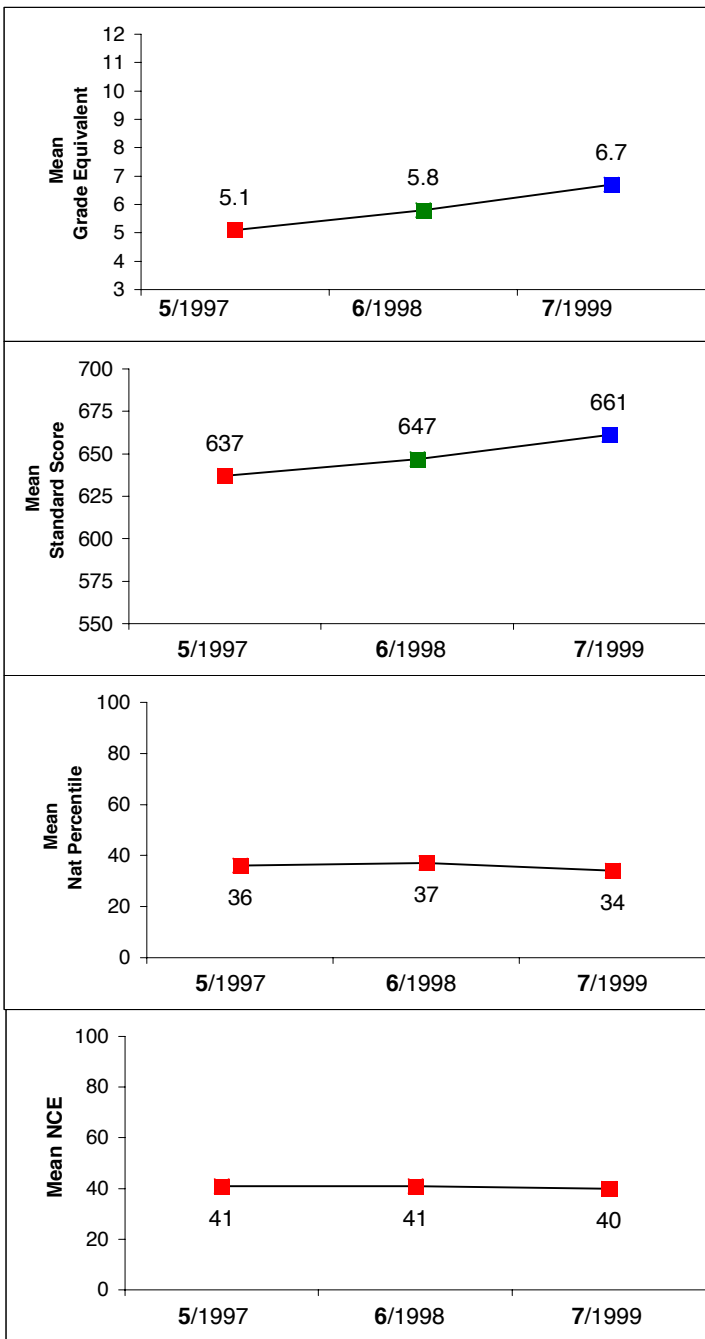


A color change represents a statistically significant change in the means. A trend with two color changes represents differences among all three means. A dual colored charting point (red and green) represents a statistically significant difference between one mean but not the other. The reader is encouraged to examine the individual table of means for these cases.

Exhibit 7:3 Boston Renaissance Charter School, Cohort C: Findings From the Analysis of Individual Student Results on the Stanford Achievement Test

Cohort C	READING			
n=55	1997	1998	1999	
Grade	5th	6th	7th	
GE Reading	5.1	5.8	6.7	$F(2,108)=30.29, p<.0001$
SS Reading	637	647	661	$F(2,108)=38.72, p<.0001$
PR Reading	36	37	34	$F(2,108)=0.61, p=.5479$
NCE Reading	41	41	40	$F(2,108)=0.25, p=.7766$

Cohort C	MATH			
n=60	1997	1998	1999	
Grade	5th	6th	7th	
GE Math	5.3	5.5	7.1	$F(2,118)=38.92, p<.0001$
SS Math	629	632	656	$F(2,118)=51.82, p<.0001$
NP Math	35	30	36	$F(2,118)=3.06, p=.0506$
NCE Math	40	36	41	$F(2,118)=4.76, p=.0103$



A color change represents a statistically significant change in the means. A trend with two color changes represents differences among all three means. A dual colored charting point (red and green) represents a statistically significant difference between one mean but not the other. The reader is encouraged to examine the individual table of means for these cases.

Exhibit 7:4 Boston Renaissance Charter School, Cohort D: Findings From the Analysis of Individual Student Results on the Stanford Achievement Test

Cohort D

READING

n=74

Grade

GE Reading

SS Reading

PR Reading

NCE Reading

1997 1998 1999

6th 7th 8th

6.7 7.4 9.4

661 671 693

48 42 52

49 45 51

$F(2,146)=92.40, p<.0001$

$F(2,146)=114.46, p<.0001$

$F(2,146)=14.43, p<.0001$

$F(2,146)=12.92, p<.0001$

Cohort D

MATH

n=73

Grade

GE Math

SS Math

NP Math

NCE Math

1997 1998 1999

6th 7th 8th

5.9 6.2 8.0

639 646 668

34 28 39

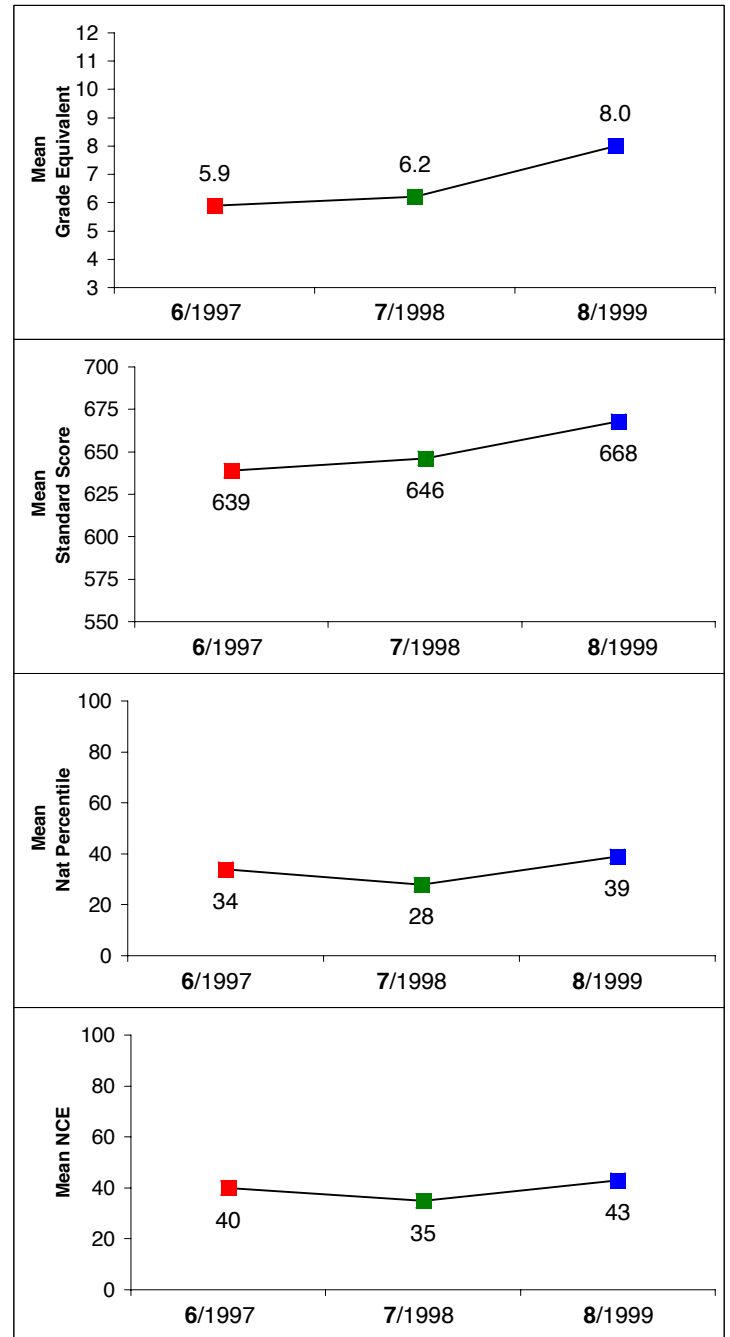
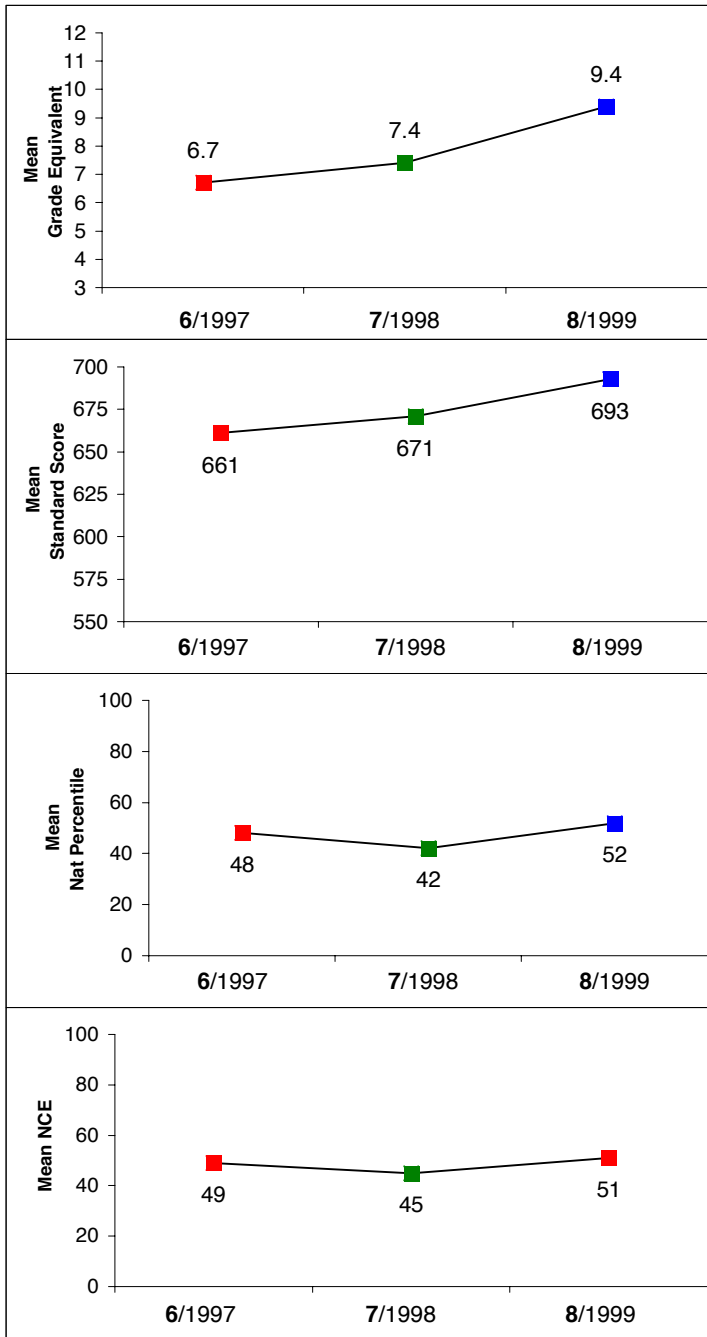
40 35 43

$F(2,144)=73.85, p<.0001$

$F(2,144)=75.93, p<.0001$

$F(2,144)=19.17, p<.0001$

$F(2,144)=16.00, p<.0001$

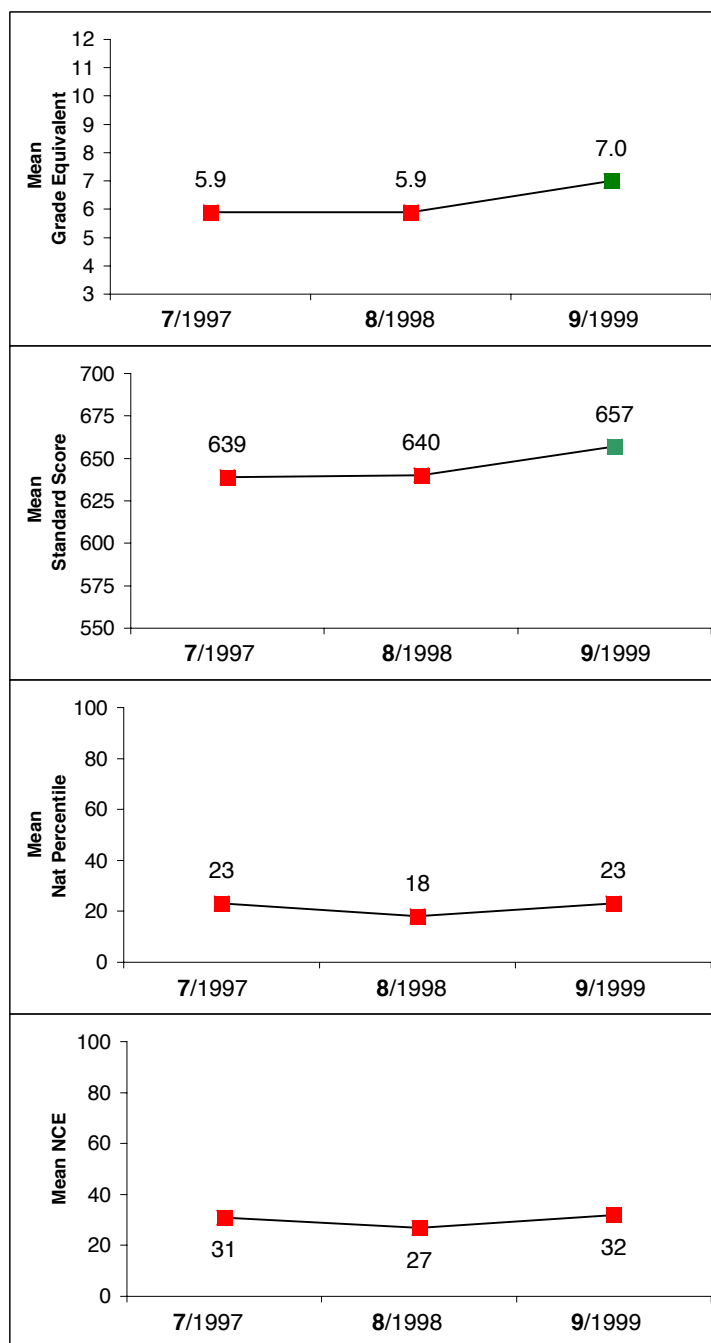
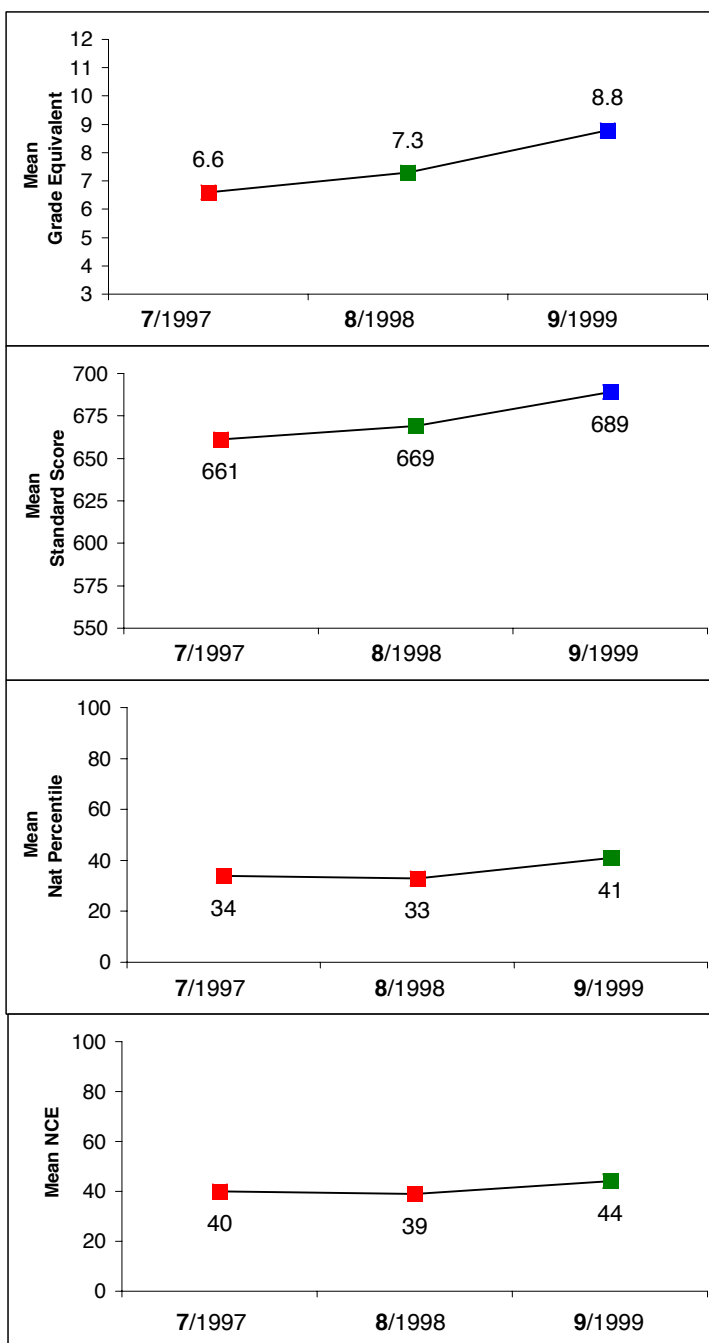


A color change represents a statistically significant change in the means. A trend with two color changes represents differences among all three means. A dual colored charting point (red and green) represents a statistically significant difference between one mean but not the other. The reader is encouraged to examine the individual table of means for these cases.

Exhibit 7:5 Boston Renaissance Charter School, Cohort E: Findings From the Analysis of Individual Student Results on the Stanford Achievement Test

Cohort E	READING			
n=33	1997	1998	1999	
Grade	7th	8th	9th	
GE Reading	6.6	7.3	8.8	$F(2,64)=35.57, p<.0001$
SS Reading	661	669	689	$F(2,64)=42.23, p<.0001$
PR Reading	34	33	41	$F(2,64)=6.74, p=.0022$
NCE Reading	40	39	44	$F(2,64)=5.42, p=.0067$

Cohort E	MATH			
n=31	1997	1998	1999	
Grade	7th	8th	9th	
GE Math	5.9	5.9	7.0	$F(2,60)=12.51, p=.0002$
SS Math	639	640	657	$F(2,60)=14.86, p<.0001$
NP Math	23	18	23	$F(2,60)=2.70, p=.0898$
NCE Math	31	27	32	$F(2,60)=2.75, p=.0723$

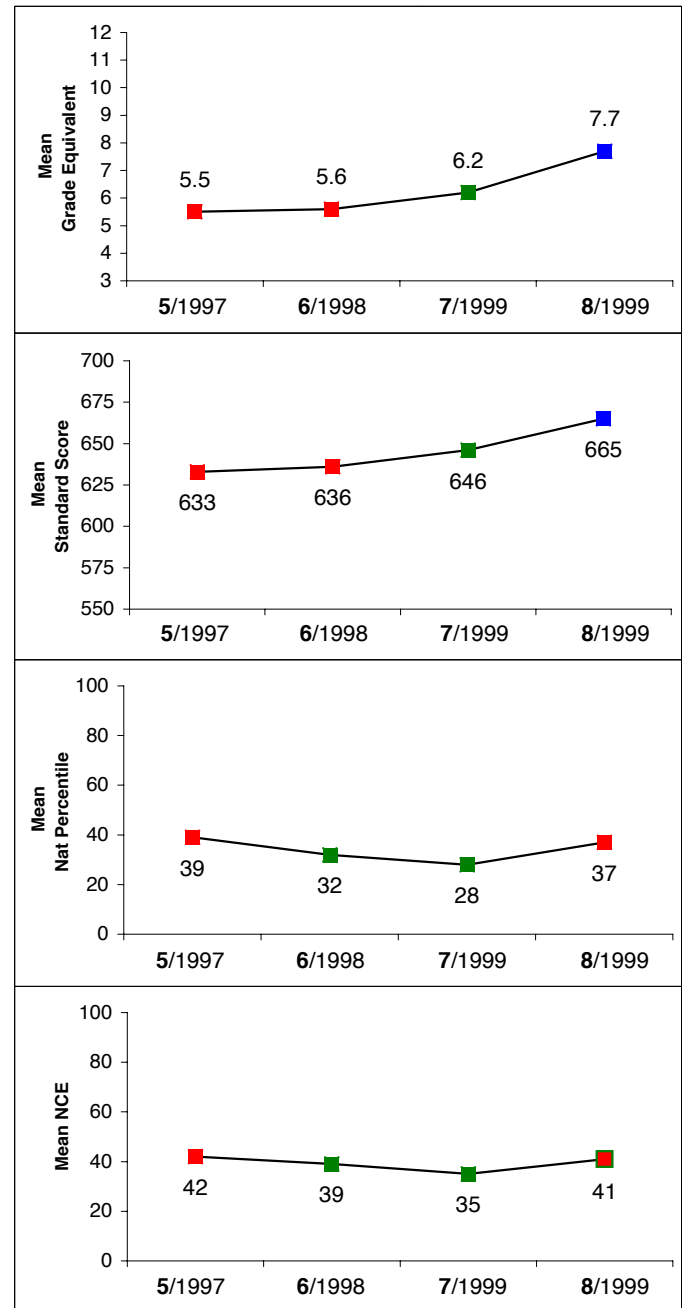
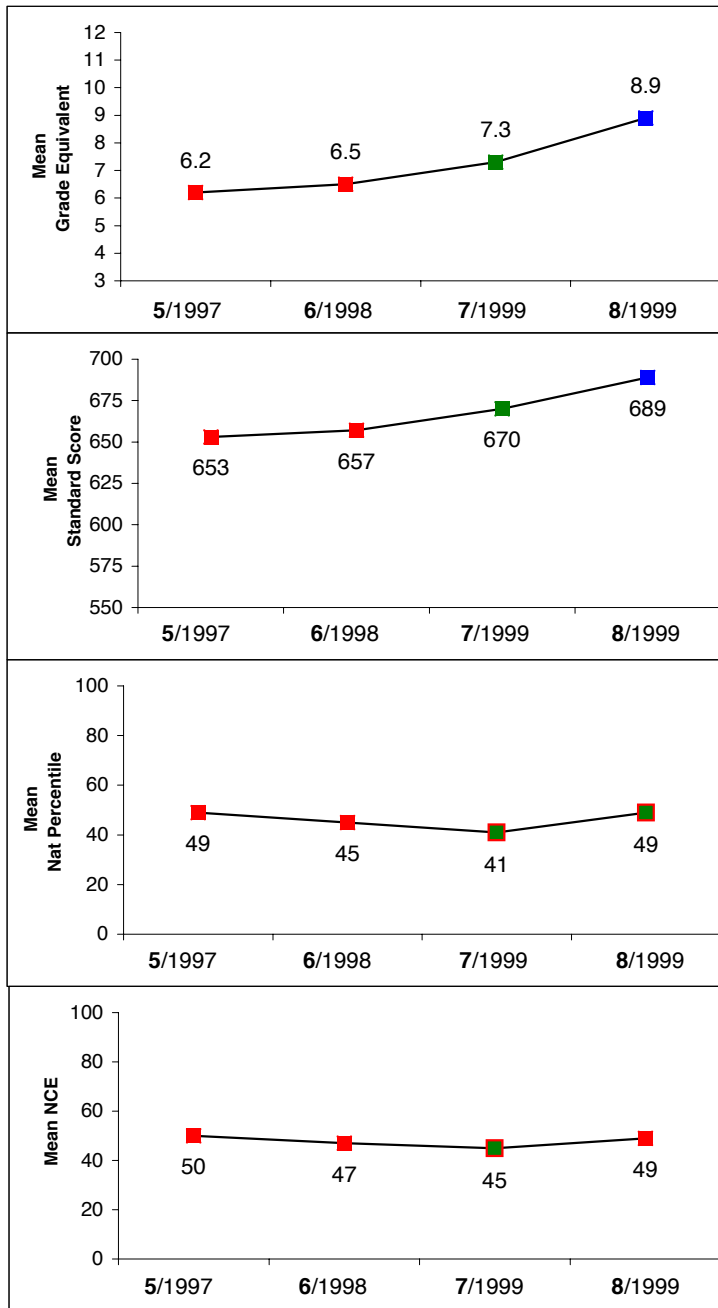


A color change represents a statistically significant change in the means. A trend with two color changes represents differences among all three means. A dual colored charting point (red and green) represents a statistically significant difference between one mean but not the other. The reader is encouraged to examine the individual table of means for these cases.

Exhibit 7:6 Boston Renaissance Charter School, Cohort F: Findings From the Analysis of Individual Student Results on the Stanford Achievement Test

Cohort F	READING				
n=41	1996	1997	1998	1999	
Grade	5th	6th	7th	8th	
GE	6.2	6.5	7.3	8.9	$F(3,120)=43.78, p<.0001$
SS	653	657	670	689	$F(3,120)=51.70, p<.0001$
PR	49	45	41	49	$F(3,120)=3.82, p=.0118$
NCE	50	47	45	49	$F(3,120)=3.40, p=.0202$

Cohort F	MATH				
n=43	1996	1997	1998	1999	
Grade	5th	6th	7th	8th	
GE	5.5	5.6	6.2	7.7	$F(3,126)=42.85, p<.0001$
SS	633	636	646	665	$F(3,126)=40.15, p<.0001$
NP	39	32	28	37	$F(3,126)=7.36, p<.0001$
NCE	42	39	35	41	$F(3,126)=6.12, p=.0006$



A color change represents a statistically significant change in the means. A trend with two color changes represents differences among all three means. A dual colored charting point (red and green) represents a statistically significant difference between one mean but not the other. The reader is encouraged to examine the individual table of means for these cases.

7.4 Chi-Square Analysis of MCAS Data

A chi-square analysis was initiated on data made available by the state of Massachusetts on the outcomes of the Massachusetts Comprehensive Assessment Systems (MCAS), the state-mandated criterion-referenced test. The MCAS is administered in grades 4 (English, mathematics, and science and technology) and grade 8 (English, mathematics, history, and science and technology). However, we did not select the grade 8 history subject test because there was only one year of data reported on the Department of Education Web site. The MCAS is scored along an ordinal four-category scale: Fail, Needs Improvement, Proficient, and Advanced. Additional information on the MCAS is provided in Appendix A.

Construction of the comparison groups

We constructed two different comparison groups for the chi-square analyses. Since we were interested in examining the number/proportion of students who met state standards (“passing”) or conversely the number/proportion of students who did not meet state standards (“failing”) on the MCAS, we needed to define a suitable comparison group. Our first comparison is with the district in which BRCS resides. The second comparison group we selected was the state passing/failing rates. While the state demographics vary from those at BRCS, we believe that comparisons with state averages can yield further information regarding the relative gains of this Edison school. Also, since Edison claims that advances in other district schools is—in part—due to its presence, we use the state as a more distant point of comparison that cannot be easily influenced by the presence of Edison’s schools.

General procedure

Utilizing published data from the state of Massachusetts, we made two-year comparisons (1998 and 1999) for both grade 4 and grade 8. Percentage data (students in each scoring category) were converted to raw frequency data prior to chi-square analysis. To insure independence of the rows in the chi-square tables, the raw frequencies for each scoring category of the MCAS in the district and state comparisons were down-weighted by subtracting the number of BRCS students in that category. Thus, both the district and state numbers reflect all students in the district or state exclusive of those in Boston Renaissance Charter School.

Four chi-square analyses were evaluated for each subtest nested within year and grade level. Two of these analyses were on uncollapsed data; that is, all scoring levels were represented in the contingency table (e.g., a 2x4) for the district and state comparisons. Two follow-up analyses were conducted on the data after collapsing the multilevel scoring into a dichotomy (pass, fail), thus producing 2x2 contingency tables. According to the Massachusetts Department of Education, a score in the “proficient or advanced” category constitutes “passing” or meeting the state standards for that particular grade and subject. On the other hand, the “fail and needs improvement” categories mean that the student has not meet the state passing standards. Students who have scores in these two categories have not met state standards and fall into the “fail” category in our 2x2 chi-square and odds-ratio analyses.

Chi-square findings

These chi-square analyses are testing the null hypothesis that the relative frequency (of students) in the four (or two) scoring categories are the same for BRCS and the comparison group (either district or the state).

Results of the chi-square analyses for grade 4 covered the MCAS administration for the two years from 1998 to 1999 for English, mathematics, and science and technology. Individual contingency tables are presented in Appendix E.

The first set of comparisons were made against district data as the comparison group, the second set against the state. Six separate chi-square statistics were evaluated from 2x4 contingency tables and six from 2x2 contingency tables (see Table 7:3) for each year. Results among the three subtests taken at grade 4 closely parallel one another. Statistically significant differences were observed against the state proportions, but not against the district proportions in each year and subtest. Fewer students fell into the “Proficient” and “Advanced” categories in BRCS relative to the overall state, but not relative to the district. This pattern was also replicated in the collapsed analyses (2x2).

Results of the grade 8 MCAS chi-square analysis are summarized in Table 7:4. Similar to the grade 4 chi-square analyses, the analyses for grade 8 covered the MCAS administration for the two years from 1998 to 1999 for English, mathematics, and science and technology. Individual contingency table results are presented in Appendix E.

The first set of comparisons were made against district data as the comparison group, the second against the state. Six separate chi-square statistics were evaluated from 2x4 contingency tables and

Table 7:3 Summary of Chi-Square Findings for Boston Renaissance Charter Schools Grade 4

	1998	1999
English		
BRCS vs. District	ns/ns	ns/ns
BRCS vs. State	sig/sig	sig/sig
Mathematics		
BRCS vs. District	ns/ns	ns/ns
BRCS vs. State	sig/sig	sig/sig
Science & Technology		
BRCS vs. District	ns/ns	ns/ns
BRCS vs. State	sig/sig	sig/sig

Note: Each result cell in the matrix is divided, with the results for the 2x4 analysis on the left-hand side and the results for 2x2 analysis on the right-hand side (i.e., 2x4/2x2).

six from 2x2 contingency tables (see Table 7:4) for each year. Results among the three subtests taken at grade 8 closely parallel one another. In all chi-square analyses statistically significant differences were observed against both the district and state comparisons. However, a notable exception is in the district comparison for the English subtest. BRCS students had comparable proportions among the four scoring levels in both 1998 and 1999. In all other comparisons (state and district) a greater proportion of BRCS students tended to meet either the “Fail” or “Needs Improvement” scoring levels. This pattern was replicated in the collapsed (2x2) analyses.

Table 7:4 Summary of Chi-Square Findings for Boston Renaissance, Grade 8

	1998	1999
English		
BRCS vs. District	ns/ns	ns/ns
BRCS vs. State	sig/sig	sig/sig
Mathematics		
BRCS vs. District	ns/ns	ns/ns
BRCS vs. State	sig/sig	sig/sig
Science & Technology		
BRCS vs. District	ns/ns	ns/ns
BRCS vs. State	sig/sig	sig/sig

Note: Each result cell in the matrix is divided with the results for the 2x4 analysis on the left-hand side, and the results for 2x2 analysis on the right-hand side.

7.5 Odds Ratio Analysis of the MCAS Data

One of the many possible statistics that can be derived from a 2x2 contingency table is the odds ratio statistic and corresponding 1- α confidence interval. As presented in Section 2.4 of this report, the 2x2 tables analyzed in the previous section can be thought of as representing consecutive class cohorts in a prospective design. From a classical epidemiological perspective, the students in the Edison school can be thought of as the “exposed” group, that is, exposed to the “Edison-effect,” and students in the comparison group as the unexposed group. From this perspective, each yearly comparison is a new cohort, measured over a period of years. There is a minimal possibility for cohort contamination if a number of students in one group are not promoted to the next grade level. However, we think this represents a very small number of possible cases and therefore has minimal impact on the validity of these analyses. Section 2.4 details the OR statistic and corresponding 1- α confidence interval. We calculated and charted OR for each of the 2x2 tables constructed from the chi-square analyses presented above. Exhibits 7:7 and 7:8 present the odds ratio findings.

Grade 4 findings

Comparison against the district. Three OR analyses were evaluated, one for each subject test on the MCAS. In Grade 4 English, the OR for a BRCS student failing the 1998 English MCAS relative to the students in the district as a whole was 2.257, and in 1999 it was 1.883. The Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the two years was not statistically significant, indicating there is no real (statistically significant) change in OR. Thus, a common

OR=2.034 with a lower bound (LB)=0.833 and an upper bound (UB)=4.966 indicates that the BRCS students were at no greater odds for failing the English MCAS than students in the rest of the district.

The OR analysis of the grade 4 mathematics component of the MCAS presents a slightly different picture. The Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the two years was not statistically significant, indicating there is no real (statistically significant) change in the OR over the two years with a common OR=2.056 and the LB=1.166 and UB=3.634. Thus, the students at BRCS do evidence an increase in odds of failing the mathematics subtest of the MCAS relative to students in the rest of the district.

Regarding the MCAS subtest of science and technology, the Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the two years was not statistically significant, indicating there is no real (statistically significant) change in OR. Thus, a common OR=1.264 with LB=0.846 and UB=1.889 indicates that the BRCS students were at no greater odds for failing the science and technology MCAS than students in the rest of the district.

Comparison against state. In grade 4 English, the OR for a BRCS student failing the 1998 English MCAS relative to students in the state as a whole was 13.108, and in 1999 it was 9.316. In both years the $1-\alpha$ CI did not eclipse 1 and thus can be interpreted as statistically significant. The Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the two years was not statistically significant, indicating there is no real (statistically significant) change in OR. Thus, a common OR=10.848 with LB=4.467 and UB=26.344 indicates that the BRCS students were about 10 times more likely to fail the MCAS English subtest as compared with students in the state as a whole.

The OR analysis of the grade 4 mathematics component of the MCAS presents a similar picture. The Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the two years was not statistically significant, indicating there is no real (statistically significant) change in the OR over the two years with a common O=8.448 and the LB=4.816 and UB=14.821. Thus, the students at BRCS do evidence an increase in odds of failing (about $8\frac{1}{2}$ times) the mathematics subtest of the MCAS relative to students in the rest of the state. Although the Breslow-Day statistic failed to identify a statistically significant change in the OR in the two years, there was a substantial decrease in OR in 1999, dropping by more than half.

Regarding the MCAS subtest of science and technology, the Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the two years was not statistically significant, indicating there is no real (statistically significant) change in OR. Thus, the common OR=7.207 with LB=4.839 and UB=10.734 indicates that the BRCS students had substantially greater odds for failing (about 7 times) than students in the rest of the state.

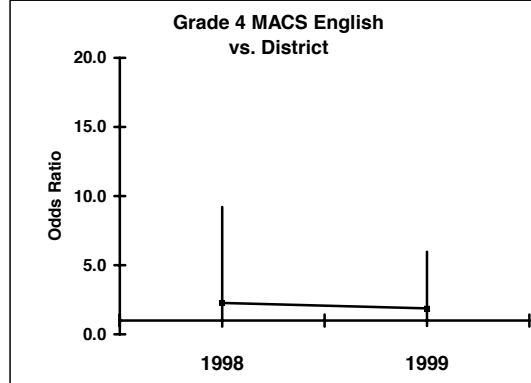
Exhibit 7:7 Results of the Odds Ratio Analysis for Boston Renaissance (Grade 4)

Grade 4 MCAS English

Year	U CI	L CI	OR
1998	9.219	0.553	2.257
1999	5.980	0.593	1.883

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=9,004) = 0.038, p = .845

OR	2.034
LB	0.833
UB	4.966

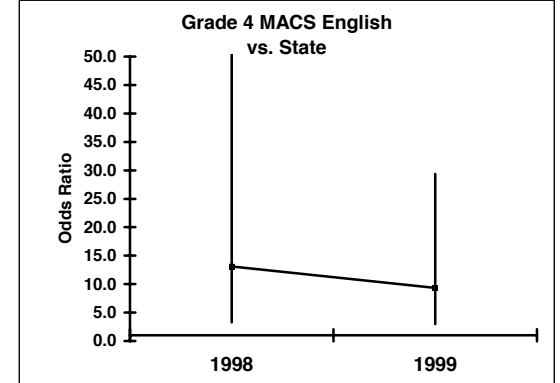


Grade 4 MCAS English

Year	U CI	L CI	OR
1998	53.105	3.236	13.108
1999	29.364	2.957	9.319

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=151,334) = 0.138, p = .710

OR	10.848
LB	4.467
UB	26.344

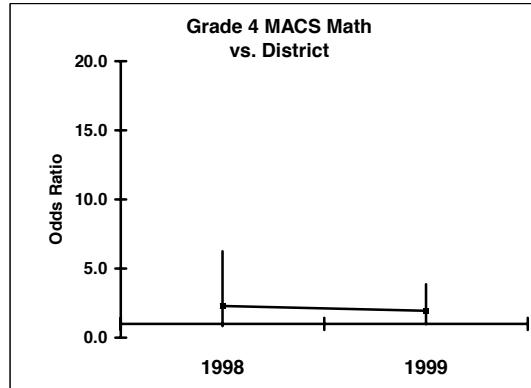


Grade 4 MCAS Math

Year	U CI	L CI	OR
1998	6.260	0.840	2.293
1999	3.861	0.977	1.943

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=9,517) = 0.071, p = .789

OR	2.056
LB	1.166
UB	3.634

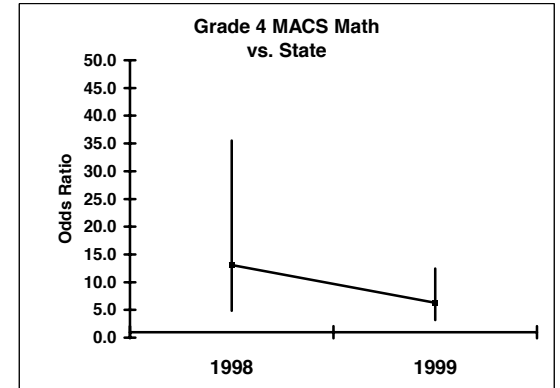


Grade 4 MCAS Math

Year	U CI	L CI	OR
1998	35.550	4.821	13.092
1999	12.460	3.182	6.300

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=152,225) = 1.468, p = .226

OR	8.448
LB	4.816
UB	14.821

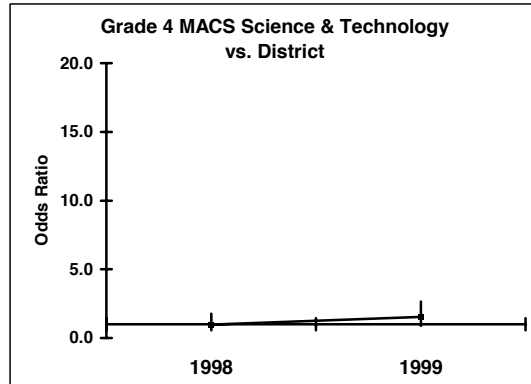


Grade 4 MCAS Science & Technology

Year	U CI	L CI	OR
1998	1.773	0.548	0.986
1999	2.652	0.881	1.529

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=9,502) = 1.152, p = .283

OR	1.264
LB	0.846
UB	1.889



Grade 4 MCAS Science & Technology

Year	U CI	L CI	OR
1998	11.947	3.744	6.688
1999	13.365	4.481	7.739

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=152,235) = 0.129, p = .719

OR	7.207
LB	4.839
UB	10.734

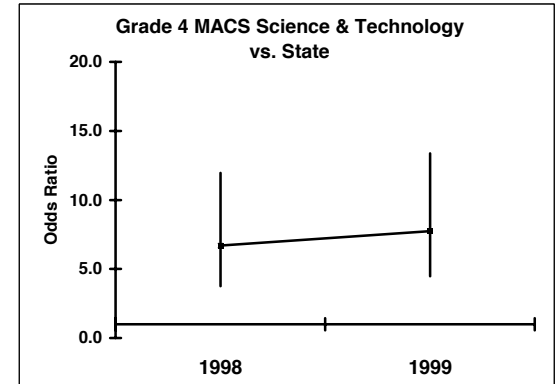


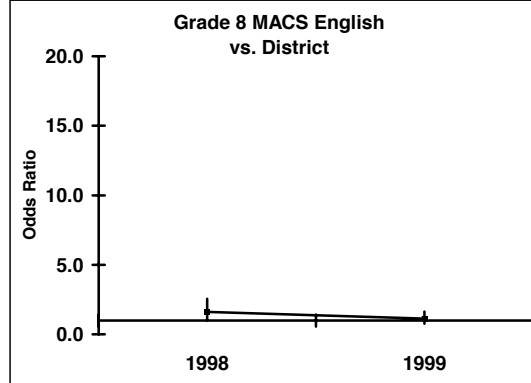
Exhibit 7:8 Results of the Odds Ratio Analysis for Boston Renaissance (Grade 8)

Grade 8 MCAS English

Year	U CI	L CI	OR
1998	2.550	1.023	1.616
1999	1.626	0.768	1.117

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=7,981) = 1.502, p = .220

OR	1.307
LB	0.98
UB	1.745

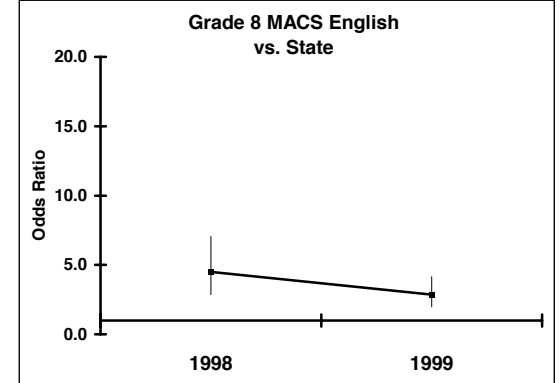


Grade 8 MCAS English

Year	U CI	L CI	OR
1998	7.056	2.859	4.492
1999	4.145	1.979	2.864

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=138,019) = 2.302, p = .129

OR	3.479
LB	2.617
UB	4.624

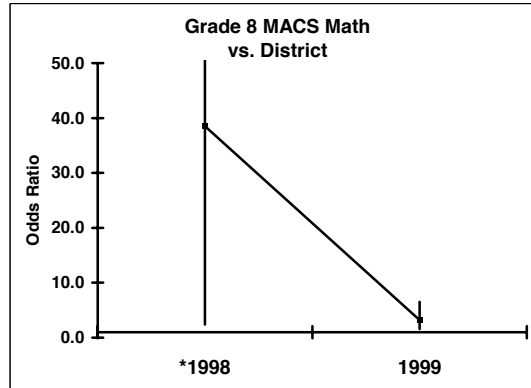


Grade 8 MCAS Math

Year	U CI	L CI	OR
*1998	620.787	2.393	38.541
1999	6.547	1.552	3.188

* Logit OR
Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=8,349) = 5.911, p = .015

OR	5.672
LB	2.787
UB	11.544

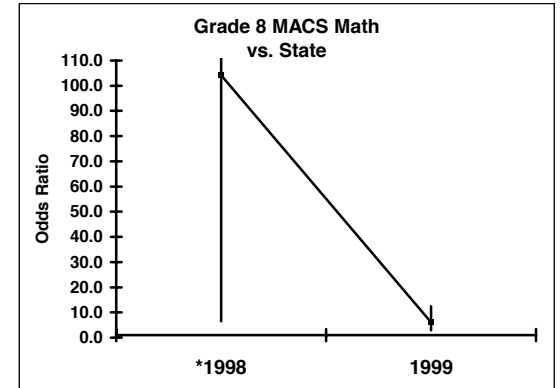


Grade 8 MCAS Math

Year	U CI	L CI	OR
*1998	#####	6.479	#####
1999	12.424	2.971	6.076

* Logit OR
Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=138,847) = 8.106, p = .004

OR	12.295
LB	6.1
UB	24.78

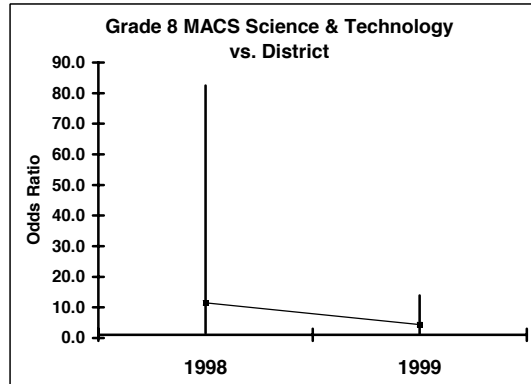


Grade 8 MCAS Science & Technology

Year	U CI	L CI	OR
1998	82.521	1.600	11.490
1999	13.948	1.400	4.418

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=8,279) = 0.724, p = .395

OR	6.193
LB	2.298
UB	16.687

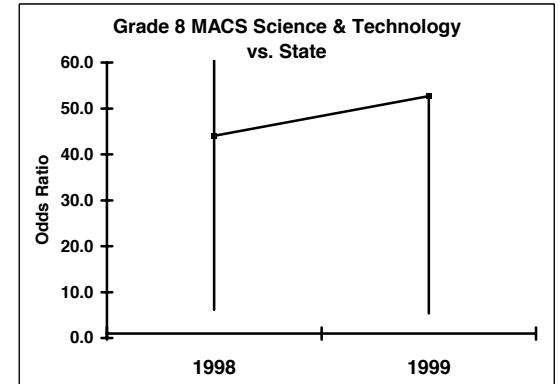


Grade 8 MCAS Science & Technology

Year	U CI	L CI	OR
1998	315.408	6.150	44.044
1999	38.771	5.335	52.664

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=140,307) = 0.744, p = .388

OR	23.584
LB	8.78
UB	63.352



Grade 8 findings

Comparison against the district. In grade 8, the OR for BRCS students shows a decreasing trend for odds of failure on the English component of the MCAS relative to students in the district. The ORs in 1998 indicated a statistically significant increase in odds of failure, but decreased to even odds in 1999. The Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the two years is not statistically significant, indicating a common OR over the two years is meaningful, $OR=1.307$, $LB=0.980$, $UB=1.745$. For the two-year period, the BRCS students were at even odds for failure relative to students in the rest of the district.

The OR for grade 8 mathematics suffered from insufficient data in 1998. No students from BRCS passed this test in 1998. The data presented in Exhibit 7:8 present the Logit OR. The Logit OR fills in a small number (0.5) in a cell with zero counts (no people) so the OR and CI can be estimated.

Regarding the Grade 8 MCAS subtest of science and technology, the Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the two years was not statistically significant, indicating there is no real (statistically significant) change in OR. Thus, a common $OR=6.193$ with $LB=2.298$ and $UB=16.687$ indicates that the BRCS students were at substantially greater odds for failing the science and technology MCAS (about 6 times) relative to students in the rest of the district.

Comparison against state. In grade 8, the ORs for BRCS students shows increased odds for failure, but a decreasing trend in these odds on the English component of the MCAS relative to students in state. Unfortunately for the BRCS students, the Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the two years is not statistically significant, indicating a common OR over the two years is meaningful, $OR=3.479$, $LB=2.617$, $UB=4.624$. For the two- year period, the odds were that a BRCS student was about 3 1/2 times more likely to fail this subtest than students in the rest of the state.

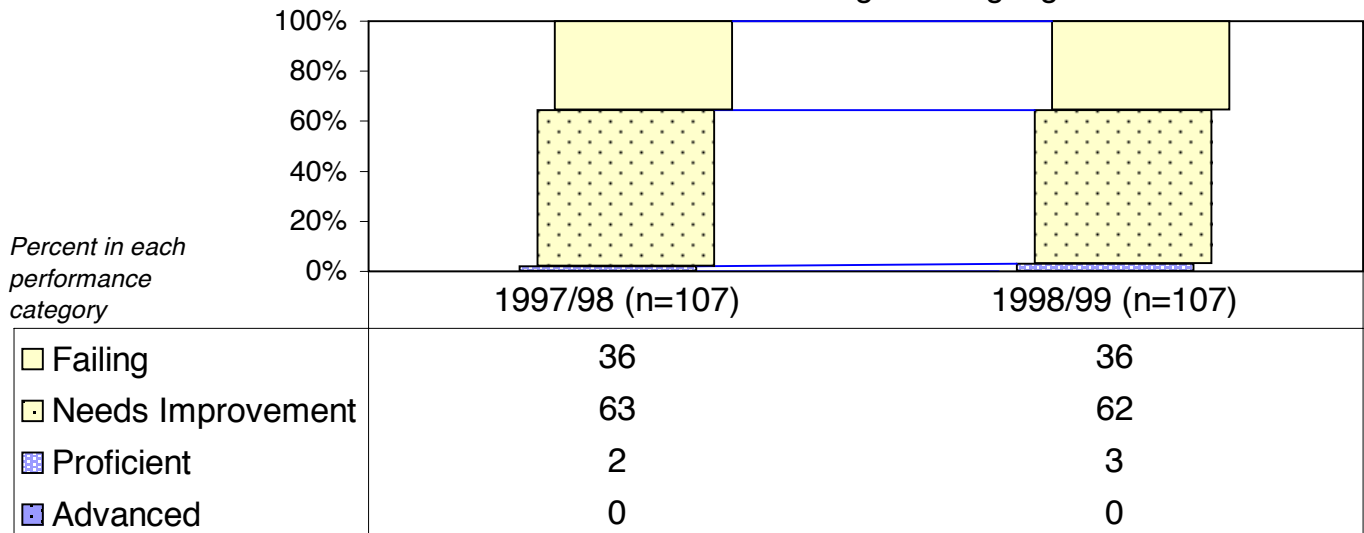
Just as the case for the district, the OR for grade 8 mathematics suffered from insufficient data in 1998. No students from BRCS passed this test that year. The data presented in Exhibit 7:8 present the Logit OR. The Logit OR fills in a small number (0.5) in a cell with zero counts (no people) so the OR and CI can be estimated.

Regarding the grade 8 MCAS subtest of science and technology, the Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the two years was not statistically significant, indicating there is no real (statistically significant) change in OR. Thus, a common $OR=23.584$ with $LB=8.780$ and $UB=63.352$ indicates that the BRCS students were at substantially greater odds for failing the science and technology MCAS (about 23 times) relative to students in the rest of the state.

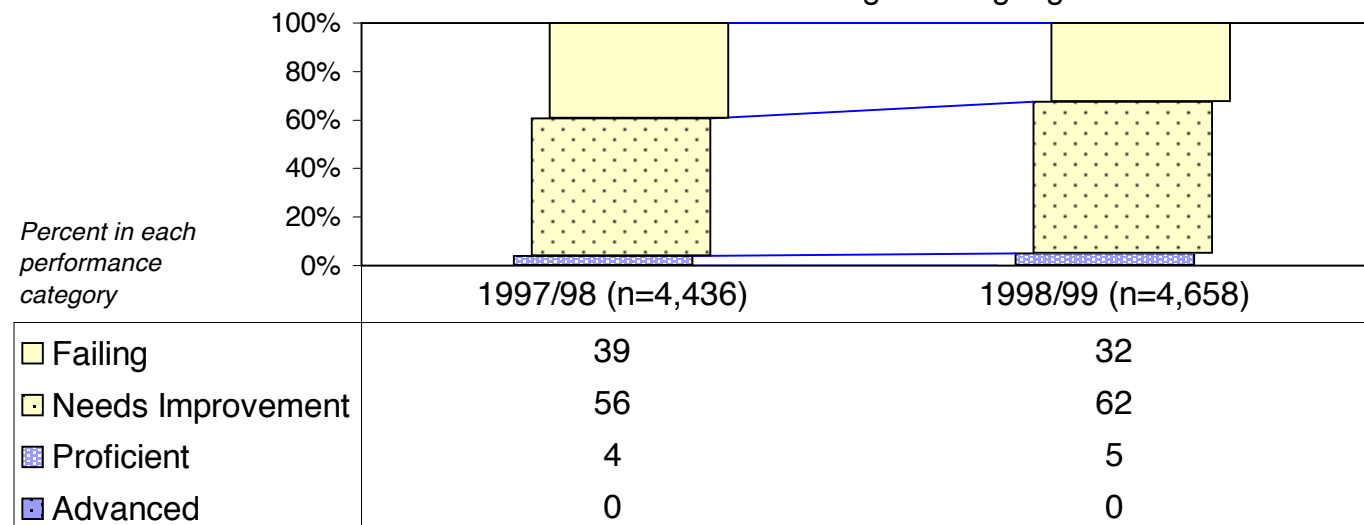
The following pages contain Exhibits 7:9 – 7:15, which illustrate the overall performance on the MCAS for Boston Renaissance, the district, and the state. These charts illustrate relative changes in all performance categories, while the odds ratio analysis focused on only the proportion of students meeting or not meeting state standards.

Exhibit 7:9 Performance on Grade 4 English Language for Edison, District, and State

Boston Renaissance MCAS Results for 4th Grade English Language Arts



Boston Public Schools MCAS Results for 4th Grade English Language Arts



State of Massachusetts MCAS Results for 4th Grade English Language Arts

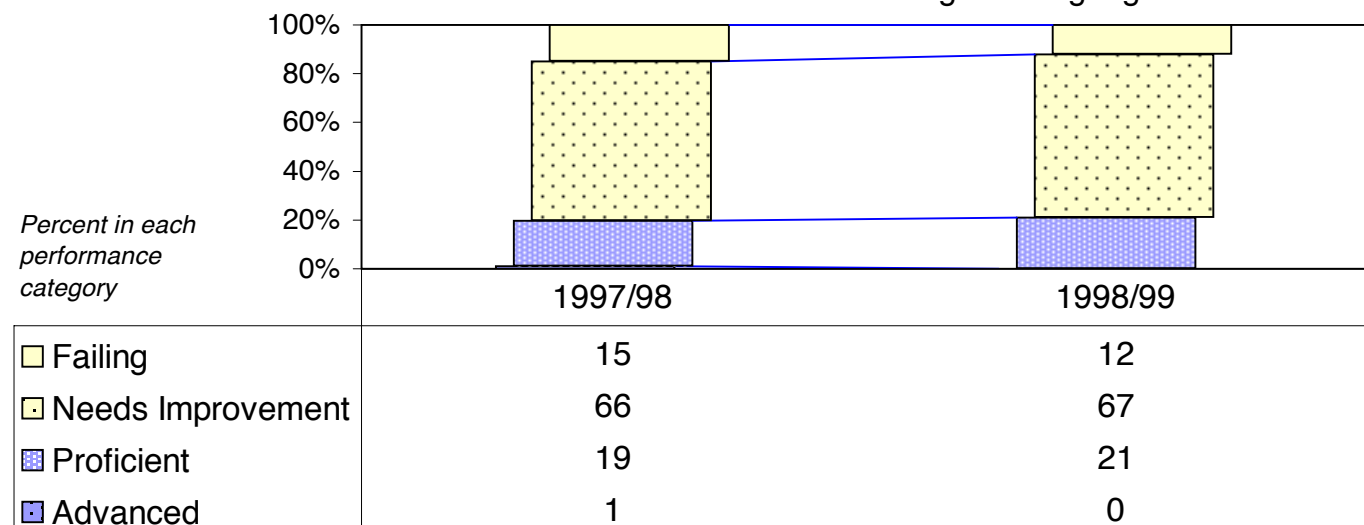


Exhibit 7:10 Performance on Grade 4 Math for Edison, District, and State

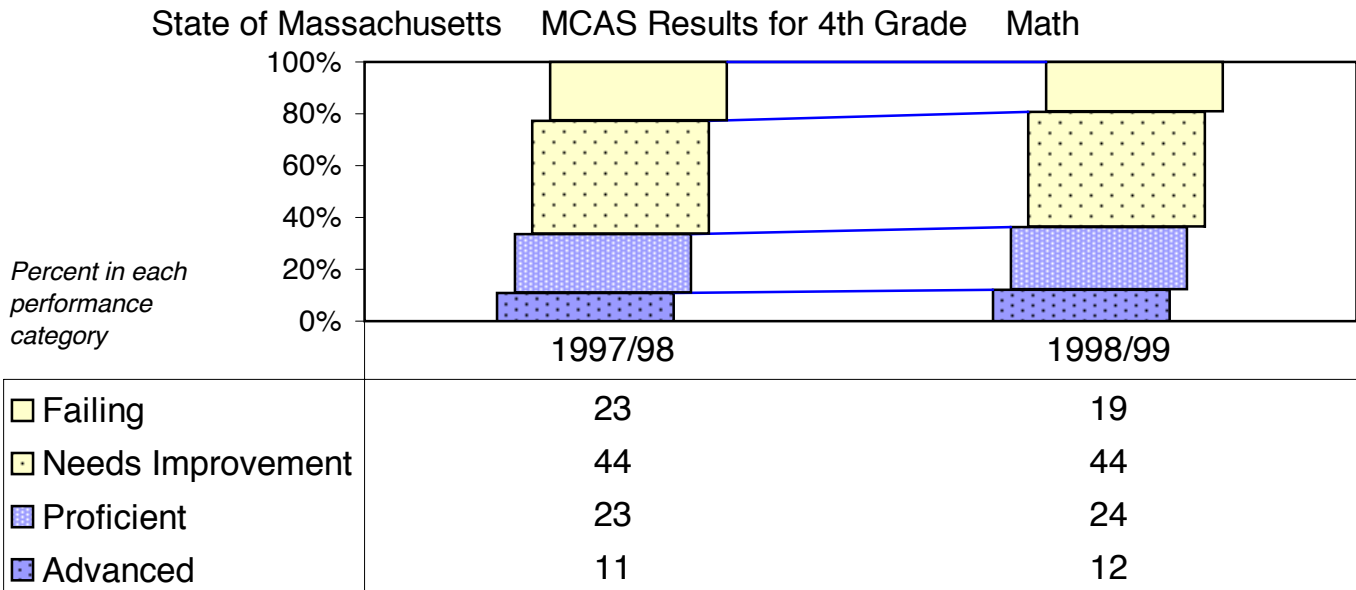
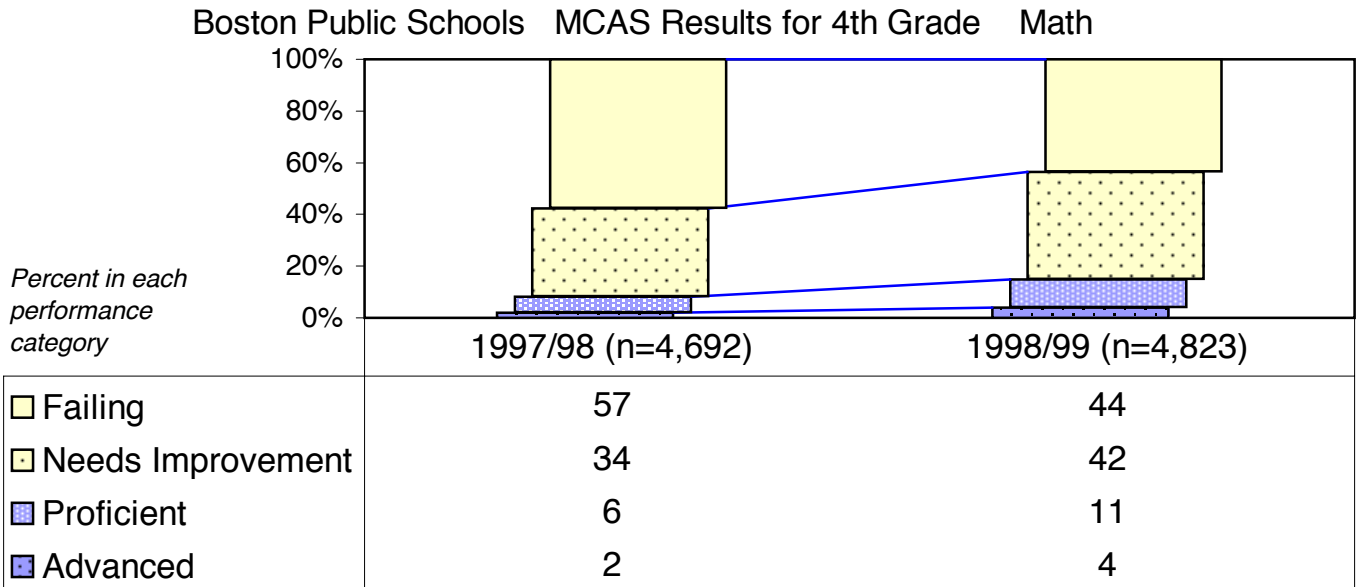
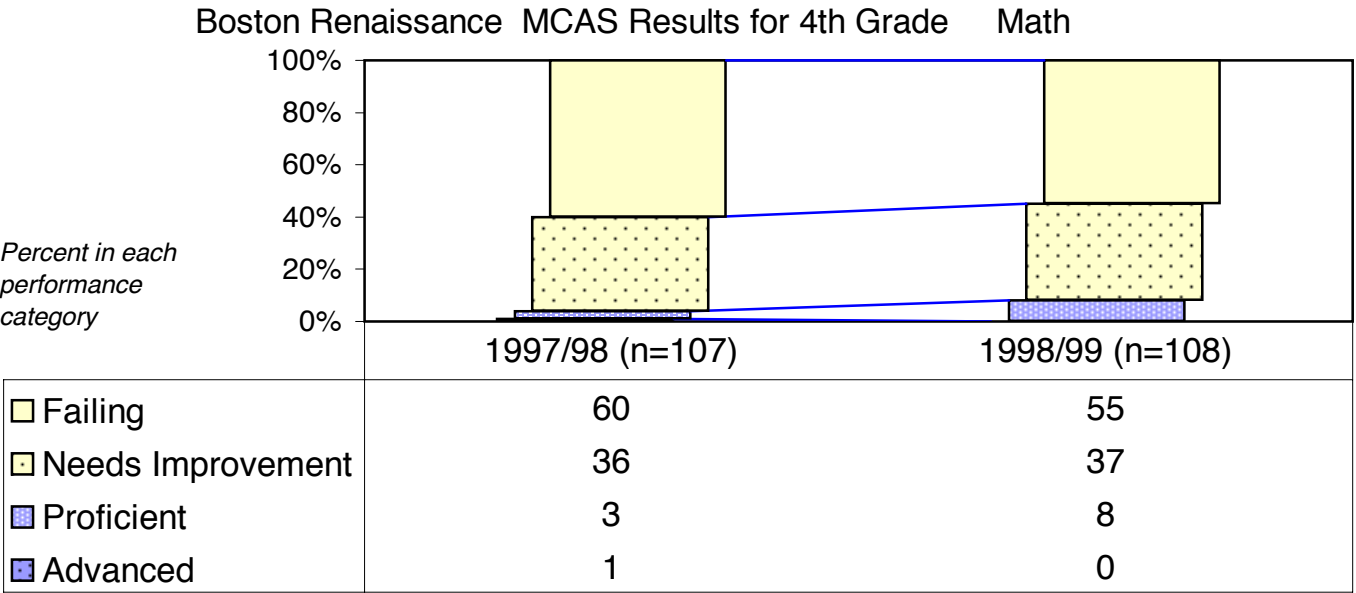
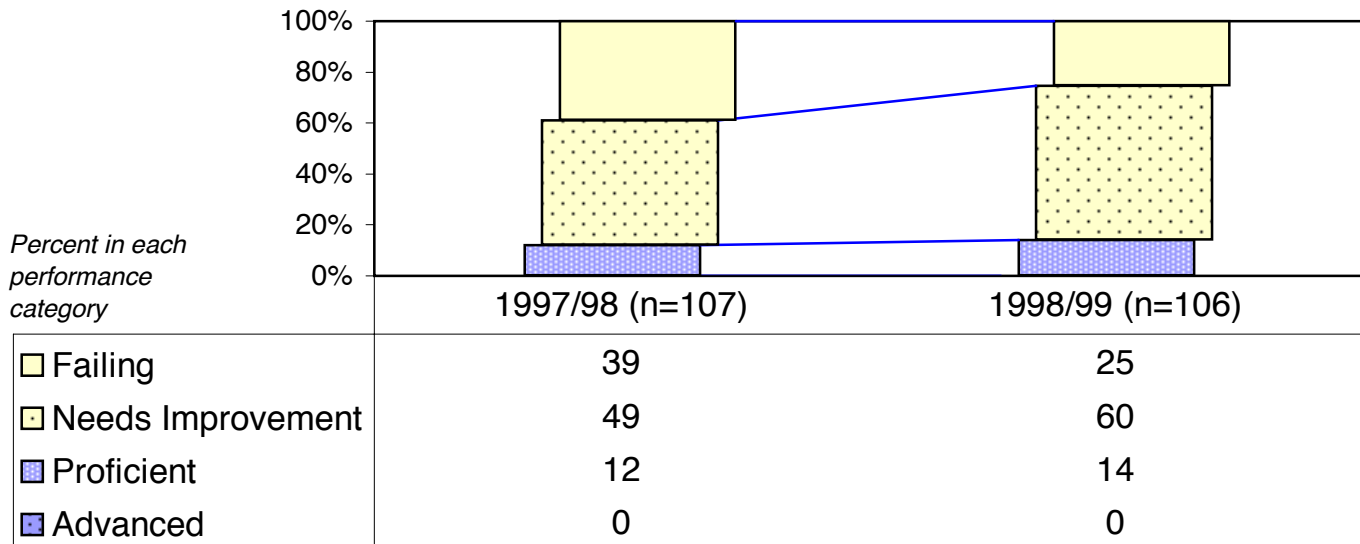
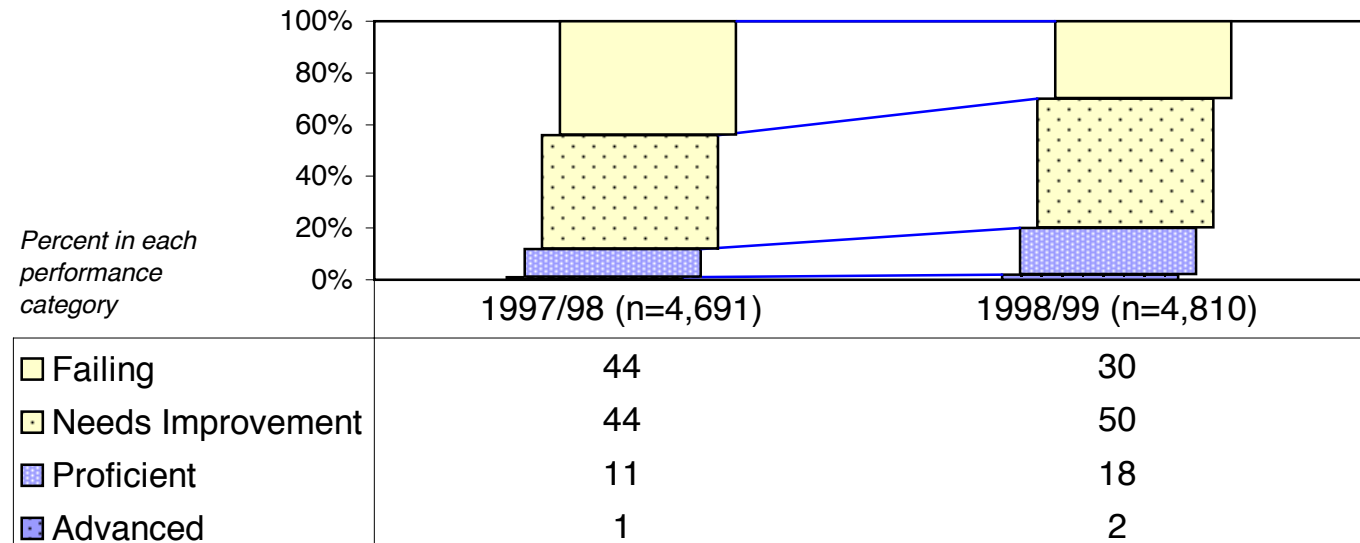


Exhibit 7:11 Performance on Grade 4 Science & Tech. for Edison, District, and State

Boston Renaissance MCAS Results for 4th Grade Science & Tech.



Boston Public Schools MCAS Results for 4th Grade Science & Tech.



State of Massachusetts MCAS Results for 4th Grade Science & Tech.

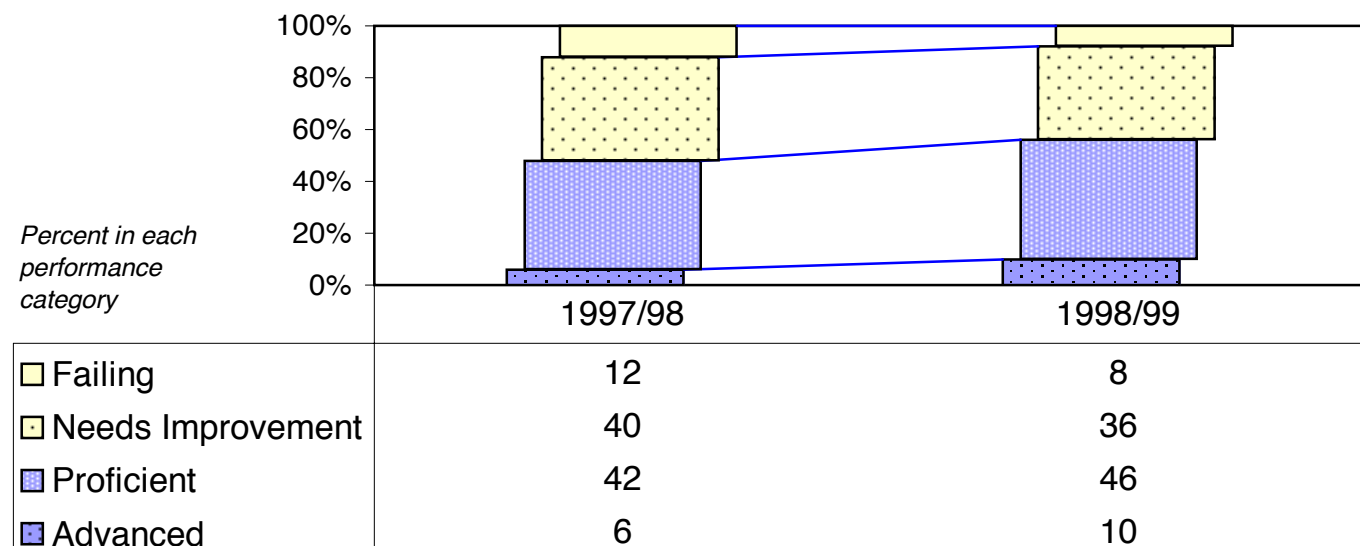
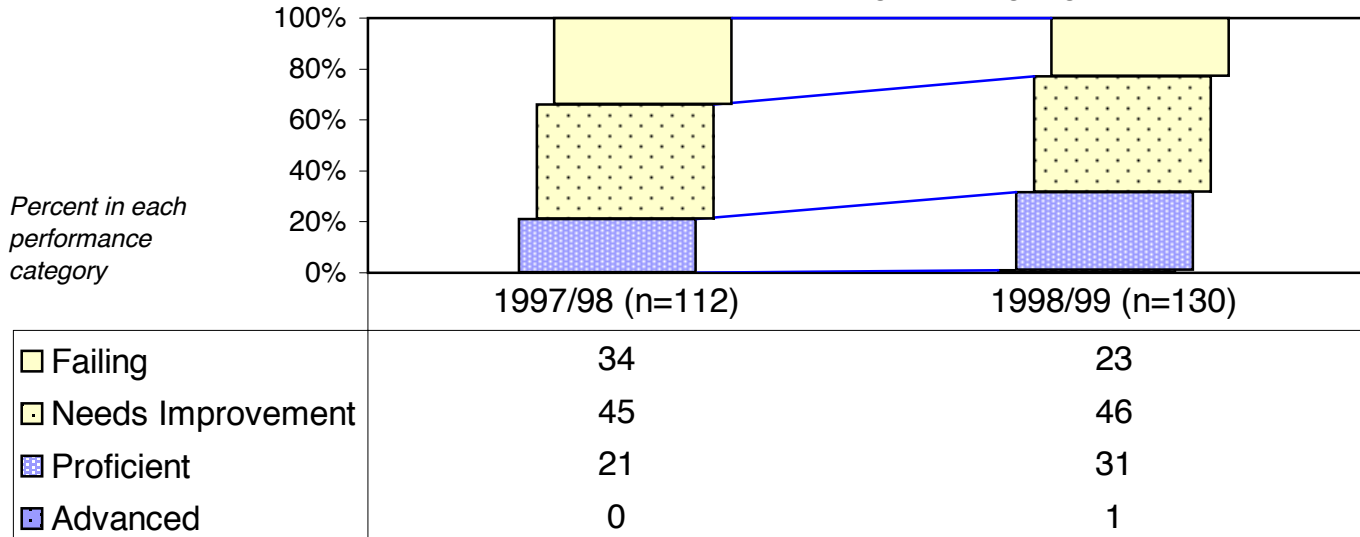
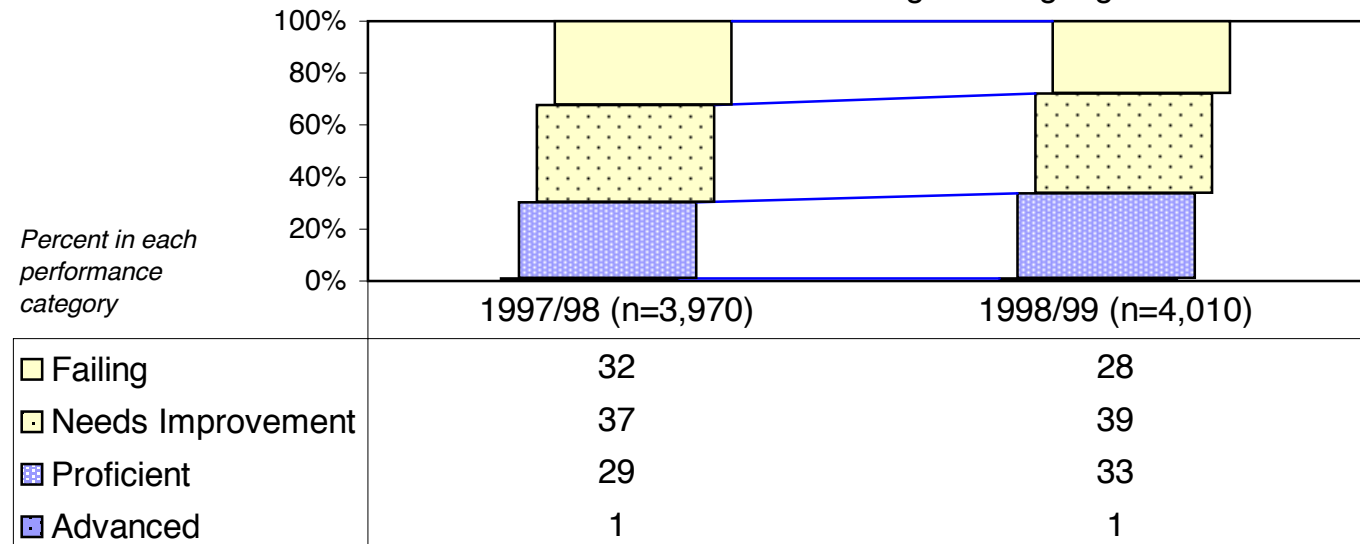


Exhibit 7:12 Performance on Grade 8 English Language for Edison, District, and State

Boston Renaissance MCAS Results for 8th Grade English Language Arts



Boston Public Schools MCAS Results for 8th Grade English Language Arts



State of Massachusetts MCAS Results for 8th Grade English Language Arts

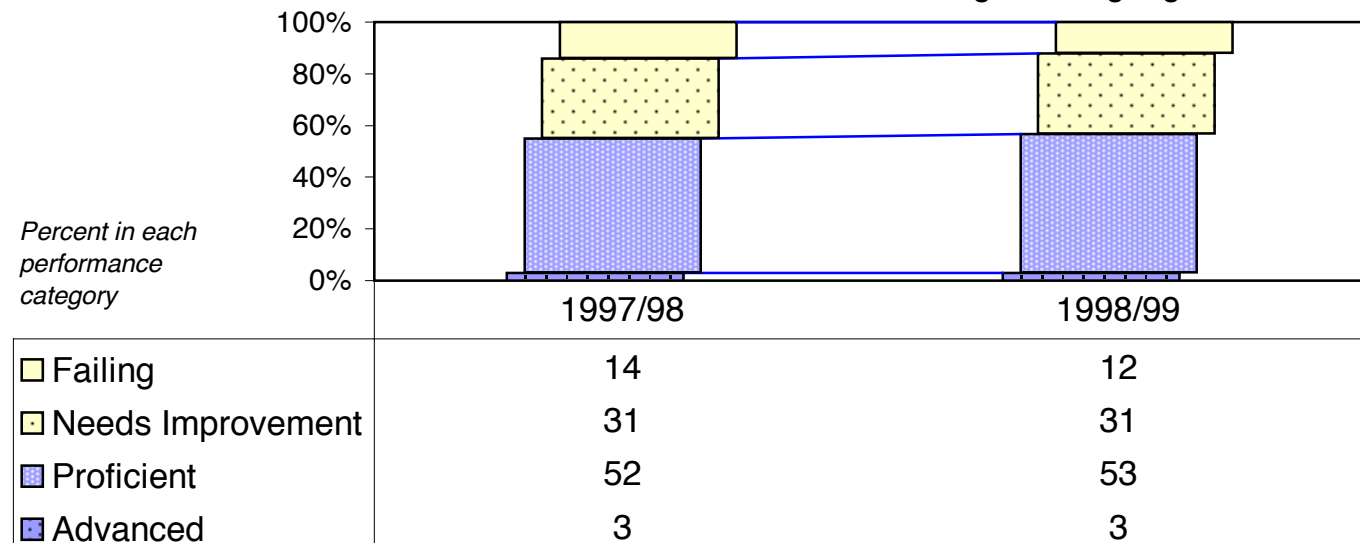


Exhibit 7:13 Performance on Grade 8 Math Language for Edison, District, and State

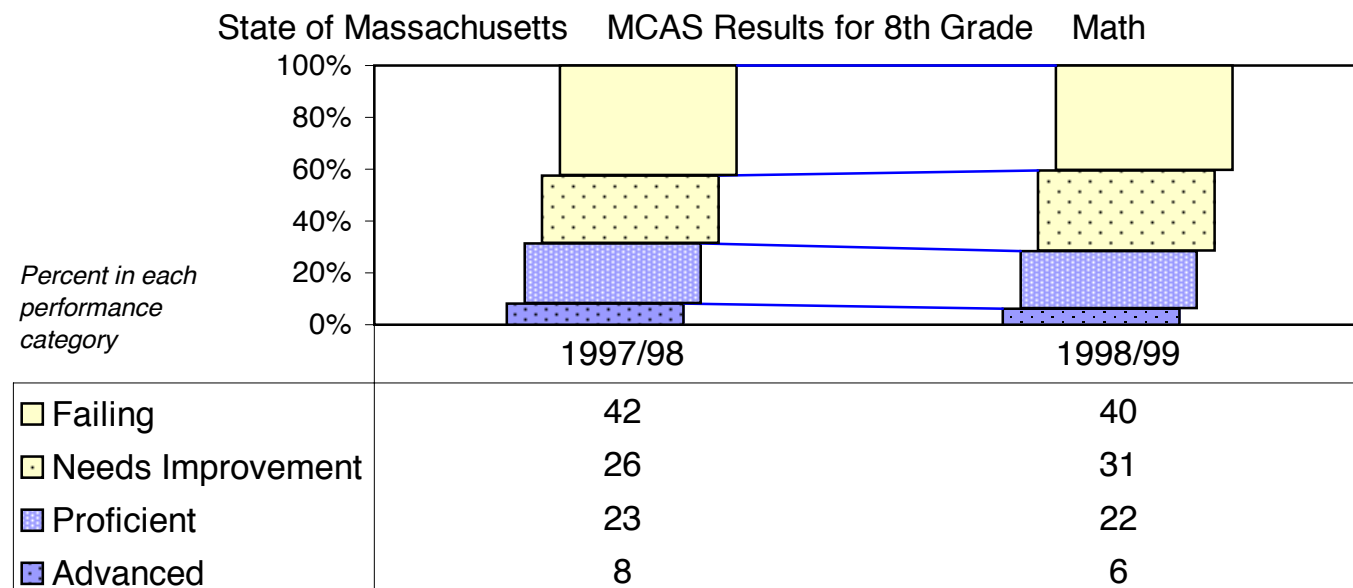
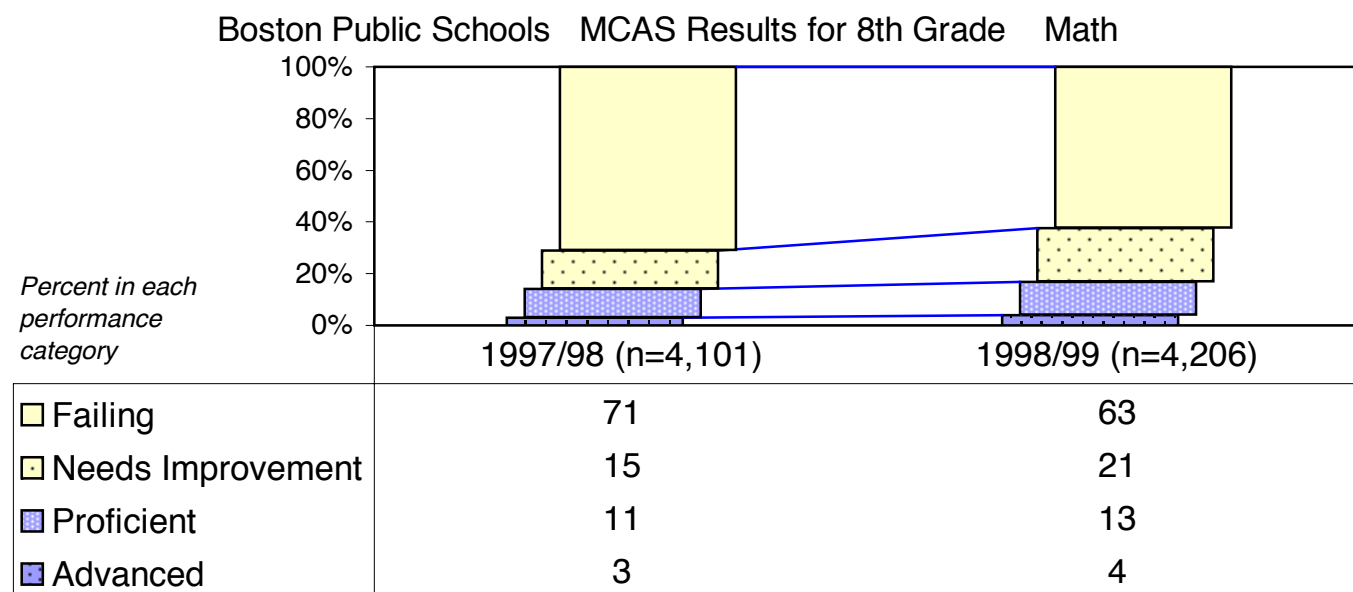
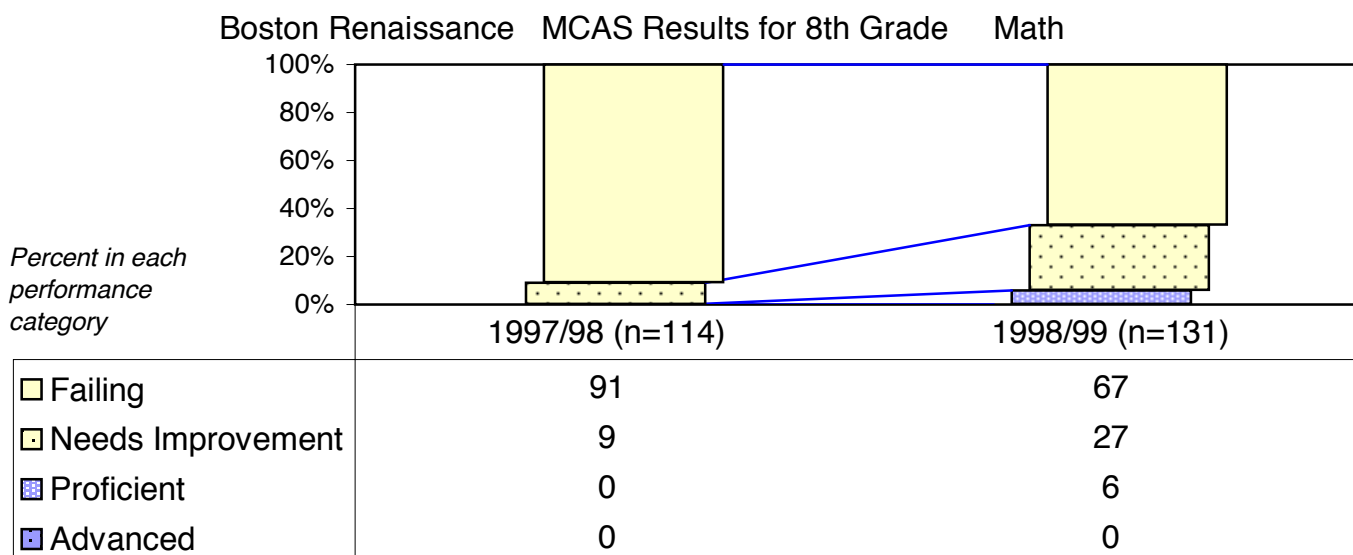
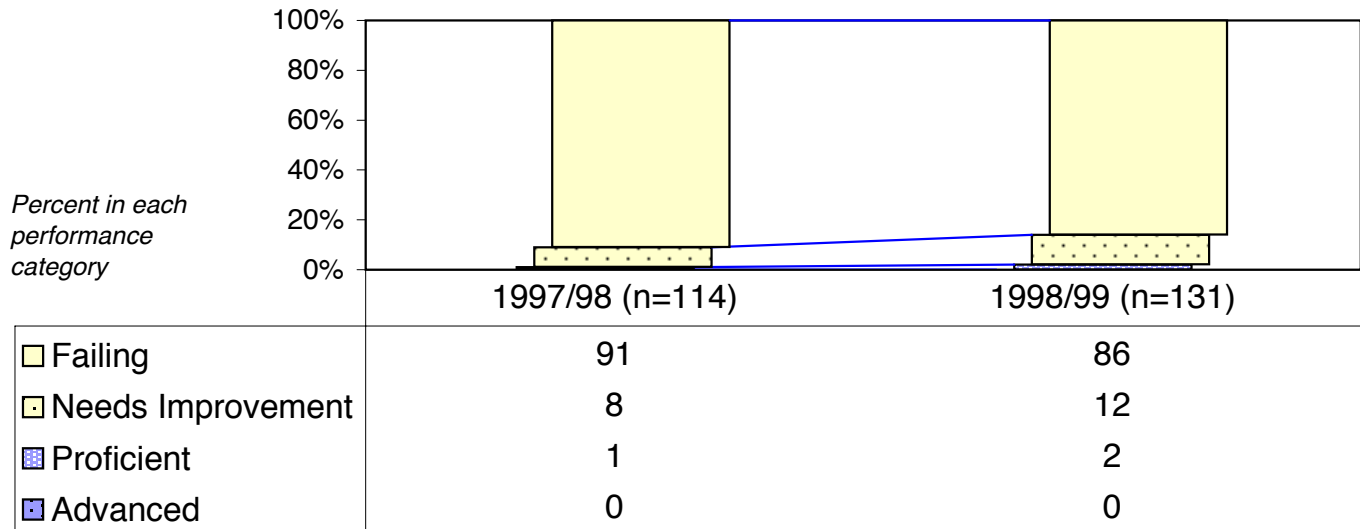
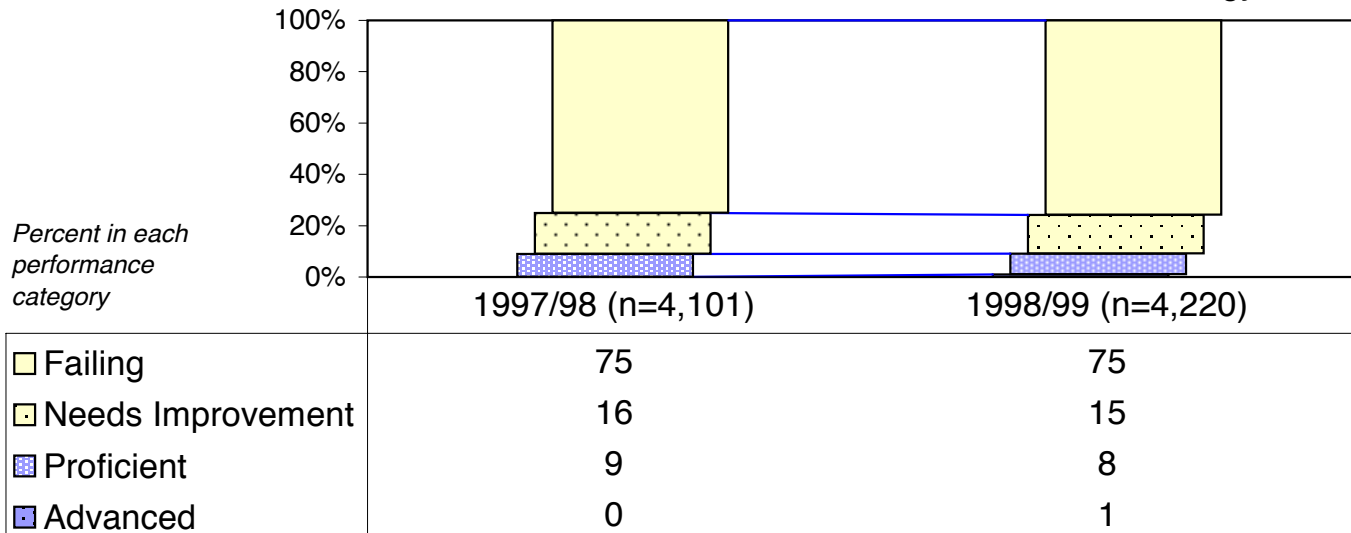


Exhibit 7:14 Performance on Grade 8 Science & Tech for Edison, District, and State

Boston Renaissance MCAS Results for 8th Grade Science & Tech



Boston Public Schools MCAS Results for 8th Grade Science & Technology



State of Massachusetts MCAS Results for 8th Grade Science & Technology

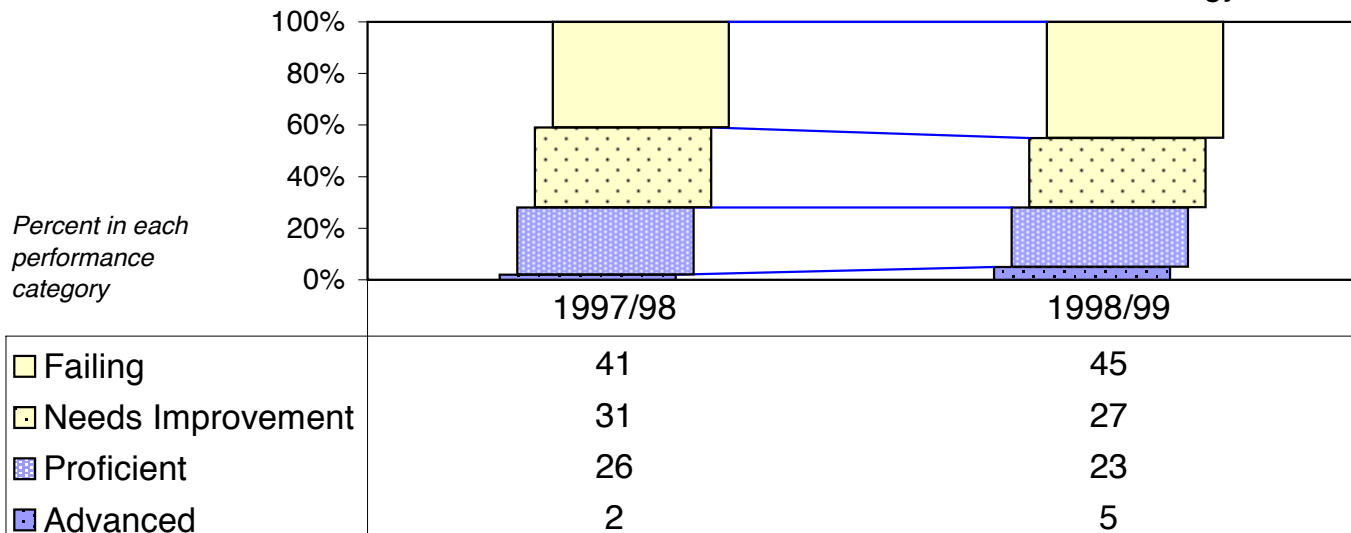
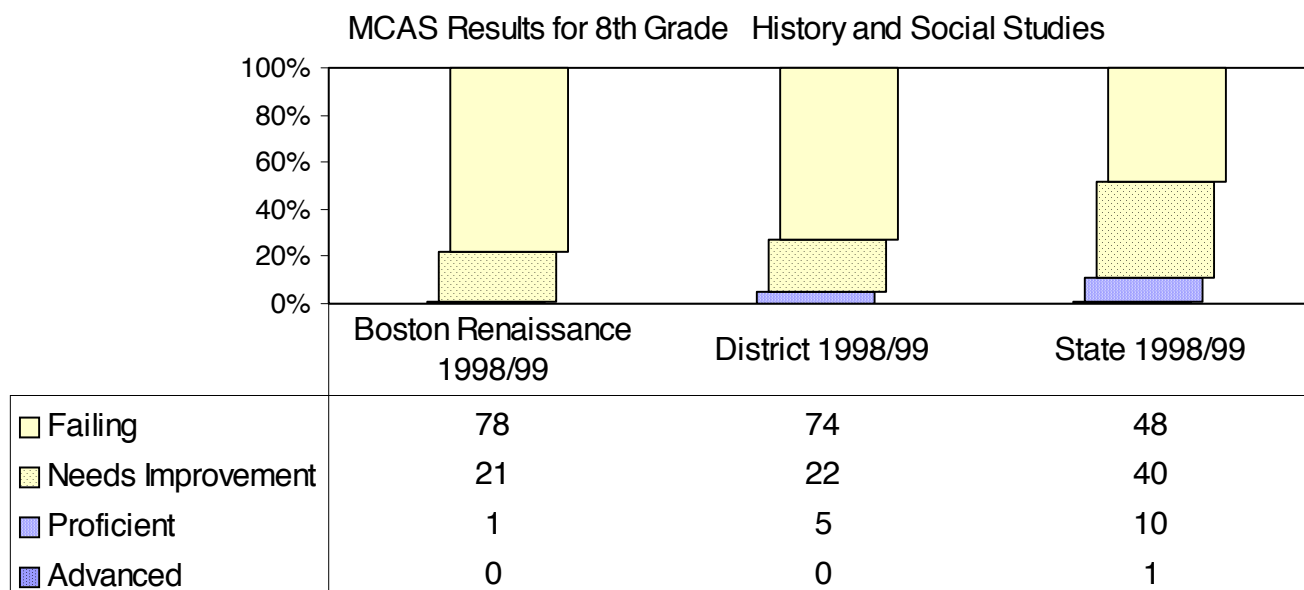


Exhibit 7:15 Performance on Grade 8 History & Social Studies for Edison, District, and the State



7.6 Summary

While the students, on the whole, made gains on the norm-referenced test (see Table 7:5), on the whole, they did not gain as much as the district and state on the criterion-referenced test mandated by the state (see Table 7:6).

Norm-referenced test findings

A summary score of -1 indicates a result that is unfavorable toward the sample school, a score of 0 indicates a neutral finding, and a score of 1 indicates a favorable result according to the criteria specified in Section 2.5. The effect size (ES) is the omega squared (ω^2) for a one-way repeated measures ANOVA.

Table 7:5 Summary of Results on Norm-Referenced Student Achievement Tests

Cohort A SAT-9, Grades 3, 4, 5 (1997-99)	Std Score		GE			PR			NCE			Trend
	p-value	ES	p-value	Δ	ES	p-value	Δ	ES	p-value	Δ	ES	
Math	<.0001	.449	<.0001	2.3	.370	.0059	-5.0	.044	.0075	-3.0 (0)	.040	mixed (0)
Reading	<.0001	.400	<.0001	1.8	.317	.0018	-4.0	.049	.0084	-2.0 (0)	.035	mixed (0)
Cohort B												
Grades 4, 5, 6 (1997-99)												
Math	<.0001	.562	<.0001	2.5	.414	<.0001	5.0	.102	<.0001	4.0 (0)	.119	mixed (0)
Reading	<.0001	.497	<.0001	2.0	.407	.0010	7.0	.060	.0003	5.0 (0)	.066	mixed (0)

Cohort C SAT-9 Grades 5, 6, 7 (1997-99)	Std Score		GE			PR			NCE			Trend
	p-value	ES	p-value	Δ	ES	p-value	Δ	ES	p-value	Δ	ES	
Math	<.0001	.361	<.0001	1.8	.296	.0506	1.0	.022	.0103	1.0 (0)	.040	mixed (0)
Reading	<.0001	.314	<.0001	1.6	.262	.5479	-2.0	-.005	.7594	-1.0 (0)	-.009	mixed (0)
Cohort D Grades 6, 7, 8 (1997-99)												
Math	<.0001	.406	<.0001	2.1	.399	<.0001	5	.142	<.0001	3 (0)	.120	mixed (0)
Reading	<.0001	.505	<.0001	2.7	.451	<.0001	4	.108	<.0001	2 (0)	.097	mixed (0)
Cohort E Grades 7, 8, 9 (1997-99)												
Math	<.0001	.230	.0002	1.1	.198	.0898	0.0	.035	.0723	1 (0)	.036	mixed (0)
Reading	<.0001	.454	<.0001	2.2	.412	.0022	7	.104	.0067	4 (0)	.082	mixed (0)
Cohort F Grades 5, 6, 7, 8 (1997-99)												
Math	<.0001	.408	<.0001	2.2	.421	<.0001	-2	.100	.0006	-1 (0)	.082	mixed (0)
Reading	<.0001	.481	<.0001	2.7	.439	.0118	0	.049	.0202	-1 (0)	.042	mixed (0)

Criterion-referenced test findings

Decisions regarding the OR were based on whether or not the $(1-\alpha)$ C.I. included 1.0. If the $(1-\alpha)$ C.I. completely fell below 1.0, this was interpreted as a protective odds ratio (1) and thus favoring the Edison school. If the $(1-\alpha)$ C.I. included 1.0 (0), this was interpreted as an equal odds situation. If the $(1-\alpha)$ C.I. fell completely above 1.0 (-1), this was interpreted as an increase in odds for failing the state CRT relative to the comparison sample.

Table 7:6 Summary Results on Criterion-Referenced Tests

Edison vs. District	1998	1999	B-D	Trend
Grade 4 English			0	mixed (0)
Grade 4 Math			-1	negative (-1)
Grade 4 Science/Technology			0	mixed (0)
Grade 8 English			0	mixed (0)
Grade 8 Math	-1	-1		negative (-1)
Grade 8 Science/Technology			-1	negative (-1)

Note: All comparisons with the state as a comparison group were negative

Interpretation of the OR tables: If the Breslow-Day statistic (B-D) is not significant, one overall OR and $(1-\alpha)$ C.I can be used to represent the odds for failing the CRT relative to the comparison group. Thus, there are no trends reported for each specific year in the tables, only a rating in the B-D column. If the B-D statistic is found to be statistically significant, then an overall common OR cannot be meaningfully interpreted; that is, there is a statistically significant change in the OR over years and yearly OR are necessary. Our summary ratings appear for each year of data and not in the B-D column.

Combined ratings

Given the total ratings for the trends that are highlighted in Table 7:7, we rate this school as Mixed with a mean rating of -0.167. In its 1999 annual report, Edison rated this school as Mixed. In its 2000 annual report, it rated the 1999-2000 school year as Strongly Positive and the achievement gains since opening as Positive.

Table 7:7 Combined Overall Trends for Boston Renaissance Charter School

	Positive	Mixed	Negative
Norm Referenced	0 of 12	12 of 12	0 of 12
Criterion Referenced	0 of 6	3 of 6	3 of 6
TOTALS	0 of 18	15 of 18	3 of 18

The design behind the trends in the norm-referenced results is based on tracing individual students over three or more years. While some would argue that this is a better design than was used with the criterion-referenced results (tracing consecutive cohorts of students), one also has to consider the sample sizes. The sample size for the norm-referenced test was very small. While we received a database from Edison with more than 1,100 records to perform this analysis, we were only able to trace a much smaller number of students (see Table 7:2). On the other hand, the test administration procedures are likely to be more regulated for the state-mandated MCAS; and the results we have for these trends include a much larger proportion of the total enrollment at the school.

The results to date indicate that this Edison school does not differ substantially from other district schools.

Chapter Eight

Seven Hills Charter School

Worcester, Massachusetts

8.1 Descriptive Summary of School

Seven Hills Charter School is an independent charter school established in 1996 to serve grades K-8. The school opened with K-6, and in 1997 grades 7 and 8 were added. The mission of the Seven Hills

Charter School is to prepare a diverse cross section of Worcester children for success with a high quality education at prevailing public school costs.

During the 1998/99 school year, 662 students were enrolled in the school. Of these, 19.3 percent were African American, 0.6 percent were of Asian descent, 25.1 percent were Hispanic, 1.4 percent were Native American, and 53.6 percent were white. The total proportion of minorities is similar to the district of Worcester, although the district has fewer African-American students and more students of Hispanic and Asian descent.

Edison's second annual report of student achievement (Edison, 1999), noted that there were 61 instructional staff members, with 10.9 pupils per staff member. The district's school profiles, as well as the Seven Hills' profile in the state's annual report (Massachusetts Department of Education [MDOE], 1999) on charter schools, indicated that the number of instructional staff was 36, with an average of 19.8 pupils per staff member. The latter figure is slightly higher than district and state averages, which are 16.6 and 18.1 pupils per staff member, respectively.

According to data obtained from the school district profiles (MDOE, 1999), Seven Hills Charter School has a slightly lower proportion of students with special educational needs (15.5 percent in 1998/99) than the local school district (17.8 percent) and the state average (16.6 percent). Edison (1999) reports that student mobility in 1997-98 was 6.7 percent. The school has slightly fewer students qualifying for free or reduced lunch (46.4 percent) than the district (50.2 percent) (MDOE, 1999), but more than the state as a whole (25.8 percent).

In terms of students with limited English proficiency, it was difficult to determine a clear figure for Seven Hills Charter School. In the 1997/98 school year, the school had 11.5 percent of its students with limited English proficiency (LEP) while in the following year, the school reportedly had 0.5 percent (MDOE, 1999). Meanwhile, the district reported that 7 percent of its students are LEP whereas the state reported 4.7 percent (MDOE, 1999). Edison (1999) reported that 10.4 percent of its students were categorized as LEP during the 1997/98 school year. Since the district has a higher proportion of Hispanic students, the proportion of Edison's students labeled as LEP most likely does not exceed the district average, although there are incomplete and conflicting figures.

The school profile in the state charter school report (MDOE, 1998) indicated that the size of the waiting list was reported to be 418 students. The report on charter schools included two figures for the total number of days of instruction (190 and 200). The estimated total hours of structured learning time was 1,239 hours per year.

The expenditures per pupil were nearly identical for Seven Hills (\$6,446) and the district (\$6,433) in 1997/98 (MDOE, 1999). The state average was slightly lower at \$6,361 per pupil in that same year. Likewise, average attendance rates were nearly identical between Seven Hills (94.6 percent), the district (93.5 percent), and the state average (93.9 percent) (MDOE, 1999).

From the school's descriptive data that we reviewed, we can see that the size and composition of the school have not altered much since it opened in 1996. In later sections of this chapter we will be comparing the performance of students in this school with the district and the state. Therefore, it is important to consider the differences in key background characteristics between the school, district, and the state as a whole. As an urban district, both the Edison school and the district have a higher proportion of minorities and students qualifying for free or reduced lunches than the state average.

8.2 Past Studies and Evaluations and Available Data for Analysis

Tests administered at Seven Hills include the Massachusetts Comprehensive Assessment System (MCAS) in Spring 1998 and Spring 1999 for grades 3,4,8,10; the Stanford Achievement Test Series-9th Edition (SAT-9) in Spring 1997 for grades 3-4, in Spring 1998 for grades 3-5; the Metropolitan Achievement Test-7th Edition (MAT-7) in Spring 1997 for grades 5-7; and the Iowa Test for Basic Skills (ITBS) in Spring 1997 for grade 3 and in Spring 1998 for grade 3.

A comparison of results on the ITBS, conducted by the Worcester School Department (1997) indicated that third grade students in a district school performed considerably better on the ITBS reading subtest than did students at Seven Hills. The same study compares scores on the ITBS reading subtest among charter schools in the state. The percentile ranking of Seven Hills was second from the bottom (9th out of 10) with a percentile rank of 40, compared with the highest ranking charter school with a percentile rank of 78.

Edison provided us with an SPSS data set containing SAT-9 test data for two years (1998-1999) and three years of student achievement data on the MAT-7 (1997-1999). Similar to all Massachusetts public schools, this school is also required to take part in the Massachusetts Comprehensive Assessment System (MCAS). We secured test result data for only two years, 1997-98 and 1998-99. This matches the CRT results reported by Edison in its 2000 annual school performance report.

8.3 Longitudinal Analysis of Individual Student Data

Individual student achievement data on SAT-9 and MAT-7 test results were available for our evaluation. The MAT-7 data set represented a longer time span for our longitudinal analysis (i.e., three academic years rather than two); thus, we place greater confidence in these analyses. We have summarized both the MAT-7 and SAT-9 results below.

MAT-7

The MAT-7 scores were provided in four separate scales: grade equivalent scores (GE), standard scores (SS), national percentile rank scores (PR), and normal curve equivalent scores (NCE) for language, math, and reading. Parallel analyses are reported for each subtest and scale.

Table 8:1 presents the various sample size breakdowns by subject tested, grade, and group for 324 students covering the 1997-1999 academic years. It was possible to trace 2 different cohorts. Cohort A included 44 students over 3 consecutive academic years (i.e., grade 5 in 1996/97, grade 6 in 1997/98, and grade 7 in 1998/99). Cohort B was much smaller in size and contained only 24 students across 3 academic years (grade 6 in 1996/97, grade 7 in 1997/98, and grade 8 in 1998/99). We debated whether or not we should report the results from Cohort B because of its small size, but have opted to do so in order to include—at this level of the analysis—a larger picture of the results. Nonetheless, the results from Cohort B should be interpreted more cautiously.

Similar to many of the other Edison schools in this study, there appears to be either a high rate of attrition or substantial inconsistencies in who is taking the tests at Seven Hills. In regard to Cohort A, 69 possible students could be followed in 1997. By 1998 about 87 percent remained (60/69 pairwise analysis), and by 1999 only about 61 percent of the students remained (longitudinal trend analysis). This represents a 39 percent drop in students over 3 years in the cohort of students that progressed from grade 5 to grade 7 between the 1997/98 school year and the 1998/99 school year. The pattern for Cohort B was similar, with 40 students starting in sixth grade in 1997, dropping to 28 seventh graders (70 percent) and to 24 eighth graders in 1999 (60 percent), or a 40 percent drop in the three years.

Table 8:1 Table of Sample Sizes for Individual Student Data on the MAT-7 by Grade and Year

Grade 5	1997			1998			1999		
	Language	Math	Reading	Language	Math	Reading	Language	Math	Reading
GE	69	69	69						
SS	69	69	69						
PR	69	69	69						
NCE	69	69	69						
Grade 6	1997			1998			1999		
	Language	Math	Reading	Language	Math	Reading	Language	Math	Reading
GE	40	40	41	78	78	78	55	55	55
SS	40	40	41	78	78	78	55	55	55
PR	40	40	41	78	75	77	55	55	55
NCE	40	40	41	78	75	77	55	55	55
Grade 7	1997			1998			1999		
	Language	Math	Reading	Language	Math	Reading	Language	Math	Reading
GE	46	46	46	78	78	78	81	81	81
SS	46	46	46	78	78	78	81	81	81
PR	46	46	46	74	77	74	81	81	81
NCE	46	46	46	74	77	74	81	81	81
Grade 8	1997			1998			1999		
	Language	Math	Reading	Language	Math	Reading	Language	Math	Reading
GE				51	51	51	78	78	78
SS				51	51	51	78	78	78
PR				51	50	51	78	77	78
NCE				51	50	51	78	77	78
Sample sizes for 1 year gains (grade 5 to grade 6)									
	1997/98								
	Language	Math	Reading						
GE	58	55	57						
SS	58	55	57						
PR	58	55	57						
NCE	58	55	57						
Sample sizes for 1 year gains (grade 6 to grade 7)									
	1997/98			1998/99					
	Language	Math	Reading	Language	Math	Reading			
GE	28	29	30	62	59	61			
SS	28	29	30	62	59	61			
PR	28	29	30	62	59	61			
NCE	28	29	30	62	59	61			

Sample sizes for 1 year gains (grade 7 to grade 8)						
1997/98			1998/99			
	Language	Math	Reading	Language	Math	Reading
GE	37	37	37	61	62	61
SS	37	37	37	61	62	61
PR	37	37	37	61	62	61
NCE	37	37	37	61	62	61

Cohort A Sample Sizes for Longitudinal Analysis (Grades 5 to 6 to 7)

1997/98/99			
	Language	Math	Reading
GE	44	41	43
SS	44	41	43
PR	44	41	43
NCE	44	41	43

Cohort B Sample Sizes for Longitudinal Analysis (Grades 6 to 7 to 8)

1997/98/99			
	Language	Math	Reading
GE	24	24	24
SS	24	24	24
PR	24	24	24
NCE	24	24	24

Longitudinal analysis findings

Repeated measures ANOVAs were examined for longitudinal trends over a three-year period for Seven Hills. Outcome data represented individual student data on the MAT-7 subtests: language, mathematics “concepts/problem solving,” and reading comprehension. Parallel analyses are reported for all types of scores reported by Edison, e.g., GE, SS, PR, and the NCE scores. In all models, the assumption of sphericity was evaluated, and if found to be violated, the Huynh-Feldt adjusted p-values are reported. If the overall linear model was found to be statistically significant, unadjusted (alpha) pair-wise comparisons were examined to identify where a difference in the means might be located.

The tables and charts in Exhibits 8:1 and 8:2 illustrate the results in terms of grade level equivalents, standard scores, national percentiles, and normal curve equivalents for Cohorts A and B, respectively. The results are grouped by subject area tests: language, math/problem solving, and reading comprehension.

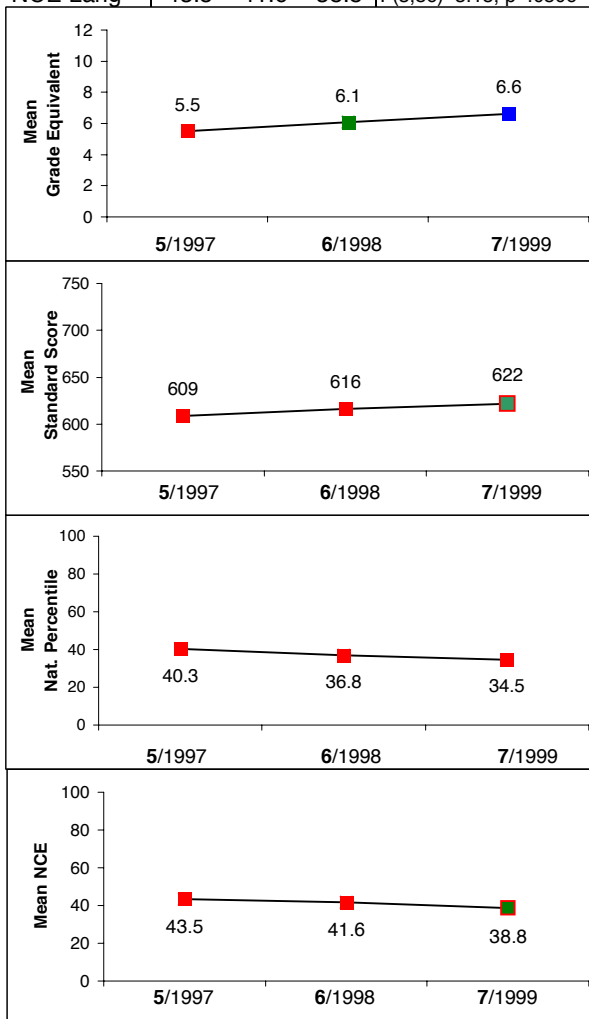
The students in Cohort A who had been enrolled at this school since it opened cannot be characterized as high performing students. On the whole they performed at a level that would be expected for students in their grade, although in comparison with national norms, the students were typically in the 35th to 43rd percentile. What is of particular importance for our analysis is not where they were at one point in time; rather, we are interested in the rate of learning or in the relative size of the gains they made each year.

Exhibit 8:1 Seven Hills Charter School, Cohort A: Analysis of Individual Student Results on MAT-7

COHORT A

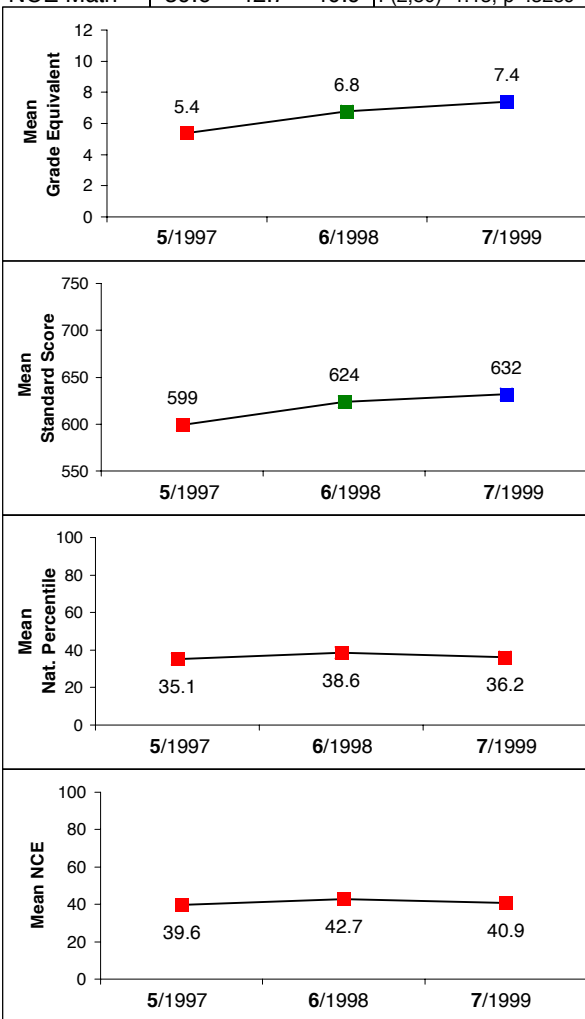
LANGUAGE

n=44	1997	1998	1999	
Grade	5th	6th	7th	
GE Lang	5.5	6.1	6.6	$F(3,86)=9.14, p=.0003$
SS Lang	609	616	622	$F(3,86)=8.31, p=.0005$
NP Lang	40.3	36.8	34.5	$F(3,86)=2.62, p=.0788$
NCE Lang	43.5	41.6	38.8	$F(3,86)=3.15, p=.0506$



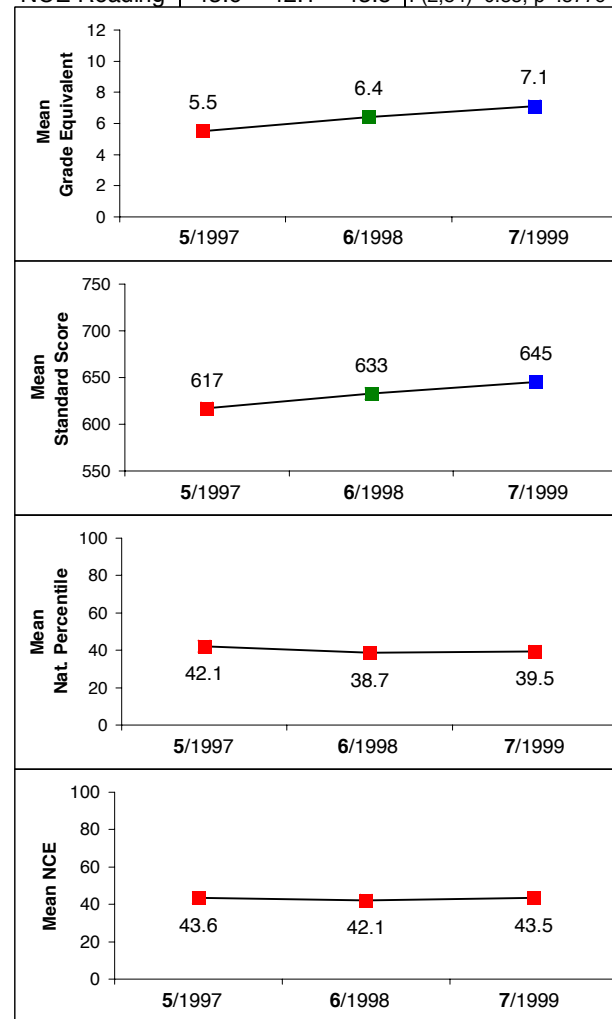
MATH

n=41	1997	1998	1999	
Grade	5th	6th	7th	
GE Math	5.4	6.8	7.4	$F(2,80)=31.13, p<.0001$
SS Math	599	624	632	$F(2,80)=60.22, p<.0001$
NP Math	35.1	38.6	36.2	$F(2,80)=0.85, p=.4311$
NCE Math	39.6	42.7	40.9	$F(2,80)=1.13, p=.3289$



READING

n=43	1997	1998	1999	
Grade	5th	6th	7th	
GE Reading	5.5	6.4	7.1	$F(2,84)=31.65, p<.0001$
SS Reading	617	633	645	$F(2,84)=41.07, p<.0001$
NP Reading	42.1	38.7	39.5	$F(2,84)=1.33, p=.2690$
NCE Reading	43.6	42.1	43.5	$F(2,84)=0.55, p=.5776$



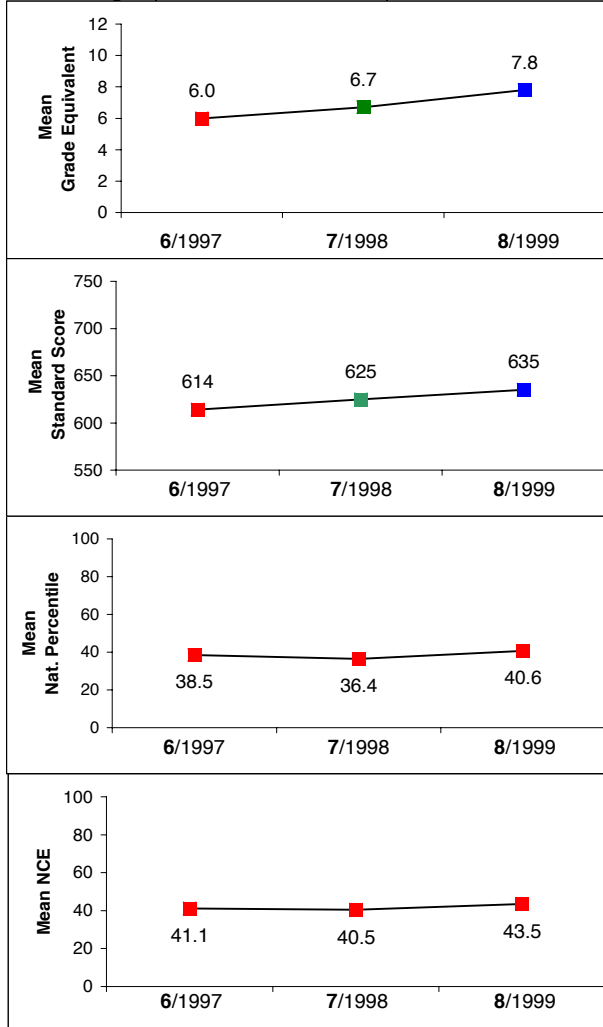
A color change represents a statistically significant change in the means. A trend with two color changes represents differences among all three means. A dual colored charting point (red and green) represents a statistically significant difference between one mean but not the other. The reader is encouraged to examine the individual table of means for these cases.

Exhibit 8:2 Seven Hills Charter School, Cohort B: Analysis of Individual Student Results on MAT-7

COHORT B

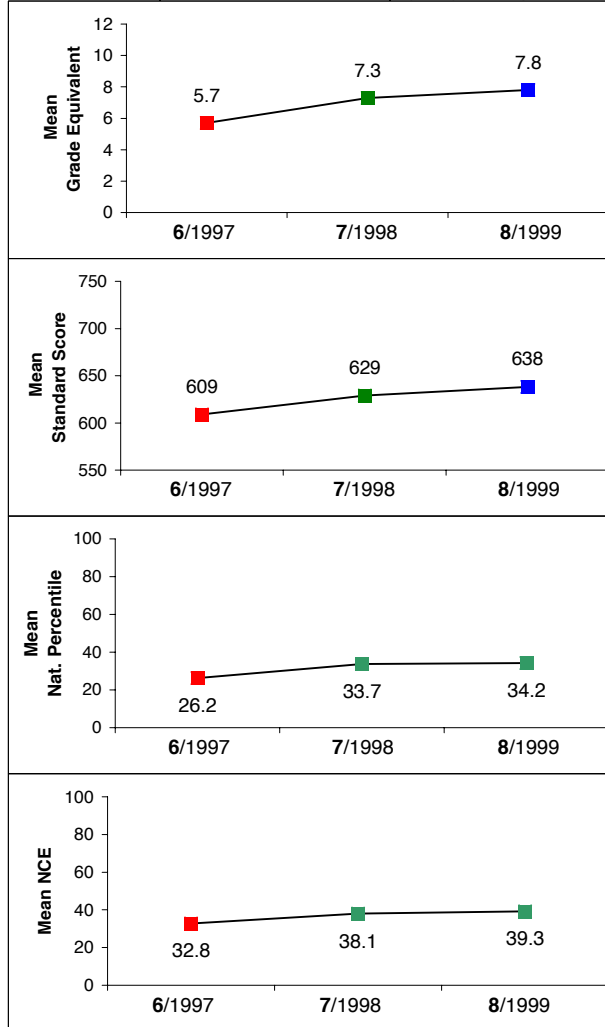
LANGUAGE

n=24	1997	1998	1999	
Grade	6th	7th	8th	
GE Lang	6.0	6.7	7.8	$F(2,46)=12.82, p<.0001$
SS Lang	614	625	635	$F(2,46)=13.69, p<.0001$
NP Lang	38.5	36.4	40.6	$F(2,46)=0.92, p=.4056$
NCE Lang	41.1	40.5	43.5	$F(2,46)=0.99, p=.3777$



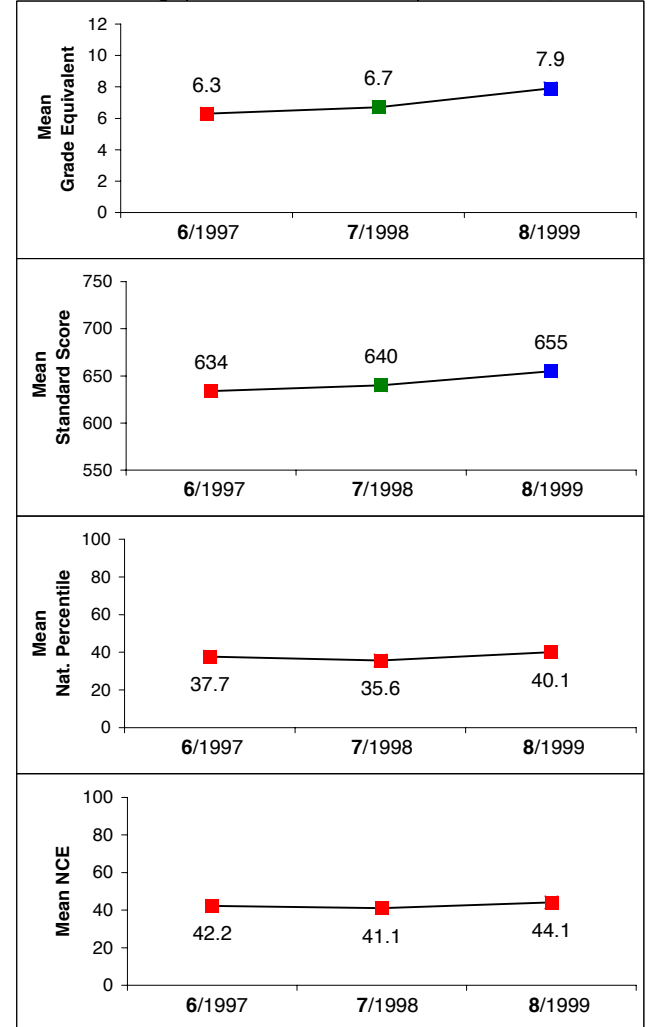
MATH

n=24	1997	1998	1999	
Grade	6th	7th	8th	
GE Math	5.7	7.3	7.8	$F(2,46)=24.30, p<.0001$
SS Math	609	629	638	$F(2,46)=33.06, p<.0001$
NP Math	26.2	33.7	34.2	$F(2,46)=5.34, p=.0082$
NCE Math	32.8	38.1	39.3	$F(2,46)=4.48, p=.0167$



READING

n=24	1997	1998	1999	
Grade	6th	7th	8th	
GE Reading	6.3	6.7	7.9	$F(2,46)=21.02, p<.0001$
SS Reading	634	640	655	$F(2,46)=18.70, p<.0001$
NP Reading	37.7	35.6	40.1	$F(2,46)=1.97, p=.1515$
NCE Reading	42.2	41.1	44.1	$F(2,46)=1.18, p=.3174$



A color change represents a statistically significant change in the means. A trend with two color changes represents differences among all three means. A dual colored charting point (red and green) represents a statistically significant difference between one mean but not the other. The reader is encouraged to examine the individual table of means for these cases.

In terms of grade level equivalents, students are expected to progress at least one grade level per year. Over the two complete calendar years, the average progress of students in Cohort A was 1.1 grade equivalents in language, 2 grade equivalents in math/problem solving, and 1.6 grade equivalents in reading. Thus, it seems that these students were making the expected progress only in math, but falling further behind the national comparison sample of the MAT-7 in language arts and reading.

In Cohort B, students progressed 1.8 grade equivalents in language, 2.1 grade equivalents in math/problem solving, and 1.6 grade equivalents in reading. The pattern in both cohorts was similar, indicating that the students were not progressing as quickly as might be expected in language and reading, although they were progressing as expected in math over two academic years. On the standard score scale, students in both cohorts increased each year in each of the three subjects except Cohort A on the language subtest where there was only an overall statistically significant gain from 1997 to 1999.

Although the grade equivalents and standard score analyses indicate that the students are progressing each year, the national percentile and the normal curve equivalent score analyses allow us to compare the growth of this cohort of students with national norms. The students in Cohort A failed to make significant gains relative to these national norms. In language, there was actually a decreasing trend in the mean national percentile scores, although not statistically significant. Mean NCEs, however, did show a statistically significant decrease in language for Cohort A between 1997 and 1999.

Cohort B performed slightly better relative to the national norms for mean national percentile and normal curve equivalents. There were no statistically significant gains for either language or reading, but there was a statistically significant gain on the math subtest between 1997 and 1998 (6th grade to 7th grade).

SAT-9

Since only two years of data were made available to us on the SAT-9, our analyses and conclusions are limited. SAT-9 data was provided on three subtests in four scales: language, mathematics, and reading in grade equivalent (GE), standard scores (SS), percentile rank (PR), and National Curve Equivalent (NCE). We defined two cohorts based on the available data; Cohort C constituted those students at Seven Hills who progressed from third to fourth grade from 1997/98 to 1998/99, and Cohort D included those students in Seven Hills who progressed from fourth to fifth grade from 1997/98 to 1998/99. Sample sizes varied slightly over the subtests. In Cohort C, there were 63, 62, and 58 students in the language, mathematics and reading subtest analyses respectively. In Cohort D there were 71, 72, and 72 students in the language, mathematics, and reading subtest analyses respectively.

Longitudinal analysis findings

Dependent t-test results for both cohorts indicated a clear pattern of increasing performance for the Seven Hills students. All pairwise differences were statistically significant except for Cohort C in reading on the PR and NCE scales. Table 8:2 presents a summary of these findings.

Table 8:2 Dependent t-Test Summary Findings for Cohort C and D

	Cohort C		Cohort D	
	Mean Diff	Std	Mean Diff	Std
GE Lang	2.56*	2.21	1.58*	2.91
GE Math	1.97*	1.87	2.45*	1.98
GE Reading	1.36*	1.95	1.23*	1.48
SS Lang	38.52*	27.40	21.89*	33.99
SS Math	46.34*	31.25	43.35*	29.44
SS Reading	25.52*	28.80	23.43*	23.94
PR Lang	12.76*	17.69	8.68*	22.26
PR Math	16.40*	21.04	16.89*	22.00
PR Reading	2.02	18.56	5.19*	17.79
NCE Lang	9.46*	13.94	5.46*	16.37
NCE Math	11.01*	15.42	11.09*	15.25
NCE Math	1.59	14.20	3.91*	11.94

* $p < .05$

8.4 Chi-Square Analysis of MCAS Data

A chi-square analysis was initiated on data available from the state of Massachusetts on the outcomes of the Massachusetts Comprehensive Assessment Systems (MCAS), the state-mandated criterion-referenced test. While the chi-square can help us distinguish the size and strength of the differences between Seven Hills Charter School and the two control groups we compared it with, our main interest is in the change over time at this Edison school. The chi-square analysis was completed because it is one step in preparation for the odds ratio analysis, which allows us to examine relative changes over time in this school or in the control groups.

The MCAS is administered in grade 4 (English, mathematics, and science and technology) and grade 8 (English, mathematics, history, and science and technology). We did not include grade 8 history, since there was only one year of data available. The MCAS is scored along an ordinal four category scale: Fail, Needs Improvement, Proficient, and Advanced. Additional information on the MCAS is provided in Appendix A.

Construction of the comparison groups

We constructed two different comparison groups for the chi-square analyses. Since we were interested in examining the number/proportion of students who met state standards (“passing”) or conversely the number/proportion of students who did not meet state standards (“failing”) on the MCAS, we needed to define a suitable comparison group. Our first comparison was with the local public school district in which Seven Hills Charter School resides (i.e., Worcester Public Schools). The second comparison group we selected was the state average passing/failing rates. While the state demographics differ from Seven Hills, we believe that comparisons with state averages can yield information regarding the relative gains of this Edison school. Also, since Edison claims that advances in other district schools are—in part—due to its presence, we use the state as a more distant point of comparison that cannot be easily influenced by the presence of Edison schools.

General procedure

Utilizing published data from the state of Massachusetts, we made comparisons over two years (Spring 1998 and Spring 1999) for both grade 4 and grade 8. Percentage data (students in each scoring category) were converted to raw frequency data prior to chi-square analysis. To insure independence of the rows in the chi-square tables, the raw frequencies for each scoring category of the MCAS in the district and state comparisons were down-weighted by subtracting the number of students in that category from Seven Hills. Thus, both the district and state numbers reflect all students in the district or state exclusive of those in Seven Hills Charter School.

Four chi-square analyses were evaluated for each subtest nested within year and grade level. Two of these analyses were on uncollapsed data; that is, all scoring levels were represented in the contingency table (e.g., a 2x4) for the district and state comparisons. Two follow-up analyses were conducted on the data after collapsing the multilevel scoring into a dichotomy (pass, fail), thus producing 2x2 contingency tables. According to the Massachusetts Department of Education, a score in the “proficient” or “advanced” category constitutes “passing,” or meeting the state standards for that particular grade and subject. On the other hand, the “fail” and “needs improvement” categories mean that the student has not met the state passing standards. Thus, for the collapsed chi-square and odds ratio analyses, we grouped the “proficient” and “advanced” levels into a passing or meeting state standards and the “needs improvement” and “failing” categories into failing or not meeting state standards.

Chi-square findings

The chi-square analyses test the null hypothesis that the relative frequency (of students) in the four (or two) scoring categories are the same for Seven Hills and the comparison group (either the district or the state). Results of the chi-square analyses for grade 4 covered the MCAS administration for the

two years from 1998 to 1999 for English, mathematics, and science and technology. Individual contingency tables are presented in Appendix E.

The first set of comparisons was made against district data as the comparison group. Six separate chi-square statistics were evaluated from 2x4 contingency tables and six from 2x2 contingency tables (see Table 8:3) for each year. Results of the three subtests administered in grade 4 closely parallel each other. Statistically significant differences were observed relative to the state proportions and district proportions in each year and subtest. Fewer students fell into the “proficient” and “advanced” categories in Seven Hills relative to the overall state and district. This pattern was also replicated in the collapsed analyses (2x2) except on the 1998 English and 1999 mathematics tests.

Results of the grade 8 MCAS chi-square analysis are summarized in Table 8:4. Similar to the grade 4 chi-square analyses, the analyses for grade 8 covered the MCAS administration for the two years from 1998 to 1999 for English, mathematics, and science and technology. Individual contingency table results are presented in Appendix E. The first set of comparisons was made against district data as the comparison group. Six separate chi-square statistics were evaluated from 2x4 contingency tables and six from 2x2 contingency tables (see Table 8:4) for each year. Results among the three subtests taken at grade 8 closely paralleled one another in 1998. In general, students at Seven Hills performed at levels commensurate with students at the district

Table 8:3 Summary of Chi-Square Findings for Seven Hills Charter School, Grade 4

	1998	1999
<i>English</i>		
Seven Hills vs. District	sig/ns	sig/sig
Seven Hills vs. State	sig/sig	sig/sig
<i>Mathematics</i>		
Seven Hills vs. District	sig/sig	sig/ns
Seven Hills vs. State	sig/sig	sig/sig
<i>Science & Technology</i>		
Seven Hills vs. District	sig/sig	sig/sig
Seven Hills vs. State	sig/sig	sig/sig

Note: Each result cell in the matrix is divided with the results for the 2x4 analysis on the left-hand side and results for 2x2 analysis on the right-hand side.

Table 8:4 Summary of Chi-Square Findings for Seven Hills Charter School, Grade 8

	1998	1999
<i>English</i>		
Seven Hills vs. District	ns/ns	ns/sig
Seven Hills vs. State	ns/sig	sig/sig
<i>Mathematics</i>		
Seven Hills vs. District	ns/ns	sig/sig
Seven Hills vs. State	sig/ns	sig/sig
<i>Science & Technology</i>		
Seven Hills vs. District	ns/ns	ns/sig
Seven Hills vs. State	ns/sig	sig/sig

Note: Each result cell in the matrix is divided with the results for the 2x4 analysis on the left-hand side and results for 2x2 analysis on the right-hand side.

and state except in three analyses. In these three exceptions, a greater proportion of students were failing at Seven Hills than in the comparison group.

8.5 Odds Ratio Analysis of the MCAS Data

One of the many possible statistics that can be derived from a 2x2 contingency table is the odds ratio statistic (OR) and corresponding $1-\alpha$ confidence interval. As presented in Section 2.4 of this report, the 2x2 tables analyzed in the previous section can be thought of as representing consecutive class cohorts in a prospective design. From a classical epidemiological perspective, the students in the Edison school can be thought of as the “exposed” group, that is, exposed to the “Edison-effect,” and students in the comparison group as the unexposed group. From this perspective, each yearly comparison is a new cohort, measured over a period of years. There is a minimal possibility for cohort contamination if a number of students in one group are not promoted to the next grade level. However, we think this represents a very small number of possible cases and therefore has minimal impact on the validity of these analyses. Section 2.4 details the OR statistic and corresponding $1-\alpha$ confidence interval.

We calculated and charted OR for each of the 2x2 tables constructed from the chi-square analyses presented above. Tables 8:5 and 8:6 and Exhibits 8:3 and 8:4 present these findings. Exhibits 8:3 and 8:4 graphically illustrate the overall performance of the Seven Hills Charter School on the MCAS in comparison with district and state performance levels at the 4th and 8th grades, respectively. The charts in Exhibits 8:5 - 8:11 illustrate the distribution of scores across all four performance categories. It is important to remember that the odds ratio findings consider only two categories, with “Advanced” and “Proficient” combined for the pass category and “Needs Improvement” and “Fail” combined for the fail category in the odds ratio analysis.

Odd ratio findings, grade 4

Comparison against the district. Three OR analyses were evaluated, one for each subject test on the MCAS. In grade 4 English, the OR for a Seven Hills student failing the 1998 English MCAS relative to the students in the district as a whole was 2.535; in 1999 it was 3.893. In other words, students attending Seven Hills Charter School were 2.5 times more likely to not meet state standards than students enrolled in other district schools in 1998. Since the odds ratio increased to 3.8 in 1999, this indicates that the performance at the Edison school on the MCAS was not as good as the performance of other grade 4 students in the district. The Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the two years was not statistically significant, indicating there was no real (statistically significant) change in OR. Thus, a common OR = 3.194 with a lower bound (LB) = 1.623 and an upper bound (UB) = 6.286 indicated that the Seven Hills’ grade 4 students were about three times more likely to fail (i.e., score in either the “needs improvement” or “fail” categories) the English MCAS than students in the rest of the district. The OR analysis of the grade 4 mathematics component of the MCAS presented a different picture. The Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the two years just barely missed conventional levels of statistically significant ($p = .053$). If one were to interpret this that a common

OR cannot be reasonably interpreted, ORs need to be examined for each year. In 1998 the OR was statistically significant—OR = 3.421, LB = 1.709, UB = 6.846—indicating that Seven Hills’ students were much more likely to fail relative to students in the rest of the district. In 1999, however, the OR decreased to a nonstatistically significant value of 1.422. Thus, the students at Seven Hills showed a rather large improvement in passing the MCAS math test in 1999.

Regarding the MCAS subtest of science and technology, the Breslow-Day chi-square for testing the hypothesis of homogeneity of OR

over the two years was not statistically significant, indicating there was no real (statistically significant) change in OR. Thus, a common OR = 3.154 with LB = 2.175 and UB = 4.574 indicated that the Seven Hills students were more than three times as likely to fail this test relative to the rest of the district. Table 8:5 contains a summary of the odds ratio findings for grade 4.

Comparison against state. In grade 4 English, the OR for a Seven Hills student failing the 1998 English MCAS relative to students in the state as a whole was 4.882, and in 1999 it was 5.256. In both years the 1- α CI did not eclipse 1 and thus can be interpreted as statistically significant. The Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the two years was not statistically significant, indicating there was no real (statistically significant) change in OR. Thus, a common OR = 5.067 with LB = 2.677 and UB = 9.591 indicated that the Seven Hills’ students were about 5 times more likely to fail the MCAS English subtest as compared with students in the state as a whole.

The OR analysis of the grade 4 mathematics component of the MCAS presented a similar picture. The Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the two years was not statistically significant ($p = .062$). Thus, the common OR was 2.919, LB = 1.954, UB = 4.361. Although students at Seven Hills showed a rather large improvement on OR, they were still about three times more likely to fail the math MCAS.

Regarding the MCAS subtest of science and technology, the Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the two years was not statistically significant, indicating there was no real (statistically significant) change in OR. Thus, the common OR = 4.525 with LB = 3.191 and UB = 6.416 indicated that the Seven Hills students had a substantially greater odds for failing (about 4½ times) than students in the rest of the state.

Table 8:5 Summary of Odds Ratio Findings for Seven Hills Charter School, Grade 4

	1998	1999	2-year OR
<i>Odds of not meeting standard compared with district</i>			
English	2.535	3.893	3.194
Mathematics	3.421	1.422	2.045
Science & Tech.	2.799	3.572	3.154
<i>Odds of not meeting standard compared with state</i>			
English	4.882	5.256	5.067
Mathematics	4.763	2.065	2.919
Science & Tech.	3.588	5.779	4.525

Odds ratio findings, grade 8

Comparison against the district. In grade 8, the OR for Seven Hills' students showed an increasing, but not statistically significant, trend for odds of failure on the English component of the MCAS relative to students in the district. The Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the two years was not statistically significant, indicating that a common OR over the two years was meaningful, OR = 1.417, LB = 0.967, UB = 2.076. Thus, for the two-year period the Seven Hills' students were at even odds for failure relative to students in the rest of the district.

On the math subtest, students at Seven Hills showed a statistically significant rise in the OR, e.g., the Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the two years was statistically significant. In 1998 the OR was 0.770 and rose to 3.485 in 1999.

The grade 8 MCAS subtest of science and technology evidenced a pattern similar to math. The Seven Hills' students started out in 1998 at even odds relative to district students but fell behind in 1999, OR = 2.729, LB = 1.132, UB = 6.580. Table 8:6 contains a summary of the odds ratio findings for grade 8.

Comparison against state. In grade 8, the ORs for Seven Hills' students revealed an increasing odds for failure on both the English and math MCAS relative to other students in the state. In both analyses the Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the two years was statistically significant. For English, the 1998 OR = 1.863, LB = 1.084, UB = 3.204, indicated that Seven Hills' students were slightly more likely to fail. However, in 1999 the OR = 4.566, LB = 2.834, UB = 7.357, indicating that Seven Hills' students lost ground relative to students in the rest of the state. On the math subtest, students at Seven Hills showed a statistically significant rise in the OR, e.g., in 1998 the OR=1.741 and rose to 7.503 in 1999.

Regarding the grade 8 MCAS subtest of science and technology, the Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the two years was not statistically significant, indicating there was no real (statistically significant) change in OR. Thus, a common OR = 3.593 with LB = 2.104 and UB = 6.135 indicated that the Seven Hills' students were about 3 ½ times more likely to fail the science and technology MCAS relative to students in the rest of the state.

Table 8:6 Summary of Odds Ratio Findings for Seven Hills Charter School, Grade 8

	1998	1999	2-year OR
<i>Odds of not meeting standard compared to district</i>			
English	1.034	1.828	1.417
Mathematics	0.770	3.485	* p=.010
Science & Tech.	0.085	2.729	* p=.046
<i>Odds of not meeting standard compared to state</i>			
English	1.863	4.566	* p=.019
Mathematics	1.741	7.503	* p=.011
Science & Tech.	2.189	5.840	3.593

Note: The asterisks indicate that the difference in OR over two years was statistically significant, thus making it impossible to calculate an overall OR. The p-value is listed in place of the 2-year OR.

Exhibit 8:3 Results of the Odds Ratio Analysis for Seven Hills Charter School (Grade 4)

Grade 4 MCAS English

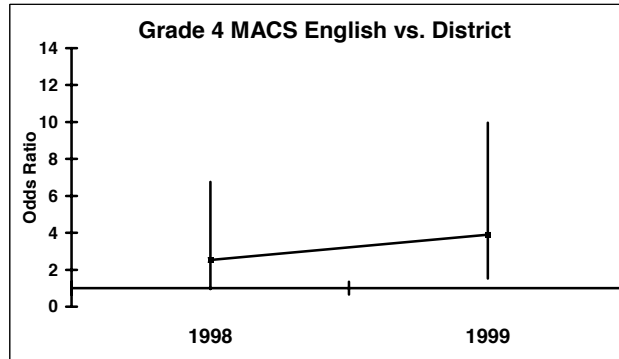
Year	U CI	L CI	OR
1998	6.756	0.951	2.535
1999	9.966	1.521	3.893

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=3913) = 0.347, p = .556

OR = 3.194

UB = 6.286

LB = 1.623



Grade 4 MCAS English

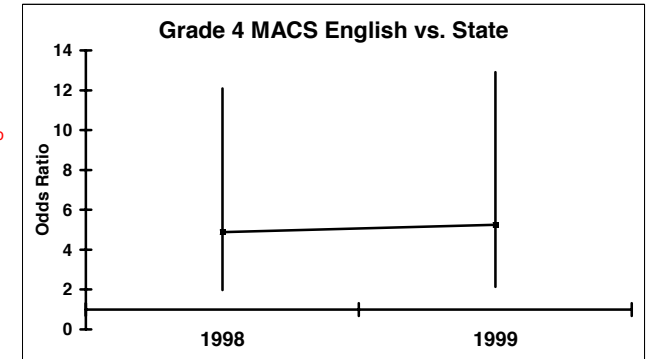
Year	U CI	L CI	OR
1998	12.088	1.971	4.882
1999	12.902	2.141	5.256

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=151,335) = 0.10, p = .919

OR = 5.067

UB = 9.591

LB = 2.677



Grade 4 MCAS Math

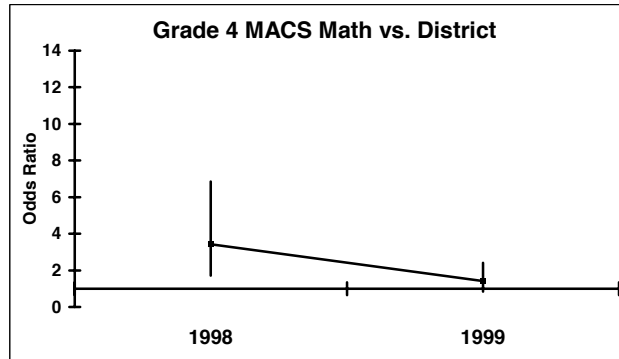
Year	U CI	L CI	OR
1998	6.846	1.709	3.421
1999	2.413	0.838	1.422

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=4,032) = 3.736, p = .053

OR = 2.045

UB = 3.105

LB = 1.348



Grade 4 MCAS Math

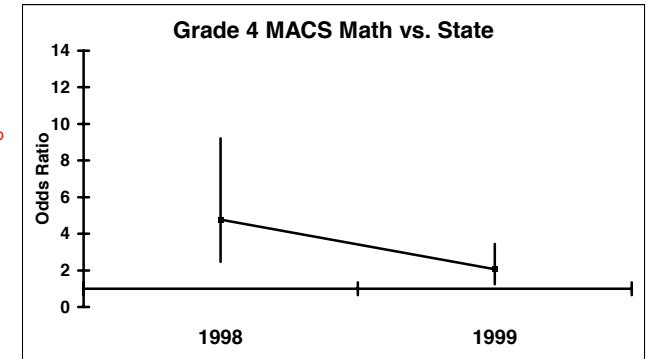
Year	U CI	L CI	OR
1998	9.217	2.461	4.763
1999	3.442	1.239	2.065

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=152,225) = 3.472, p = .062

OR = 2.919

UB = 4.361

LB = 1.954



Grade 4 MCAS Science & Tech.

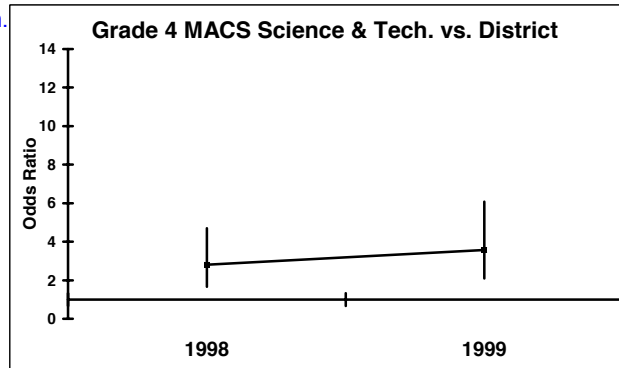
Year	U CI	L CI	OR
1998	4.707	1.665	2.799
1999	6.085	2.097	3.572

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=4,032) = 0.373, p = .542

OR = 3.154

UB = 2.175

LB = 4.574



Grade 4 MCAS Science & Tech.

Year	U CI	L CI	OR
1998	5.908	2.179	3.588
1999	9.469	3.528	5.779

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=152,234) = 1.469, p = .225

OR = 4.525

UB = 6.416

LB = 3.191

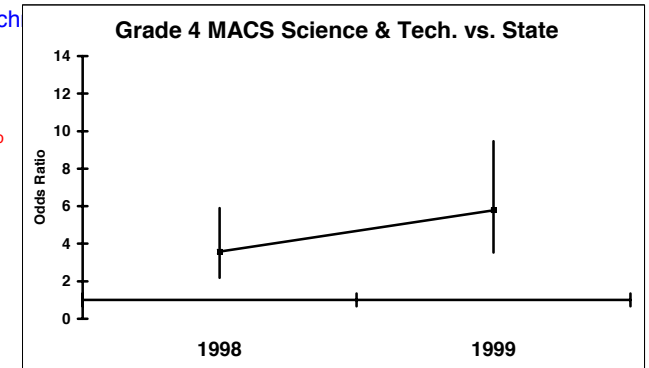


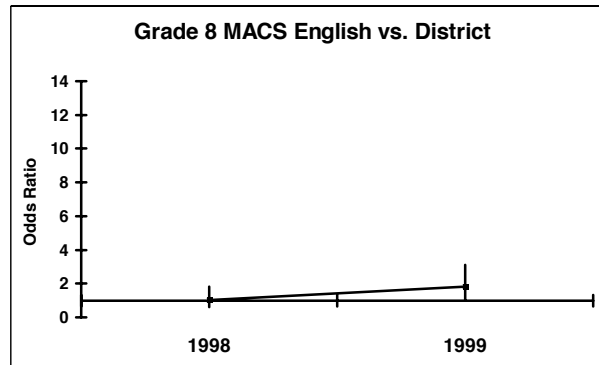
Exhibit 8:4 Results of the Odds Ratio Analysis for Seven Hills Charter School (Grade 8)

Grade 8 MCAS English

Year	U CI	L CI	OR
1998	1.811	0.591	1.034
1999	3.097	1.079	1.828

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=3,189) = 2.100, p = .147

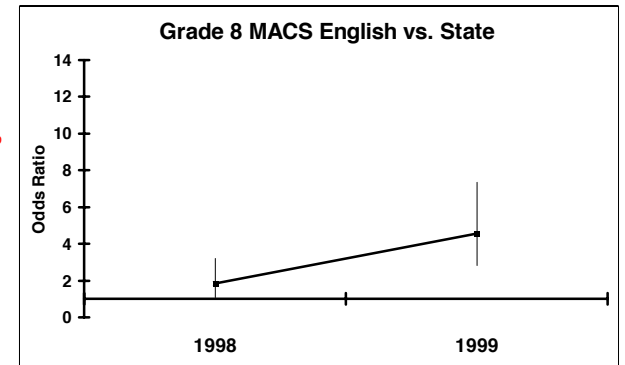
OR = 1.417
UB = 2.076
LB = 0.967



Grade 8 MCAS English

Year	U CI	L CI	OR
1998	3.204	1.084	1.863
1999	7.357	2.834	4.566

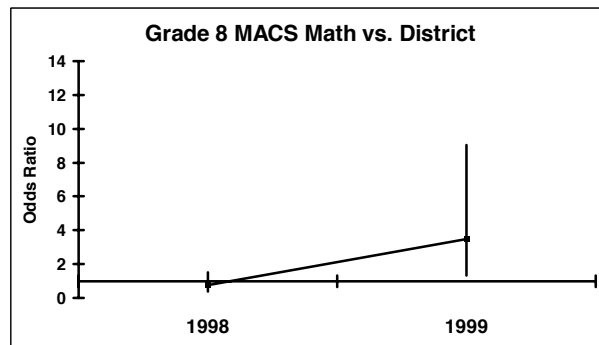
Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=138,021) = 5.468, p = .019



Grade 8 MCAS Math

Year	U CI	L CI	OR
1998	1.520	0.394	0.770
1999	9.046	1.343	3.485

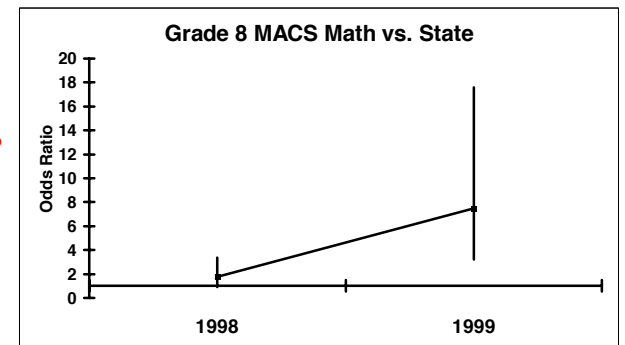
Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=3,275) = 6.604, p = .010



Grade 8 MCAS Math

Year	U CI	L CI	OR
1998	3.354	0.904	1.741
1999	17.620	3.195	7.503

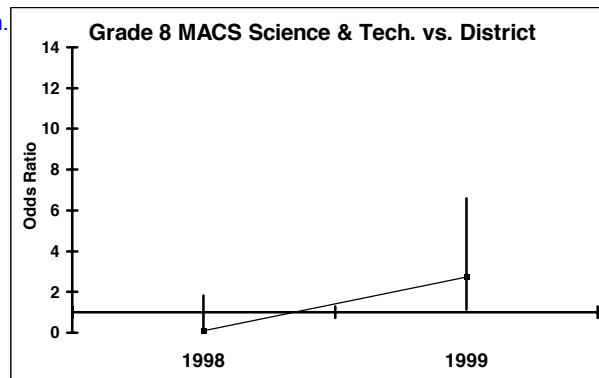
Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=138,937) = 6.434, p = .011



Grade 8 MCAS Science & Tech.

Year	U CI	L CI	OR
1998	1.818	0.393	0.085
1999	6.580	1.132	2.729

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=3,231) = 3.985, p = .046



Grade 8 MCAS Science & Tech

Year	U CI	L CI	OR
1998	4.557	1.051	2.189
1999	12.970	2.630	5.840

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (1, N=140,306) = 2.830, p = .093

OR = 3.593
UB = 6.135
LB = 2.104

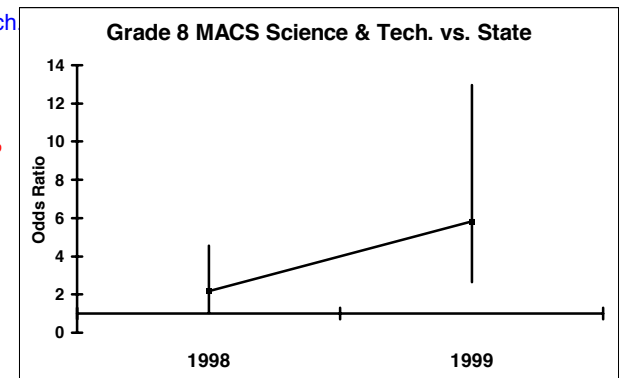
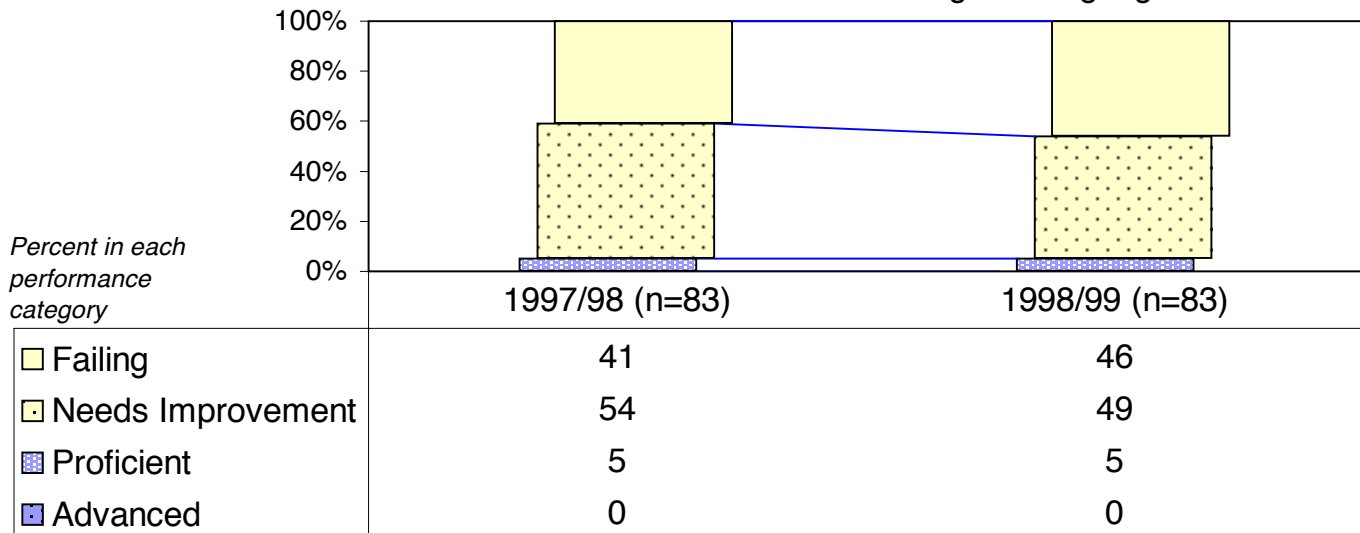
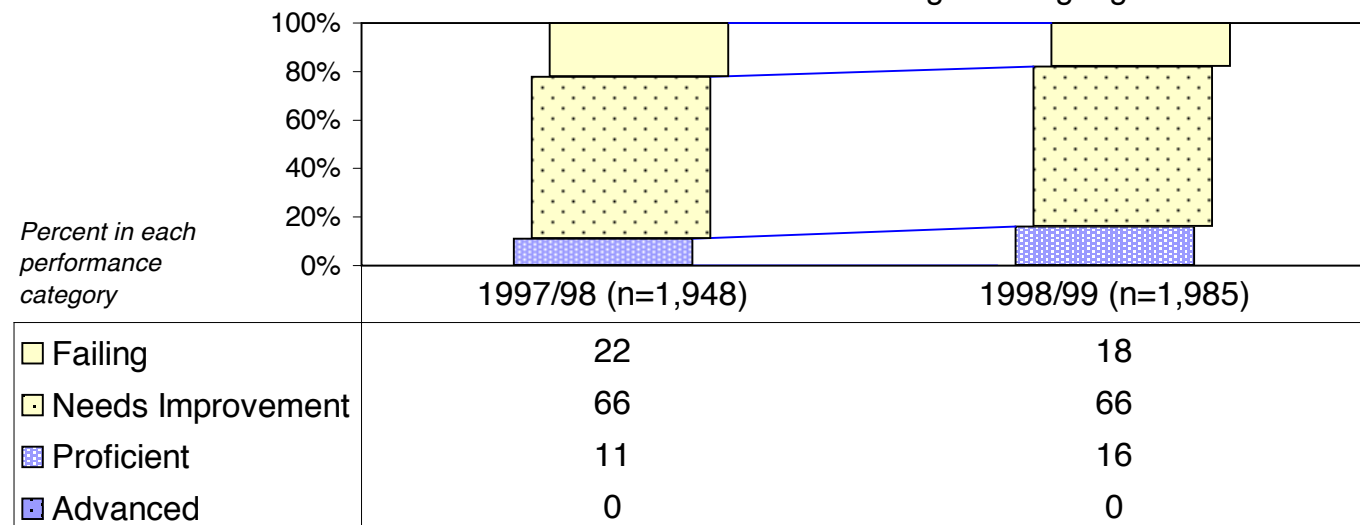


Exhibit 8:5 Performance on Grade 4 English Lang. Arts for Edison, District, and State

Seven Hills Charter School MCAS Results for 4th Grade English Language Arts



Worcester Public Schools MCAS Results for 4th Grade English Language Arts



State of Massachusetts MCAS Results for 4th Grade English Language Arts

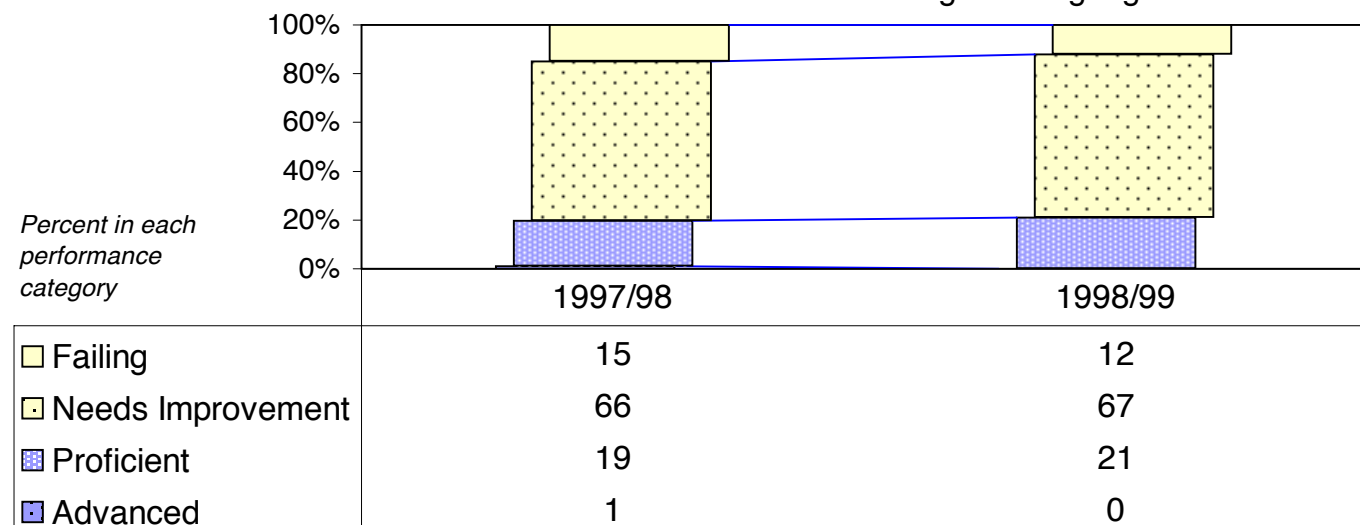


Exhibit 8:6 Performance on Grade 4 Math for Edison, District, and State

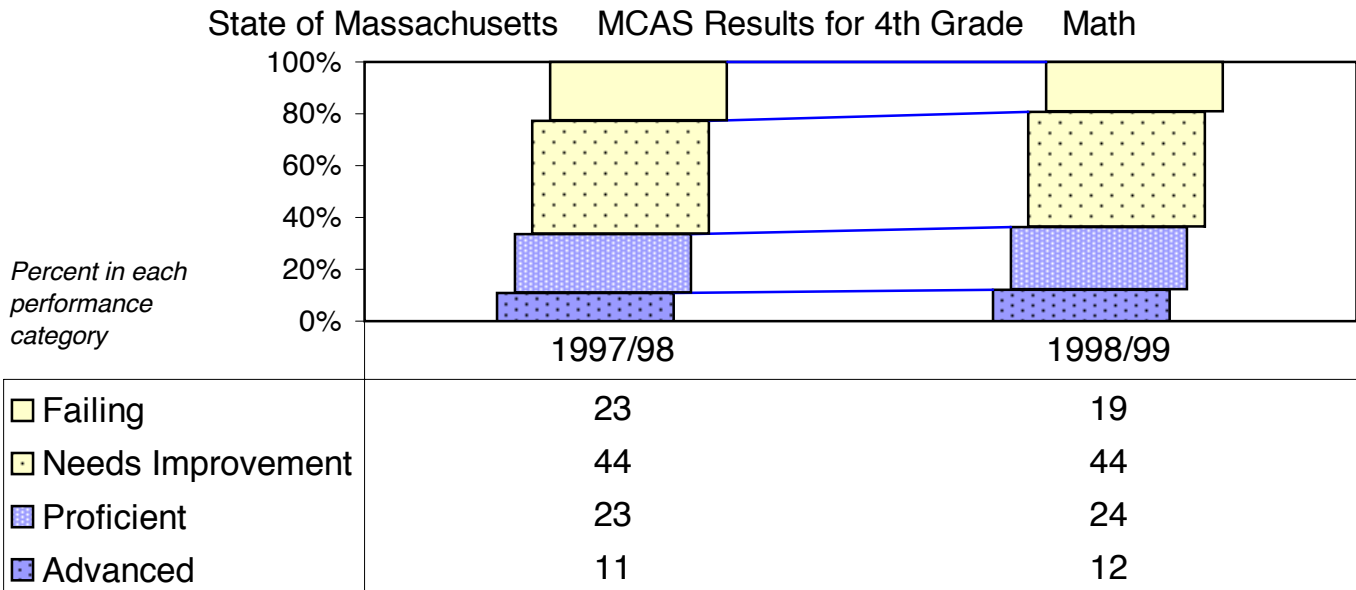
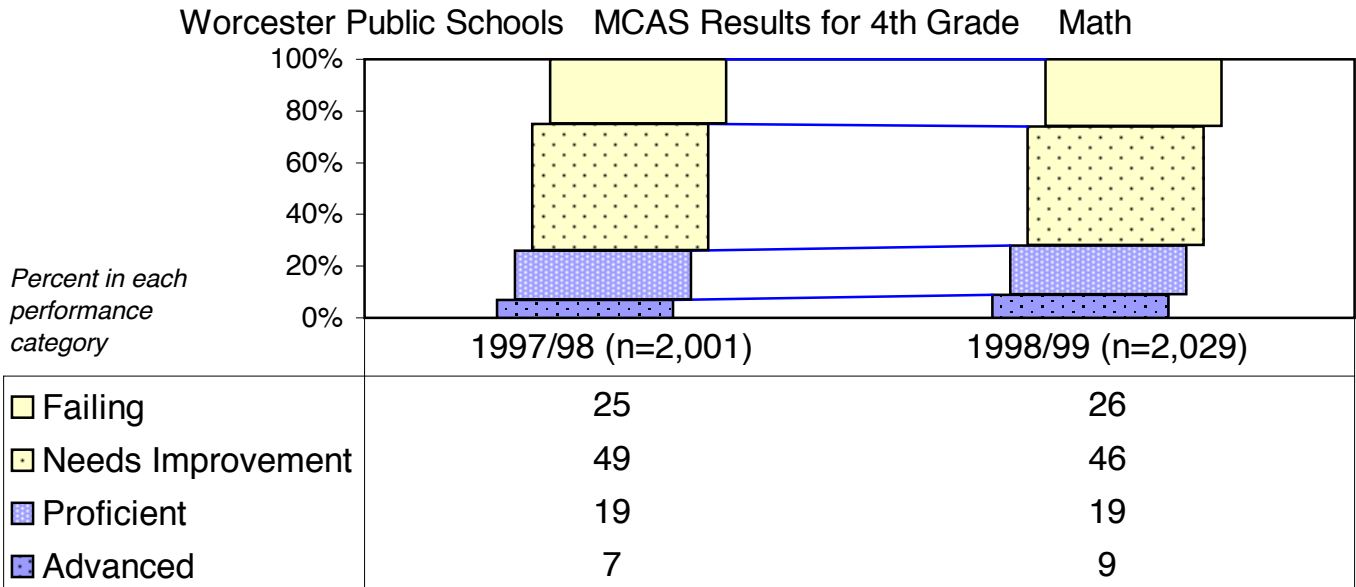
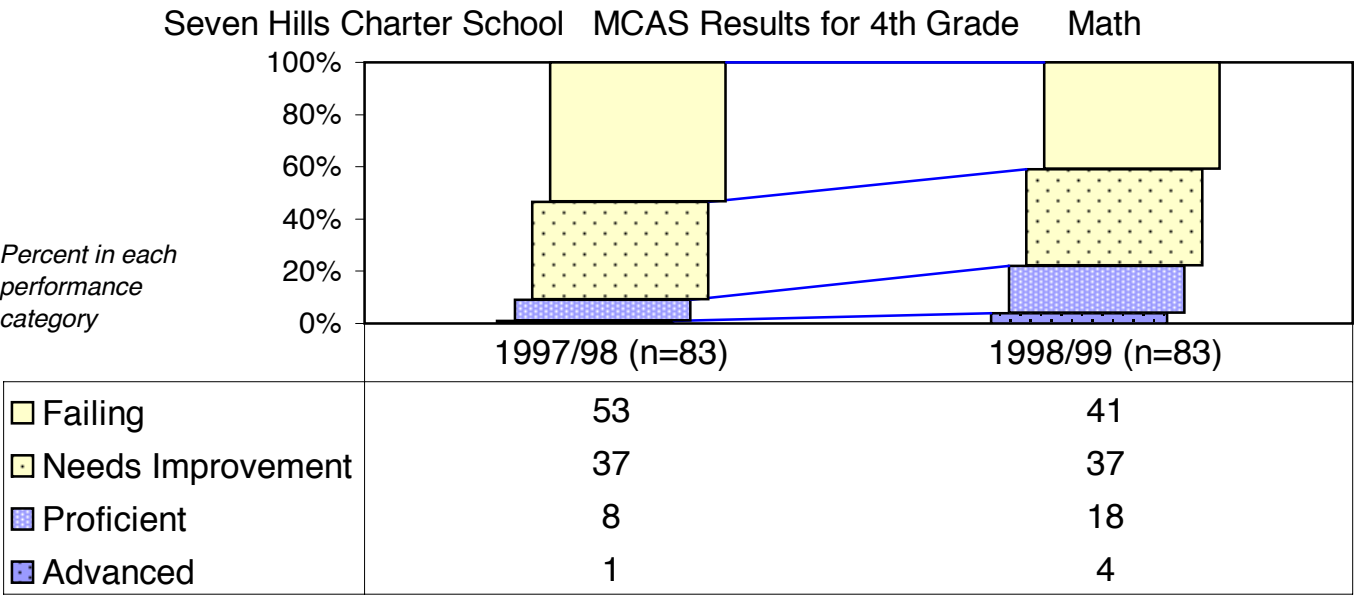
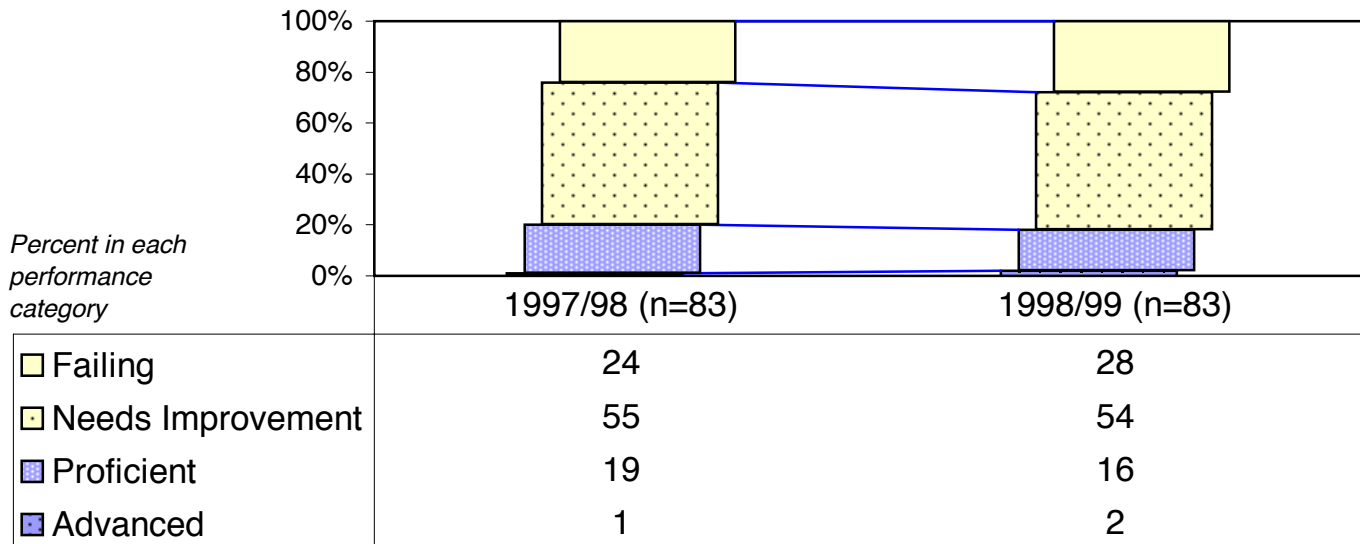
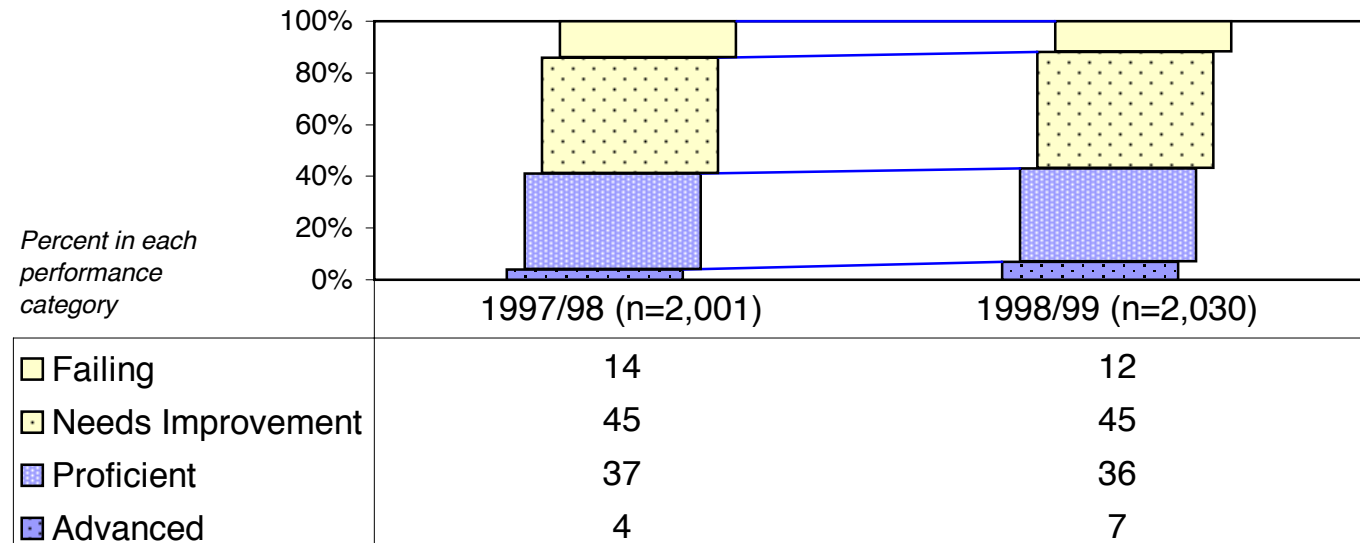


Exhibit 8:7 Performance on Grade 4 Science & Tech. for Edison, District, and State

Seven Hills Charter School MCAS Results for 4th Grade Science & Tech



Worcester Public Schools MCAS Results for 4th Grade Science & Tech.



State of Massachusetts MCAS Results for 4th Grade Science & Tech.

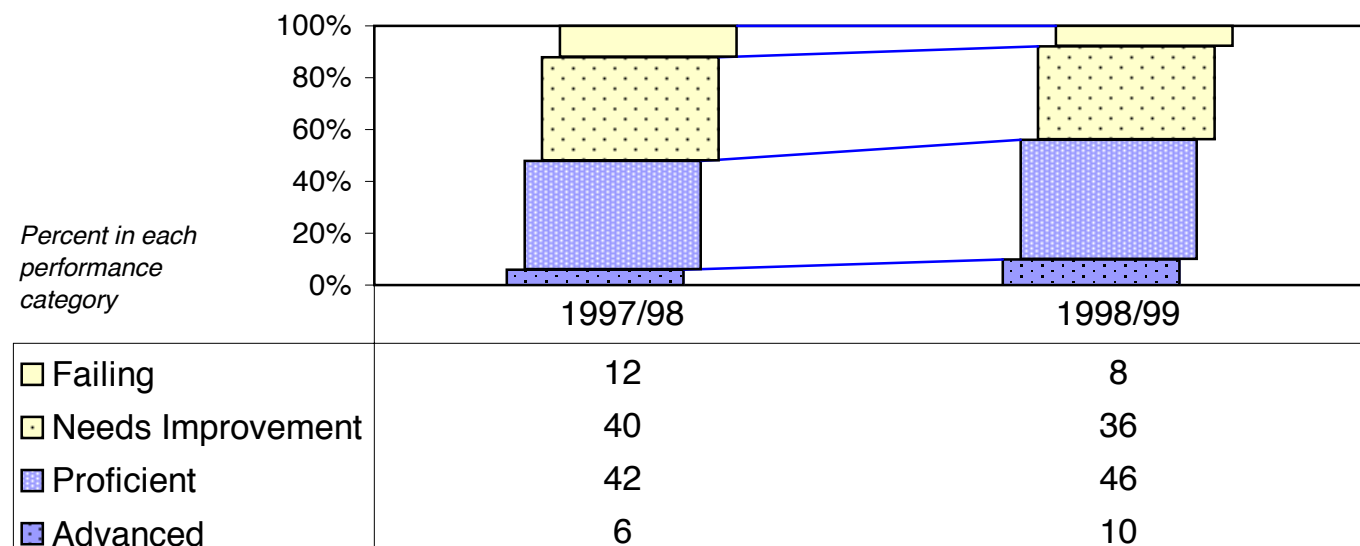
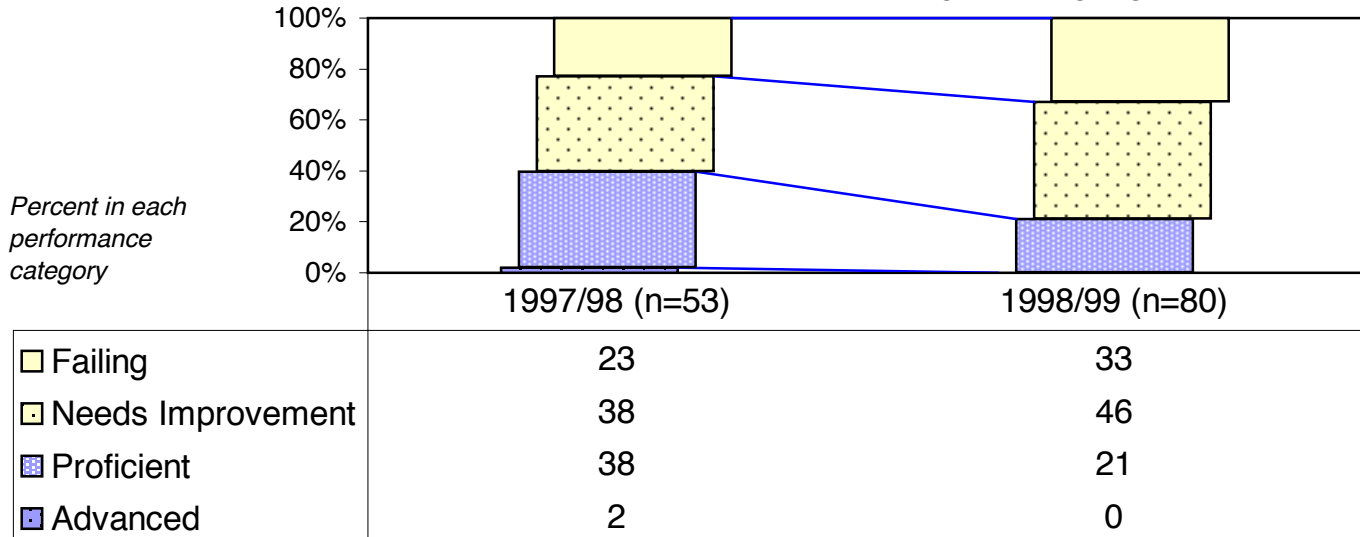
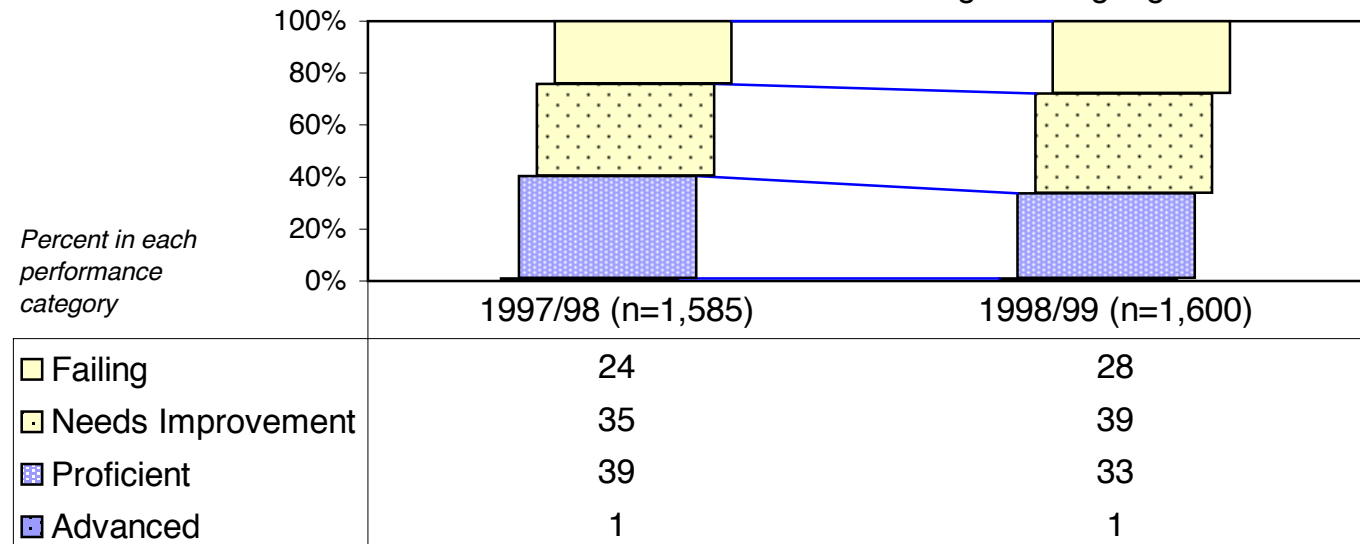


Exhibit 8:8 Performance on Grade 8 English Lang. Arts for Edison, District, and State

Seven Hills Charter School MCAS Results for 8th Grade English Language Arts



Worcester Public Schools MCAS Results for 8th Grade English Language Arts



State of Massachusetts MCAS Results for 8th Grade English Language Arts

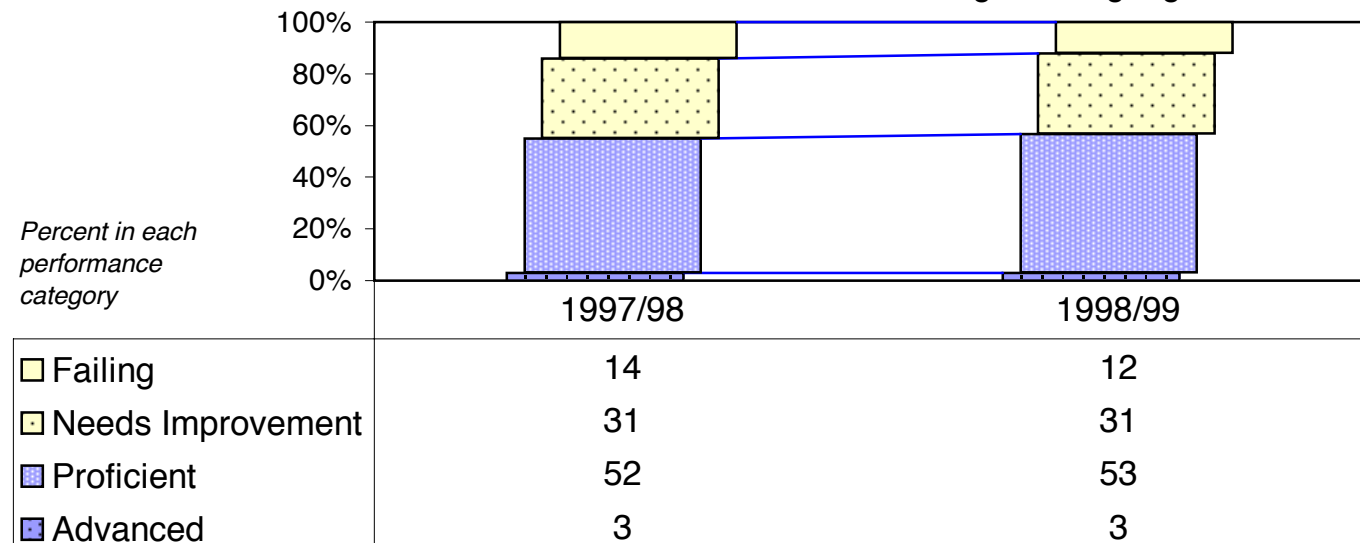
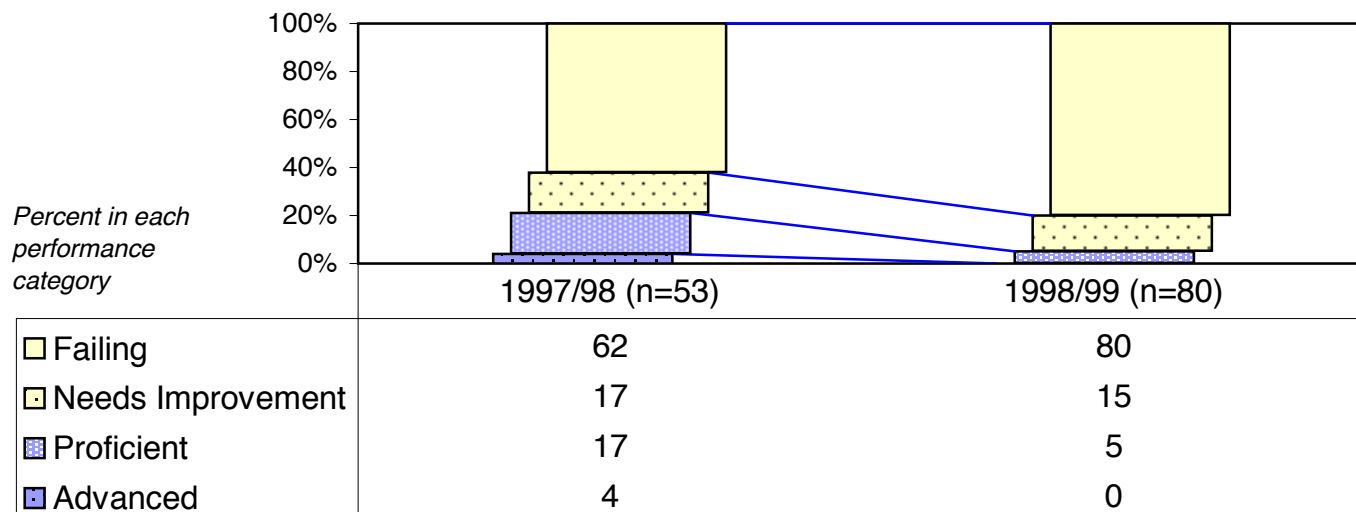
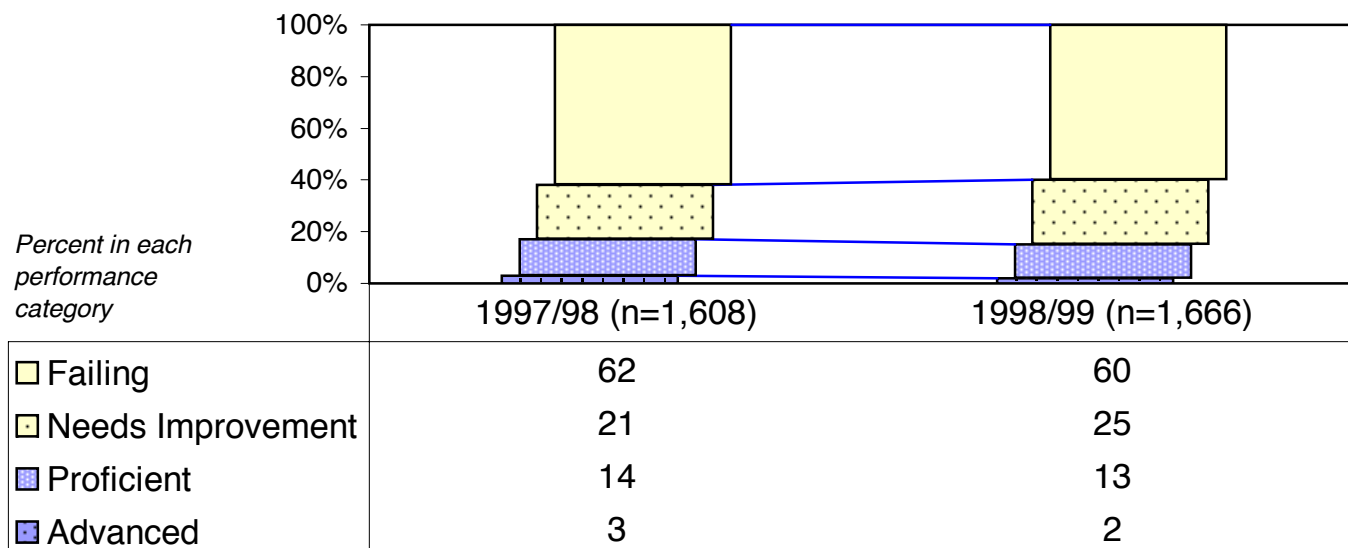


Exhibit 8:9 Performance on Grade 8 Math for Edison, District, and State

Seven Hills Charter School MCAS Results for 8th Grade Math



Worcester Public Schools MCAS Results for 8th Grade Math



State of Massachusetts MCAS Results for 8th Grade Math

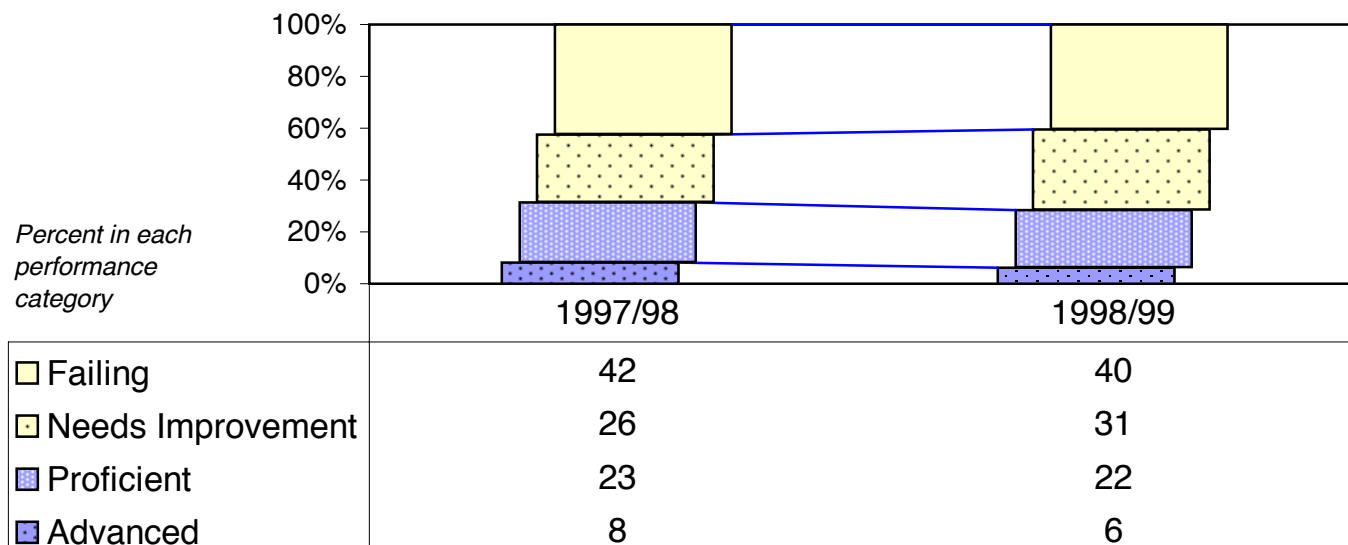
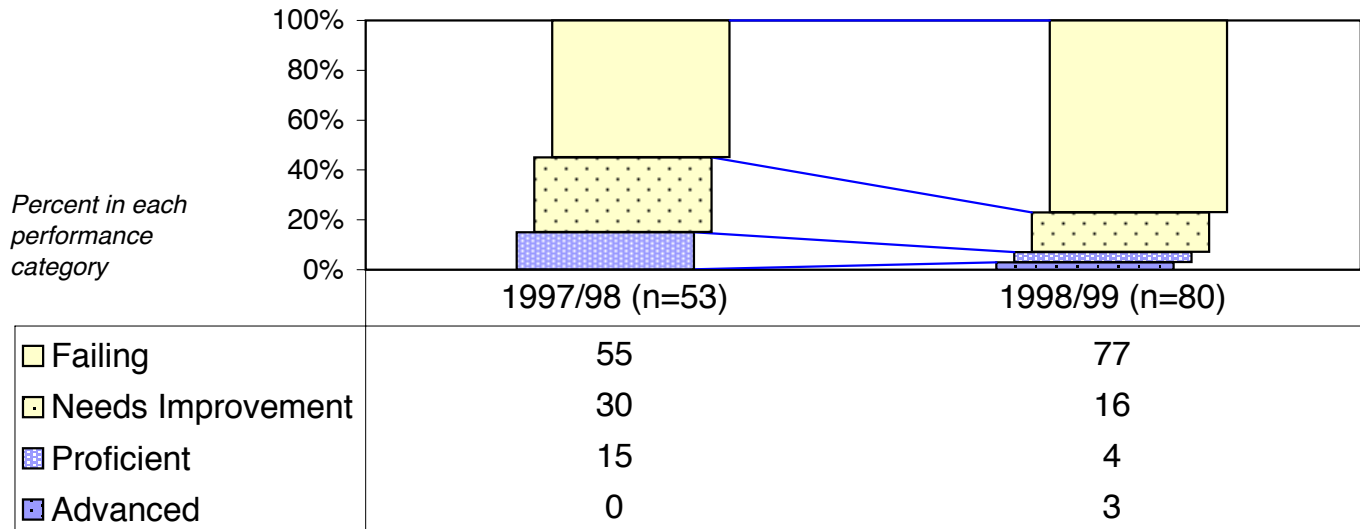
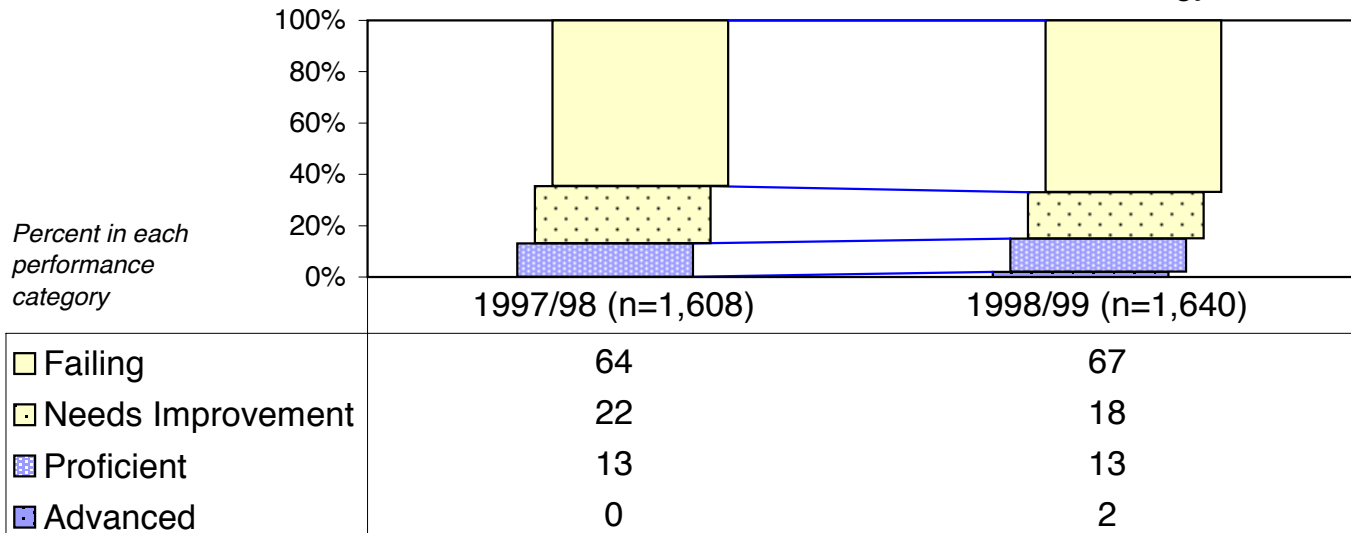


Exhibit 8:10 Performance on Grade 8 Science & Tech for Edison, District, and State

Seven Hills Charter School MCAS Results for 8th Grade Science & Tech



Worcester Public Schools MCAS Results for 8th Grade Science & Technology



State of Massachusetts MCAS Results for 8th Grade Science & Technology

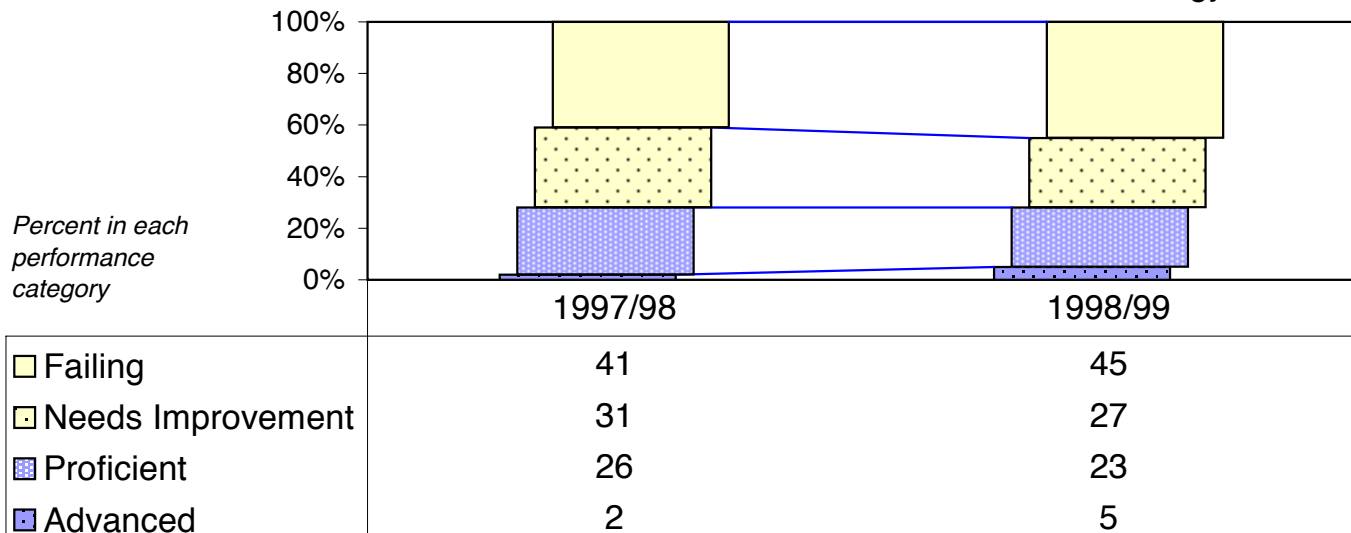
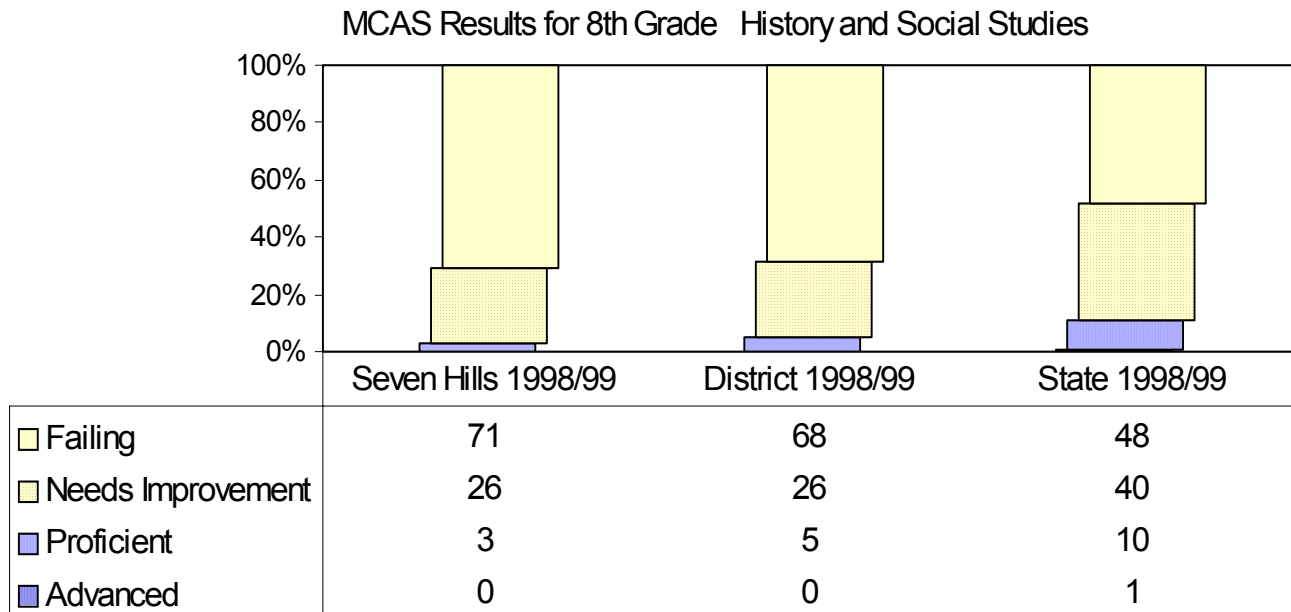


Exhibit 8:11 Performance on Grade 8 History and Social Studies for Edison, District, and State



8.6 Summary

While the students, on the whole, made some gains on the norm-referenced test, the gain was mixed and often depended on the achievement test.

Norm-referenced test findings

We constructed summary tables for aggregating our findings on the MAT-7 (Table 8:6) and SAT-9 (Table 8:7). A summary score of -1 indicates a result that is unfavorable toward the sample school, a score of 0 indicates a neutral finding, and a score of 1 indicates a favorable result according to the criteria specified in Section 2.5. The effect size (ES) is the omega squared (ω^2) for a one-way repeated measures ANOVA (Kepple, 1991).

Tables 8:6 and 8:7 highlight the main findings and trend ratings for the two norm-referenced test results considered. It appears that Seven Hills' students evidenced mild achievement growth over the three academic years, but did not make any noteworthy gains in comparison with the national norm. The Edison effect, as measured by the relative growth of student learning as measured by the MAT-7, indicated primarily no gains or losses relative to national norms over the three years we traced the students. Two exceptions to this are when the larger Cohort A showed a statistically significant decrease relative to the national norm for NCE over two years in language and the smaller Cohort B showed a statistically significant gain on the NCE over one year in math.

Table 8:6 Summary of Results on the Metropolitan Achievement Test (MAT-7)

Cohort A Grades 5,6,7 (1997-99)	Std Score		GE			PR			NCE			Trend
	p-value	ES	p-value	Δ	ES	p-value	Δ	ES	p-value	Δ	ES	
Language	.0005	.100	.0003	1.1	.110	.0788	-5.8	.024	.0476	-4.7 (0)	.0476	mixed (0)
Math	<.0001	.491	<.0001	2.0	.329	.4311	1.1	-.002	.3289	1.3 (0)	.002	mixed (0)
Reading	<.0001	.383	<.0001	1.6	.321	.2690	-2.6	.005	.5776	-0.1 (0)	-.007	mixed (0)
Cohort B Grades 6,7,8 (1997-99)	Std Score		GE			PR			NCE			Trend
	p-value	ES	p-value	Δ	ES	p-value	Δ	ES	p-value	Δ	ES	
Language	<.0001	.261	<.0001	1.8	.247	.4056	2.1	-.002	.3777	2.4 (0)	0.0	mixed (0)
Math	<.0001	.471	<.0001	2.1	.393	.0088	8.0	.108	.0167	6.5 (0)	.088	mixed (0)
Reading	<.0001	.330	<.0001	1.6	.359	.1515	2.4	.026	.3174	1.9 (0)	.005	mixed (0)

Table 8:7 Summary of Results on the Stanford Achievement Test (SAT-9)

Cohort C Grades 3&4 (1998-99)	Std Score		GE			PR			NCE			Trend
	p-value	ES	p-value	Δ	ES	p-value	Δ	ES	p-value	Δ	ES	
Language	<.0001	.495	<.0001	2.6	.400	<.0001	12.8	.201	<.0001	9.5 (+1)	.182	positive (+1)
Math	<.0001	.522	<.0001	1.9	.354	<.0001	16.4	.226	<.0001	11.0 (+1)	.198	positive (+1)
Reading	<.0001	.277	<.0001	1.4	.189	.4113	2.0	-.003	.3960	1.6 (0)	-.002	mixed (0)
Cohort D Grades 4&5 (1998-99)	Std Score		GE			PR			NCE			Trend
	p-value	ES	p-value	Δ	ES	p-value	Δ	ES	p-value	Δ	ES	
Language	<.0001	.167	<.0001	1.6	.123	.0016	8.7	.064	.0064	5.5 (0)	.046	mixed (0)
Math	<.0001	.519	<.0001	2.5	.430	<.0001	16.9	.223	<.0001	11.1 (1)	.205	positive (+1)
Reading	<.0001	.321	<.0001	1.2	.254	.0156	5.2	.034	.0070	3.9 (0)	.044	mixed (0)

Criterion-referenced test findings

Decisions regarding the OR were based on whether or not the $(1-\alpha)$ C.I. included 1.0. If the $(1-\alpha)$ C.I. fell completely below 1.0, this was interpreted as a protective odds ratio (1), thus favoring the Edison School. If the $(1-\alpha)$ C.I. included 1.0 (0), this was interpreted as an equal odds situation. If the $(1-\alpha)$ C.I. fell completely above 1.0 (-1), this was interpreted as an increase in odds for failing the state CRT relative to the comparison sample.

Table 8:8 Summary Results on MCAS Criterion-Referenced Tests

Grade 4	1998	1999	B-D	Trend
District - English			-1	negative (-1)
District - Math			-1	negative (-1)
District - Science/Tech			-1	negative (-1)
Grade 8	1998	1999	B-D	Trend
District - English			0	mixed (0)
District - Math	0	-1		negative (-1)
District - Science/Tech	0	-1		negative (-1)

Note: All comparisons with the state as a comparison group were negative

If the Breslow-Day statistic (B-D) is nonsignificant, one overall OR and $(1-\alpha)$ C.I can be used to represent the odds for failing the CRT relative to the comparison group. Thus, there are no trends reported for each specific year in the tables, only a rating in the B-D column. If the B-D statistic is found to be statistically significant, then an overall common OR cannot be meaningfully interpreted; that is, there is a statistically significant change in the OR over years and yearly ORs are necessary. Thus, our summary ratings appear for each year of data and not in the B-D column (see Table 8:8).

The odds ratio analysis of the MCAS results indicated that the odds of failing (i.e., scoring in the “needs improvement” or “failure” categories) the state assessment test were higher at Seven Hills Charter School than in the district and in the state as a whole. What is most disconcerting is that the odds of Seven Hills’ students failing increased over the two years in all subject areas but one in both grades 4 and 8. The one exception was grade 4 math where the odds of failing decreased at the Edison school compared with both district and state.

Combined ratings

In terms of absolute scores, the school performed below national norms on the MAT-7 and below district and state levels on the state assessment test (MCAS). The overall achievement gains made by this school were slightly smaller than comparison groups.

Table 8:9 Combined Overall Trends for Seven Hills Charter School

	Positive	Mixed	Negative
Norm Referenced	3 of 12	9 of 12	0 of 12
Criterion Referenced	0 of 6	1 of 6	5 of 6
TOTALS	3 of 18	10 of 18	5 of 18

Table 8:9 contains the combined overall trends for Seven Hills Charter School. Our findings based upon the analysis of individual student results on the MAT-7, SAT-9, and the odds ratio analysis of consecutive cohorts on the MCAS, indicate that the performance of this school—in terms of student achievement—could best be characterized as Mixed with a mean rating of -0.11 because nearly all the trends were mixed although 3 NRT trends were positive and 5 of the CRT trends were moving in the wrong direction. In its 1999 annual report, Edison rated this school as Strongly Positive. In its 2000 annual report it rated the 1999-2000 school year as Mixed and the achievement gains since opening also as Mixed. This is the only school in this study where the rating given by Edison and The Evaluation Center are the same.

Chapter Nine

Dr. Martin Luther King Jr. Academy

Mt. Clemens, Michigan

9.1 Descriptive Summary of School

Dr. Martin Luther King Jr. Academy (MLK) is a district contract school established in 1995 to serve grades K-5. During the 1999-2000 school year Edison reported that the school had 110 instructional staff and enrolled 863 students (Edison, 2000).

Enrollment at MLK has increased steadily from 484 students during the 1994/1995 school year (pre-Edison), to 558 students during the 1997/1998 school year. The student mobility in 1997/1998 was reported at 1 percent (Edison, 1999), and the enrollment in 1998/1999 dramatically increased to 877 students. The enrollment of 863 students for the 1999/2000 school year indicated a change of less than 2 percent. George Washington Elementary, a comparison school, steadily declined in enrollment since MLK became an Edison contract school. George Washington's change in enrollment from 529 students in 1995/1996 to 475 students in 1998/1999 showed an opposite trend from the total enrollment for the Mt. Clemens School District, which increased from 3,198 students during the 1996/1997 school year to 3,448 students during the 1998/1999 school year. While Edison reports that student mobility was minimal, we found quite the opposite when we looked at the high attrition rate from the cohorts in our longitudinal analysis. Mislevy (1998, p.11) also noted in his study of reading levels that attrition had "reduced sample sizes for matched comparisons from about 60 per grade down to about 20-25, leading to a loss of about half of the original study's power for test of statistical significance."

The student-teacher ratio for the 1996/1997 school year was 15.4 to 1 at MLK, lower than the 21.4 to 1 at George Washington; however, by the 1998/1999 school year, the student-teacher ratio at MLK increased to 36.5 students per teacher, rising above the ratio of 22.6 students per teacher at George Washington. The average student-teacher ratio for the Mt. Clemens School District remained steady at about 20 students per teacher from the 1996/1997 school year to the 1998/1999 school year. The percent of minorities for the 1998/1999 school year at MLK was 49.4, percent while at George Washington it was 35.6 percent.

The 558 students enrolled in 1997/1998 represented a mix of ethnic backgrounds: 44.3 percent were African American, 0.9 percent Asian/Pacific, 0.7 percent Hispanic, and 53.3 percent Caucasian. In addition, enrolled students had various compensatory needs with slightly more than 6 percent of the students receiving special education services and 28.9 percent qualifying for free/reduced lunch

(Edison, 1999). The percentage of students in the state of Michigan who were eligible for free or reduced price lunch remained at about 31 percent over the years between 1995 and 1999. The eligibility for students in the Mt. Clemens School District, however, increased from 39.7 percent during the 1996/1997 school year to 51.6 percent during the 1998/1999 school year. The eligibility for free or reduced price lunch for students at George Washington was much greater. During the 1995/1996 school year, 52 percent of the students were eligible, rising to 61.9 percent during the 1998/1999 school year. The percentage of eligible students at MLK has fluctuated over the years from a high of 59.3 percent during the 1994/1995 school year, decreasing to a low of 39.9 percent during the 1997/1998 school year, and increasing back to 59.3 percent in 1998/1999.

The total expenditures per pupil for the Mt. Clemens School District increased from \$7,824 during the 1996/1997 school year to \$8,045 the following year. For 1996/1997 expenditures per pupil in the state of Michigan were \$6,507. Total expenditures at George Washington decreased from \$5,862 in 1995/1996 to \$4,827 in 1997/1998. In contrast, the total expenditures for MLK increased from \$4,499 during the 1994/1995 school year (pre-Edison) to \$5,789 the following year (Edison) and steadily increasing to \$6,078 for the 1997/1998 school year.

The average teacher salary in the state of Michigan increased from \$46,570 during the 1996/1997 school year to \$47,009 the following year. Mt. Clemens School District's average teacher salary increased from \$45,554 in 1996/1997 to \$47,777 the following year. Similarly, the average teacher salary at George Washington increased from \$46,472 during the 1996/1997 school year to \$65,330 in the 1997/1998 school year. MLK had a significantly lower average teacher salary during the 1996/1997 school year at \$36,632, but it increased to \$56,264 in 1997/1998.

9.2 Past Studies and Evaluation and Available Data for Analysis

The 1995/1996 Mislevy study of King-Edison vs. Control reported “moderately large significant effects in favor of King-Edison at the kindergarten level; mixed significant and non-significant effects favoring King-Edison at first grade; and no significant differences at second grade” (Mislevy, 1996; pp. 6-7). Results from the 1996/1997 study indicated both groups making similar gains, except for that year's second graders at King-Edison who made significantly higher gains than the previous year's same grade cohort and same year control cohort. The 1997/1998 study reported, “Year-to-year comparisons of cohorts at the same point in their schooling career show increases over time in both programs, more strongly within King” (Mislevy, 1998, p. 11). Because of a high rate of attrition in study participants, the 1997/98 study suggested that this particular study be discontinued. It recommended that a new study be started with kindergarten students entering in the next two to three school years.

In its second annual report on student performance (Edison, 1999), Edison categorized the gains in this school as “Strongly Positive” (a five star rating). The following statements from its second annual report sum up Edison's findings (Edison, 1999, P. 22).

- MEAP, the state criterion-referenced test, shows steady improvement in the scores of successive cohorts of different students in all subjects, except fifth-grade writing, which fell statewide due to changes in scoring standards.
- MAT-7 shows substantial gains in reading and math achievement of successive cohorts of second graders.
- ITBS shows clear net gains for one- and two-year cohorts in all subjects, with the minor exception of grade five language during the school's first year.
- Primary reading scores from the Mislavy studies show both cohorts—students who entered school in 1995 in kindergarten and in first grade—ahead of matched control groups after three years of schooling.
- Relative achievement gains are difficult to measure in this district, except by the controlled reading study. The MEAP tests different groups of students every year, subjecting trends to the ups and downs of different annual cohorts. The only other elementary school in the district has seen its MEAP scores fall from above King's to below and now above again, while King's scores have risen steadily. Differences in students year to year at these campuses render comparisons meaningless. The school has had very low levels of student mobility.

Data available for our analyses

Edison provided us with an SPSS data set containing the Iowa Test of Basic Skills (ITBS) test data for three years (1997-1999). In its second annual report, Edison reported on ITBS data for the years 1995, 1996, and 1997. Unfortunately, the data set we were provided did not contain data for these earlier years. Edison also reported data on the Metropolitan Achievement Test (MAT-7), although this was for successive cohorts, indicating that it did not have individual student data on this test. Like all Michigan public schools, this school is also required to take part in the state assessment program (MEAP). We secured MEAP test results for the years 1995-2000 from the Michigan Department of Education.

Tests administered include the Michigan Educational Assessment Program (MEAP) in reading and math for grade 4 in 1995-2000 and in science and writing for grade 5 in 1996–2000; Metropolitan Achievement Test, 7th Edition (MAT-7) in 1996–1998; the Iowa Test of Basic Skills (ITBS) for grade 3 in 1996 and grade 4 in 1997; and the Mislavy Reading Study in 1995/1996 for grades K-2, in 1996/1997 for grades 1-3, and in 1997/1998 for grades 2-4.

9.3 Longitudinal Analysis of Individual Student Data

Edison Schools Inc. provided individual student test results only for the Iowa Test of Basic Skills (ITBS). ITBS scores were provided in four separate scales: grade equivalent scores (GE), standard scores (SS), national percentile rank scores (PR), and normal curve equivalent scores (NCE). Parallel analyses are reported for each scale.

Data records included 562 students for the 1997-1999 academic years, although we were able to trace only one cohort of students representing a small portion of students over 2 or more years. For example, utilizing ITBS language GE scores in 1997, there were 72 possible third graders to follow. By 1998, 82 percent remained (59/72 pair-wise analysis), and by 1999 only 67 percent of these students remained for the longitudinal trend analysis (48/72). This represents a 33 percent drop in students over the three years in the cohort of students that progressed from grades 3 to grade 5 between the 1996/97 school year and the 1998/99 school year. Table 9:1 presents sample size information for each ITBS score scale for each subtest by year and grade level. Also depicted in this table are the sample sizes for the 1 year gain analysis on this and the longitudinal analyses.

Table 9:1 Table of Sample Sizes for Individual Student Data by Grade and Year on the ITBS

	Grade 3								
	1997			1998			1999		
	Language	Reading	Mathematics	Language	Reading	Mathematics	Language	Reading	Mathematics
GE	72	78	73	77	77	74	127	131	125
SS	72	79	73	77	77	74	127	131	125
PR	72	79	73	77	77	74	127	131	125
NCE	72	79	73	77	77	74	127	131	125
	Grade 4								
	1997			1998			1999		
	Language	Reading	Mathematics	Language	Reading	Mathematics	Language	Reading	Mathematics
GE	86	87	85	84	85	81	105	105	107
SS	86	87	85	84	85	81	105	105	107
PR	86	87	85	84	85	81	105	105	107
NCE	86	87	85	84	85	81	105	105	107
	Grade 5								
	1997			1998			1999		
	Language	Reading	Mathematics	Language	Reading	Mathematics	Language	Reading	Mathematics
GE	84	85	85	85	85	83	97	97	96
SS	84	85	85	85	85	83	97	97	96
PR	84	85	85	85	85	83	97	97	96
NCE	84	85	85	85	85	83	97	97	96
Sample Sizes for 1 Year Gains									
	3rd to 4th grade								
	1997/98			1998/99					
	Language	Reading	Mathematics	Language	Reading	Mathematics			
GE	59	63	57	67	67	66			
SS	59	64	57	68	67	66			
PR	59	64	57	68	67	66			
NCE	59	64	57	68	67	66			
ITBS Subtest	4th to 5th grade								
	1997/98			1998/99					
	Language	Reading	Mathematics	Language	Reading	Mathematics			
GE	71	71	68	66	67	63			
SS	71	71	68	66	67	63			
PR	71	71	68	66	67	63			
NCE	71	71	68	66	67	63			

Sample Size for Longitudinal Trend Analysis			
ITBS Subtest	1997 to 1998 to 1999		
	Language	Reading	Mathematics
GE	48	45	50
SS	48	45	51
PR	48	45	51
NCE	48	45	51

Longitudinal analysis findings

Repeated measures ANOVAs were examined for longitudinal trends over a three-year period for MLK. Outcome data were individual student data on the ITBS mathematics, reading, and language skill subtests. Parallel analyses are reported for all types of scores reported by Edison, e.g., grade equivalent (GE), standard score (SS), percentile rank or national percentile rank (PR), and the normal curve equivalent (NCE) score. In all models, the assumption of sphericity was evaluated, and if found to be violated, the Huynh-Feldt adjusted p-values are reported. If the overall linear model was found to be statistically significant, unadjusted (alpha) pair-wise comparisons were examined to identify where a difference in the means might be located. The tables and charts in Exhibit 9:1 illustrate the results in terms of these four score scales.

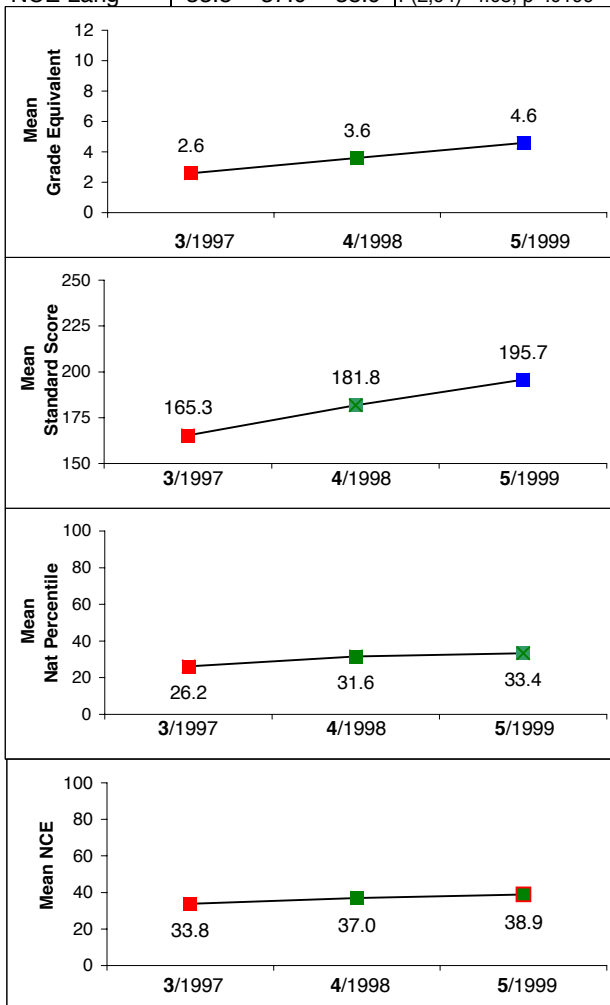
In terms of GE scores, the students are expected to progress at least one grade level per year. As seen in the results, over the two years the students progressed 2.3 grade levels in math, 2.0 grades in language, and 2.2 years on average in reading. In each subject analysis there was a statistically significant gain each year relative to the preceding year.

In terms of scaled scores on the ITBS, the students' scores increased significantly each year in each of the three subjects. Because these gains were statistically significant, one can be certain that the students were moving ahead. However, as seen in the mean PR and NCE scale scores, these students were only making marginal advances relative to the national norms. In fact, the PR score scale evidenced a statistically significant gain in year 2 (1998) from year 1 (1997) that failed to keep gaining in year 3 (1999) in all three subject areas. Regarding the NCE score scale on the language test, the only statistically significant gain was over the three years, from 1997 to 1999. The mathematics and reading test analyses for NCE scores showed large and statistically significant gains from 1997 to 1998 but no significant change between 1998 and 1999.

Exhibit 9:1 Results From the Analysis of Individual Student Results on the Iowa Test of Basic Skills Martin Luther King Jr. Academy (Mt. Clemens, Michigan)

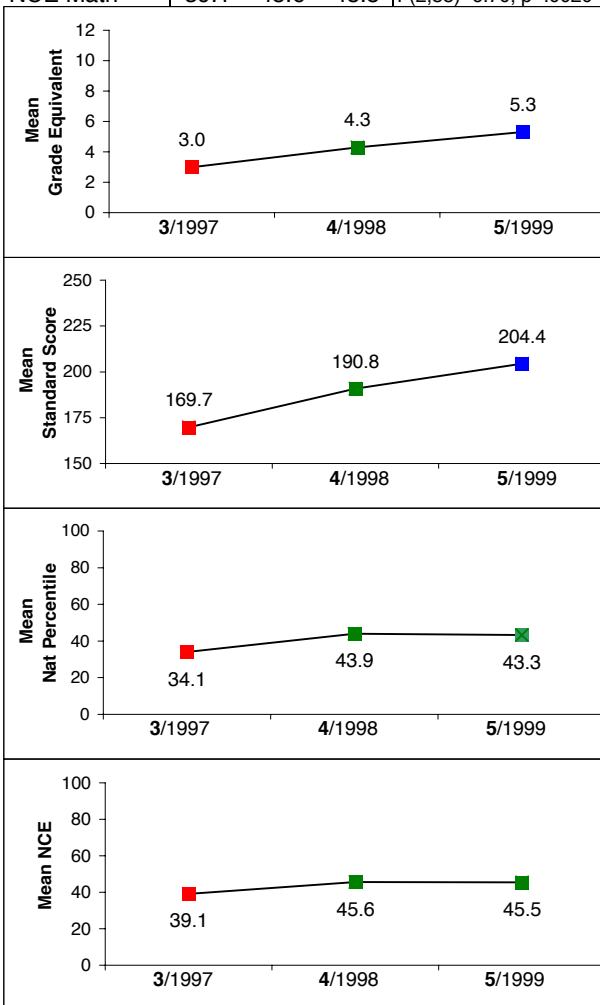
LANGUAGE

n=48	1997	1998	1999	
Grade	3rd	4th	5th	
GE Lang	2.6	3.6	4.6	$F(2,94)=91.73, p<.0001$
SS Lang	165.3	181.8	195.7	$F(2,94)=104.75, p<.0001$
NP Lang	26.2	31.6	33.4	$F(2,94)=5.50, p=.0055$
NCE Lang	33.8	37.0	38.9	$F(2,94)=4.08, p=.0199$



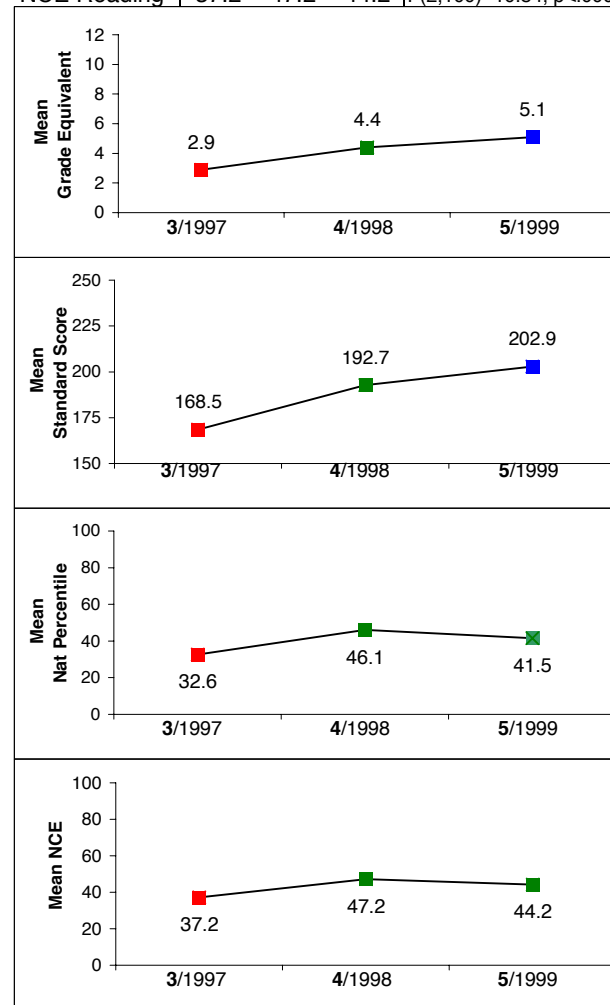
MATH

n=45	1997	1998	1999	
Grade	3rd	4th	5th	
GE Math	3.0	4.3	5.3	$F(2,88)=85.34, p<.0001$
SS Math	169.7	190.8	204.4	$F(2,88)=110.55, p<.0001$
NP Math	34.1	43.9	43.3	$F(2,88)=7.13, p=.0014$
NCE Math	39.1	45.6	45.5	$F(2,88)=6.70, p=.0020$



READING

n=50	1997	1998	1999	
Grade	3rd	4th	5th	
GE Reading	2.9	4.4	5.1	$F(2,98)=97.91, p<.0001$
SS Reading	168.5	192.7	202.9	$F(2,100)=106.88, p<.0001$
NP Reading	32.6	46.1	41.5	$F(2,100)=11.79, p<.0001$
NCE Reading	37.2	47.2	44.2	$F(2,100)=10.84, p<.0001$



A color change represents a statistically significant change in the means. A trend with two color changes represents differences among all three means. A dual colored charting point (red and green) represents a statistically significant difference between one mean but not the other. The reader is encouraged to examine the individual table of means for these cases.

9.4 Chi-Square Analysis of MEAP Data

A chi-square analysis was initiated on data available from the state of Michigan on the outcomes of the Michigan Educational Assessment Program (MEAP), the state-mandated criterion referenced test. The MEAP is administered in grades 4 (mathematics and reading) and grade 5 (science, writing and, in 1999, social studies, a new component). In grade 4, mathematics and reading components of the MEAP are scored along an ordinal three-category scale: Satisfactory, Moderate, and Low. Similarly, in grade 5 the science component of the MEAP is scored with the same scale, although different labels are used: Proficient, Novice, and Not yet novice. The fifth grade MEAP writing component had only two categories of data results: Proficient and Not Yet Novice. In 1999 social studies was added to the MEAP and was scored along a 4-point continuum: Level 1 (highest) to Level 4 (lowest). Additional information on the MEAP is provided in Appendix A.

Construction of the comparison groups

We constructed two different comparison groups for the chi-square analyses. Since we were interested in examining the number/proportion of students who meet state standards (“passing”) or conversely the number/proportion of students within MLK who do not meet state standards (“failing”) on the MEAP, we needed to define a suitable comparison group. Our first comparison was with George Washington School, which has similar demographic characteristics as MLK. In fact, this is the only other public elementary school in the district for which there are MEAP test results available. We did not make district comparisons, because George Washington School essentially is the rest of the district for the test data that are available for grades 4 and 5.

The second comparison group we selected was the state passing/failing rates. While the state demographics vary from MLK Academy and the Mt. Clemens School District, we think that comparisons with state averages can yield further information regarding the relative gains of the Edison school. Also, since Edison claims that advances in other district schools is in part due to its presence, we use the state as a more distant point of comparison that cannot be easily influenced by the presence of Edison schools.

General procedure

Utilizing published data from the state of Michigan, we made yearly comparisons (consecutive cohorts) at grade 4 from 1995 through 1999 and from 1996 to 1999 in grade 5 for each subject component of the MEAP test administered within a specific grade level. It is important to recall that 1995 was the first year the school was operated by Edison so data from this year should be viewed as baseline.

Percentage data (students in each scoring category) were converted to raw frequency data prior to chi-square analysis. To insure independence of the rows in the chi-square tables, the raw frequencies for each scoring category of the MEAP in the state comparisons were down-weighted by subtracting the number of students in that category from MLK. Thus, the state numbers reflect all students in

the state exclusive of those at MLK. The down weighting was not necessary for the George Washington comparisons.

Four chi-square analyses were evaluated for each subtest nested within year and grade level. Two of these analyses were on uncollapsed data; that is, all scoring levels were represented in the contingency table (e.g., a 2x3) for the George Washington comparison and the state comparisons. Note that the uncollapsed fifth grade social studies analyses constituted a 2x4 contingency table. Two follow-up analyses were conducted on the data after collapsing the multilevel scoring into a dichotomy (pass, fail), thus producing 2x2 contingency tables. According to the Michigan Department of Education, a score in the “satisfactory” category constitutes “passing” or meeting the state standard for that particular grade and subject. On the other hand, the “moderate” and “low” categories refer to “slightly below the state standard” and “not well prepared,” respectively. Students who have scores in the moderate and low categories have not met state standards and fall into the “fail” category in our 2x2 chi-square and odds-ratio analyses.

There was one exception to this general procedure, since only two categories of results were available for the grade 5 writing MEAP component (i.e., “proficient” and “not proficient”); thus, only the 2x2 analyses could be presented.

Chi-square findings

The chi-square analyses are testing the null hypothesis that the relative frequency (of students) in the three (or two) scoring categories are the same for MLK and the comparison group (either George Washington School or the state).

Results of the chi-square analyses for grade 4 covered the MEAP administration years from 1995 to 1999 for two subject tests: mathematics and reading. Individual contingency tables are presented in Appendix E.

The first set of comparisons were made against George Washington as the comparison group for mathematics. Ten separate chi-square statistics were evaluated, five from a 2x3 contingency table and five from a 2x2 contingency table (see Table 9:2). Results indicated that only in 1995 was there a statistically significant difference among the scoring categories’ proportions between the two schools among the five 2x3 analyses such that there were fewer students in the “satisfactory” category at MLK. This pattern was also replicated in the collapsed analyses (2x2). In these five chi-square analyses, the only statistically significant difference between the two schools’ passing/failing rates was in 1995 when a significantly higher number of students “failed” this MEAP subject test in MLK. In parallel comparisons against the state, we observed statistically significant differences among the category proportions in 1998 and 1999 in the 2x3 tables and parallel findings in the 2x2 tables. Once again, there was a statistically significant tendency for MLK students to do poorer than the students across the state.

This pattern of statistical significance differs when the MEAP reading subtest is examined. Only one statistically significant difference was noted when the comparison was with George Washington:

in the 1998 2x3 analysis. However, this finding was not replicated in the 2x2 comparison. When the comparison was made against the state, differences were noted in both the 2x3 and 2x2 tables in 1995 and 1999, such that in 1995 the results favored MLK but reversed in 1999.

Table 9:2 Summary of Chi-Square Findings for MLK, Grade 4

	1995	1996	1997	1998	1999
4 th Grade Math					
MLK vs. GW	sig/sig	ns/ns	ns/ns	ns/ns	ns/ns
MLK vs. State	ns/ns	ns/ns	ns/ns	sig/sig	sig/sig
4 th Grade Reading					
MLK vs. GW	ns/ns	ns/ns	sig/ns	ns/ns	ns/ns
MLK vs. State	sig/sig	ns/ns	ns/ns	ns/ns	sig/sig

Note: Each result cell in the matrix is divided with the results for the 2x3 analysis on the left-hand side and the results for 2x2 analysis on the right-hand side (i.e., 2x3/2x2)

For the results from the Grade 5 science chi-square analyses with George Washington as the comparison group, eight separate chi-square statistics were evaluated: four 2x3 contingency tables and four 2x2 contingency tables. Results indicated (see Table 9:3) that in 1997 there were no statistically significant differences among the scoring categories' proportions between the two schools among the four 2x3 analyses and only one difference in the collapsed 2x2 tables. In this year, fewer MLK students scored in the "satisfactory" category. In the parallel state comparison, we observed statistically significant differences among the category proportions in three of the four years: 1997, 1998, and 1999 in both the 2x3 and 2x2 tables. The general trend was for fewer students at MLK to score in the "satisfactory" category.

Due to the limited availability of data, we were able to conduct "collapsed" analyses only for the fifth grade MEAP writing component. Here the pattern of statistical significance observed with the George Washington comparison more or less paralleled the state findings. Once again, the general trend was for a larger proportion of MLK students to "fail" relative to students in George Washington or their state peers.

In 1999 a new component was added to the fifth grade MEAP: social studies. This component of the MEAP was scored along a four-point continuum, with Level 1 representing the highest level of performance and Level 4 representing the lowest level of performance. A statistically significant difference in proportions was observed relative to George Washington students but not the state for both uncollapsed (2x4) and collapsed (2x2) analyses. It should be noted that the Michigan Department of Education did not offer information on passing/failing for this test. We consequently (and somewhat arbitrarily) defined Level 1 as "passing" and Levels 2-4 as "failing" in the collapsed analyses.

Table 9:3 Summaries of the Chi-Square Findings for MLK, Grade 5

		1996	1997	1998	1999
5 th Grade Science					
	MLK vs. GW	ns/ns	ns/sig	ns/ns	ns/ns
	MLK vs. State	ns/ns	sig/sig	sig/sig	sig/sig
5 th Grade Writing					
	MLK vs. GW	ns	ns	sig	sig
	MLK vs. State	sig	ns	sig	sig
5 th Grade Social Studies					
	MLK vs. GW	–	–	–	sig/sig
	MLK vs. State	–	–	–	ns/ns

9.5 Odds Ratio Analysis of the MEAP Data

One of the many possible statistics that can be derived from a 2x2 contingency table is the odds ratio (OR) statistic and corresponding 1- α confidence interval (CI). As presented in Section 2.4 of this report, the 2x2 tables analyzed in the previous section can be thought of as representing consecutive class cohorts in a prospective design. From a classical epidemiological perspective, the students in the Edison school can be thought of as the “exposed” group—that is, exposed to the “Edison-effect”—and students in the comparison group as the unexposed group. From this perspective each yearly comparison is a new cohort; measured over a period of years there are consecutive class cohorts. There is a minimal possibility for cohort contamination if a number of students in one group are not promoted to the next grade level. However, we think this represents a very small number of possible cases and therefore has minimal impact on the validity of these analyses. Section 2.4 details the OR statistic and corresponding 1- α confidence interval. We calculated and charted OR for each of the 2x2 tables constructed from the chi-square analyses presented above. Exhibit 9:2 presents these findings. It is important to recall that 1995 was the first year the school was operated by Edison so data from this year should be viewed as baseline.

Odds ratio findings, grade 4

Comparison against George Washington School. In grade 4 mathematics, the OR for MLK shows relative stability in magnitude against George Washington. Only in 1995 does the 95 percent CI not include 1.00, indicating that the OR of 2.593 is statistically significant: if you were a student at MLK there is about 2 ½ times greater odds (chance) for failing or not meeting the state standard for the math component of the MEAP in that year. This increased risk rapidly evaporates in the following years through 1999, so that in these subsequent years students at MLK were at no greater risk (odds) for failing the MEAP math component. Moreover, the Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the five years was not statistically significant, indicating there was no real (statistically significant) change in the OR over the five years. Thus, although there was a noted increase in odds of failing for MLK students in 1995, this was not enough

to overcome the relative stability of the OR in the following years. The common OR for the five years is 1.226, and the 95 percent CI is from 0.935 to 1.606. Since the CI includes 1.00, the variation in OR was most likely due to sampling variability and no real difference in the failing rate on the grade 4 math test between MLK and George Washington.

The grade 4 reading component of the MEAP presented a marginally different picture. All the CI around the ORs included 1.00 and thus are considered not statistically significant. Likewise, the Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the five years was not statistically significant, indicating there was no real (statistically significant) change in the OR over the five years. The common OR for the six years was 1.116, and the 95 percent CI is from 0.850 to 1.466. Thus, the variation observed in the ORs was most likely due to sampling variability and no real difference in the failing rate between MLK and George Washington.

Comparison against state. In grade 4 mathematics, the OR for MLK showed changing odds for failing the mathematics component of the MEAP over the 5 years in comparison with the state data. A curvilinear trend in OR is noticed in the chart such that from 1995 to 1997 there was a decreasing trend in the OR (although not statistically significant) that reversed an increase in odds of failing in 1998 and 1999, which is also statistically significant. In 1998 students were about 2 times more likely to fail relative to the state, and in 1999 they were about 2.25 times more likely to fail. The Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the 5 years was not statistically significant, indicating there was no real (statistically significant) change in the OR over the 5 years. Thus, the common OR over the 5 years was 1.656 and the 95 percent CI was from 1.355 to 2.024, which would be considered statistically significant since the CI does not include 1.00. Thus, over the five year period the odds were that MLK students were about 1.6 times more likely to fail the MEAP mathematics component relative to students in the rest of the state.

In grade 4 reading, the OR for MLK showed generally increasing odds for failing that component of the MEAP over the five years in comparison with the state data. In 1995 MLK students started out showing a statistically significant protective effect, so that MLK students were less than half as likely to fail the MEAP reading component. However, in 1996 and thereafter, there was a noted increase in odds of failing for MLK students that became statistically significant in 1999. The Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the five years was statistically significant, indicating a statistically significant change in the OR over the five years and thus no common OR could be meaningfully interpreted. That is, the year-by-year ORs were meaningful and could not be represented by one overall OR.

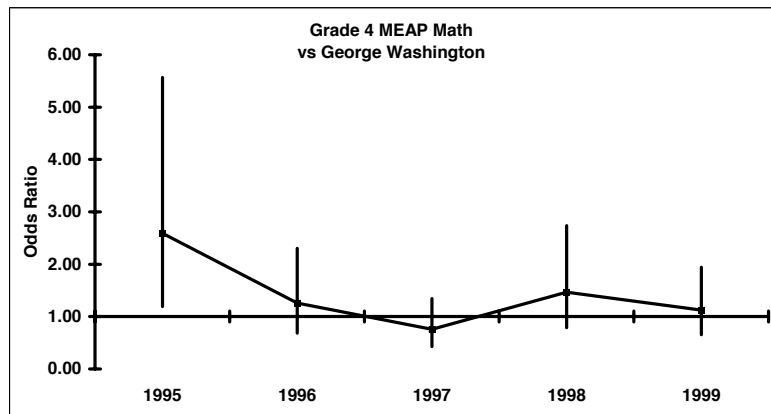
Exhibit 9:2 Grade 4 Odds Ratio Results When Compared with the District and State

Grade 4 MEAP Math vs George Washington

Year	U CL	L CI	OR
1995	5.564	1.189	2.593
1996	2.304	0.684	1.256
1997	1.346	0.430	0.760
1998	2.736	0.786	1.467
1999	1.941	0.653	1.126

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (4, N=896) = 6.680, p = .154

Common OR = 1.226
U CL = 1.606
L CL = 0.935



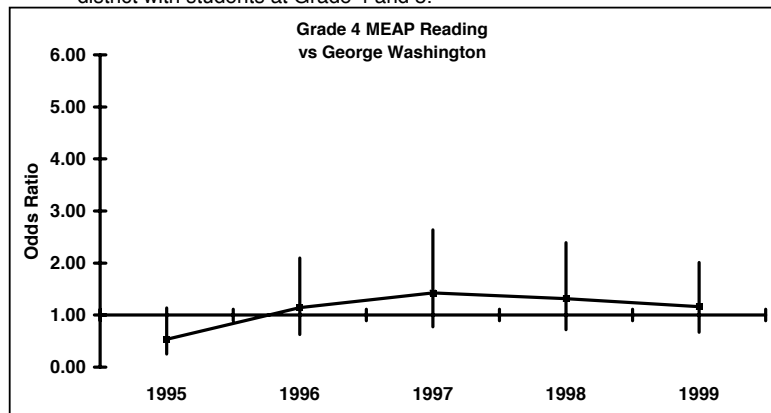
Note: George Washington Elementary is the only other Elementary School in the district with students at Grade 4 and 5.

Grade 4 MEAP Reading vs George Washington

Year	U CL	L CI	OR
1995	1.140	0.251	0.535
1996	2.098	0.624	1.144
1997	2.636	0.774	1.429
1998	2.392	0.721	1.313
1999	2.012	0.671	1.162

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (4, N=859) = 4.607, p = .330

Common OR = 1.116
U CL = 1.466
L CL = 0.850

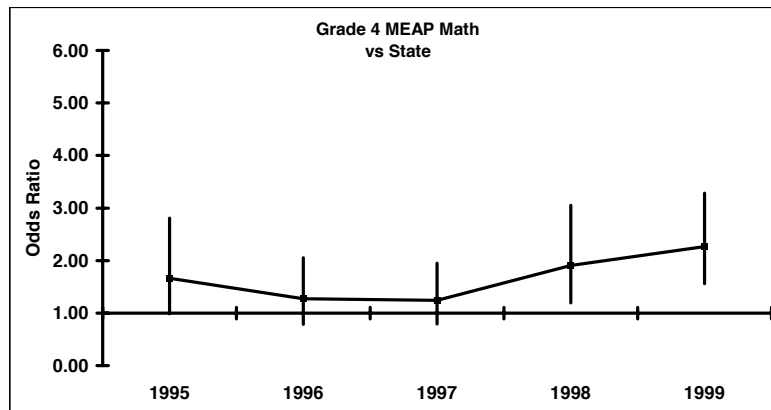


Grade 4 MEAP Math vs State

Year	U CL	L CI	OR
1995	2.808	0.986	1.664
1996	2.054	0.787	1.272
1997	1.948	0.790	1.240
1998	3.053	1.192	1.908
1999	3.282	1.562	2.264

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (4, N=578,820) = 5.685, p = .224

Common OR = 1.656
U CL = 2.024
L CL = 1.355



Grade 4 MEAP Reading vs State

Year	U CL	L CI	OR
1995	0.812	0.280	0.477
1996	2.456	0.936	1.516
1997	1.788	0.726	1.139
1998	2.388	0.941	1.499
1999	3.132	1.469	2.145

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (4, N=577,510) = 21.773, p < .001

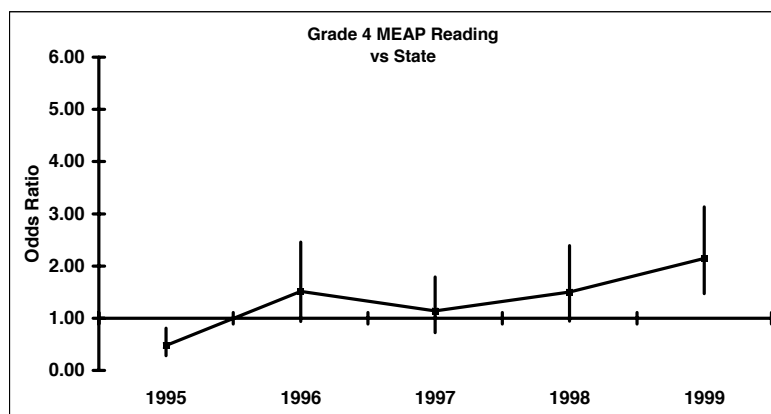
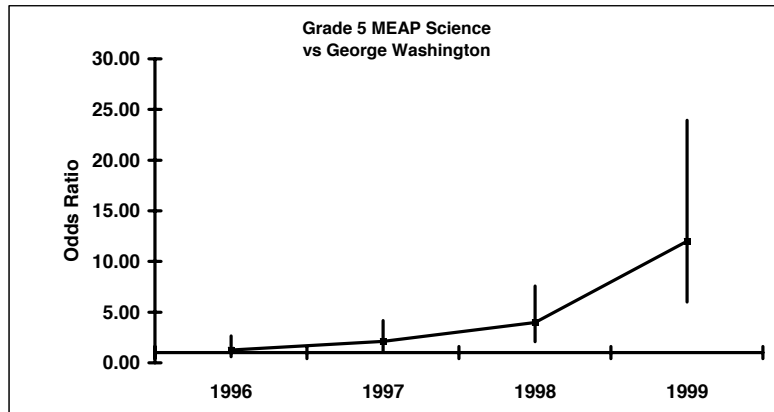


Exhibit 9:3 Grade 5 Odds Ratio Results When Compared with the District and State

Grade 5 MEAP Science vs George Washington

Year	U CL	L CI	OR
1996	2.661	0.585	1.248
1997	4.175	1.079	2.122
1998	7.570	2.099	3.986
1999	23.931	6.012	11.995

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (3, N=780) = 20.495, $p < .001$

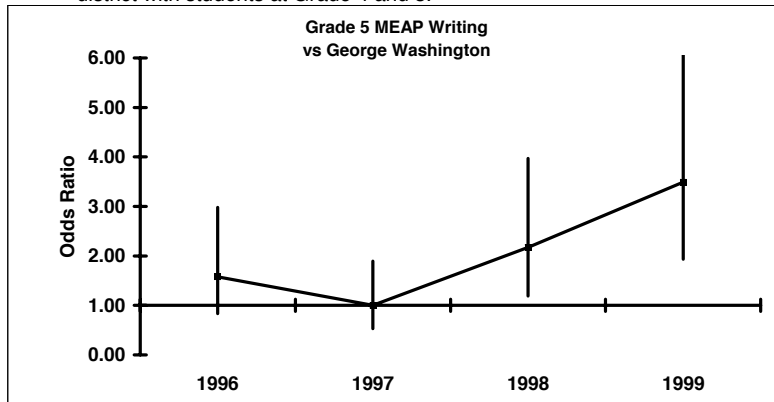


Note: George Washington Elementary is the only other Elementary School in the district with students at Grade 4 and 5.

Grade 5 MEAP Writing vs George Washington

Year	U CL	L CI	OR
1996	2.981	0.833	1.576
1997	1.897	0.532	1.005
1998	3.970	1.189	2.170
1999	6.310	1.934	3.494

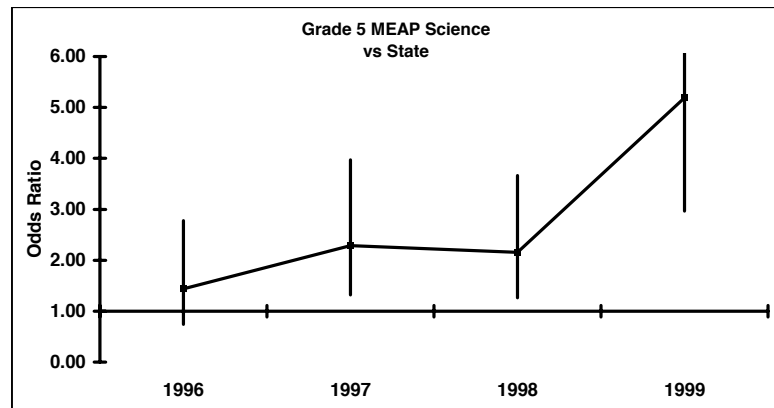
Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (3, N=771) = 8.434, $p = .038$



Grade 5 MEAP Science vs State

Year	U CL	L CI	OR
1996	2.776	0.743	1.436
1997	3.968	1.318	2.287
1998	3.663	1.264	2.152
1999	9.074	2.966	5.188

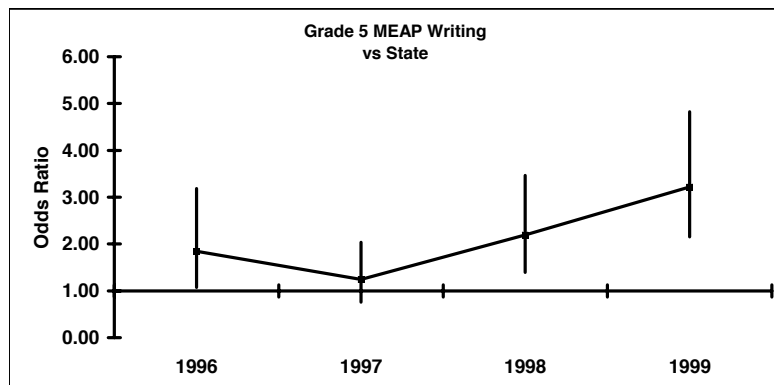
Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (3, N=458,782) = 8.775, $p = .032$



Grade 5 MEAP Writing vs State

Year	U CL	L CI	OR
1996	3.186	1.070	1.846
1997	2.035	0.761	1.245
1998	3.465	1.392	2.196
1999	4.826	2.152	3.222

Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (3, N=456,133) = 8.627, $p = .035$



Odds ratio findings, grade 5

Comparison against George Washington. In grade 5, the OR for MLK showed an increasing trend toward greater odds of failure on the science component of the MEAP relative to students at George Washington. The ORs started out close to 1.00 and nonsignificant, but gradually increased in 1997 and 1998. Both years represent a statistically significant increase in odds for failure, and in 1999 there was a rather large jump in the OR, such that MLK students were almost 12 times more likely to fail the fifth grade MEAP science component. Not surprisingly, the Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the 4 years was statistically significant, indicating a statistically significant change in the OR over the 4 years. Thus, no common OR could be meaningfully interpreted, so interpretation focuses on the OR corresponding to each year.

The grade 5 writing component of the MEAP presented an encouraging picture that unfortunately reversed in 1999. Although not statistically significant, from 1996 through 1998 there was a developing protective effect for being a student at MLK relative to George Washington. This protective trend, however, was replaced in 1999 with a dramatic and statistically significant increase in odds for failure. In 1999, MLK students were almost 3.5 times more likely to fail, whereas in 1998, they were about a third less likely to fail. The Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the 4 years was statistically significant, indicating there was a statistically significant change in the OR over the 4 years. Thus, no common OR could be meaningfully interpreted.

Although the MEAP social studies component was administered in 1999, we did not include these analyses in the OR analyses since the state had not defined passing and failing levels and because only one year of data was available. The results for the social studies test is included in Exhibit 9:8 for MLK, the district, and the state.

Comparison against state. In grade 5, the OR for MLK showed an increasing trend toward greater odds of failure on the science component of the MEAP relative to students across the state. Only in 1995 was the OR close to 1.00 and nonsignificant. In all succeeding years the ORs were statistically greater than 1.00 and generally increasing. Not surprisingly, the Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the 4 years was statistically significant, indicating a statistically significant change in the OR over the 4 years. Thus, no common OR could be meaningfully interpreted.

In grade 5 writing, the OR for MLK showed a curvilinear trend with the OR moving from a statistically significant increase in odds for failure in 1996 to essentially equal odds of failure in 1997 back to an increase in failure odds in 1998 and 1999. The Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the six years was statistically significant, indicating a statistically significant change in the OR over the six years. Thus, no common OR could be meaningfully interpreted. Although the MEAP social studies component was administered in 1999, we did not include these analyses in the OR analyses since the state had not defined passing and failing levels.

9.6 Overall Performance on the MEAP Assessments

Exhibits 9:4 to 9:8 illustrate the performance trends for Martin Luther King Academy, George Washington Elementary, and the state of Michigan. We used Washington Elementary as a comparison group, since this is the only other school in the district with students in grades 4 and 5. Therefore, the population of students taking the MEAP tests at Washington represents all non-Edison students in the district taking the grade 4 and 5 MEAP tests.

MLK made large gains in math and reading, approaching state performance levels in math, while Washington went down during the last two years. Exhibit 9:4 contains the results for grade 4 math, and Exhibit 9:5 contains the grade 4 reading results.

In grade 5 science Washington Elementary gained more than the Edison school. In fact, MLK students dropped sharply on the science component of the MEAP in 1998-99, but made a big gain in the 1999-00 school year; however, MLK still did not match gains made by the district (see Exhibit 9:6). The grade 5 writing test only had two performance levels, Proficient or Not Yet Proficient. Gains made by students at Washington Elementary were more consistent, but the students at MLK showed overall larger gains. As with the grade 5 science test, the MLK students dropped sharply on the science component of the MEAP in 1998-99, but made a big gain in the 1999-00 school year; however, MLK still did not match gains made by the district (see exhibit 9:7)

Exhibit 9:8 contains the results on the MEAP social studies test. This was administered for the first time in the 1998-99 school year, so we had two years of data to chart. In terms of the proportion of students meeting or exceeding state standards, MLK performed lower than district and state levels in both years this test was administered. Nevertheless, the proportion of students scoring in the highest category (i.e., exceeding state standards) was higher at MLK than at the district and state in 1999-00.

The blue components of the bar charts indicate the proportion of students meeting or exceeding state standards, while the yellow parts of bar charts indicates the proportion of students not meeting state standards. The blue and yellow demarcation indicates the proportions used when we conducted the odds ratio. The results from the odds ratio analysis are presented in the previous section.

Exhibit 9:4 Performance on Grade 4 Math for Edison, District, and State

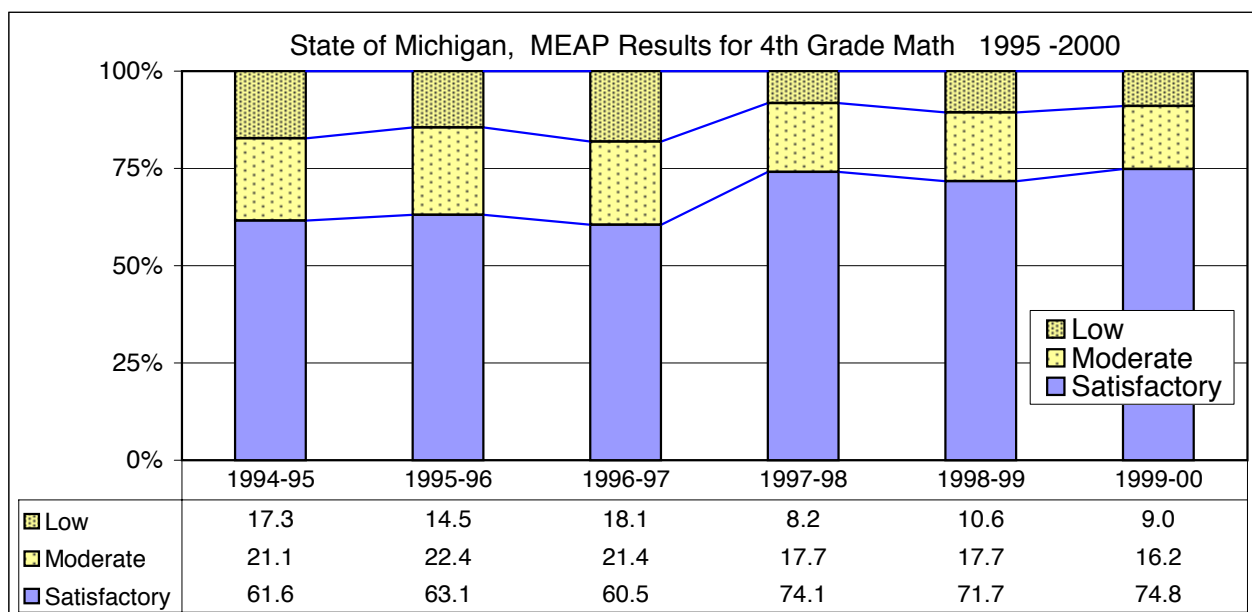
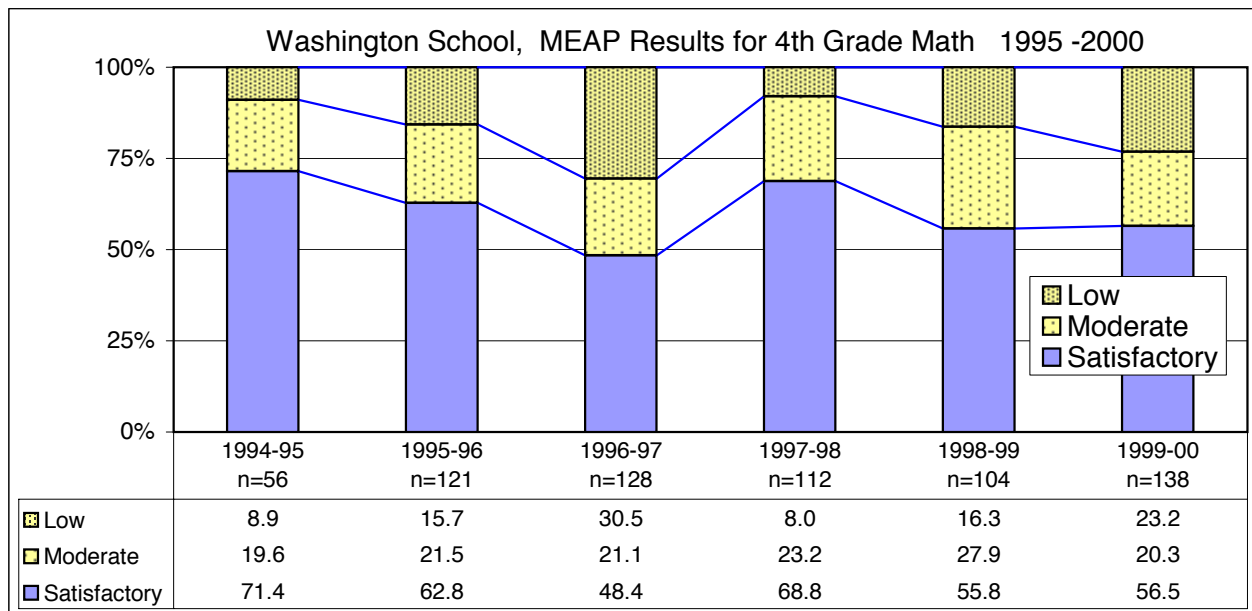
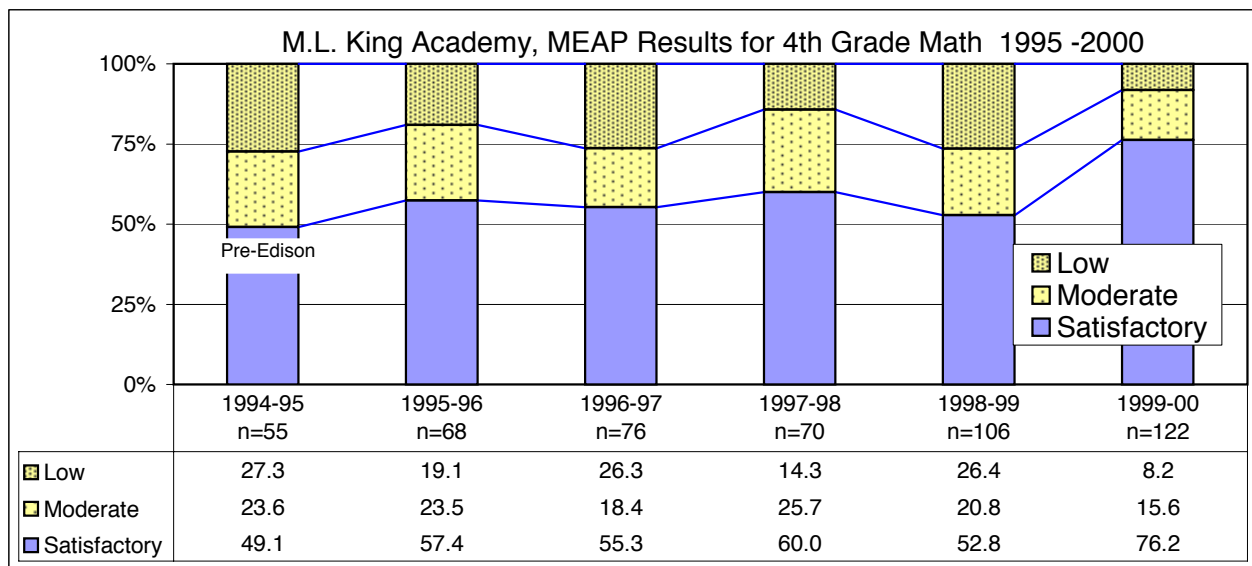


Exhibit 9:5 Performance on Grade 4 Reading for Edison, District, and State

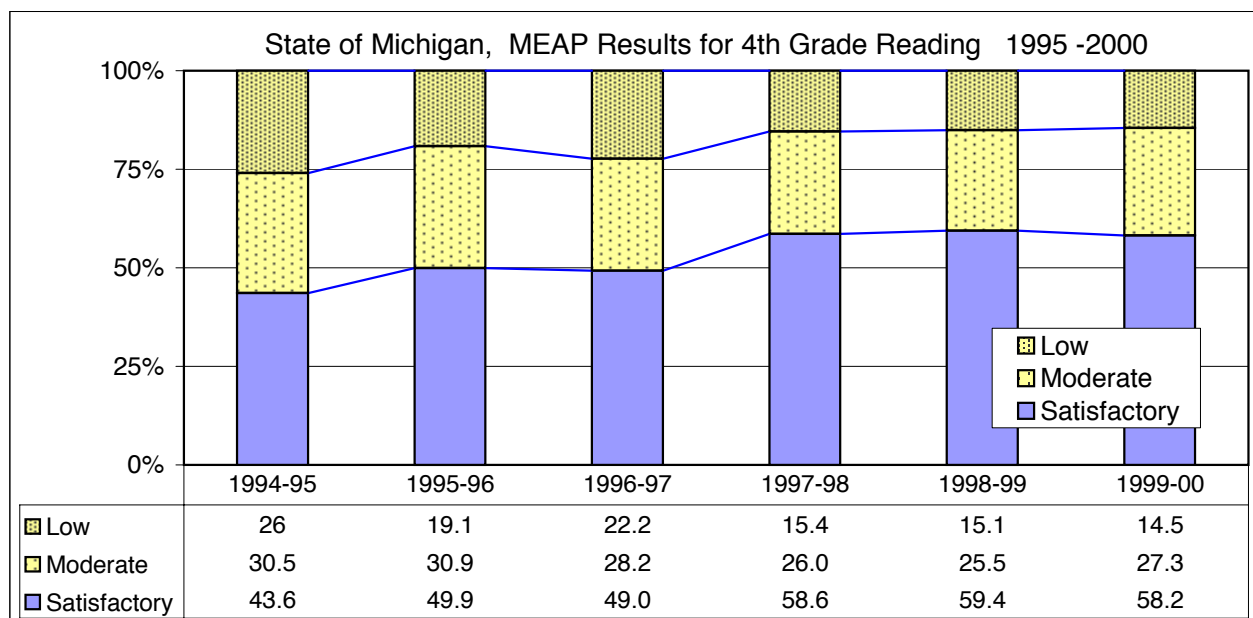
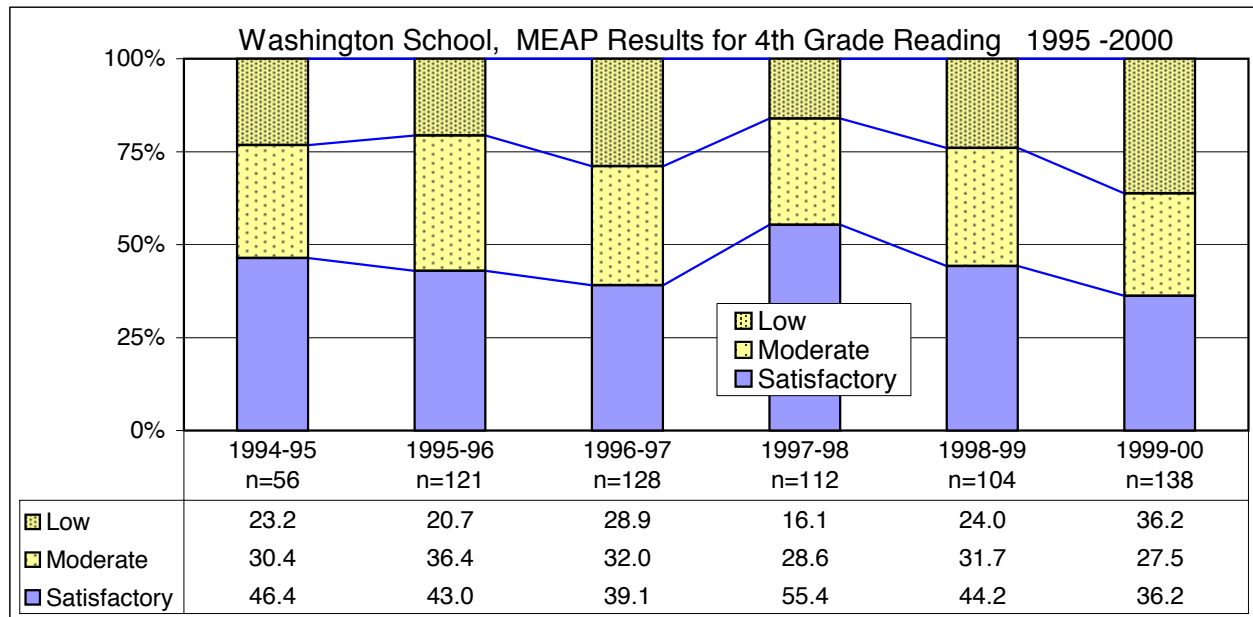
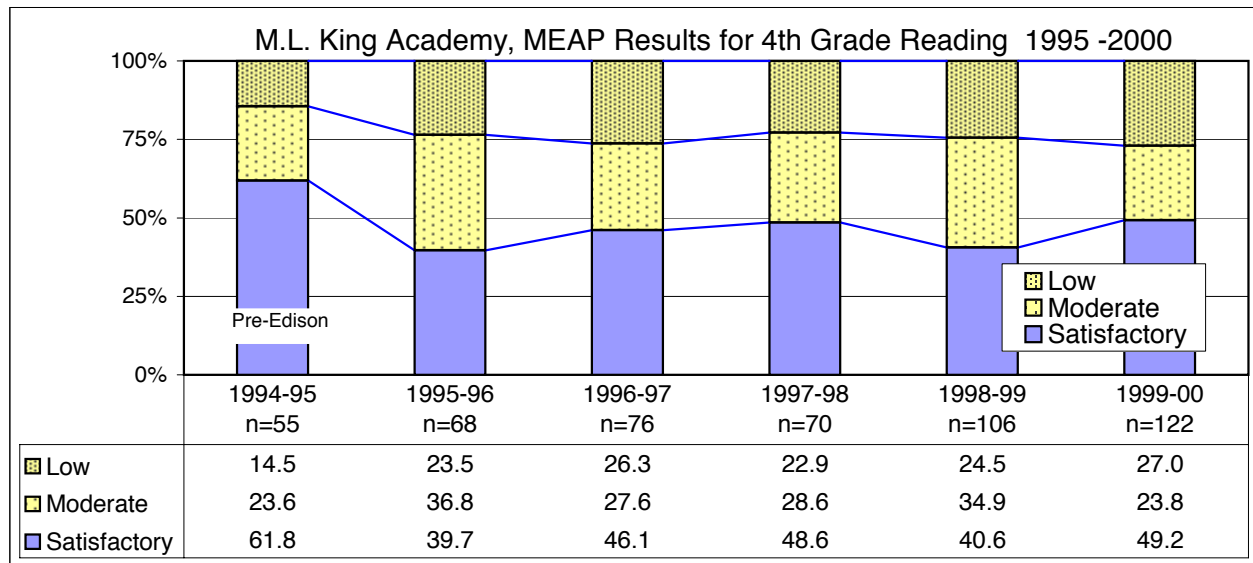


Exhibit 9:6 Performance on Grade 5 Science for Edison, District, and State

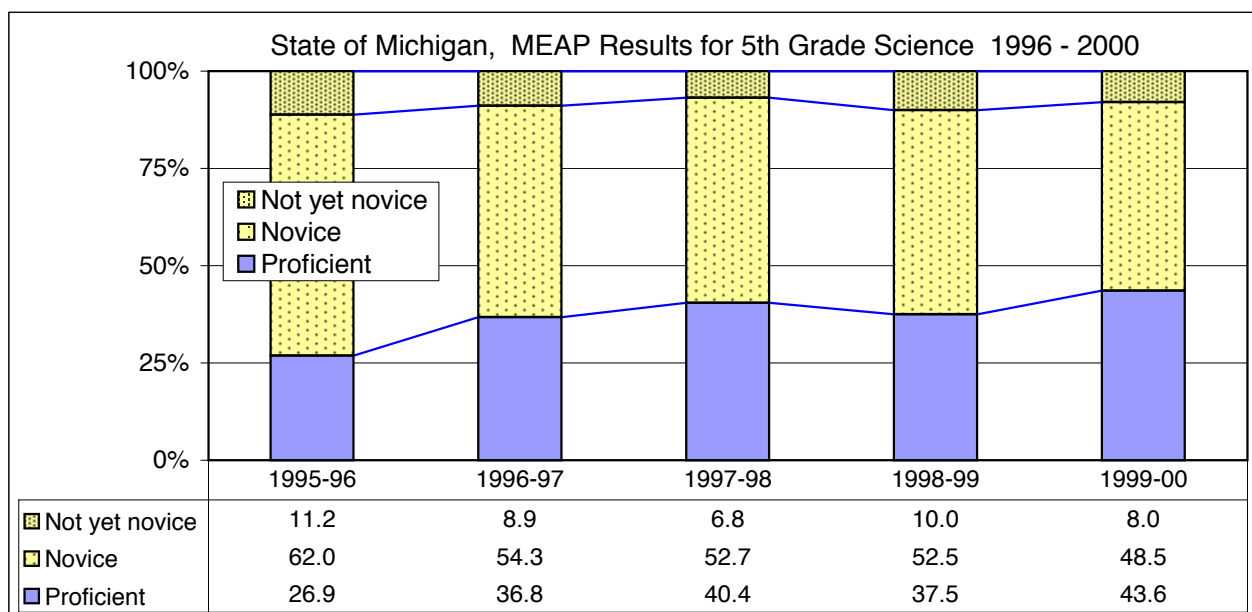
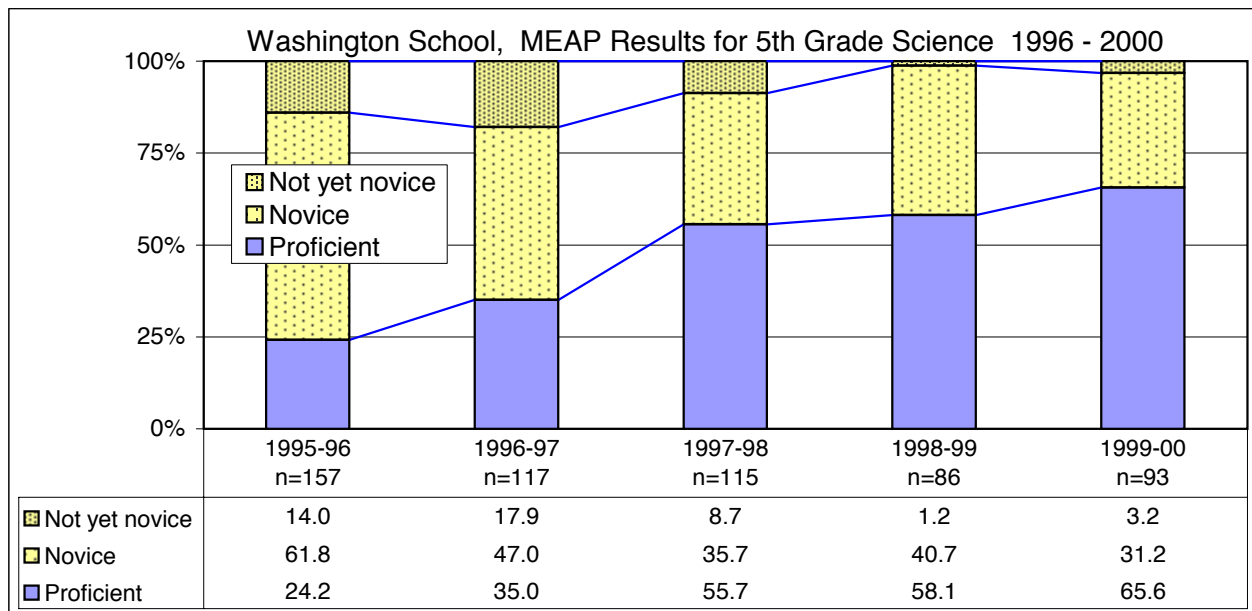
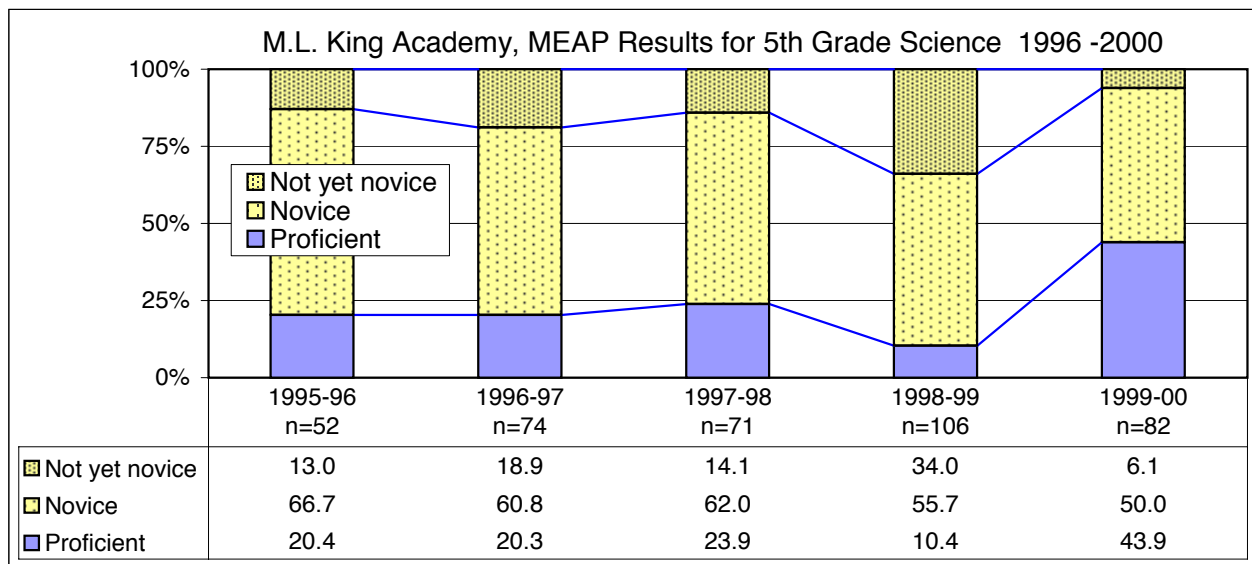


Exhibit 9:7 Performance on Grade 5 Writing for Edison, District, and State

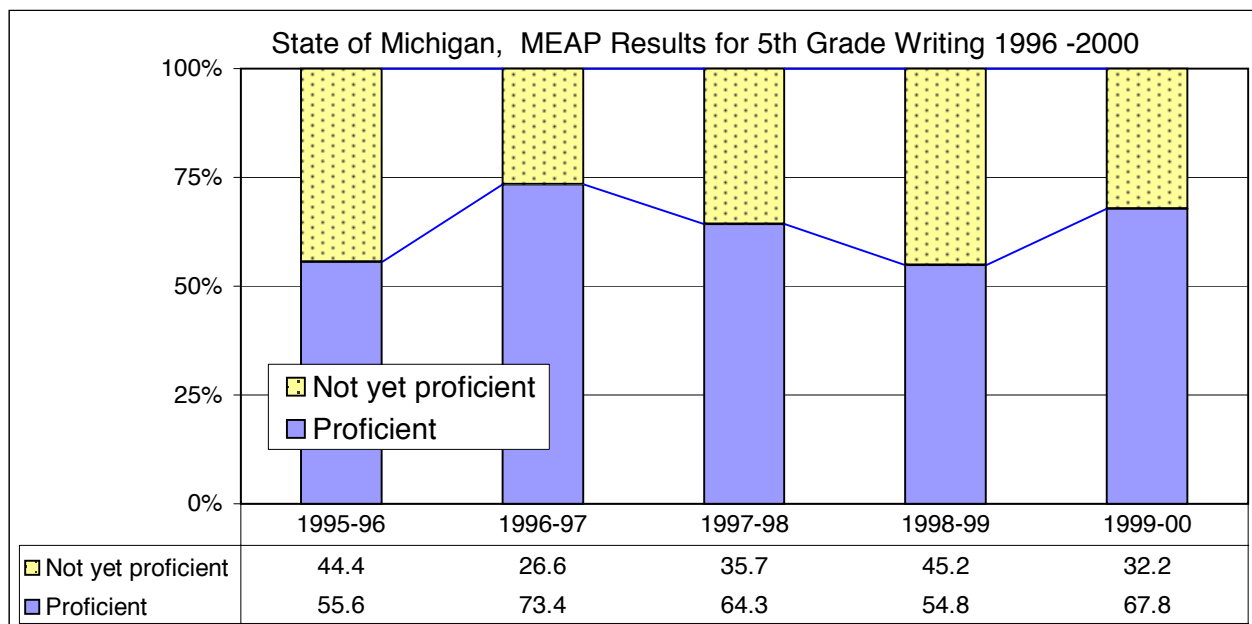
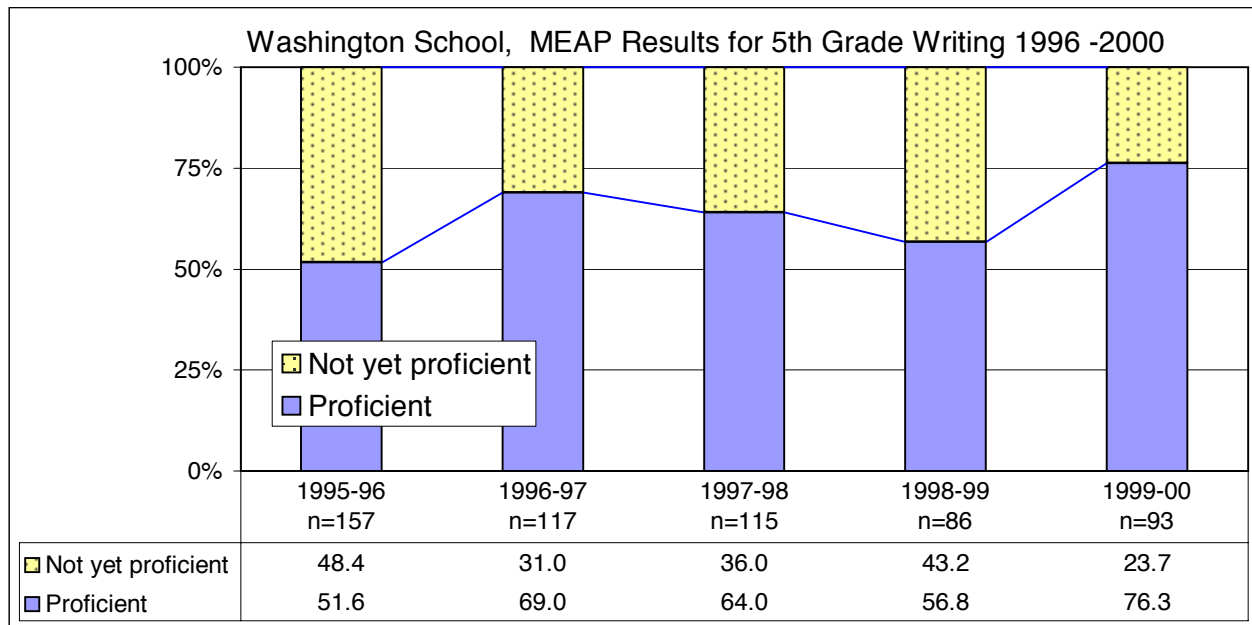
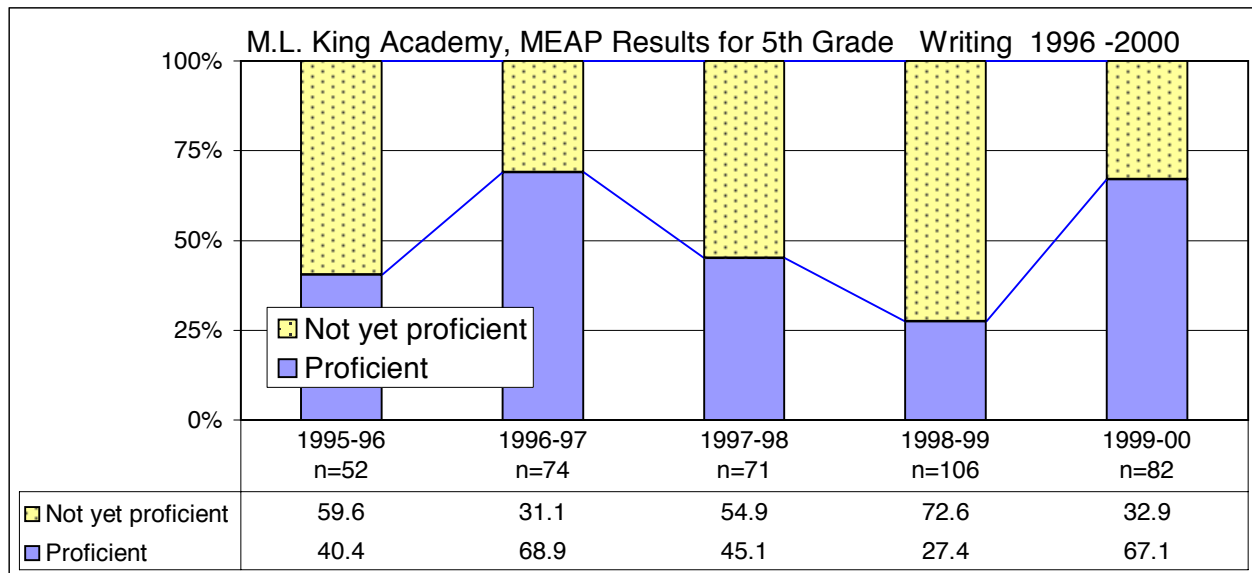
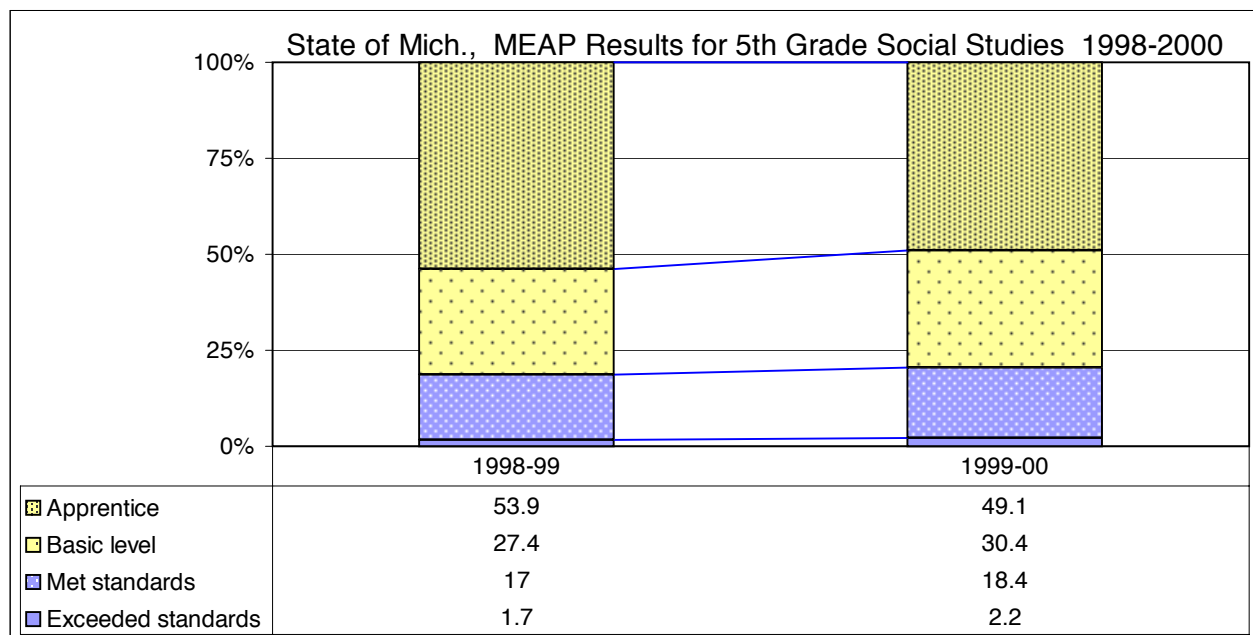
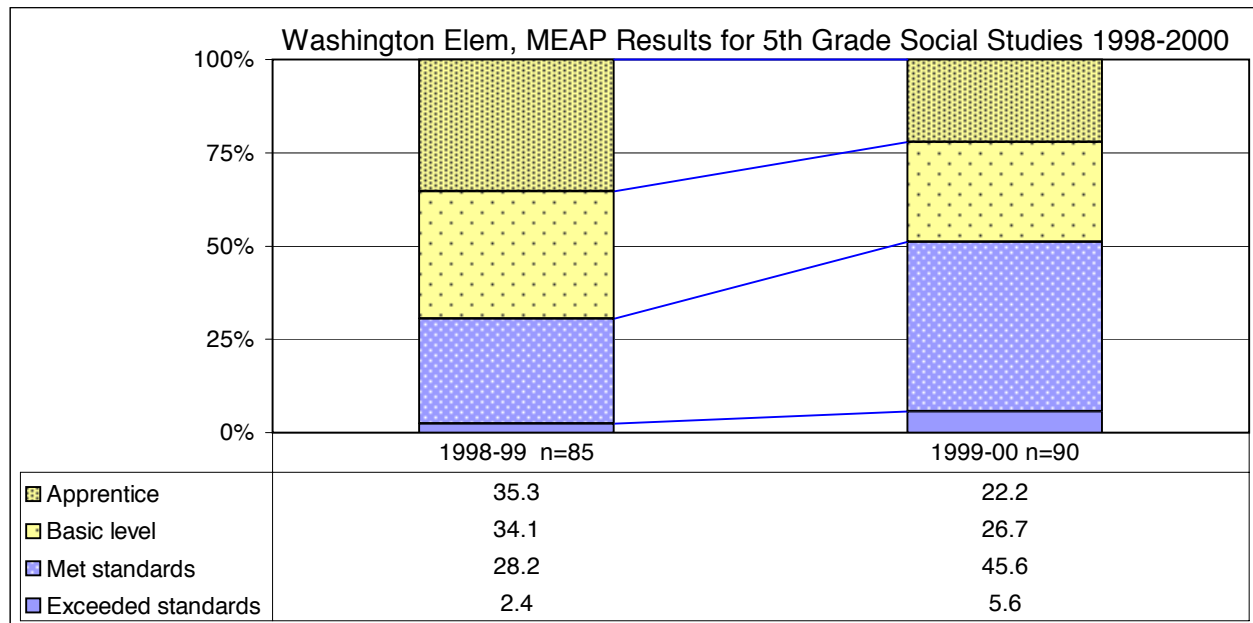
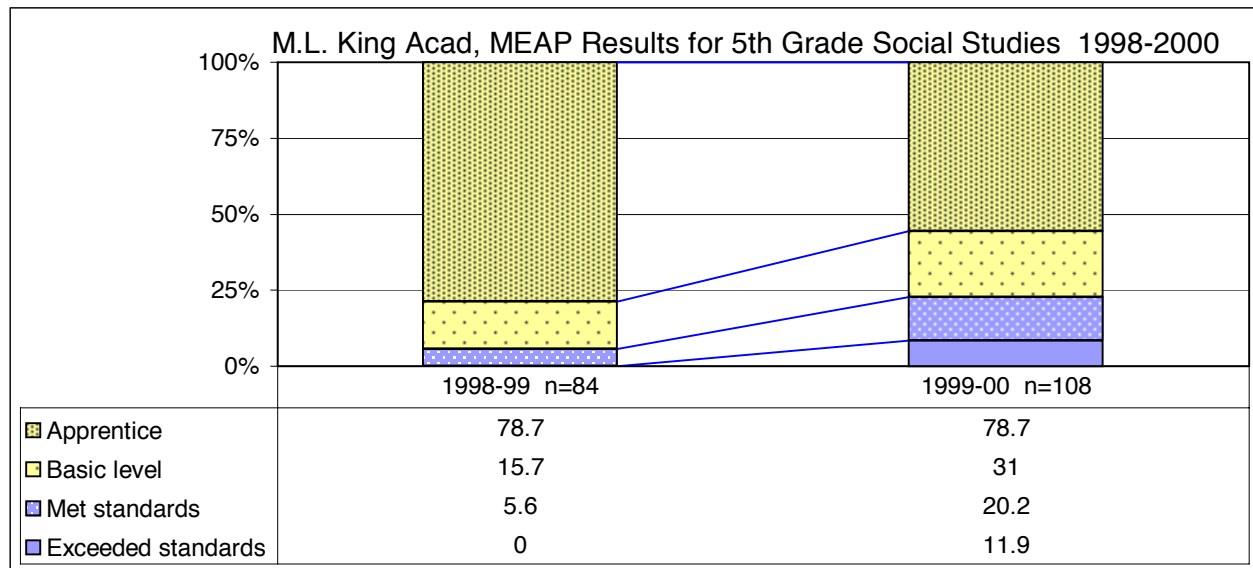


Exhibit 9:8 Performance on Grade 5 Social Studies for Edison, District, and State



9.7 Summary

Norm-referenced test findings

While Edison maintains that the gains at this school are Strongly Positive, based on our own analysis we found that the school has Negative results. While some gains could be seen in the results from the Iowa Test of Basic Skills, they were not sustained over two years. There were gains in terms of normal curve equivalents (NCE) in the first year and then minor decreases in the NCE in the second year of the longitudinal analysis. One exception to this was in language, where the cohort of students made gradual gains over two years, with the NCE ranking in 1998/99 being statistically significant from the NCE in the 1996/97 school year. According to the results on the ITBS, the students gained the equivalent of 2.0, 2.2, and 2.3 academic years over two calendar years in language, reading, and math, respectively. This is certainly a promising result. One possible factor affecting these gains was the sharp drop in the percent of students taking the test.

Edison claimed that MLK also made large gains in most subjects measured by the Iowa Test of Basic Skills between 1995 and 1997. Since the data provided to us by Edison was only for 1997-1999, we could not verify these findings. Table 9:4 summarizes our analysis of individual student achievement data on the ITBS. A summary score of -1 indicates a result that is unfavorable toward the sample school, a score of 0 indicates a neutral finding, and a score of 1 indicates a favorable result according to the criteria specified in Section 2.5.

Table 9:4 Summary of Results on Norm-Referenced Student Achievement Tests

Grades 3, 4, and 5 (1997-99)	Std Score		GE			PR			NCE			Trend
	p-value	ES	p-value	Δ	ES	p-value	Δ	ES	p-value	Δ	ES	
Language	<.0001	.590	<.0001	2.0	.559	.0055	7.2	.059	.0199	5.1 (0)	.041	mixed (0)
Math	<.0001	.619	<.0001	2.3	.555	.0014	9.2	.083	.0020	6.4 (0)	.078	mixed (0)
Reading	<.0001	.581	<.0001	2.2	.565	<.0001	8.9	.124	<.0001	7.0 (0)	.114	mixed (0)

In its second annual report, Edison claimed that students were also making great gains that were measured by the Metropolitan Achievement Test. No MAT results were reported for this school in Edison's 2000 annual school performance report. Unfortunately, we could not confirm or refute this claim, since we did not receive these data from Edison.

Criterion-referenced test findings

Results from the state assessment were not promising, however. The odds of not meeting state standards on the subject tests at MLK were very good. The odds of failing or not meeting state standards remained largely the same over time in comparison with the state total and with George Washington Elementary, the only other school in the district with students enrolled in grades 4-5, (see Table 9:5).

Decisions regarding the OR were based on whether or not the $(1-\alpha)$ C.I. included 1.0. If the $(1-\alpha)$ C.I. fell completely below 1.0, this was interpreted as a protective odds ratio (1) and thus favored the Edison School. If the $(1-\alpha)$ C.I. included 1.0 (0), this was interpreted as an equal odds situation. If the $(1-\alpha)$ C.I. fell completely above 1.0 (-1), this was interpreted as an increase in odds for failing the state CRT relative to the comparison sample.

Interpretation of the OR tables: If the Breslow-Day statistic (B-D) is nonsignificant, one overall OR and $(1-\alpha)$ C.I. can be used to represent the odds for failing the CRT relative to the comparison group. Thus, there are no trends reported for each specific year in the tables, only a rating in the B-D column. If the B-D statistic is found to be statistically significant, then an overall common OR cannot be meaningfully interpreted; that is, there is a statistically significant change in the OR over years and thus yearly OR are necessary. Thus, our summary ratings appear for each year of data and not in the B-D column.

Table 9:5 Summary Results on Criterion-Referenced Tests

MLK vs. G. Washington	1995	1996	1997	1998	1999	B-D	Trend
Grade 4 Math						0	mixed (0)
Grade 4 Reading						0	mixed (0)
Grade 5 Science		0	-1	-1	-1		negative (-1)
Grade 5 Writing		0	0	-1	-1		negative (-1)

Note: All comparisons against the state were negative.

Combined ratings

Because we were able to trace individual students in our longitudinal analysis of the ITBS results, this is clearly a stronger design than our analysis of the MEAP results, which was based on consecutive groups of students. Nevertheless, the very limited number of students we could include in the longitudinal analysis undermined the strength of this design. While we received data records for 562 students between 1997-1999 (this presumably relates to all third, fourth, and fifth graders taking the test during the 1998/99 school year), our results are based on sample sizes of 45, 48, and 50 for reading, language, and mathematics, respectively. In language, there was a 33 percent drop in students over the 3 years in the cohort of students that progressed from grades 3 to 5 between the 1996-97 school year and the 1998-99 school year. Mislevy (1998), in his earlier research on reading at this school, also ran into this problem when his sample sizes for matched comparisons were cut from 60 per grade down to 20-25 per grade. There is clearly a need for a closer examination of the reasons/explanations for large groups of students not taking part in the testing. This information would help us weigh the relative strength of contradictory results when they appear.

Table 9:6 Combined Overall Trends for Martin Luther King Jr. Academy

	Positive	Mixed	Negative
Norm Referenced	0 of 3	3 of 3	0 of 3
Criterion Referenced	0 of 4	2 of 4	2 of 4
TOTALS	0 of 7	5 of 7	2 of 7

Based on our analyses, we rate this school to be Negative with a mean rating of -0.29. Both the ITBS and the MEAP results indicated that students at MLK were below national and district norms, and this is essentially how the school performed before Edison took over operation. In its 1999 annual report, Edison rated this school as Strongly Positive. In its 2000 annual report it rated the 1999-2000 school year as Strongly Positive and the achievement gains since opening as Strongly Positive also.

Chapter Ten

Mt. Clemens Secondary Academies

Mt. Clemens, Michigan

10.1 Descriptive Summary of School

Edison began operating schools in Mt. Clemens in 1995. Mt. Clemens is a small Title I district that lies just north of Detroit. It is largely a low income community, although there are some pockets of expensive housing as well as a developed business section in the community. Martin Luther King Jr. Academy was the first school in the district to be operated by Edison. The following year, Edison began operating a junior academy within the district middle school. The junior and senior academies, referred to as the Mt. Clemens Secondary Academies, are the focus of this case study. The academies are essentially “schools within schools” since they are housed in the local secondary education complex. Nevertheless, the Edison-run academies have their own administration and operational budgets. Edison reported that the total enrollment in the secondary academies totaled 481 for the 1999-00 school year. These schools are operated by Edison under contract with the school district.

These two Edison academies may be considered as two separate schools since the Michigan Department of Education has assigned different building codes for them. During the first few years of operation, however, there were no separate building codes designated for the Edison academies and their middle and high school counterparts in the district. Therefore, it was impossible to disaggregate the test data as well as much of the school data for the first few years these academies were in operation. In its last annual report, Edison (1999) reported on these two academies as a separate school/case entitled “Mt. Clemens Secondary Academies.” In this report, we will also consider the two academies as one entity, although the test data we report will be grade specific.

The Mt. Clemens Secondary Academies have a combined staff of approximately 40 persons including teachers and administrators. In its second annual report, Edison (1999, p. 32) listed the following information about the academies for the 1997-98 school year: student mobility was 4.7 percent; the student/staff ratio was 15.1:1; total enrollment was 603, of which 31.7 percent were African Americans, 1.8 percent were Asian/Pacific, 2.3 percent were Hispanic, and 62.9 percent were Caucasian. According to data presented in Edison’s annual reports, the proportion of students receiving special education has dropped in recent years (11.4 percent in 1997-98 and 7.9 percent in 1999-00). The proportion of students qualifying for free or reduced lunches has dropped even more sharply (27 percent in 1997-98 down to 12.5 percent in 1999-00).

The total enrollment for the school district has steadily increased from 3,132 students during the 1994/1995 school year to 3,448 students during the 1998/1999 school year. Enrollments in the

secondary schools, both district- and Edison-operated, have fluctuated extensively during the same period, however. Enrollment at the academies is reflected in the decline of total students enrolled at Mt. Clemens Middle and High Schools. Though in its second year of operation, the Junior Academy enrollment declined from 425 (1997/1998) students to 375 (1998/1999) students. In contrast, the Senior Academy increased from 127 (1997/1998) students to 172 (1998/1999) students.

The student-teacher ratio for the Junior Academy went from 17.7 students per teacher during the 1997/1998 school year to 15 students per teacher the following year, while the student-teacher ratio for the Senior Academy increased from 12.7 students during the 1997/1998 school year to 21.5 students. These figures, collected from the Michigan Department of Education (MDE) K-12 Database, are close, but not identical to the figures reported by Edison (1999). The changes in student-teacher ratios may reflect the decrease in enrollments at the junior academy and the increase in enrollments at the senior academy, if few or no new teachers were hired or let go for that year (1998/1999). The average student-teacher ratios for Mt. Clemens Middle and High Schools were higher than their Edison counterparts.

Expenditures per pupil at the Junior Academy during the 1997/1998 school year were \$5,556, which was higher than the other middle school in the district, which spent \$5,193 per pupil in the same year. Nevertheless, Edison's senior academy spent far less than the district high school: \$5,242 compared with \$7,299. This may be because the district high school has vocational programs that are more expensive than other academic programs.

Table 10:1 Descriptive Information for Mt. Clemens Secondary Academies and District Middle and High Schools (1994/95 - 1999/2000)

		Pre-Edison				
		Grades	1994-95	1995-96	1996-97	1997-98 1998-99
Enrollment						
Mt. Clemens Junior Acad. (Edison)	6-8	-	-	-	425	375
Mt. Clemens Senior Acad. (Edison)	9-11	-	-	-	127	172
Mt. Clemens Middle School	6-8	642	-	689	354	340
Mt. Clemens High School	9-12	739	1,344	1,395	578	536
Mt. Clemens Comm. School District	K-12	3,132	3,162	3,198	3,378	3,448
Pupil/Teacher Ratio						
Mt. Clemens Junior Acad. (Edison)		-	-	-	17.7	15.0
Mt. Clemens Senior Acad. (Edison)		-	-	-	12.7	21.5
Mt. Clemens Middle School		24.7	-	17.6	15.7	18.9
Mt. Clemens High School		28.0	53.3	54.2	16.8	32.1
Mt. Clemens Comm. School District		24.4	25.2	20.4	20.6	19.8

According to the Edison Annual School Performance Report for the 1997/1998 school year, 27 percent of its students qualified for free and reduced lunch, which is lower than the district middle school but similar to district high school figures in the same year. Overall, the Edison academies have a lower proportion of their students qualifying for free or reduced lunches than the district average of 41 percent in 1997-98.

10.2 Data Available for Our Analyses

Tests administered at the Mt. Clemens academies include the Michigan Educational Assessment Program (MEAP) as well as the Iowa Test of Basic Skills (ITBS). The Iowa Test of Basic Skills was administered at the school, at least during the 1996-97 and 1997-98 school years, although the results from this norm-referenced test were not reported in Edison's second annual report (1999). We were unable to secure individual student achievement data on the ITBS from Edison for this school, although we did receive copies of the summaries of school results for 1996/97 that were supplied to the school by the test company. It was unclear whether or not the school summaries included both the Edison academy and the district middle school that share the building.

For the secondary academies, the MEAP has tests for grade 7 students in reading and math; for grade 8 students in science, writing, and social studies; and for grade 11 in reading, math, science, and writing. Unfortunately, we had difficulty securing data disaggregated from the Michigan Department of Education. We were able to identify partially complete MEAP data (percentage data only) for the Junior Academy for 1996-97 and 1997-98 since Edison reported disaggregated figures in its second annual report (Edison, 1999). Complete data for the grade 7 and 8 tests should have been available, but the results for the school were merged with the results for the district's middle school for 1996-97 and 1997-98 (the Edison school and the district middle school had the same building code until 1998-99). Separate building codes were used in 1999, so we could more easily separate the data between the two schools that occupied the same building. In addition to the partially complete 96/97 and 97/98 data, we were able to extract data from the 1998 through the 2000 MEAP assessments in grade 7 math and reading tests; and from the 1999 and 2000 grade 8 science, writing, and social studies tests. We secured one year of data for the MEAP High School Test (HST), which measures students' academic performance in grade 11, since the Edison senior academy did not enroll grade 11 students until 1998-99.

The grade 7 math and reading and the grade 8 science and writing tests are scored along a 3-point ordinal scale: Satisfactory, Moderate, and Low. The grade 8 science test is scored as Proficient, Novice, and Not Yet Novice, while the writing test is scored on a 2-point scale: Proficient or Not Yet Proficient. The grade 8 social studies test is scored on a 4-point ordinal scale: (1) exceeded standards, (2) met standards, (3) basic level, and (4) apprentice. The grade 11 test is scored as (1) exceeded standards, (2) met standards, (3) basic, and (4) not endorsed. Additional information on the MEAP is provided in Appendix A.

10.3 Chi-Square Analysis of MEAP Data

A chi-square analysis was initiated on data available from the Michigan Department of Education on the outcomes of the Michigan Educational Assessment Program (MEAP), which is the state-mandated criterion-referenced test. As previously discussed, Mt. Clemens Secondary Academies present a unique case since the two academies that comprise the school are actually housed within the district's middle and high school. During the first two years that the Junior Academy was operated by Edison, the test results were aggregated with the district's middle school since they shared the same building code assigned by the Michigan Department of Education until separate building codes were used in 1999. Thus, for the first two years, it was more difficult to disaggregate the data. We were, however, able to calculate the number of students taking the grade 7 MEAP Math and Reading tests for 1998-99 from district and Edison numbers. Thus, for the grade 7 MEAP we included three years of data: 1997/98, 1998/99 and 1999/00; for grade 8 we secured data for only two years: 1998/99 and 1999/00; and for the HST, grade 11, we had only one year of data: 1998-99.

Construction of the comparison groups

We constructed two different comparison groups for our chi-square analyses. Since we were interested in examining the number/proportion of students who met state standards ("passing") or conversely the number/proportion of students who did not meet state standards ("failing") on the MEAP within Mt. Clemens Secondary Academies, we needed to define a suitable comparison group. In the grades 7 and 8 chi-square analyses, our first comparison was with Mt. Clemens Middle School. This comparison was essentially equivalent to the district comparison in the other case studies, since this is the only other middle school in the district. The state performance constituted our second comparison group. In the HST analyses we utilized Mt. Clemens High School (district) as the comparison group. As was the case at the middle school level, Mt. Clemens High School is the only other high school in the district.

Because the state demographics vary from those of Mt. Clemens Secondary Academies and Mt. Clemens Middle School, we believe that comparisons with state averages can yield further information regarding the relative gains of the Edison school. Also, since Edison claims that advances in other district schools are—in part—due to its presence, we use the state as a more distant point of comparison that cannot be easily influenced by the presence of Edison's schools.

General procedure

Utilizing published data from the Michigan Department of Education (MDE), we made comparisons at grades 7, 8, and 11. Percentage data (students in each scoring category) were converted to raw frequency data prior to chi-square analysis. To insure independence of the rows in the chi-square tables, the raw frequencies for each scoring category of the MEAP in the state comparisons were down-weighted by subtracting the number of students in that category from Mt. Clemens Secondary Academies. Thus, the state numbers reflect all students in the state exclusive of those in Mt. Clemens Secondary Academies.

We constructed four chi-square analyses for each subtest nested within year and grade level. Two of these analyses were on uncollapsed data; that is, all scoring levels were represented in the contingency table (e.g., a 2x3) for the district and the state comparisons. Note that the uncollapsed grade 8 social studies analyses constituted a 2x4 contingency table. Two follow-up analyses were conducted on the data after collapsing the multilevel scoring into a dichotomy (pass, fail), thus producing 2x2 contingency tables. According to the Michigan Department of Education, a score in the “satisfactory” category constitutes “passing” or meeting the state standard for that particular grade and subject. On the other hand, the “moderate” and “low” categories refer to “slightly below the state standard” and “not well prepared,” respectively. Students who have scores in the moderate and low categories have not met state standards and fall into the “fail” category in our 2x2 chi-square and odds-ratio analyses.

Chi-square findings

Individual contingency tables for the chi-square analyses of the 1999 MEAP administration are presented in Appendix E. A summary of the chi-square results on the grade 7 MEAP math and reading subtests are presented in Table 10:2. Overall, there were statistically significant differences in the cell proportions in both the 2x3 and 2x2 tables such that there were lower proportions of students in the higher categories from the Junior Academy relative to both the district and state.

Relative to the district, Mt. Clemens Jr. Academy students significantly outperformed students from the district in both 1999 and 2000. However, Mt. Clemens students fell significantly below the performance level of the remainder of the state. In terms of performance on the reading test, Mt. Clemens students performed at comparable levels relative to the district but significantly below students in the rest of the state.

Table 10:2 Summary of Chi-Square Findings for
Mt. Clemens Academies, Grade 7

	1998	1999	2000
<i>Mathematics</i>			
Mt. Clemens Jr. Acad. vs. District	ns/ns	sig/sig	sig/sig
Mt. Clemens Jr. Acad vs. State	ns/ns	sig/sig	ns/ns
<i>Reading</i>			
Mt. Clemens Jr. Acad vs. District	ns/ns	ns/ns	ns/ns
Mt. Clemens Jr. Acad vs. State	sig/sig	sig/sig	sig/sig
Note: Each result cell in the matrix is divided, with the results for the 2x3 analysis on the left-hand side and the results for 2x2 analysis on the right-hand side.			

Table 10:3 Summary of Chi-Square Findings for
Mt. Clemens Academies, Grade 8

	1999	2000
<i>Science</i>		
Mt. Clemens Jr. Acad. vs. District	ns/ns	ns/ns
Mt. Clemens Jr. Acad. vs. State	sig/sig	sig/sig
<i>Writing</i>		
Mt. Clemens Jr. Acad. vs. District	--/ns	--/sig
Mt. Clemens Jr. Acad. vs. State	--/sig	--/sig
<i>Social Studies</i>		
Mt. Clemens Jr. Acad. vs. District	sig/ns	sig/sig
Mt. Clemens Jr. Acad. vs. State	sig/sig	sig/sig
Note: Each result cell in the matrix is divided with the results for the 2x3 analysis on the left-hand side (2x4 for social studies), and the results for 2x2 analysis on the right-hand side.		

Grade 8 chi-square analysis included the MEAP tests of science, writing, and social studies for the 1999 and 2000 assessments. We secured data for a 2x2 analysis of the writing results only. As can be seen from Table 10:3, on the MEAP science test, the Mt. Clemens Junior Academy students performed at a level comparable to students in the district, but again fell below students in the rest of the state. This was also true on the writing test, although in 2000, students from Mt. Clemens Jr. Academy also fell significantly below district students.

Regarding the MEAP social studies test, students at Mt. Clemens Jr. Academy tended to score significantly better than students in the district but worse than students in the rest of the state.

Sixteen chi-square analyses were evaluated in the MEAP High School Test (HST) in which four subtests are administered: mathematics, reading, science, and writing. Table 10:4 presents the summary findings for these analyses. These results are strikingly similar and consistent. All sixteen analyses are not significantly different, providing compelling evidence that the students in Edison's Senior Academy are meeting the minimum state standards in comparable proportions to students in the district and in the state.

10.4 Odds Ratio Analysis of the MEAP Data

One of the many possible statistics that can be derived from a 2x2 contingency table is the odds ratio statistic (OR) and corresponding 1- α confidence interval. As presented in Section 2.4 of this report, the 2x2 tables analyzed in the previous section can be thought of as representing consecutive class cohorts in a prospective design. From a classical epidemiological perspective, the students in the Edison school can be thought of as the "exposed" group, that is, exposed to the "Edison-effect," and students in the comparison group as the unexposed group. From this perspective, each yearly comparison is a new cohort, measured over a period of years. There is a minimal possibility for cohort contamination if a number of students in one group are not promoted to the next grade level. However, we think this represents a very small number of possible cases and therefore has minimal impact on the validity of these analyses. Section 2.4 in the report details the OR statistic and corresponding 1- α confidence interval (CI). We calculated and charted OR for each of the 2x2 tables constructed from the chi-square analyses presented above. Note that the Breslow-Day statistic cannot be calculated since there is only one time point. Exhibit 10:1 presents these findings.

Table 10:4 Summary of Chi-Square Findings for Mt. Clemens Academies, HST

	1999
<i>Math</i>	
Mt. Clemens Sr. Acad. vs. District	ns/ns
Mt. Clemens Sr. Acad. vs. State	ns/ns
<i>Reading</i>	
Mt. Clemens Sr. Acad. vs. District	ns/ns
Mt. Clemens Sr. Acad. vs. State	ns/ns
<i>Science</i>	
Mt. Clemens Sr. Acad. vs. District	ns/ns
Mt. Clemens Sr. Acad. vs. State	ns/ns
<i>Writing</i>	
Mt. Clemens Sr. Acad. vs. District	ns/ns
Mt. Clemens Sr. Acad. vs. State	ns/ns

Odds ratio findings, grade 7

Four OR analyses were evaluated at grade 7, two for each subject test on the MEAP. Table 10:5 presents summary OR findings for the grade 7 math and reading tests. Students at Mt. Clemens Jr. Academy performed significantly better than students in the district; the OR is below 1.0 and the 95 percent CI does not eclipse 1.0. The Breslow-Day statistic was not statistically significant; thus, one OR and CI can represent this protective effect. However, the picture is the opposite when one considers performance against the state where students at Mt. Clemens Jr. Academy had slightly higher odds for failure relative to students in the rest of the state. In reading, Mt. Clemens students performed at par over the three year period relative to the district students, but showed a slightly higher odds for failure relative to the students in the rest of the state.

Table 10:5 Summary of Odds Ratio Findings for Mt. Clemens Academies, Grade 7

	B-D OR	LB	UB
<i>Odds of not meeting standard compared with district</i>			
Mathematics	0.543	0.392	0.753
Reading	0.753	0.536	1.057
<i>Odds of not meeting standard compared with state</i>			
Mathematics	1.318	1.052	1.650
Reading	1.817	1.440	2.292

Odds ratio findings, grade 8

Four OR analyses were evaluated, two for each subject test on the MEAP. Table 10:6 presents summary OR findings for the grade 8 science and writing tests. Students at Mt. Clemens Jr. Academy evidenced striking improvements relative to district students over the two year period on both tests. In both analyses, the Breslow-Day statistics were statistically significant, and the individual ORs showed a dramatic improvement in Mt. Clemens students' performance (see Table 10:6). In science, grade 8 students from Mt. Clemens had about twice the odds for failure in 1999, but in 2000 grade 8 students had about half the odds for failure. In writing, Mt. Clemens grade 8 students showed comparable passing rates to district students in 1999, but in 2000 Mt. Clemens grade 8 students were about one-sixth as likely to fail the writing MEAP. A similar picture emerges from the state comparison, but not as strong. In science, Mt. Clemens students were about twice as likely to fail; but in writing these students showed strong improvement, decreasing their odds of failure by about half.

Table 10:6 Summary of Odds Ratio Findings for Mt. Clemens Academies, Grade 8

	1999	2000	B-D OR
<i>Odds of not meeting standard compared with district</i>			
Science	2.078	0.495	p<.05
Writing	1.104	0.152	p<.0001
<i>Odds of not meeting standard compared with state</i>			
Science			2.383
Writing	2.998	1.526	p<.05

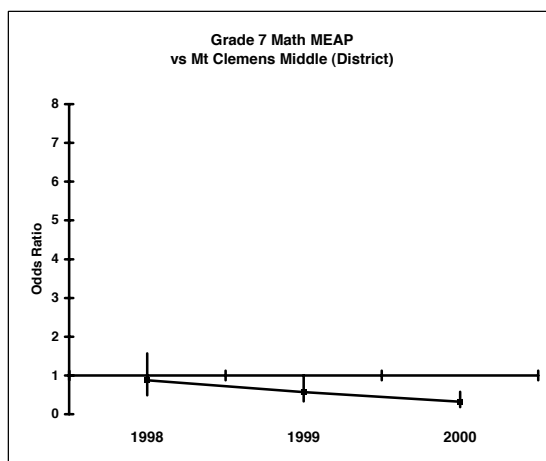
Exhibit 10:1 Results of the Grade 7 Odds Ratio Analysis for Mt. Clemens Secondary Academies 1998-00

Grade 7 MEAP Math vs. Mt. Clemens Middle School (district)

Year	U CI	L CI	OR
1998	1.57	0.488	0.875
1999	0.997	0.329	0.573
2000	0.575	0.182	0.323

Breslow-Day for Homogeneity of Odds Ratio
Chi-square (2, N=600) = 5.765, p = .0560

Common OR = 0.543
U CL = 0.753
L CL = 0.392

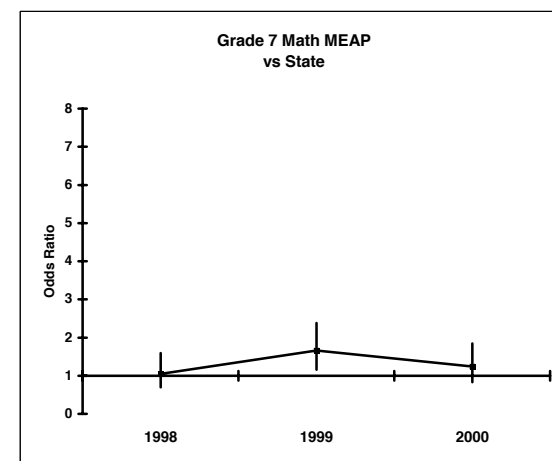


Grade 7 MEAP Math vs. State

Year	U CI	L CI	OR
1998	1.592	0.694	1.051
1999	2.383	1.159	1.662
2000	1.843	0.83	1.237

Breslow-Day for Homogeneity of Odds Ratio
Chi-square (2, N=345,457) = 2.806, p = .2458

Common OR = 1.318
U CL = 1.650
L CL = 1.052

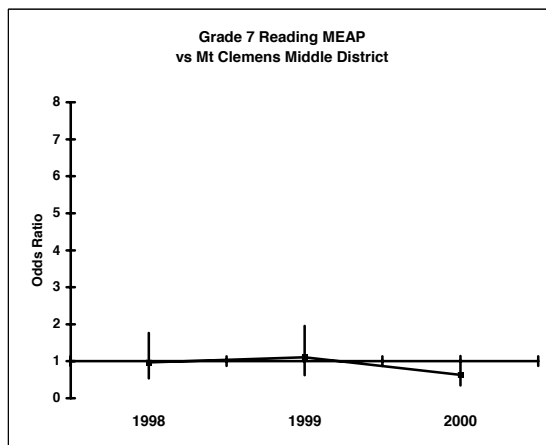


Grade 7 MEAP Reading vs. Mt. Clemens Middle School (district)

Year	U CI	L CI	OR
1998	1.764	0.533	0.970
1999	1.958	0.623	1.104
2000	1.139	0.35	0.631

Breslow-Day for Homogeneity of Odds Ratio
Chi-square (2, N=613) = 1.082, p = .5823

Common OR = 0.753
U CL = 1.057
L CL = 0.536



Grade 7 MEAP Reading vs. State

Year	U CI	L CI	OR
1998	2.91	1.245	1.904
1999	2.904	1.361	1.988
2000	2.354	1.042	1.566

Breslow-Day for Homogeneity of Odds Ratio
Chi-square (2, N=336,208) = 0.775, p = .6787

Common OR = 1.817
U CL = 2.292
L CL = 1.440

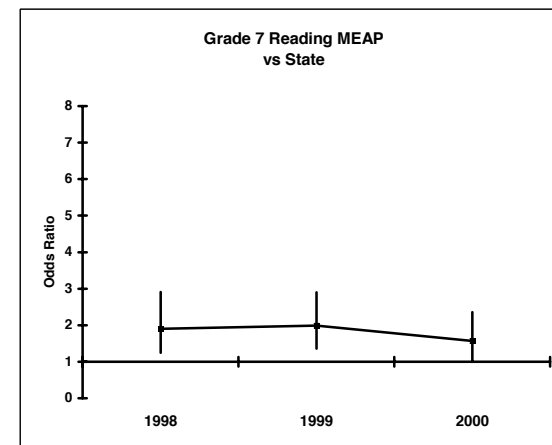
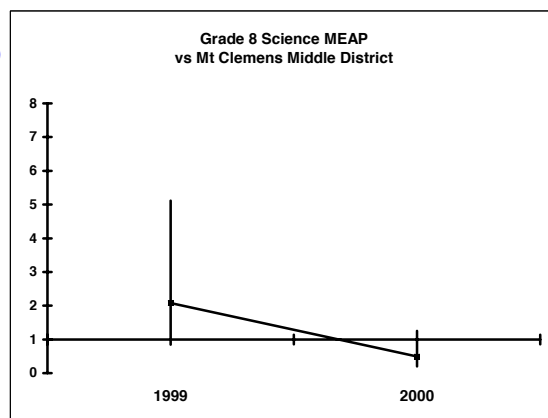


Exhibit 10:2 Results of the Grade 8 Odds Ratio Analysis for Mt. Clemens Secondary Academies 1998-00

Grade 8 MEAP Science vs. Mt. Clemens Middle School (district)

Year	U CI	L CI	OR
1999	5.115	0.844	2.078
2000	1.254	0.196	0.495

Breslow-Day for Homogeneity of Odds Ratio
Chi-square (2, N=398) = 4.867, p = .0274

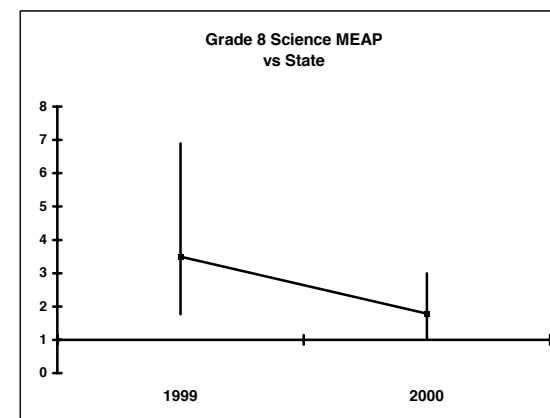


Grade 8 MEAP Science vs. State

Year	U CI	L CI	OR
1999	6.899	1.768	3.492
2000	2.994	1.066	1.787

Breslow-Day for Homogeneity of Odds Ratio
Chi-square (2, N=226,064) = 2.429, p = .1191

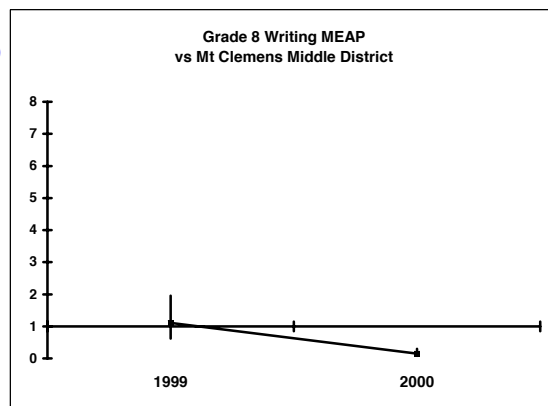
Common OR = 2.383
L CL = 1.583



Grade 8 MEAP Writing vs. Mt. Clemens Middle School (district)

Year	U CI	L CI	OR
1999	1.955	0.623	1.104
2000	0.297	0.078	0.152

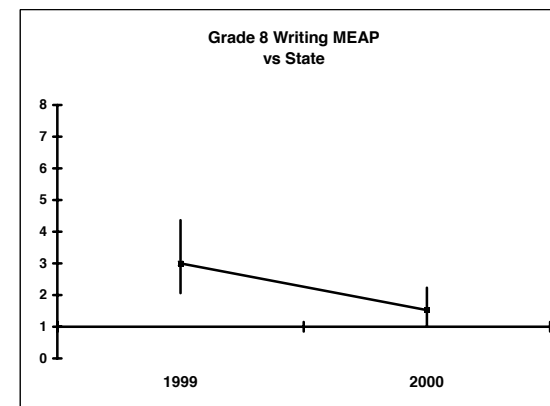
Breslow-Day for Homogeneity of Odds Ratio
Chi-square (2, N=403) = 20.119, p < .0001



Grade 8 MEAP Writing vs. State

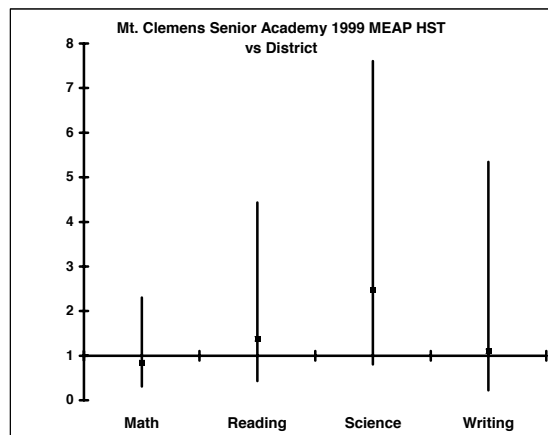
Year	U CI	L CI	OR
1999	4.366	2.058	2.998
2000	2.23	1.044	1.526

Breslow-Day for Homogeneity of Odds Ratio
Chi-square (2, N=223,936) = 6.201, p = .0128



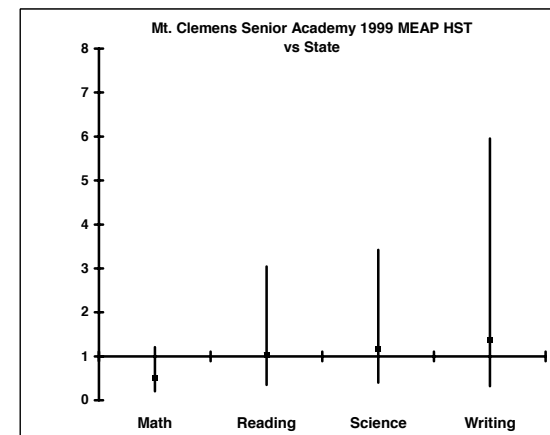
Mt. Clemens Senior Academy 1999 MEAP HST Grade 11 MEAP vs. District

Subject	U CI	L CI	OR
Math	2.301	0.31	0.845
Reading	4.436	0.43	1.381
Science	7.607	0.804	2.473
Writing	5.347	0.224	1.095



Mt. Clemens Senior Academy 1999 Grade 11 MEAP vs. State

Subject	U CI	L CI	OR
Math	1.205	0.207	0.500
Reading	3.041	0.352	1.034
Science	3.422	0.397	1.165
Writing	5.953	0.319	1.377



Odds ratio findings, grade 11 MEAP HST

The HST test is administered in grade 11, covering writing, science, reading, and math. Eight OR analyses (See Table 10:7) were examined for the 1999 MEAP administration, and the results are all quite similar. Thus, students are generally at even odds for failing (or passing) the HST relative to students in either the district or the state.

10.5 Overall Performance on the MEAP Assessments

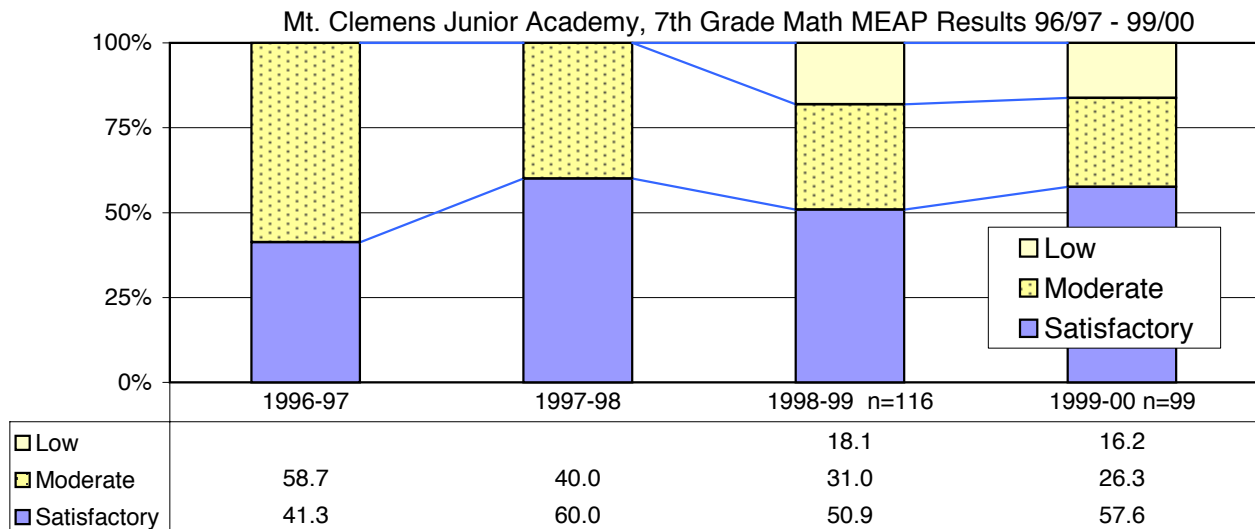
Exhibits 10:3 and 10:4 contain charts that illustrate the relative growth of consecutive groups of grade 7 students at the Edison Junior Academy in relation to the district middle school as well as the average performance in the state. According to Edison (1999), its Mt. Clemens Junior Academy students showed a large gain in math between 1996-97 and 1997-98 (see Exhibit 10:3) when the proportion of students meeting state expectations grew from 41.3 percent to 60 percent. The following year, however, the proportion of grade 7 students meeting state expectations in math dropped by about 9 percent. Then in 1999-00 it rebounded up to about 58 percent of the students meeting the state standards. Edison grade 7 students made small but consecutive gains in reading (Exhibit 10:4), and these gains were about the same as in the state. While gains were being made in the Edison half of the school, the scores for the district middle declined.

Like the grade 7 students, the grade 8 Edison students showed gains in science, writing, and social studies. The district students, housed in the other half of the building, lost ground on the state tests. In fact, the proportion of district students meeting state standards in science and writing dropped sharply between the 1998-99 school year and the 1999-00 school year. Exhibit 10:5 illustrates the relative performance on the MEAP science, writing, and social studies tests for the Edison junior academy, the district middle school, and the state of Michigan. The charts in Exhibit 10:6 cover only grade 8 students for the 1998-99 school year, but illustrate the breakdown of students by performance. From these charts we can see that both Mt. Clemens Middle School and the Edison Junior Academy are still below the average performance levels in the state. Because this is an urban and largely low-income district, the latter finding is not surprising.

Table 10:7 Summary of Odds Ratio Findings for
Mt. Clemens Academics, HST

	B-D OR	LB	UB
<i>Odds of not meeting standards compared with district</i>			
Writing	1.095	0.224	5.347
Science	2.473	0.804	7.607
Reading	1.381	0.430	4.436
Mathematics	0.845	0.310	2.301
<i>Odds of not meeting standards compared with state</i>			
Writing	1.377	0.319	5.953
Science	1.165	0.397	3.422
Reading	1.034	0.352	3.041
Mathematics	0.500	0.207	1.205

Exhibit 10:3 Performance on Grade 7 Math for Edison, District, and State



Note: The MEAP results available from the Michigan Department of Education for 1996-97 and 1997-98 were aggregated with the local middle school that shared the same building. The results for 1996-97 and 1997-98 were derived from the figures reported by Edison and only include figures for satisfactory or unsatisfactory (i.e., moderate plus low).

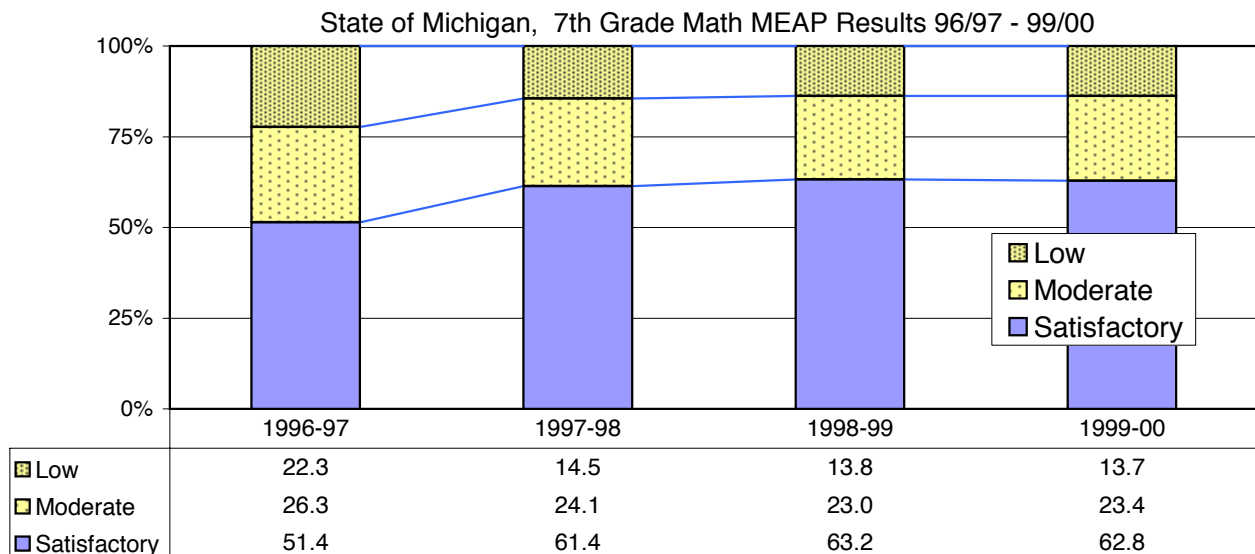
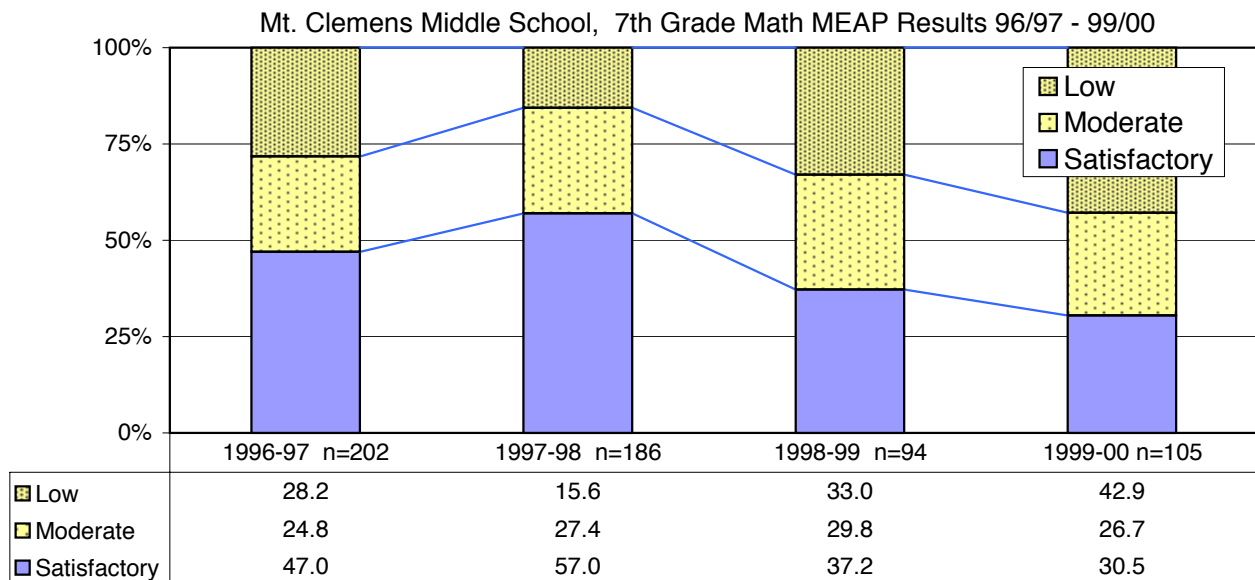
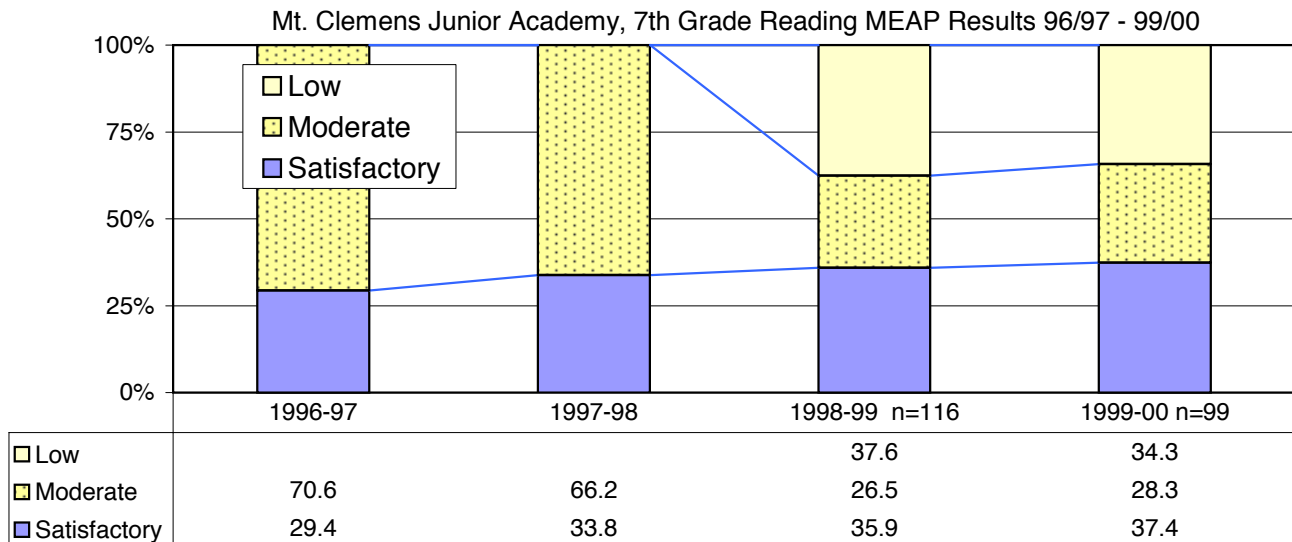


Exhibit 10:4 Performance on Grade 7 Reading for Edison, District, and State



Note: The MEAP results available from the Michigan Department of Education for 1996-97 and 1997-98 were aggregated with the local middle school that shared the same building. The results for 1996-97 and 1997-98 were derived from the figures reported by Edison and only include figures for satisfactory or unsatisfactory (i.e., moderate plus low).

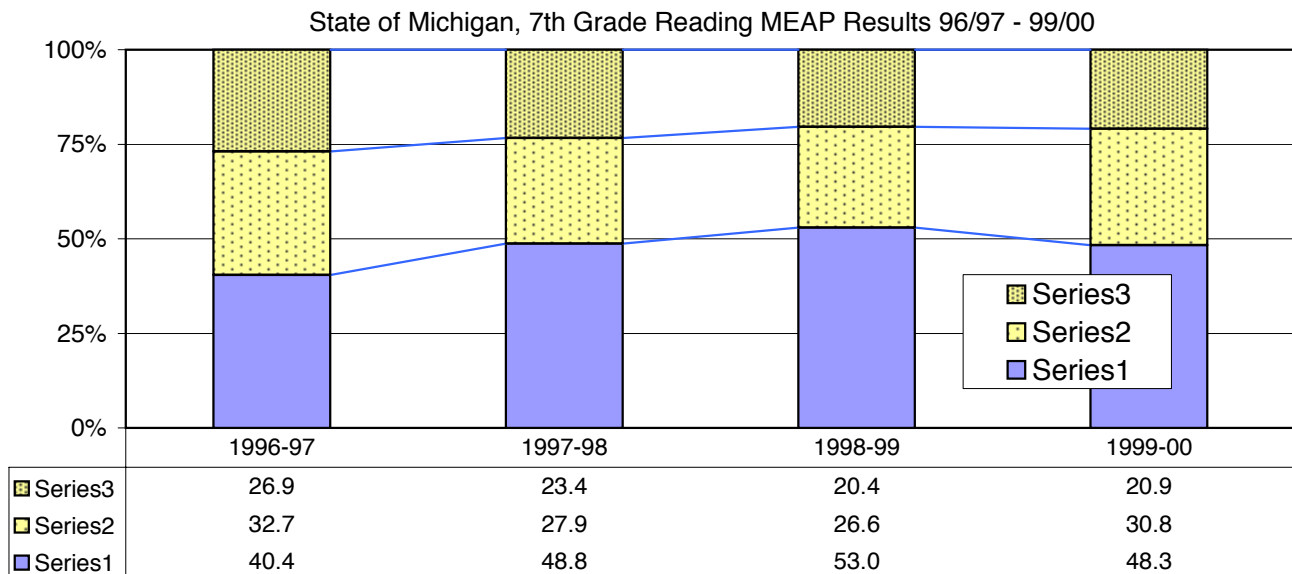
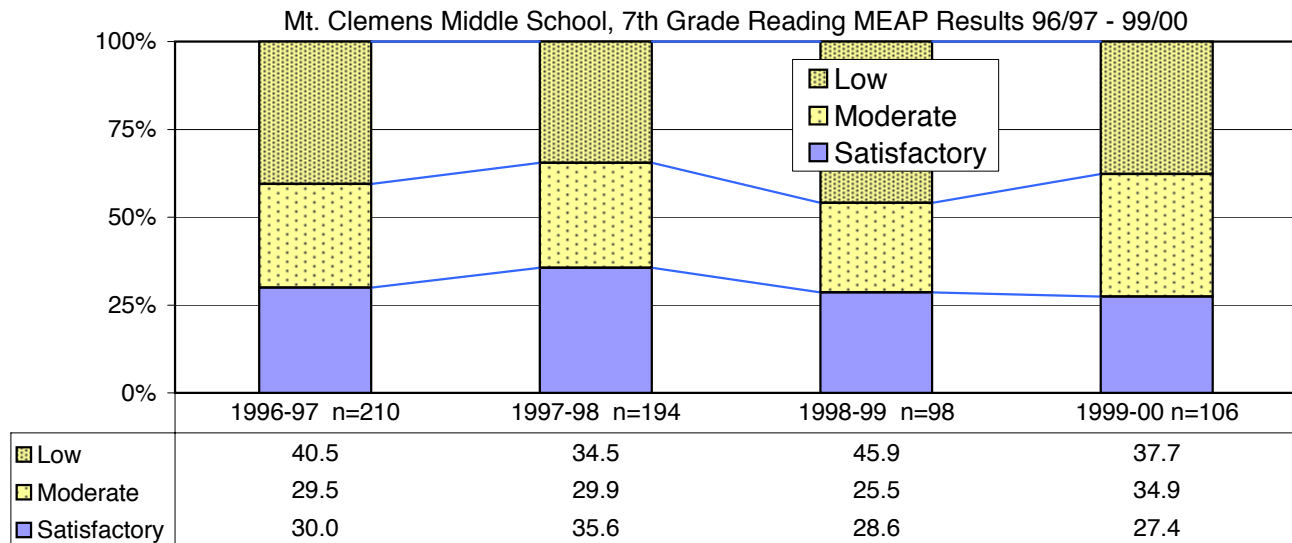
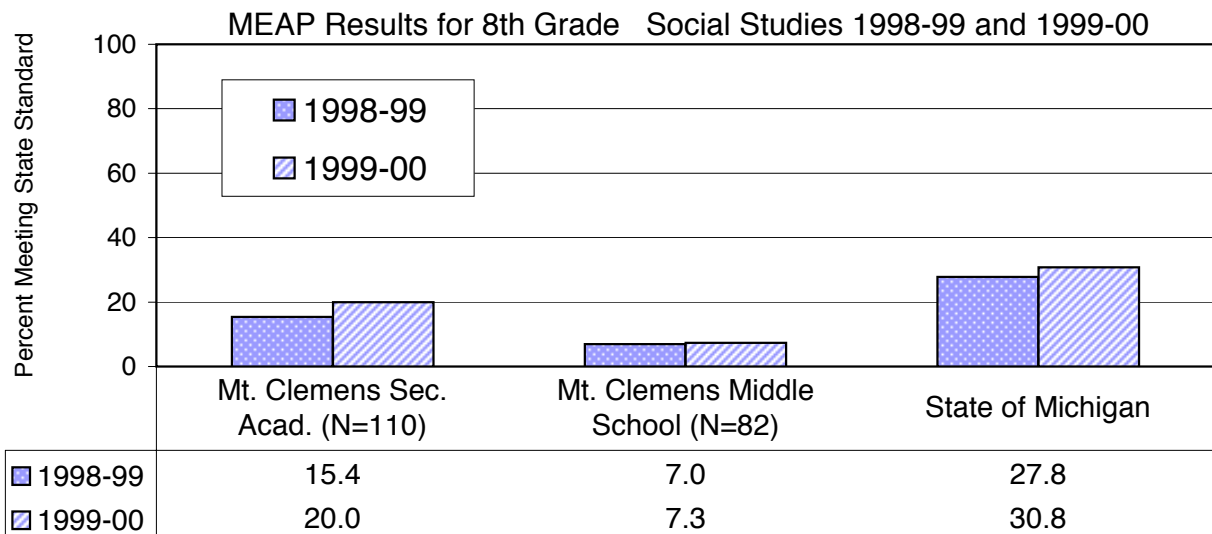
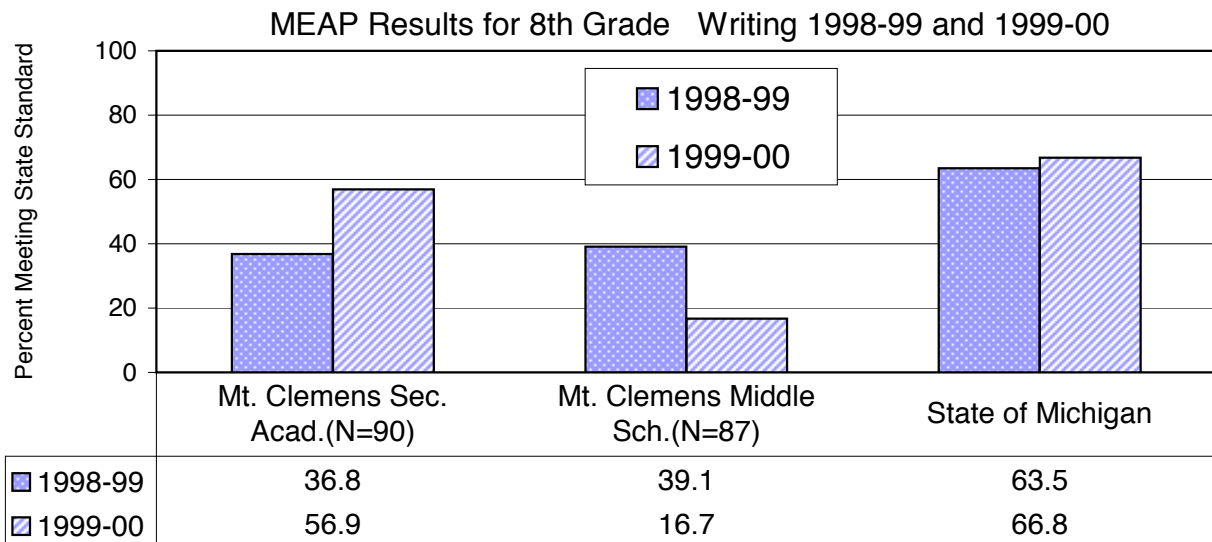
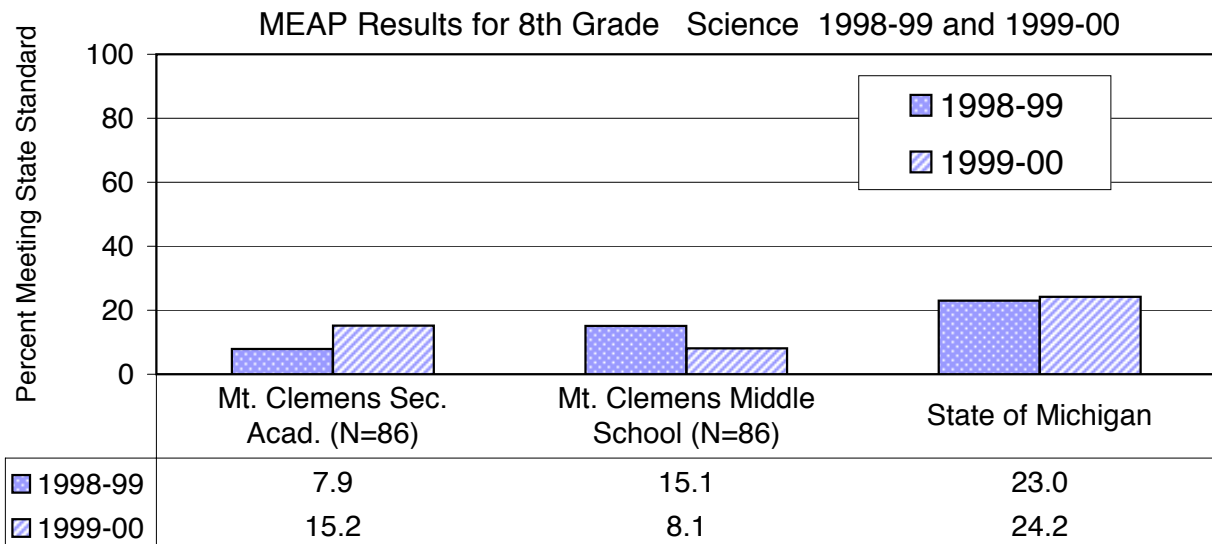


Exhibit 10:5 Performance on Grade 8 Science, Writing, and Social Studies for Edison, District, and State



N refers to the number of test takers in 1999-00

Exhibit 10:6 Performance on Grade 8 Science, Writing, and Social Studies for Edison, District and State

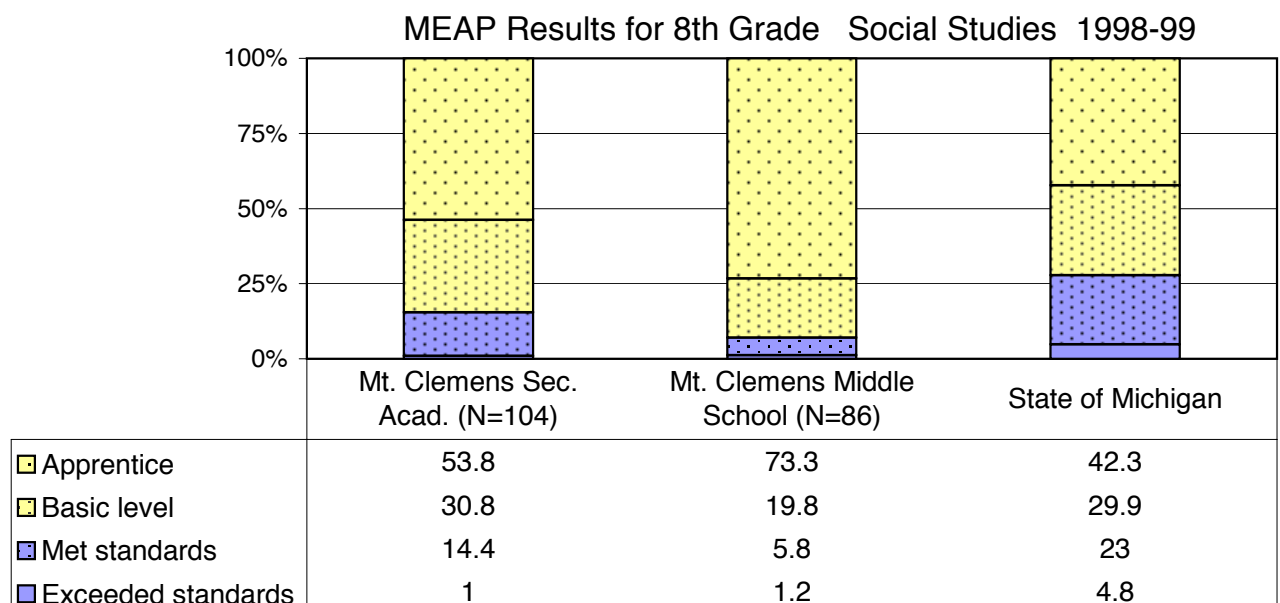
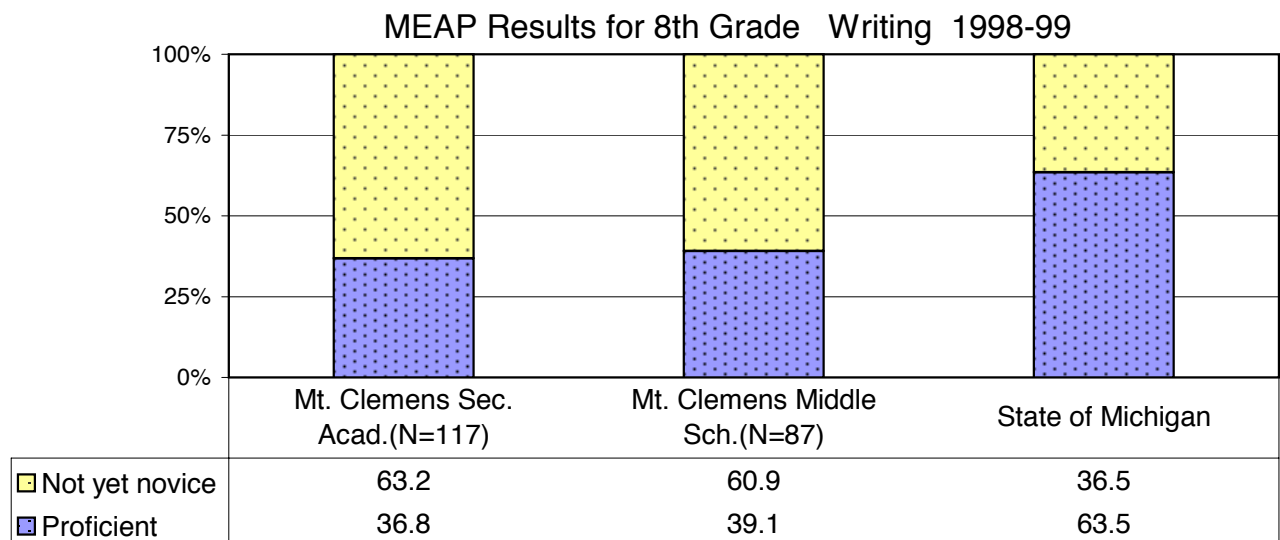
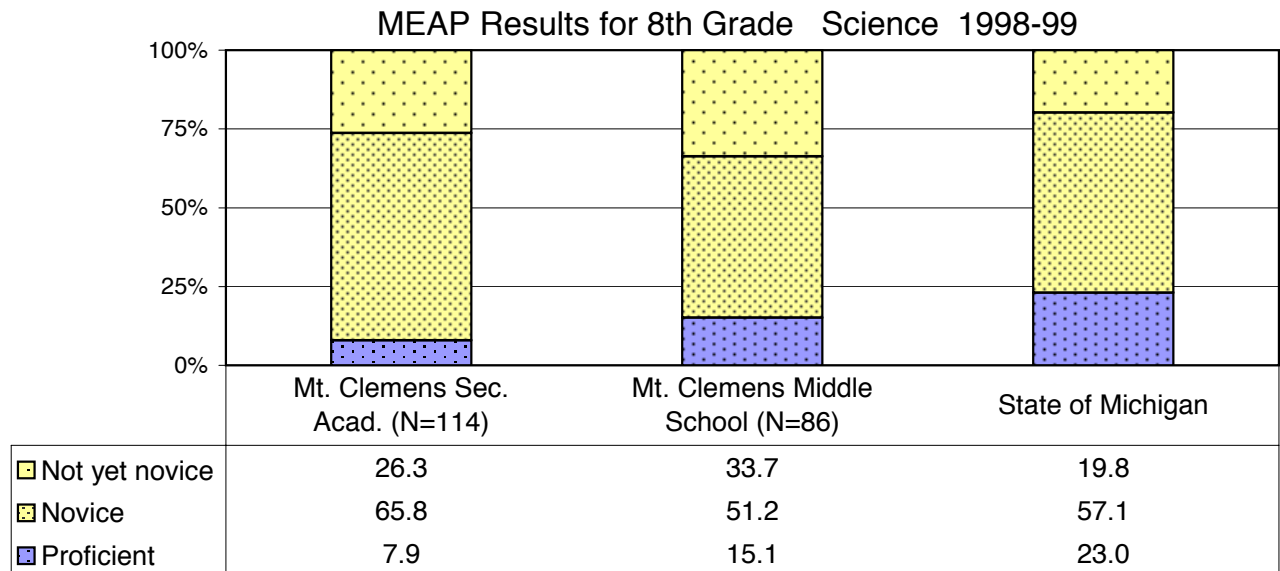


Exhibit 10:7 Performance on Grade 11 Math and Reading for Edison, District, and State

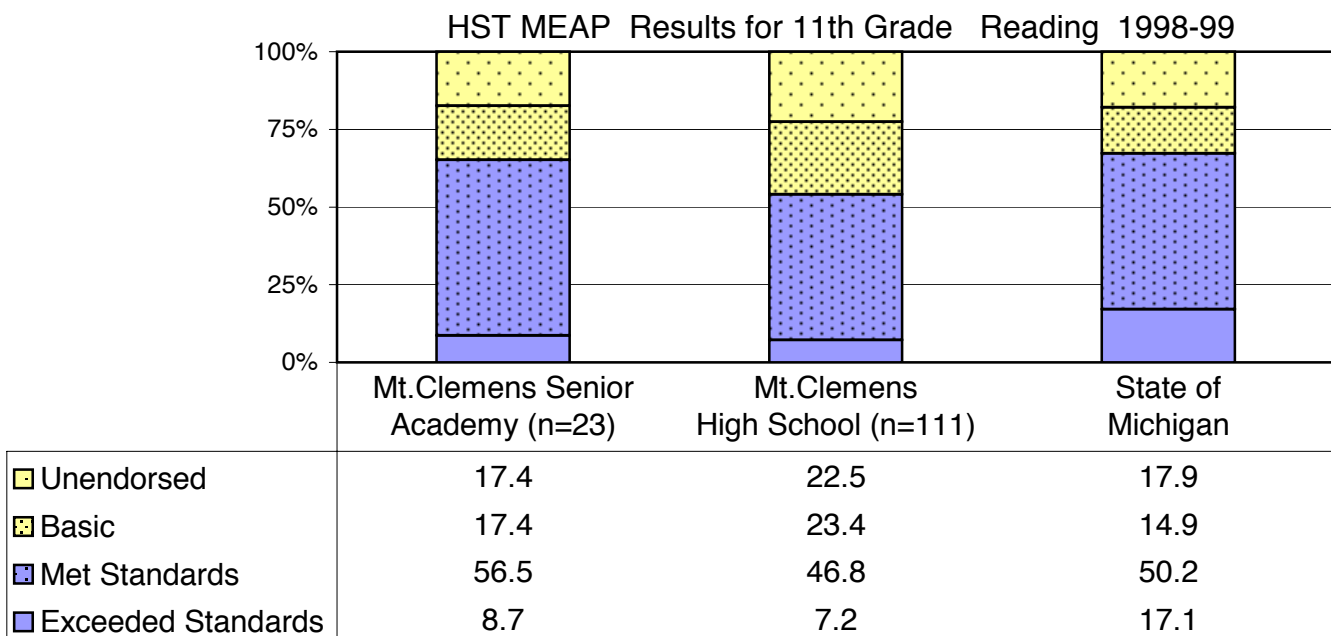
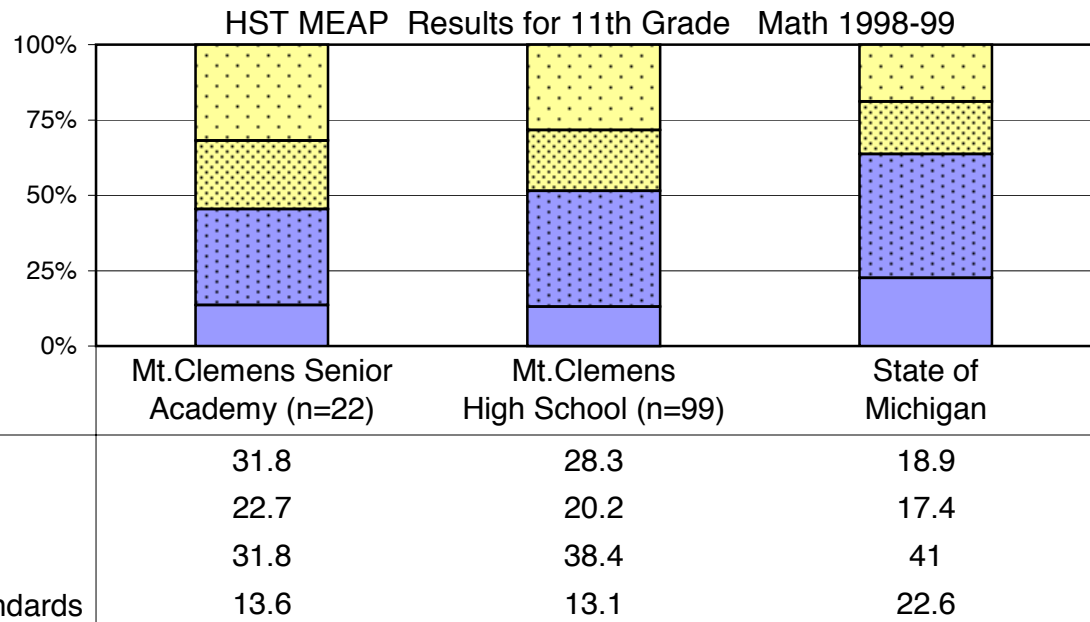
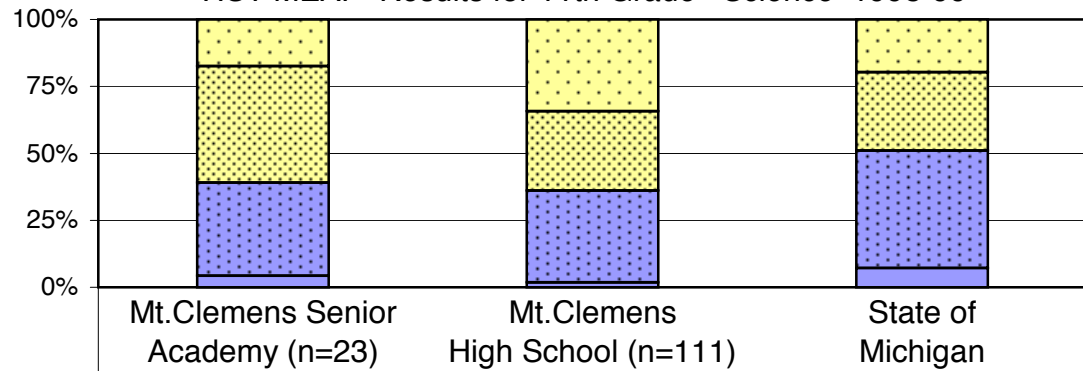


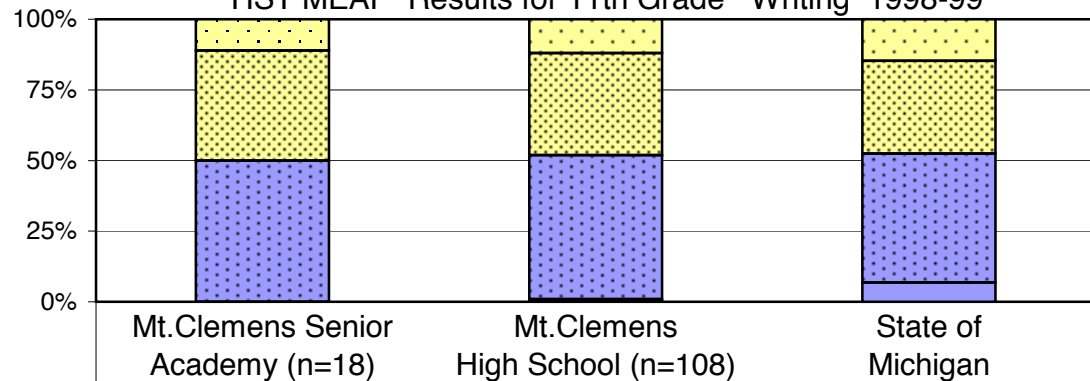
Exhibit 10:8 Performance on Grade 11 Science and Writing for Edison, District, and State

HST MEAP Results for 11th Grade Science 1998-99



Unendorsed	17.4	34.2	19.7
Basic	43.5	29.7	29.3
Met Standards	34.8	34.2	43.8
Exceeded Standards	4.3	1.8	7.2

HST MEAP Results for 11th Grade Writing 1998-99



Unendorsed	11.1	12	14.7
Basic	38.9	36.1	32.9
Met Standards	50	50.9	45.7
Exceeded Standards	0	0.9	6.8

Exhibits 10:7 and 10:8 illustrate the results on the high school component of the MEAP. The grade 11 students are tested in math, reading, science, and writing. The Mt. Clemens Senior Academy formed within the Mt. Clemens High School in 1997-98 school year, but it wasn't until the 1998-99 school year that the academy had students at grade 11, when they are required to take the test. Most of the Edison students taking this test had, at most, participated in the Edison program for 1.5 academic years. The grade 11 students in the Edison senior academy did not perform as well as the district high school students in math, but they performed slightly better in reading and science. The two groups performed similarly in writing. Both Mt. Clemens schools performed lower than the state averages.

In Exhibits 10:3 – 10:8, we used colors to denote the proportion of students meeting state standards. For some of the MEAP tests, several performance levels can be distinguished, but only one or two of them are considered at or above state standards. The blue components of the bar charts indicate the proportion of students meeting or exceeding state standards, while the yellow parts of bar charts indicate the proportion of students not meeting state standards. The blue and yellow demarcation indicate the proportions used when we conducted the odds ratio. The results from the odds ratio analysis are presented in the previous section.

10.6 Summary

Because of the limited data available, the school could be grouped with the others that Edison opened later and which they categorized as “baseline” rather than attempting to place a label on their progress. Unfortunately, we are still faced with limited information on the secondary academies in Mt. Clemens.

Table 10:8 Summary Results on Criterion-Referenced Tests

Mt. Clemens vs. District	1998	1999	2000	B-D	Trend/effect
Grade 7 Math				+1	positive (+1)
Grade 7 Reading				0	mixed (0)
Grade 8 Science		0	0	NA	mixed (0)
Grade 8 Writing		0	+1	NA	positive (+1)
HST Writing				0	mixed (0)
HST Science				0	mixed (0)
HST Reading				0	mixed (0)
HST Math				0	mixed (0)

Note: Comparisons against the state were a mixture of neutral and negative effects

Test results should be available for the Iowa Test of Basic Skills, but no such results were presented in Edison's 2000 annual report and it did not supply us with these data. Moreover, only a limited number of trends could be examined for grade 7 and grade 8 MEAP data and only a one year time point for the MEAP HST. The MEAP results available from the Michigan Department of Education presented a possible positive picture, but limited data are available (see Table 10:8). Clearly, the grade 7 Edison students performed better than the district students on the math and reading tests; and grade 8 students, who should have spent more years in the Edison program, do show some evidence of improvement relative to the district students. Unfortunately with only one year of data for the HST exam, we cannot do much more than speculate.

Table 10:9 Overall Summary of Trends

	Positive	Mixed	Negative
Criterion Referenced	2 of 8	6 of 8	0 of 8
TOTALS	2 of 8	6 of 8	0 of 8

Given the data available for this school, it would be fair to say that the "jury is still out." Judgment on the performance of the Mt. Clemens Secondary Academies should await further years of data. If we were to make a judgment based on the limited data available (see Table 10:9), the Secondary Academies would be rated as Positive with a mean trend rating of 0.25. In its 1999 annual report, Edison rated this school as Strongly Positive. In its 2000 annual report, it rated the 1999-2000 school year as Strongly Positive and the achievement gains since opening as Strongly Positive also.

Chapter Eleven

Mid-Michigan Public School Academy

Lansing, Michigan

11.1 Descriptive Summary of School

Mid-Michigan Public School Academy is a charter school established in 1996 to serve grades K-5, with grades 6-8 added in 1997. Mid-Michigan aims to create a school that “achieves quantum gains in students’ academic performance and in the quality of their lives, a school that serves the diverse needs of all students.” Mid-Michigan is on a large, 40-acre, 10-building campus that was formerly the Michigan State School for the Blind. This school runs on a 205-day calendar with a 7-hour day for primary students (Mid-Michigan Public School Academy, 1999).

This is the largest charter school in Michigan. It began in 1996 with 691 students, but expanded to more than 1,000 by 1999 according to the Mid-Michigan Public School Academy School Profile for the 1999/2000 school year. Of 1,016 students, 85 have special education needs. The student body consists of 64 percent African-Americans, 21 percent Caucasians, 9 percent Hispanics, 4 percent Asians, and 2 percent other ethnic groups. The total proportion of minorities in 1995-1996 was different from the Lansing School District, which indicated that nearly 50 percent of the district students are Caucasian (Mid-Michigan Public School Academy, 1999).

The student enrollment for the 1999/2000 school year was significantly greater for students in grades K-4 than for the upper grades. Kindergarten through grade 4 ranged from an enrollment of 139-157 students per grade. For grades 5-8 the number of enrolled students per grade ranged from 58-88. Primary and elementary academies each have three separate “houses”; each house holds 120-150 students. The Junior Academy (grades 6-8) has two houses with approximately 100 students in each house.

By 1999, Mid-Michigan had a hot lunch program for students and was adding a cafeteria. The free or reduced lunch program in 1998/1999 in both the Mid-Michigan Academy and the Lansing School District were nearly equivalent in terms of the proportion of students qualifying (i.e., Mid-Michigan had 50.7 percent of students receiving free lunches; Lansing School District had 53.5 percent of students receiving free lunches).

The teacher/pupil ratio at both Mid-Michigan and Lansing Public Schools were similar with 19.5:1 and 19.7:1 ratios respectively. Mid-Michigan had 51 classroom teachers, 15 specialists (art, music,

physical education, and Spanish), 5 special education teachers, 6 certified tutors, 8 instructional aides, a guidance counselor, and 1 paraprofessional (Mid-Michigan Public School Academy, 1999).

In the evaluation of Michigan Public School Academies conducted by The Evaluation Center (Horn & Miron, 1999), the average salaries for Mid-Michigan teachers during the 1996/1997 school year reflected a difference of almost \$7,000 when compared with the host district. The Mid-Michigan average teacher salary was \$42,073, while the district average teacher salary was \$48,826.

The school received the maximum foundation grant for charter schools in 1997/98 with \$5,962 per pupil, whereas the Lansing School District received \$6,066. Here there is minimal difference. However, in 1996/1997, the per pupil revenues versus per pupil expense at Mid-Michigan showed a significant gap; per pupil expenditures were \$6,156, but total per pupil revenue was \$10,178. For the host district the gap was not as great. Whereas the expenditure per pupil was \$7,692, the actual revenue per pupil was \$8,320.

Tests administered include the Michigan Educational Assessment Program (MEAP) in reading and math for grade 4 in 1997/98 and 1998/99 and grade 7 in 1997/98 and 1998/99; in writing and science for grade 5 in 1997/98 and 1998/99 and for grade 8 in 1997/98 and 1998/99; and the Metropolitan Achievement Tests, 7th Edition (MAT-7) in Spring 1997 for Grades 3-4 and in Spring 1998 for grades 1-8. The 1998/99 annual report included test results for grades 2-8.

11.2 Past Studies and Evaluations and Currently Available Data

In its second annual report on student performance, Edison categorized the gains in this school as “mixed” (a three star rating). The following statements from its second annual report sum up Edison’s findings reported by the company’s central office (Edison, 1999, p. 42):

- “MEAP, the state criterion-referenced test, shows encouraging results in grade four, where the percentage of students scoring satisfactory or better (i.e., “passing”) rose 12 percentiles in reading and 15 percentiles in math. These gains are somewhat difficult to interpret because the state moved the test date in 1997–98 from fall to winter, to give students’ skills more time to develop. The state also provides no demographically similar schools for comparison. The gains at Mid-Michigan did exceed average state gains, and the state is more economically advantaged than the Mid-Michigan student body.”
- “MEAP scores for fifth grade, though not clouded by a change in the test time, are distorted by a change in writing scoring procedures that dropped scores 10 percentiles statewide. Science gains, therefore, stand as a small improvement while writing gains are unclear.”
- “The Junior Academy at Mid-Michigan began in 1997–98, so the MEAP scores merely provide baseline data—indicate a great need for improvement.”

- “The MAT-7 is voluntarily administered by Edison for diagnostic purposes at Mid-Michigan because no other nationally normed test is mandated. Those results are mixed, with two cohorts down somewhat and one cohort up.”

However, in the school’s own annual education report, Mid-Michigan Public School Academy (1999) revealed that too few students were scoring in the proficient category, especially in the area of science, on the MEAP tests. In eighth grade, 0 percent of males and 3 percent of females scored “proficient” on the science portion of the test in 1998-1999; in fifth grade the results were only slightly less alarming with 19 percent of males and 4 percent of females scoring “proficient” in science.

In the area of math for grade 4, 41 percent of males and 25 percent of females were satisfactory. These results along with those in the area of science, show that girls continue to lag behind boys by a significant margin in these two areas.

Fifth and eighth graders participated in a writing portion of the test. The Mid-Michigan fifth-grade results show that 24 percent of males and 33 percent of females were proficient. Seventeen percent of grade 8 males were proficient, while 37 percent of females were proficient.

The reading test scores that Mid-Michigan released show a smaller gap between the scores of boys and girls. The scores are low, indicating that less than half of the students performed proficiently in standard tests; however, 25 percent of fourth grade males and 23 percent of females received a “satisfactory” score.

Other studies have indicated that Mid-Michigan has not been successful in increasing student performance. Two studies conducted by The Evaluation Center (Horn & Miron, 1999, 2000) evaluated Michigan charter schools, including Mid-Michigan. Appendices from both studies include school-level data. Compared with its host district, the school consistently showed smaller gains. This is the largest charter school in Michigan; but as the statewide charter school studies show, the performance of students at this school is among the worst. Central Michigan University (CMU, 1998), which granted the charter for this school and ranks its schools in terms of performance on the MEAP, indicated that Mid-Michigan was among the poorest performing schools it authorized.

11.3 Longitudinal Analysis of Individual Student Data

Edison Schools Inc. provided individual student test results for the Metropolitan Achievement Test (MAT-7). The MAT-7 scores were provided in four separate scales: Grade equivalent scores (GE), standard scores (SS), national percentile rank scores (PR) and normal curve equivalent scores (NCE) for the language, math/problem solving, and reading comprehension subtests. Parallel analyses are reported for each subtest and scale.

Data records included 782 students covering the 1997-1999 academic years, although we were able to trace only 1 cohort of students representing a small portion over the 3 years. In 1997 there were

approximately 105 possible students to follow. By 1998 about 68 percent remained (71/105 pair wise analysis), and by 1999 only about 42 percent of the students remained (longitudinal trend analysis). This represents a 58 percent drop in students over the 3 years in the cohort of students that progressed from grade 3 to grade 5 from the 1997/98 school year to 1998/99 school year. Table 11:1 presents sample size information for each MAT-7 subtest by year and grade level. Also depicted in this table are the sample sizes for the 1-year gain analysis on this and the longitudinal analyses.

Table 11:1 Sample Sizes for Individual Student Data by Grade and Year on the MAT-7

Grade 3									
	1997			1998			1999		
	Language	Math	Reading	Language	Math	Reading	Language	Math	Reading
GE	106	106	106	98	98	98	126	126	126
SS	106	106	106	98	98	98	126	126	126
PR	106	106	106	98	98	98	126	126	126
NCE	106	106	106	97	97	95	126	125	125
Grade 4									
	1997			1998			1999		
	Language	Math	Reading	Language	Math	Reading	Language	Math	Reading
GE	106	106	106	96	96	96	93	93	93
SS	106	106	106	96	96	96	93	93	93
PR	106	106	106	96	96	96	93	93	93
NCE	104	104	106	96	96	96	89	93	91
Grade 5									
	1997			1998			1999		
	Language	Math	Reading	Language	Math	Reading	Language	Math	Reading
GE	85	85	85	96	96	96	106	106	106
SS	85	85	85	96	96	96	106	106	106
PR	85	85	85	96	96	96	106	106	106
NCE	83	83	84	95	95	95	106	105	105

Sample sizes for 1 year gains

Grade 3 to grade 4

	1997/98			1998/99		
	Language	Math	Reading	Language	Math	Reading
GE	71	71	71	49	49	47
SS	71	71	71	49	49	47
PR	71	71	71	49	49	49
NCE	71	71	71	49	49	47

Sample sizes for 1 year gains (Grades 4 to 5)						
1997/98			1998/99			
	Language	Math	Reading	Language	Math	Reading
GE	75	74	0	62	62	0
SS	75	74	76	62	62	62
PR	75	74	76	62	62	62
NCE	75	74	76	62	62	62

Sample sizes for Longitudinal Analysis

Grades 3 to 4 to 5			
1997/98/99			
	Language	Math	Reading
GE	44	44	44
SS	44	44	44
PR	44	44	44
NCE	44	44	44

Repeated measures ANOVAs were examined for longitudinal trends over a three year period for Mid-Michigan Academy. In all models, the assumption of sphericity was evaluated and, if found to be violated, the Huynh-Feldt adjusted p-values are reported. If the overall linear model was found to be statistically significant, unadjusted (alpha) pair-wise comparisons were examined to identify where a difference in the means might be located.

Students are expected to progress at least one grade level per year. Over the two years, the average progress was 2.8 years in language and 2 years in math/problem solving, but only 1.6 years in reading. Similarly, the standard scores of the students in the cohort increased each year in each of the three subjects. These were statistically significant gains relative to the previous year.

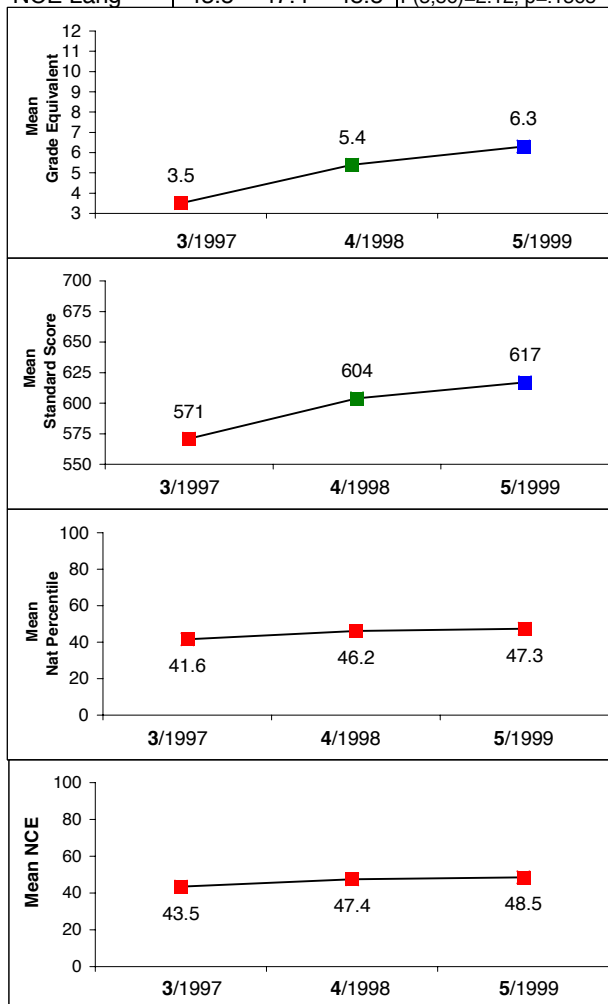
Although the grade equivalents and standard scores indicated that the students were progressing each year, the national percentile and the normal curve equivalent allowed us to compare the growth of this cohort of students with the national norms. On the whole, these students made some gains in language, increasing on the average about 5 ½ percentile points, did not make any gains in math/problem solving, and actually lost ground in reading. Specifically, none of these changes reached statistical significance. Thus, there is no real evidence to suggest that these students made any real positive or negative gains relative to the national comparison sample.

Exhibit 11:1 Mid-Michigan Public School Academy (Lansing, Michigan)

Results from the Analysis of Individual Student Results on the Metropolitan Achievement Test (MAT-7)

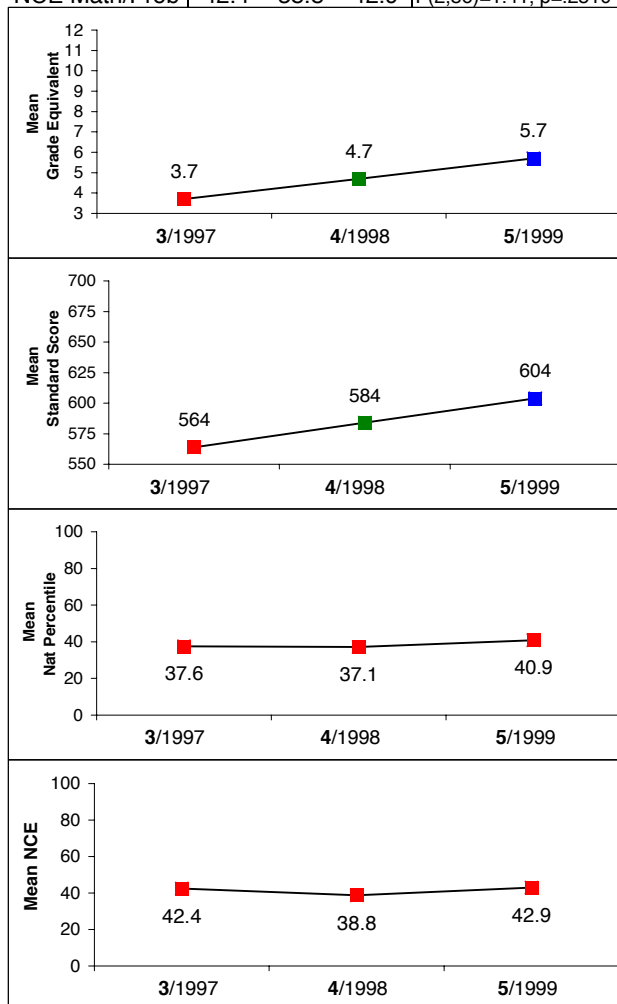
LANGUAGE

n=44	1997	1998	1999	
Grade	3rd	4th	5th	
GE Lang	3.5	5.4	6.3	$F(3,86)=34.57, p<.0001$
SS Lang	571	604	617	$F(3,86)=44.71, p<.0001$
NP Lang	41.6	46.2	47.3	$F(3,86)=1.67, p=.1995$
NCE Lang	43.5	47.4	48.5	$F(3,86)=2.12, p=.1365$



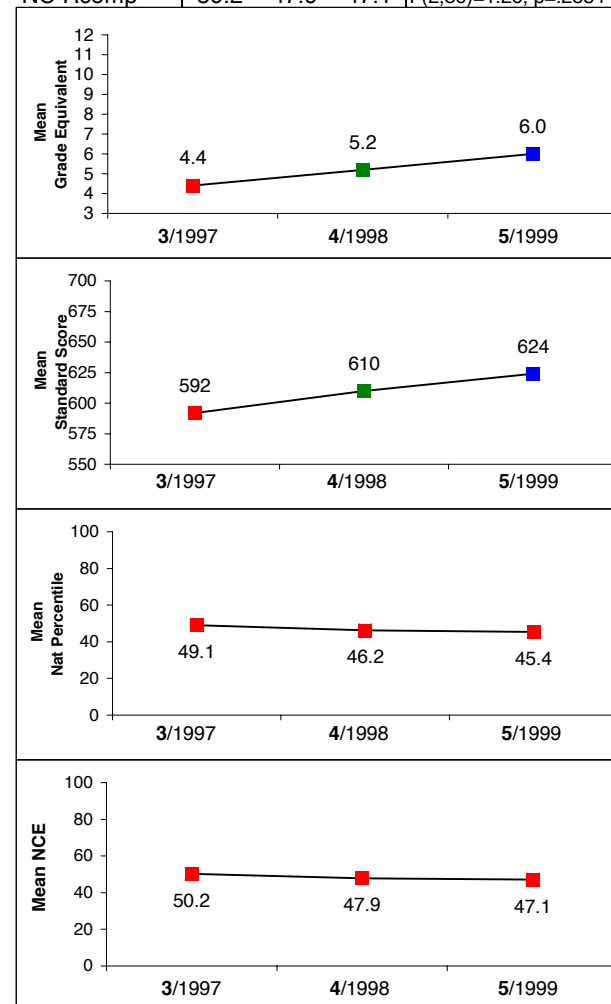
MATH / PROBLEM SOLVING

n=44	1997	1998	1999	
Grade	3rd	4th	5th	
GE Math/Prob	3.7	4.7	5.7	$F(2,86)=28.32, p<.0001$
SS Math/Prob	564	584	604	$F(2,86)=37.85, p<.0001$
NP Math/Prob	37.6	37.1	40.9	$F(2,86)=0.75, p=.4516$
NCE Math/Prob	42.4	38.8	42.9	$F(2,86)=1.41, p=.2510$



READING COMPREHENSION

n=44	1997	1998	1999	
Grade	3rd	4th	5th	
GE Rcomp	4.4	5.2	6.0	$F(2,86)=19.88, p<.0001$
SS Rcomp	592	610	624	$F(2,86)=23.97, p<.0001$
NP Rcomp	49.1	46.2	45.4	$F(2,86)=1.02, p=.3653$
NC Rcomp	50.2	47.9	47.1	$F(2,86)=1.26, p=.2884$



A color change represents a statistically significant change in the means. A trend with two color changes represents differences among all three means. A dual colored charting point (red and green) represents a statistically significant difference between one mean but not the other. The reader is encouraged to examine the individual table of means for these cases.

11.4 Chi-Square Analysis of MEAP Data

A chi-square analysis was initiated on data (1997 to 1999) available from the state of Michigan on the outcomes of the Michigan Educational Assessment Program (MEAP), the state-mandated criterion-referenced test. The MEAP is administered in grades 4 (mathematics and reading) and grade 5 (science; writing; and in 1999 social studies was added). In grade 4, mathematics and reading components of the MEAP are scored along an ordinal three category scale: Satisfactory, Moderate, and Low. Similarly, in grade 5 the science component of the MEAP is scored with the same scale, although different labels are used: “Proficient,” “Novice,” and “Not Yet Novice.” The grade 5 MEAP writing component had only two categories of data results: “Proficient” and “Not Yet Novice.” In 1999 the new component of the grade 5 MEAP, social studies, was scored along a four point continuum: Level 1 (highest) to Level 4 (lowest). Additional information on the MEAP is provided in Appendix A. We also secured two years of MEAP data (1998 and 1999) on the grade 7 reading and mathematics components and on the grade 8 writing, science, and social studies (1999) components.

Construction of the comparison groups

We constructed a variety of different comparison groups for our chi-square analyses. Since we were interested in examining the number/proportion of Mid-Michigan PSA students who met state standards (“passing”) or conversely the number/proportion of students who did not meet state standards (“failing”) on the MEAP, we needed to define a suitable comparison group. In the grade 4 and 5 analyses our first comparison was with the Lansing School District in which Mid-Michigan resides. The state performance constituted our second comparison group.

While the state demographics vary from Mid-Michigan Public School Academy and the Lansing Public School District, we believe that comparisons with state averages can yield further information regarding the relative gains of the Edison school. Also, since Edison claims that advances in other district schools is—in part—due to its presence, we used the state as a more distant point of comparison that cannot be easily influenced by the presence of Edison schools.

General procedure

Utilizing published data from the state of Michigan, we made yearly comparisons (consecutive cohorts) at grade 4 and grade 5 from 1997 through 1999 and from 1998 to 1999 in grades 7 and 8 for each subject component of the MEAP test administered at those specific grade levels.

Percentage data (students in each scoring category) were converted to raw frequency data prior to chi-square analysis. To insure independence of the rows in the chi-square tables, the raw frequencies for each scoring category of the MEAP in the district and state comparisons were down-weighted by subtracting the number of students in that category from Mid-Michigan. Thus, the state numbers reflect all students in the state exclusive of those at Mid-Michigan.

It was our intention to construct four chi-square analyses for each subtest nested within year and grade level. Two of these analyses were on uncollapsed data; that is, all scoring levels were represented in the contingency table (e.g., a 2x3) for the district and the state comparisons. Note that the uncollapsed grade 5 and 8 social studies analyses constituted 2x4 contingency tables. Two follow-up analyses were conducted on the data after collapsing the multilevel scoring into a dichotomy (pass, fail), thus producing 2x2 contingency tables. According to the Michigan Department of Education, a score in the “satisfactory” category constitutes “passing” or meeting the state standard for that particular grade and subject. On the other hand, the “moderate” and “low” categories refer to “slightly below the state standard” and “not well prepared,” respectively. Students who had scores in the moderate and low categories did not meet state standards and fell into the “fail” category in our 2x2 chi-square and odds-ratio analyses.

Unfortunately, the data as extracted from the Michigan Department of Education were partially incomplete regarding the three point scoring scale of the MEAP. Specifically in grade 4, we were able to construct the 2x3 and 2x2 tables for the state comparisons but only the 2x2 table for the district comparisons. In grade 5 we could construct 2x3 and 2x2 tables for the science MEAP, but only 2x2 tables for the writing MEAP in both sets of comparisons. For grade 5, we also constructed the 2x4 table for the social studies MEAP test. In grade 7 reading and math, we constructed both 2x3 and 2x2 tables for both comparisons (district and state). In grade 8 we could get the 2x3 and 2x2 tables for science, but only 2x2 for writing tests and the 2x4 table for the social studies test.

Chi-square findings

Results of the chi-square analyses covered the MEAP administration years from 1997 to 1999. Individual contingency tables are presented in Appendix E.

The MEAP reading and math subtests are administered in grade 4. Summary findings from the 22 separate chi-square statistics are presented in Table 11:2. Overall, there were rather substantial and statistically significant differences in the cell proportions in the Mid-Michigan to state comparison in both the 2x3 and 2x2 tables such that there were higher proportions of students in the higher categories in the state in all three years for both reading and mathematics test. Parallel findings were observed when Mid-Michigan was compared against the district on the mathematics test. There were significantly more students passing the test in the district than at Mid-Michigan.

Table 11:2 Summary of Chi-Square Findings
for Mid-Michigan Academy, Grade 4

	1997	1998	1999
4 th Grade Reading			
Mid-Michigan vs. District	NA	sig/sig	sig/sig
Mid-Michigan vs. State	sig/sig	sig/sig	NA/ns
4 th Grade Math			
Mid-Michigan vs. District	NA	sig/sig	sig/ns
Mid-Michigan vs. State	sig/sig	sig/sig	sig/sig

Note: Each result cell in the matrix is divided with the results for the 2x4 analysis on the left-hand side and the results for 2x2 analysis on the right-hand side (i.e., 2x4/2x2). Red color indicates a statistically significant difference that favors the comparison group; blue color indicates a significant difference that favors the Edison school. NA means not available.

Chi-square results in the district comparison on the reading subtest are not as consistent. In 1997 (2x2 analysis) there were no differences in the cell proportions for reading, although by 1999, the district was passing a significantly greater number of students than Mid-Michigan.

Table 11:3 summarizes the 26 grade 5 chi-square findings on the MEAP science, writing, and social studies subtests. Overall, there were rather substantial and statistically significant differences in the cell proportions in the Mid-Michigan to state comparison in both the 2x3 and 2x2 tables such that there were higher proportions of students in the higher categories in the state in all three years on all three tests: writing, science, and social studies. Parallel findings were observed when Mid-Michigan was compared against the district on the MEAP science test. Significantly more students passed the test in the district than at Mid-Michigan. More variable findings were observed in the district comparison on the MEAP writing test. Chi-square results revealed that in 1997 (2x2 analysis) there were statistically significant differences in the cell proportions with a higher proportion of students passing at Mid-Michigan relative to the district. However, the proportions equalized in 1998 and moved significantly in the failure direction for Mid-Michigan 1999. In 1999 the social studies test was administered for the first time. Thus, we examined only the 2x4 table since there were published guidelines to indicate passing or failing this test.

Table 11:4 summarizes the grade 7 chi-square findings (20 tables) for 2 years of MEAP (1998 and 1999) reading and math tests. Overall, there were rather substantial and statistically significant differences in the cell proportions in the Mid-Michigan to

Table 11:3 Summary of Chi-Square Findings for Mid-Michigan Academy, Grade 5

	1997	1998	1999
5 th Grade Writing			
Mid-Michigan vs. District	NA	sig/sig	sig/sig
Mid-Michigan vs. State	sig/sig	sig/sig	NA/ns
5 th Grade Science			
Mid-Michigan vs. District	sig/sig	sig/sig	sig/sig
Mid-Michigan vs. State	sig/sig	sig/sig	sig/sig
5 th Grade Social Studies			
Mid-Michigan vs. District			NA/sig
Mid-Michigan vs. State			NA/sig

Note: Each result cell in the matrix is divided with the results for the 2x4 analysis on the left-hand side, and the results for 2x2 analysis on the left-hand side (i.e., 2x4/2x2). Red color indicates a statistically significant difference that favors the comparison group; blue color indicates a significant difference that favors the Edison school.

Table 11:4 Summary of Chi-Square Findings for Mid-Michigan Academy, Grade 7

	1998	1999
7 th Grade Reading		
Mid-Michigan vs. District	NA/sig	NA/sig
Mid-Michigan vs. State	sig/sig	sig/sig
7 th Grade Math		
Mid-Michigan vs. District	NA/sig	NA/sig
Mid-Michigan vs. State	sig/sig	sig/sig

Note: Each result cell in the matrix is divided with the results for the 2x4 analysis on the left-hand side, and the results for 2x2 analysis on the left-hand side (i.e., 2x4/2x2). Red color indicates a statistically significant difference that favors the comparison group the comparison group; blue color indicates a significant difference that favors the Edison school.

state comparison in both the 2x3 and 2x2 tables such that there were higher proportions of students in the higher categories in the state in both years on both tests. Parallel findings were observed when Mid-Michigan was compared against the district on the MEAP math test where there were significantly more students passing the test in the district than at Mid-Michigan, with an exception in 1998.

Finally, Table 11:5 summarizes the 21 grade 8 chi-square findings on the MEAP science, writing, and social studies subtests. Overall there were rather substantial and statistically significant differences in the cell proportions at Mid-Michigan relative to both the district and the state in both the 2x3 and 2x2 tables such that there were higher proportions of students in the higher categories in the state and district in both years on both the science and writing tests.

Table 11:5 Summary of Chi-Square Findings for Mid-Michigan Academy, Grade 8

	1998	1999
<hr/>		
8 th Grade Writing		
Mid-Michigan vs. District	NA/sig	NA/sig
Mid-Michigan vs. 8 th Grade	NA/sig	NA/sig
Science		
Mid-Michigan vs. District	sig/sig	sig/sig
Mid-Michigan vs. State	sig/sig	sig/sig
8 th Grade Social Science		
Mid-Michigan vs. District		NA/sig
Mid-Michigan vs. State		NA/sig

Note: Each result cell in the matrix is divided with the results for the 2x4 analysis on the left-hand side, and the results for 2x2 analysis on the right-hand side (i.e., 2x4/2x2). Red color indicates a statistically significant difference that favors the comparison group; blue color indicates a significant difference that favors the Edison school.

11.5 Odds Ratio Analysis of the MEAP Data

One of the many possible statistics that can be derived from a 2x2 contingency table is the odds ratio statistic and corresponding $1-\alpha$ confidence interval. As presented in Section 2.4 of this report, the 2x2 tables analyzed in the previous section can be thought of as representing consecutive class cohorts in a prospective design. From a classical epidemiological perspective, the students in the “Edison” school can be thought of as the “exposed” group—that is, exposed to the “Edison-effect”—and students in the comparison group as the unexposed group. From this perspective each yearly comparison is a “new” cohort; and, measured over a period of years, there are consecutive class cohorts. There is a minimal possibility for cohort contamination if a number of students in one group are not promoted to the next grade level. However, we think this represents a very small number of possible cases and has minimal impact on the validity of these analyses. Section 2.4 details the OR statistic and corresponding $1-\alpha$ confidence interval. We calculated and charted OR for each of the 2x2 tables constructed from the chi-square analyses presented above. Exhibits 11:2 – 11:54 present these findings.

Exhibit 11:2 Results of the Odds Ratio Analyses, Grade 4 Reading and Math

Mid-Michigan Grade 4 MEAP Reading vs. Lansing

Year	UB	LB	OR
1997	2.221	0.877	1.396
1998	2.289	0.948	1.473
1999	3.680	1.433	2.296

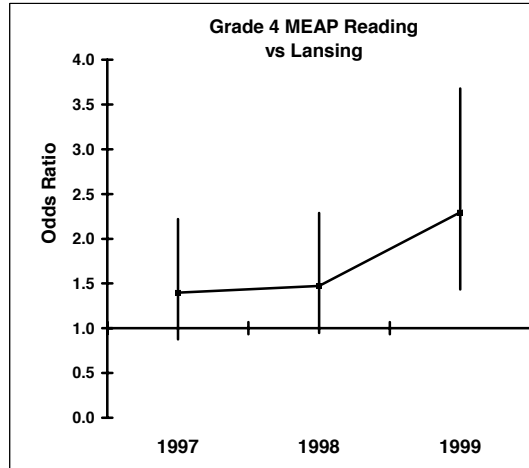
Breslow-Day for Homogeneity of Odd Ratio

Chi-Sq (2, N=3,845) = 2.564, p = .277

OR = 1.672

LB = 1.284

UB = 2.176



Mid-Michigan Grade 4 MEAP Reading vs. State

Year	UB	LB	OR
1997	4.659	1.978	3.036
1998	3.763	1.647	2.490
1999	7.117	3.037	4.649

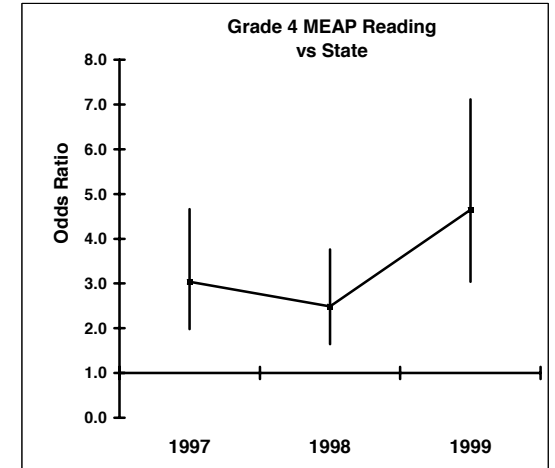
Breslow-Day for Homogeneity of Odd Ratio

Chi-Sq (2, N=345,616) = 3.854, p = .146

OR = 3.249

LB = 2.549

UB = 4.142



Mid-Michigan Grade 4 MEAP Math vs. Lansing

Year	UB	LB	OR
1997	3.163	1.302	2.029
1998	3.482	1.501	2.286
1999	3.733	1.584	2.432

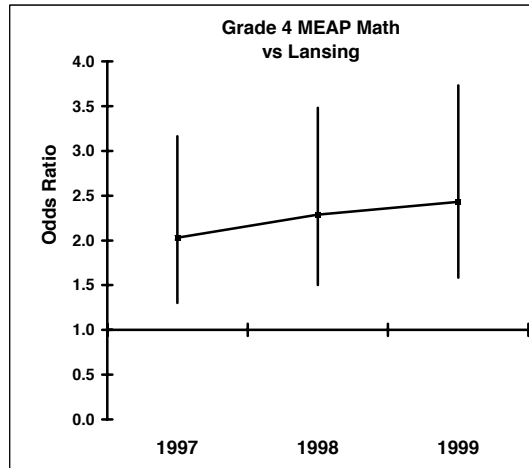
Breslow-Day for Homogeneity of Odd Ratio

Chi-Sq (2, N=3,856) = 0.327, p = .849

OR = 2.242

LB = 1.748

UB = 2.876



Mid-Michigan Grade 4 MEAP Math vs. State

Year	UB	LB	OR
1997	6.686	3.012	4.488
1998	6.067	2.832	4.145
1999	7.431	3.466	5.047

Breslow-Day for Homogeneity of Odd Ratio

Chi-Sq (2, N=346,553) = 0.453, p = .797

OR = 4.553

LB = 3.639

UB = 5.698

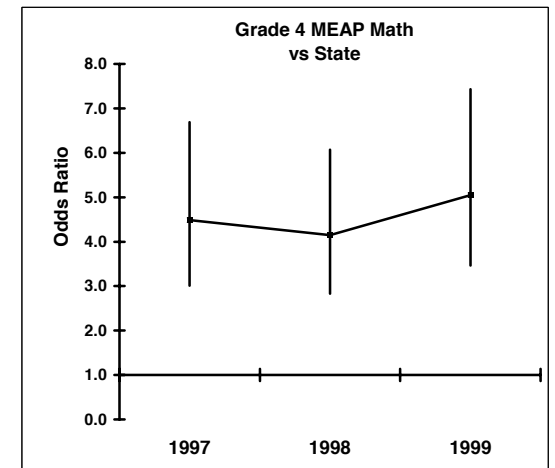


Exhibit 11:3 Results of the Odds Ratio Analyses, Grade 7 Reading and Math

Mid-Michigan Grade 7 MEAP Reading vs. Lansing

Year	UB	LB	OR
1998	1.893	0.727	1.173
1999	4.059	1.382	2.369

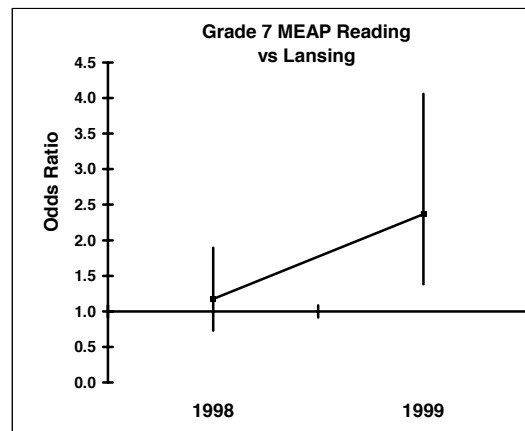
Breslow-Day for Homogeneity of Odd Ratio

Chi-Sq (1, N=2,052) = 3.597, p = .058

OR = 1.622

LB = 1.138

UB = 2.312



Mid-Michigan Grade 7 MEAP Reading vs. State

Year	UB	LB	OR
1998	3.440	1.402	2.197
1999	6.846	2.555	4.183

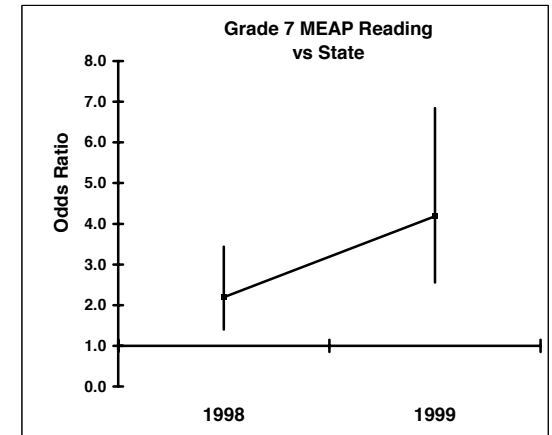
Breslow-Day for Homogeneity of Odd Ratio

Chi-Sq (1, N=230,510) = 3.232, p = .072

OR = 2.941

LB = 2.116

UB = 4.087



Mid-Michigan Grade 7 MEAP Math vs. Lansing

Year	UB	LB	OR
1998	4.674	1.662	2.787
1999	4.028	1.409	2.383

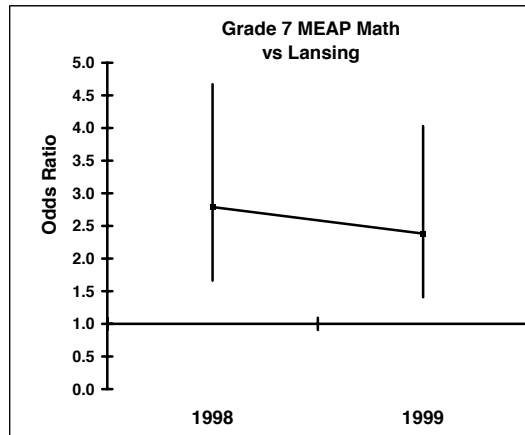
Breslow-Day for Homogeneity of Odd Ratio

Chi-Sq (1, N=2,054) = 0.164, p = .686

OR = 2.579

LB = 1.785

UB = 3.712



Mid-Michigan Grade 7 MEAP Math vs. State

Year	UB	LB	OR
1998	9.245	3.688	5.840
1999	9.933	4.003	6.305

Breslow-Day for Homogeneity of Odd Ratio

Chi-Sq (1, N=230,578) = 0.042, p = .838

OR = 6.067

LB = 4.392

UB = 8.381

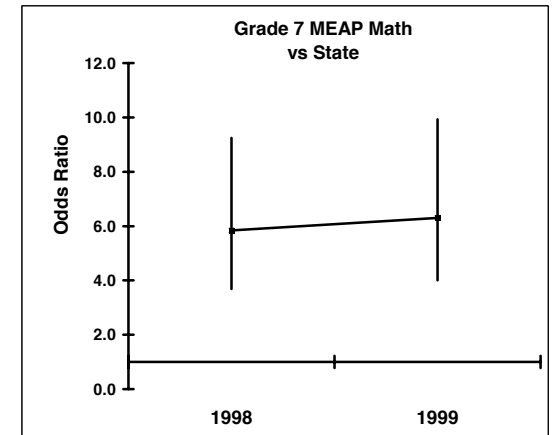


Exhibit 11:4 Results of the Odds Ratio Analyses, Grade 5 Science and Writing

Mid-Michigan Grade 5 MEAP Science vs. Lansing

Year	UB	LB	OR
1997	4.062	1.085	2.099
1998	3.890	1.287	2.238
1999	5.249	1.611	2.908

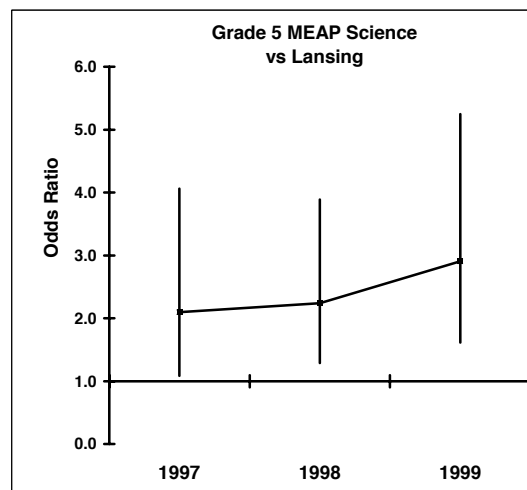
Breslow-Day for Homogeneity of Odd Ratio

Chi-Sq (2, N=3,651) = 0.587, p = .746

OR = 2.411

LB = 1.709

UB = 3.401



Mid-Michigan Grade 5 MEAP Science vs. State

Year	UB	LB	OR
1997	8.263	2.490	4.536
1998	5.983	2.129	3.569
1999	8.192	2.760	4.755

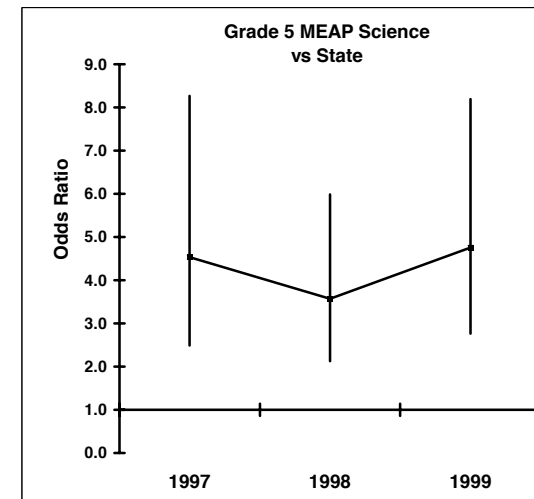
Breslow-Day for Homogeneity of Odd Ratio

Chi-Sq (2, N=341,233) = 0.554, p = .758

OR = 4.228

LB = 3.078

UB = 5.809

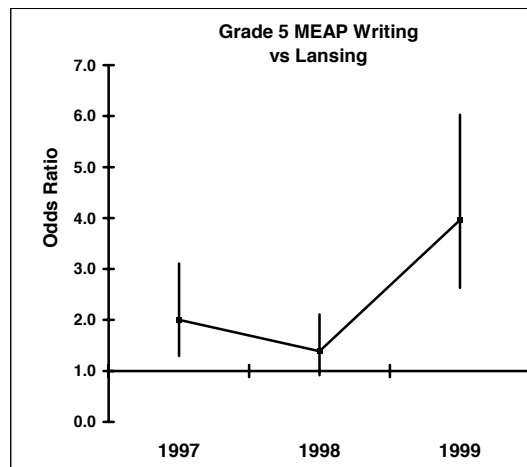


Mid-Michigan Grade 5 MEAP Writing vs. Lansing

Year	UB	LB	OR
1997	3.105	1.291	2.002
1998	2.110	0.914	1.389
1999	6.030	2.629	3.960

Breslow-Day for Homogeneity of Odd Ratio

Chi-Sq (2, N=3,594) = 11.998, p = .002



Mid-Michigan Grade 5 MEAP Writing vs. State

Year	UB	LB	OR
1997	3.311	1.443	2.186
1998	3.099	1.408	2.089
1999	4.593	2.058	3.074

Breslow-Day for Homogeneity of Odd Ratio

Chi-Sq (2, N=339,393) = 1.977, p = .372

OR = 2.436

LB = 1.930

UB = 3.076

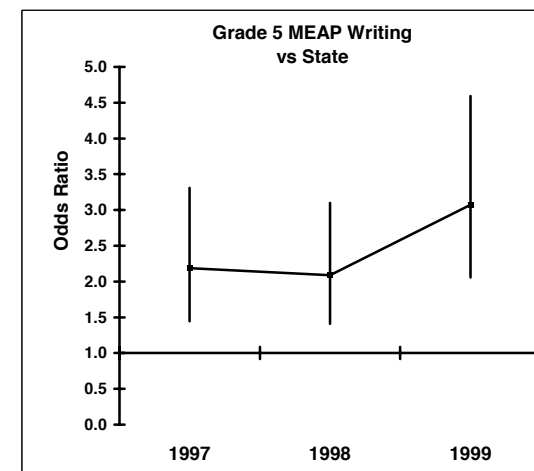


Exhibit 11:5 Results of the Odds Ratio Analyses, Grade 8 Science and Writing

Mid-Michigan Grade 8 MEAP Science vs. Lansing

Year	UB	LB	OR
1998	38.297	1.240	6.891
1999	51.643	2.097	10.407

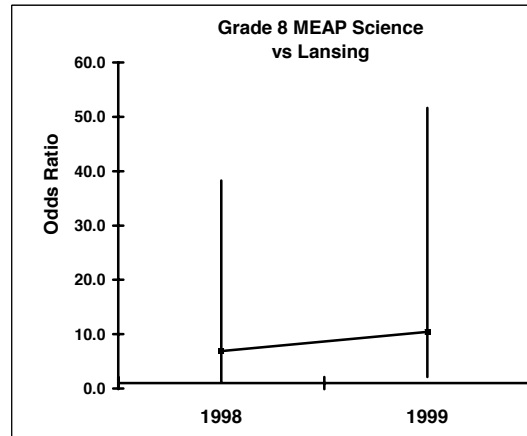
Breslow-Day for Homogeneity of Odd Ratio

Chi-Sq (1, N=2,037) = 0.084, p = .773

OR = 8.606

LB = 2.677

UB = 27.683



Mid-Michigan Grade 8 MEAP Science vs. State

Year	UB	LB	OR
1998	63.399	3.151	14.134
1999	73.911	4.218	17.658

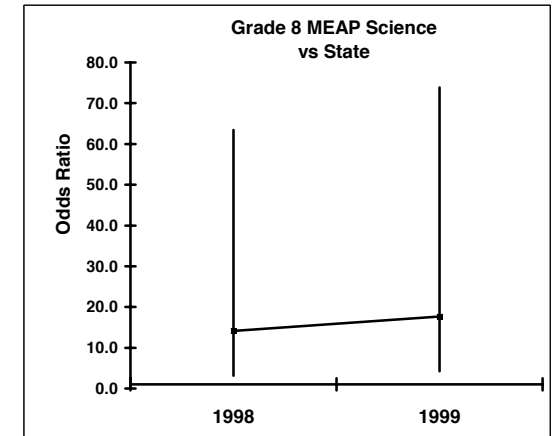
Breslow-Day for Homogeneity of Odd Ratio

Chi-Sq (1, N=224,400) = 0.024, p = .876

OR = 15.885

LB = 5.641

UB = 44.731



Mid-Michigan Grade 8 MEAP Writing vs. Lansing

Year	UB	LB	OR
1998	4.797	1.388	2.581
1999	5.460	2.432	3.644

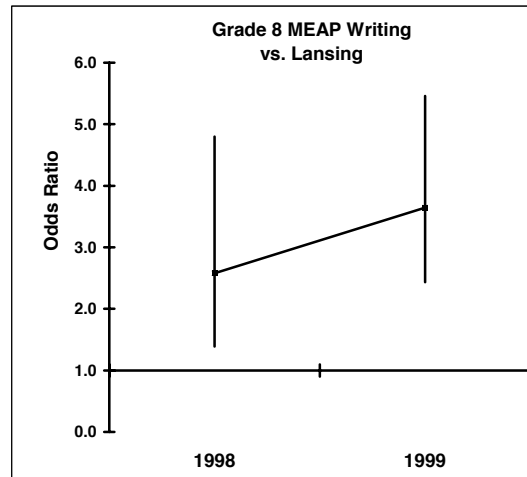
Breslow-Day for Homogeneity of Odd Ratio

Chi-Sq (1, N=1,967) = 1.999, p = .157

OR = 3.644

LB = 2.432

UB = 5.460



Mid-Michigan Grade 8 MEAP Writing vs. State

Year	UB	LB	OR
1998	7.406	2.334	4.158
1999	8.037	2.853	4.788

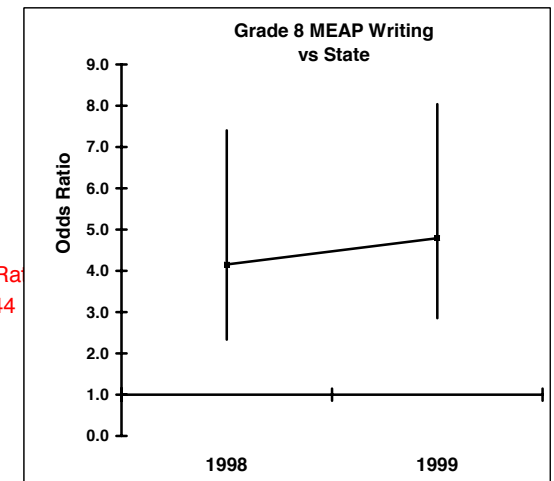
Breslow-Day for Homogeneity of Odd Ratio

Chi-Sq (1, N=221,783) = 0.107, p = .744

OR = 4.509

LB = 3.066

UB = 6.631



Odds ratio findings for grade 4

In grade 4 reading (see Exhibit 11:2), the OR for Mid-Michigan showed relative stability in magnitude against the district, revealing a slight rise in OR in 1999. The Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the three years was not statistically significant, indicating there was no real (statistically significant) change in the OR over the three years. The common OR for the three years is 1.672 and the 95 percent CI is from 1.284 to 2.176. Since the CI does not include 1.00, there was a statistically significant increase in odds for a Mid-Michigan student to fail the grade 4 MEAP reading test relative to students in the district. That is, the Mid-Michigan students were about 1.6 times more likely to fail this test relative to students in the district. A similar pattern in OR is observed relative to the state except that the magnitude is larger, indicating that the Mid-Michigan students were even more likely to fail the reading MEAP test relative to the state. The Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the three years was not statistically significant, indicating there was no real (statistically significant) change in the OR over the three years. The common OR is 3.249 and the 95 percent CI is from 2.549 to 4.142, indicating that Mid-Michigan students were about 3¼ times more likely to fail this test.

The grade 4 math component of the MEAP presents a similar picture. All the CI around the individual ORs exclude 1.00 and thus are considered statistically significant. Likewise, the Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the three years was not statistically significant, indicating there was no real (statistically significant) change in the OR. The common OR relative to the district is 2.242 and the 95 percent CI is from 1.748 to 2.876. For the state comparison the common OR is 4.553 and the 95 percent CI is from 3.639 to 5.698. The Mid-Michigan students were about 2¼ times more likely than district students to fail the math test over the three years and about 4½ times more likely to fail relative to the rest of the state.

Odds ratio findings for grade 5

In grade 5 science (see Exhibits 11:4), the OR for Mid-Michigan showed relative stability in magnitude against the district. The Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the three years was not statistically significant, indicating there was no real (statistically significant) change in the OR. The common OR for the three year period is 2.411 and the 95 percent CI is from 1.709 to 3.401. Thus, there was a statistically significant increase in odds for a Mid-Michigan student to fail the Grade 5 MEAP science test relative to students in the district in this three year period. Mid-Michigan students were almost 2.5 times more likely to fail this test relative to students in the district. A similar pattern in OR is observed relative to the state except that the magnitude is larger, indicating that the Mid-Michigan students were even more likely to fail the science MEAP test relative to the state. The Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the three years was not statistically significant, indicating there was no real (statistically significant) change in the OR over the three years. The common OR is 4.228 and the 95 percent CI is from 3.078 to 5.809, indicating that Mid-Michigan students were about 4¼ times more likely to fail this test as compared with the students in the state.

The grade 5 writing component of the MEAP presented a slightly different picture. In the district comparison the Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the three years was statistically significant, indicating that the OR needs to be examined each year due to its changing value. As can be seen in Exhibit 11:4, the OR in 1999 took a substantial jump. This rising pattern is also apparent in the state analysis, although it did not reach statistical significance. The Breslow-Day chi-square for testing the hypothesis of homogeneity of OR over the three years was not statistically significant, and the common OR for the three year period was 2.436 and the 95 percent CI was from 1.930 to 3.076.

Odds ratio findings for grade 7

The OR analyses for the grade 7 MEAP were based on only two years of data (see Exhibit 11:3). Nevertheless, students at Mid-Michigan showed an increasing trend in failure relative to the district on the reading test, although the Breslow-Day chi-square for testing the homogeneity of OR was not statistically significant. The common OR for the two year period was 1.622 and the 95 percent CI was from 1.138 to 2.312. Thus, students were about 1½ times more likely to fail the grade 7 MEAP reading test relative to students in the district. The OR pattern relative to the state showed the same pattern but with a larger magnitude. The Breslow-Day chi-square was not statistically significant; thus, the common OR is 2.941 and the 95 percent CI is from 2.116 to 4.087. Thus, the Mid-Michigan students were about 3 times more likely to fail the reading test relative to students in the rest of the state.

The grade 7 math component of the MEAP presents a similar picture. In the district comparison, Mid-Michigan students failed to perform at levels consistent with the district students. The Breslow-Day chi-square was not statistically significant. Thus, the common OR is 2.579 and the 95 percent CI is from 1.785 to 3.712, indicating that the Mid-Michigan students were about 2½ times more likely to fail. Relative to the state the Breslow-Day chi-square was not statistically significant. Thus, the common OR is 6.067 and the 95 percent CI is from 4.392 to 8.381.

Odds ratio findings for grade 8

In grade 8 science (Exhibit 11:5), the OR for Mid-Michigan showed relative stability in magnitude against the district. The Breslow-Day chi-square was not statistically significant. The common OR is 8.606 and the 95 percent CI is from 2.677 to 27.683. Thus, there was a statistically significant increase in odds for a Mid-Michigan student to fail the grade 8 MEAP science test relative to students in the district. A similar pattern in OR was observed relative to the state except that the magnitude is larger, indicating that the Mid-Michigan students were even more likely to fail the science MEAP test relative to the state. The Breslow-Day chi-square was not statistically significant. The common OR is 15.885, and the 95 percent CI is from 5.641 to 44.731.

The grade 8 writing component of the MEAP presented a similar picture. In the district comparison, the Breslow-Day chi-square was not statistically significant. The common OR is 3.644, and the 95 percent CI is from 2.432 to 5.460. Relative to the state, the Breslow-Day chi-square for testing the hypothesis of homogeneity of OR is not statistically significant. The common OR is 4.509 and the

95 percent CI is from 3.066 to 6.63, indicating that the Mid-Michigan students were about 4½ times more likely to fail.

11.6 Overall Performance on the MEAP Assessments

Exhibits 11:6 and 11:7 contain charts that illustrate the relative growth of consecutive groups of grade 4 and grade 7 students in reading at Mid-Michigan PSA in relation to the district and state average performance. Performance levels are similar to the district average, but the district is making gains while the Edison school has a decreasing proportion of students meeting state standards.

Performance in math for grade 4 and 7 students is charted in Exhibits 11:8 and 11:9. As with the reading results at these same grade levels, the performance level is similar to the district but below the state levels. Over time, however, the proportion of district students meeting state standards remained rather constant, while the proportion of Edison students meeting state standards declined sharply at grade 4 and slightly at grade 7.

Exhibits 11:10 and 11:11 contain the results in science for grade 5 and grade 8 students. The Edison data for grade 8 science began in the 1997-98 school year, so we could only make a 2-year comparison for this group. Science results at both grade levels were rather poor, both for the Edison school and for the district. District performance increased at grade 4 over time, while the Edison results declined. At the grade 8 level, the district performance remained stable, while less than 2 percent of the Edison students were meeting state standards in grade 8 science.

Writing results for grade 5 and grade 8 students are presented in Exhibits 11:12 and 11:13. The Edison school performed very poorly, both in terms of absolute passing rates and in terms of change over time, which has been negative. Exhibit 11:14 illustrates the performance levels on the relatively new state social studies test. We included the first year of data for this test. Because this was a new test, students across the state performed very poorly. In the years to come, schools will be adjusting their curricula to the state standards, and we can expect passing levels to increase. Much of the growth in the next few years will depend on how quickly schools can modify their curricula so that students are better prepared for this test.

In Exhibits 11:6 to 11:14, we used colors to denote the proportion of students meeting state standards. For some of the MEAP tests, several performance levels can be distinguished, but only one or two of them are considered at or above state standards. The blue components of the bar charts indicate the proportion of students meeting or exceeding state standards, while the yellow parts of bar charts indicate the proportion of students not meeting state standards. The blue and yellow demarcation indicates the proportions used when we conducted the odds ratio. The results from the odds ratio analysis are presented in the previous section.

Exhibit 11:6 Grade 4 Reading , MEAP Results for Mid-Michigan, District, and State

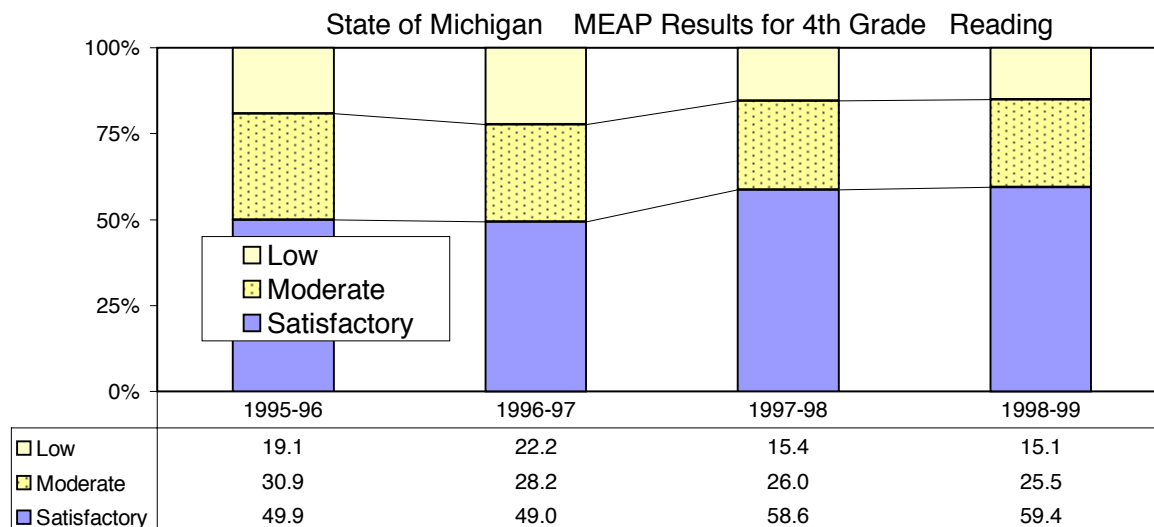
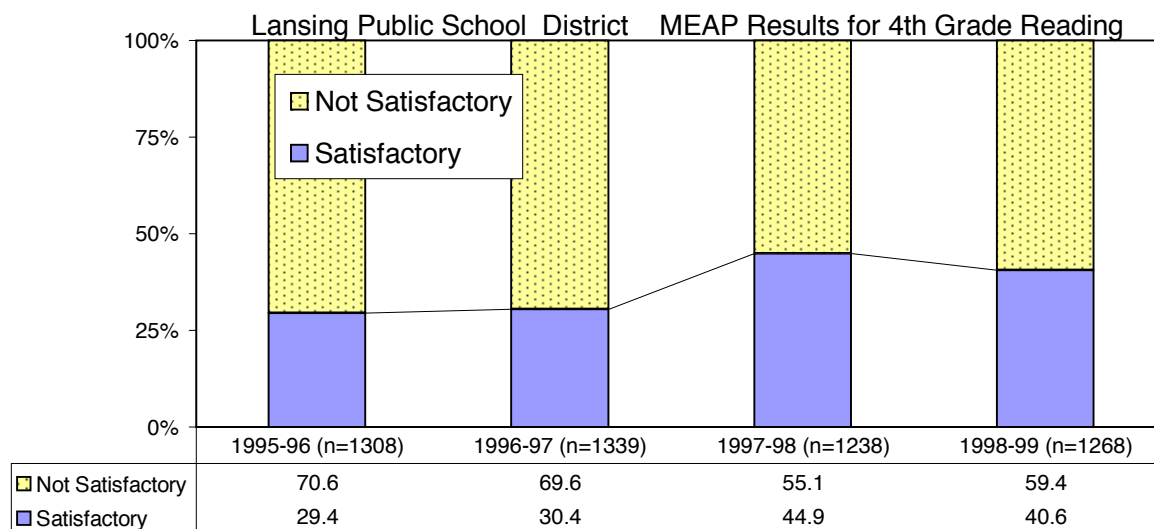
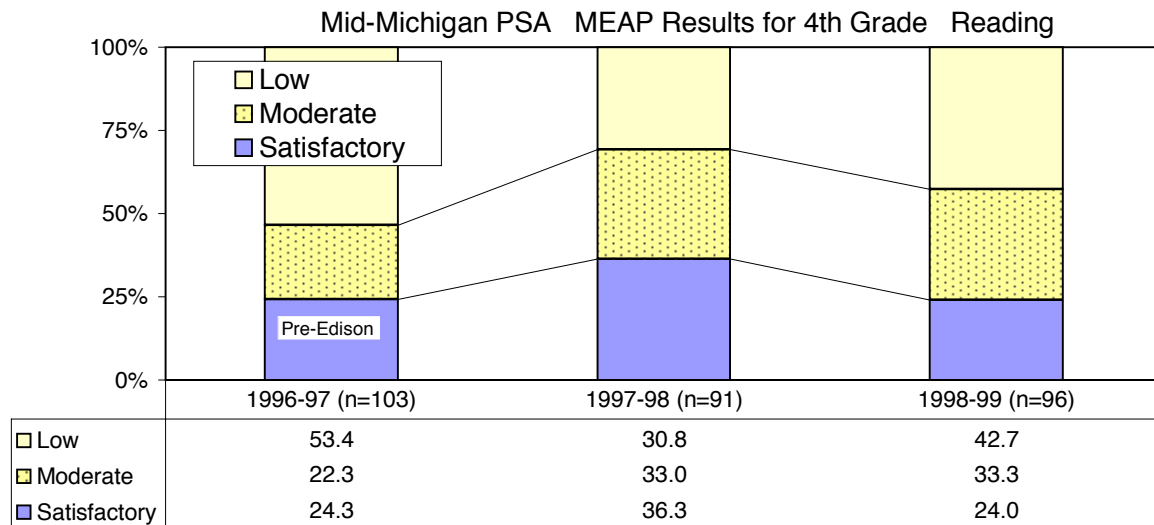


Exhibit 11:7 Grade 7 Reading, MEAP Results for Mid-Michigan, District, and State

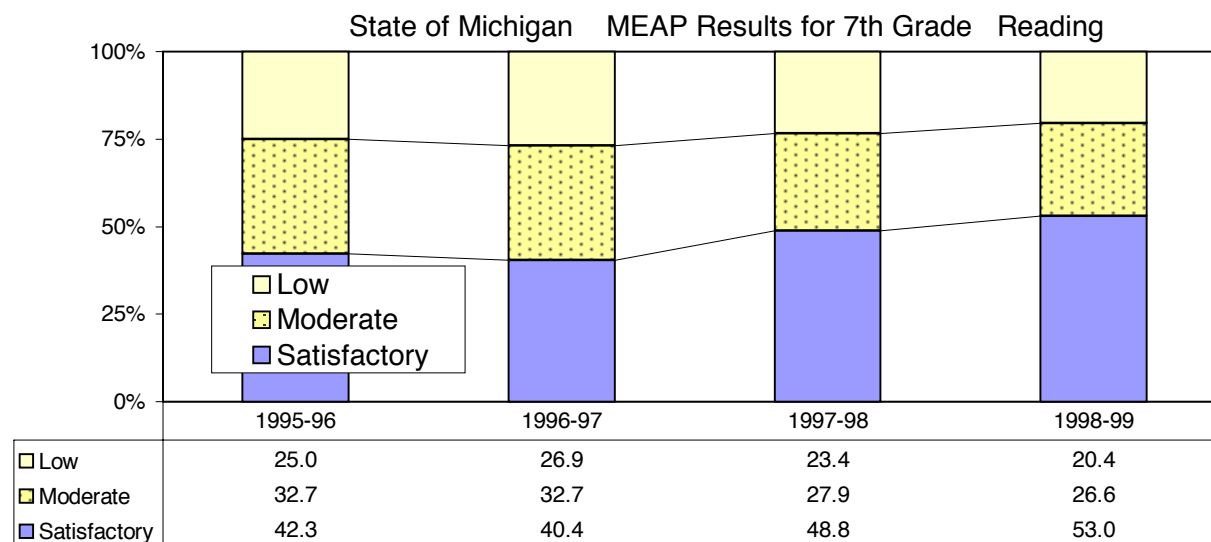
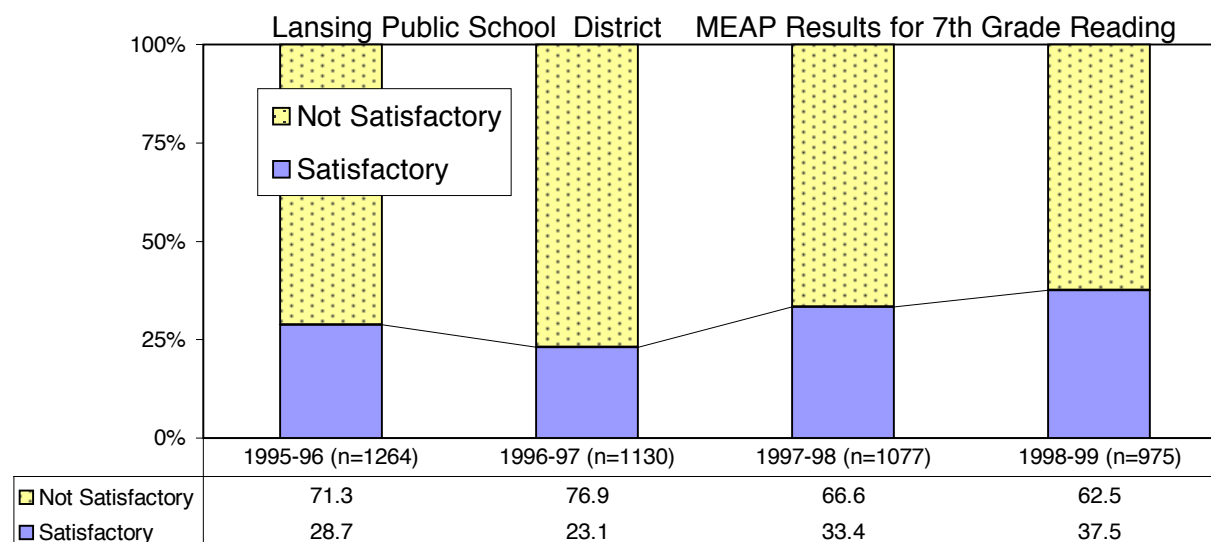
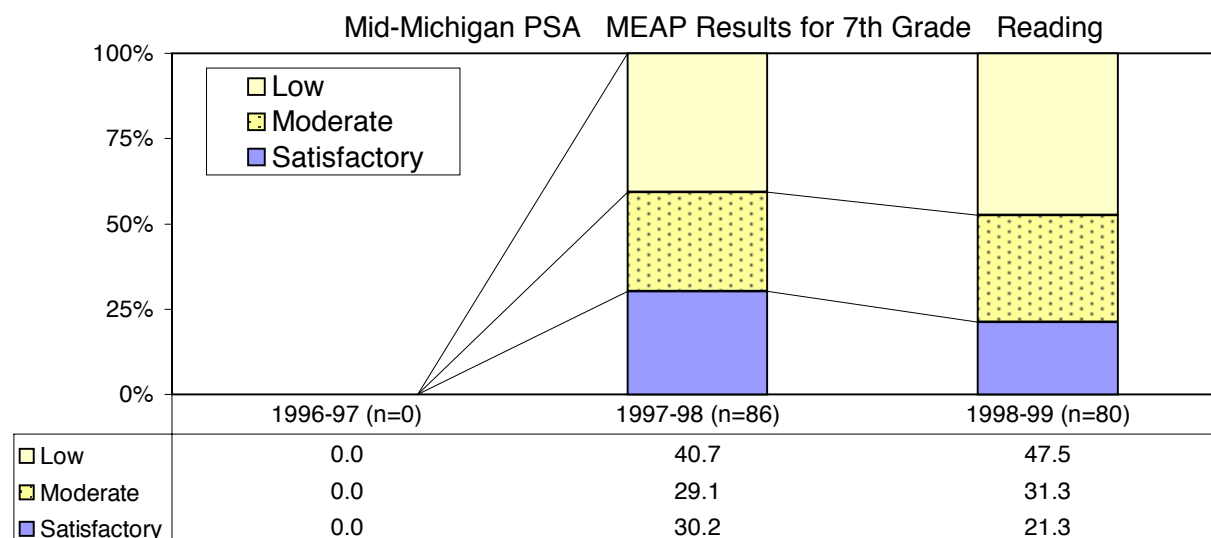


Exhibit 11:8 Grade 4 Math, MEAP Results for Mid-Michigan, District, and State

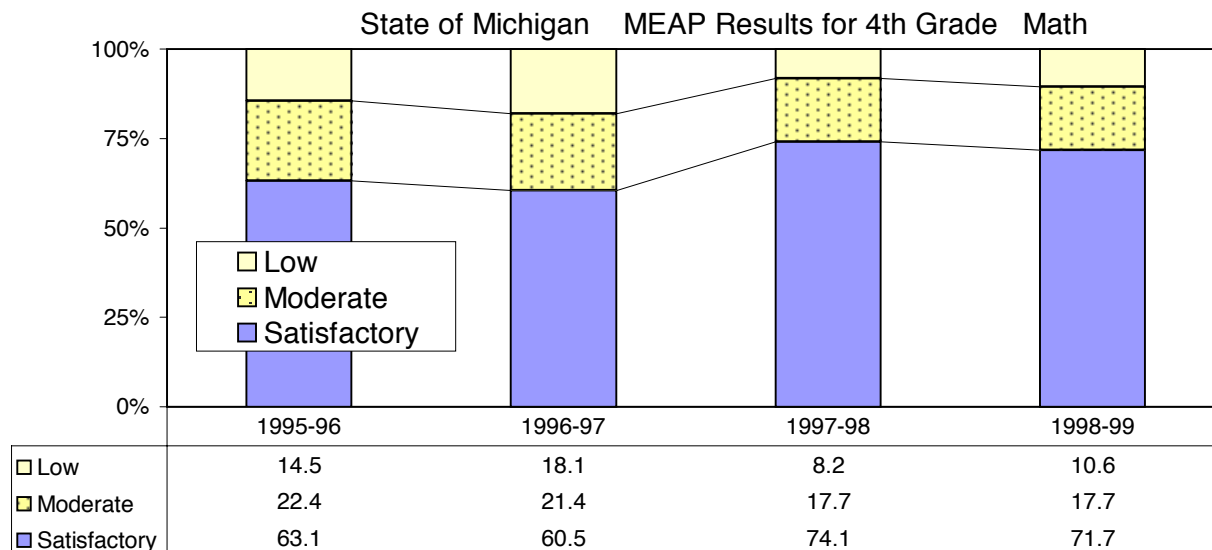
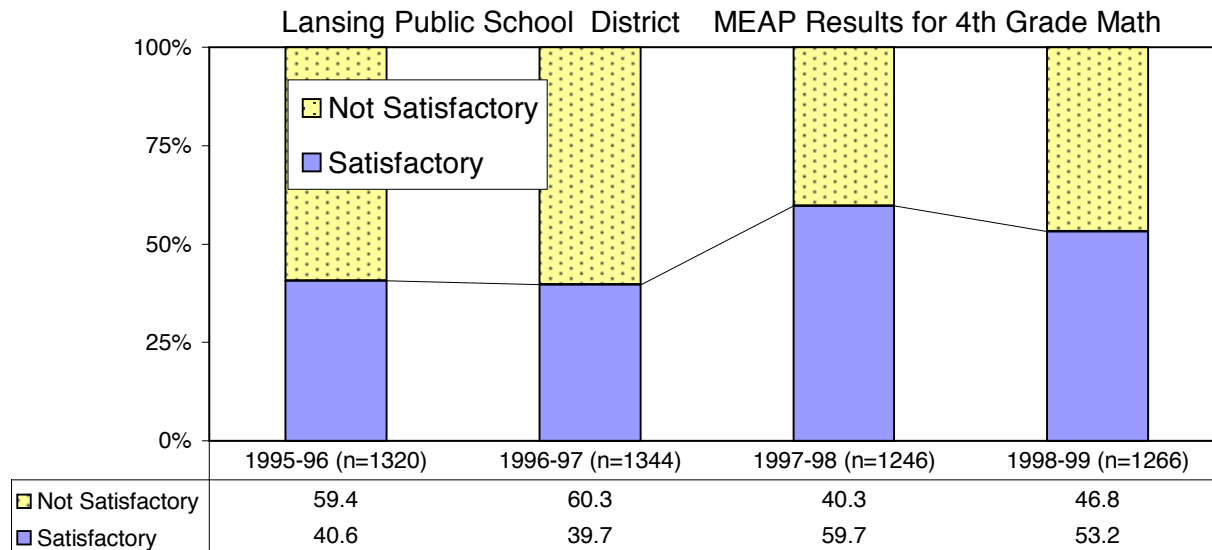
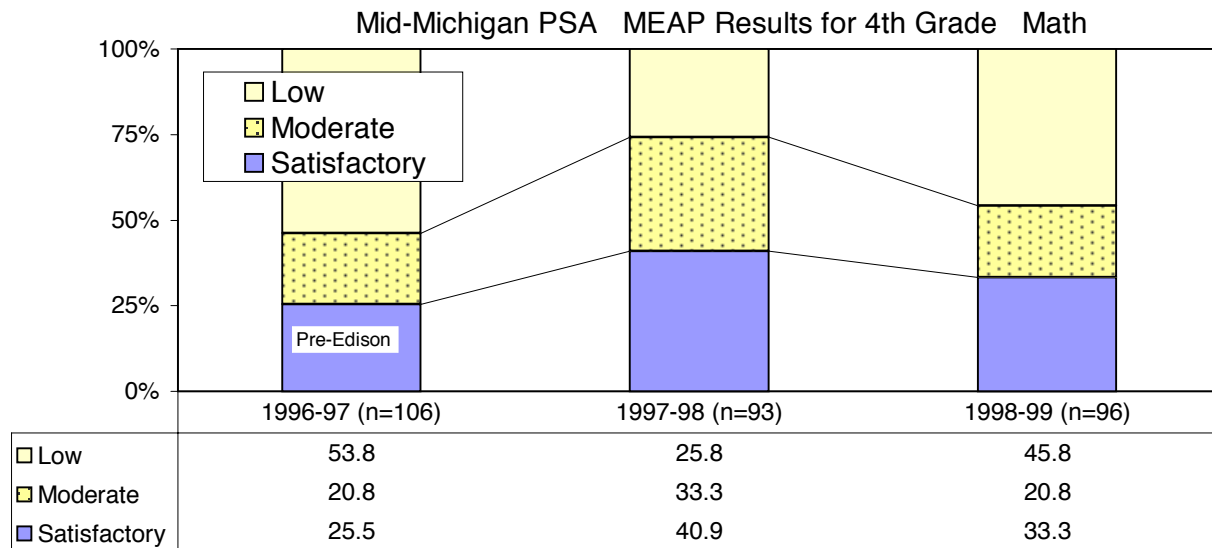


Exhibit 11:9 Grade 7 Math, MEAP Results for Mid-Michigan, District, and State

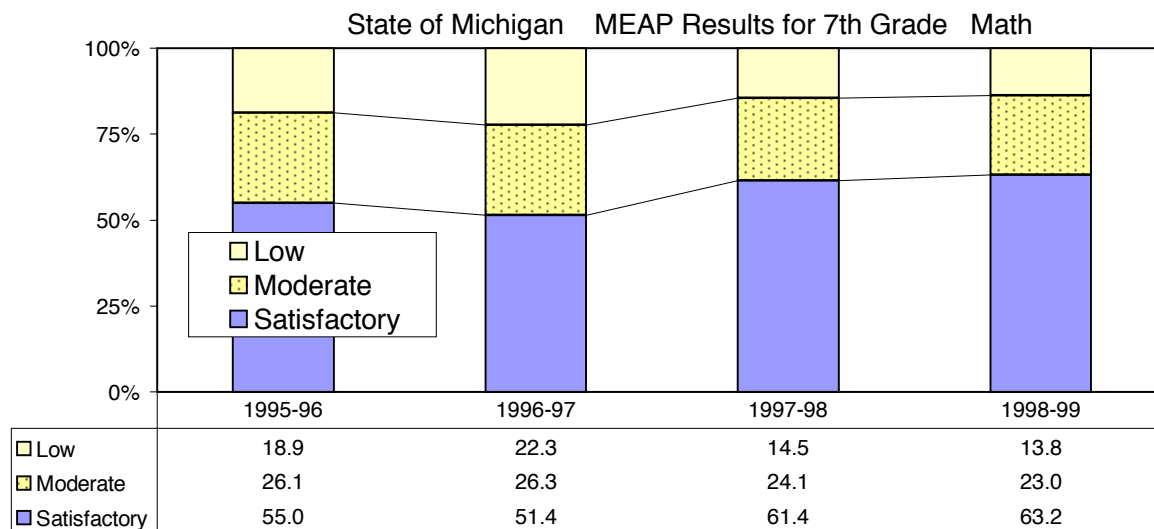
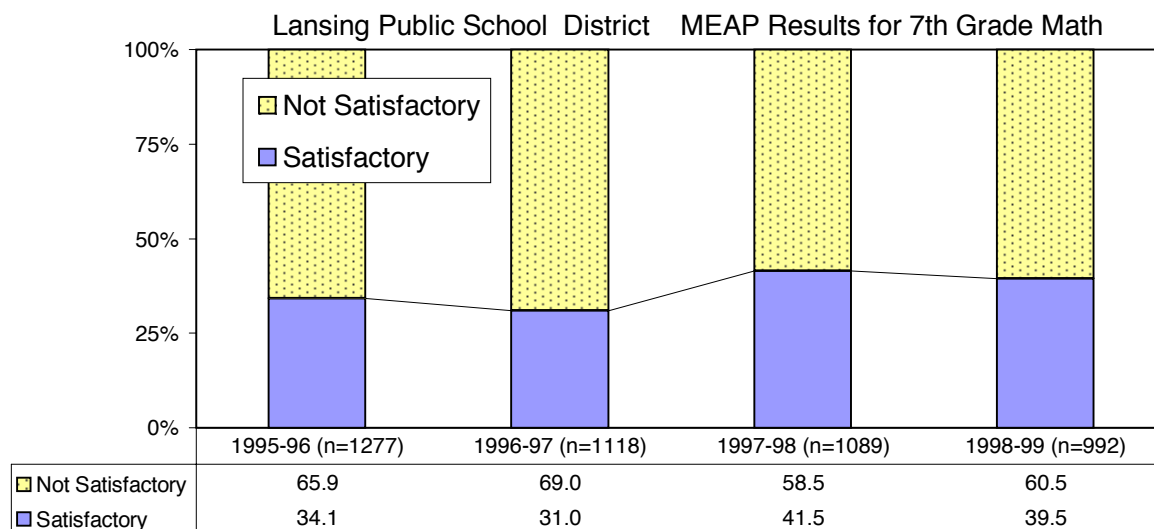
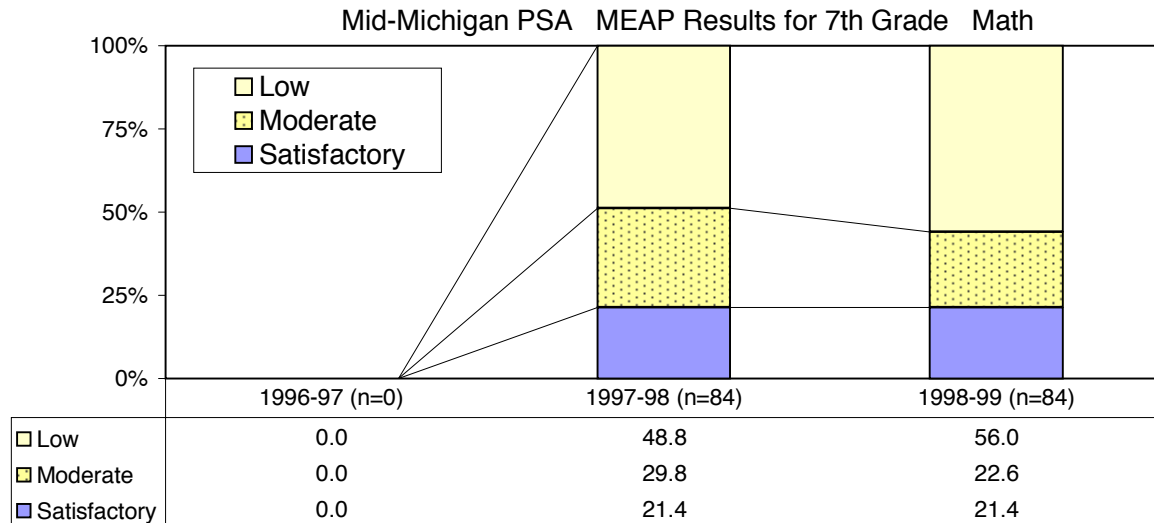


Exhibit 11:10 Grade 5 Science, MEAP Results for Mid-Michigan, District, and State

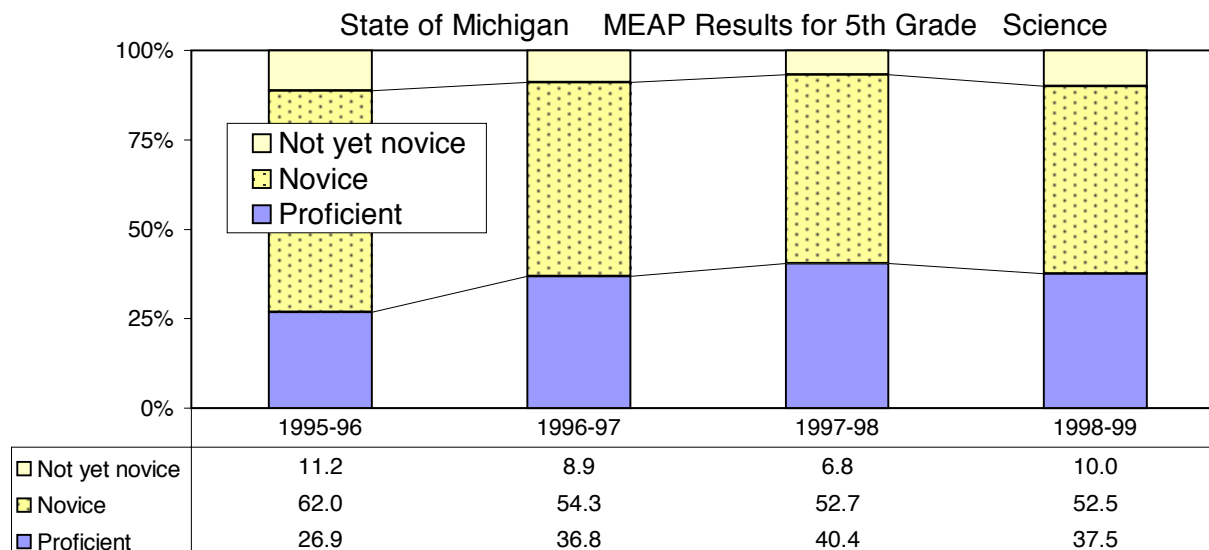
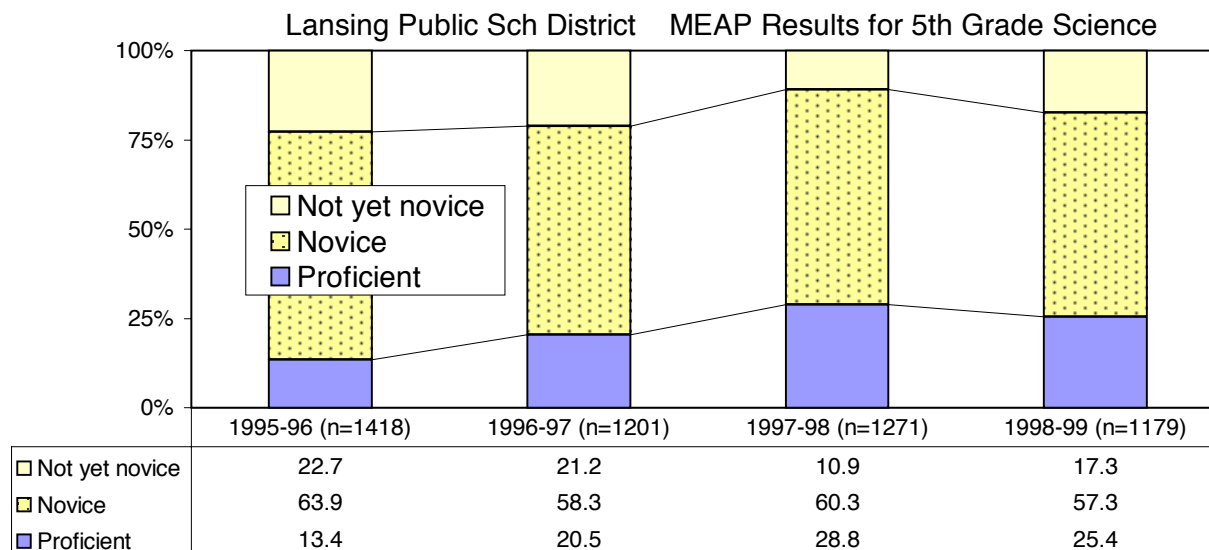
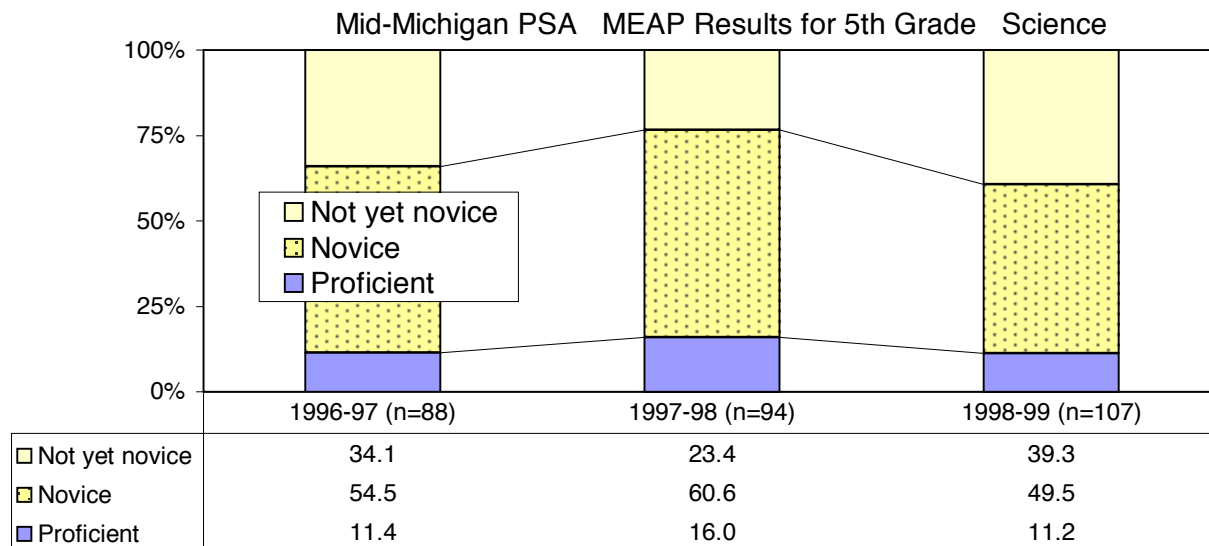


Exhibit 11:11 Grade 8 Science, MEAP Results for Mid-Michigan, District, and State

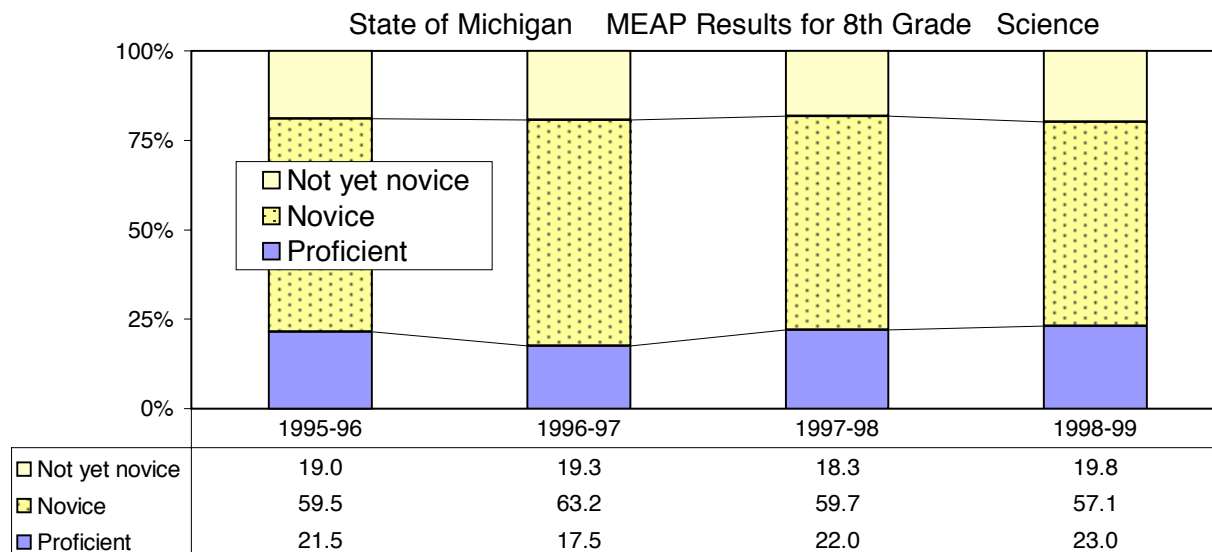
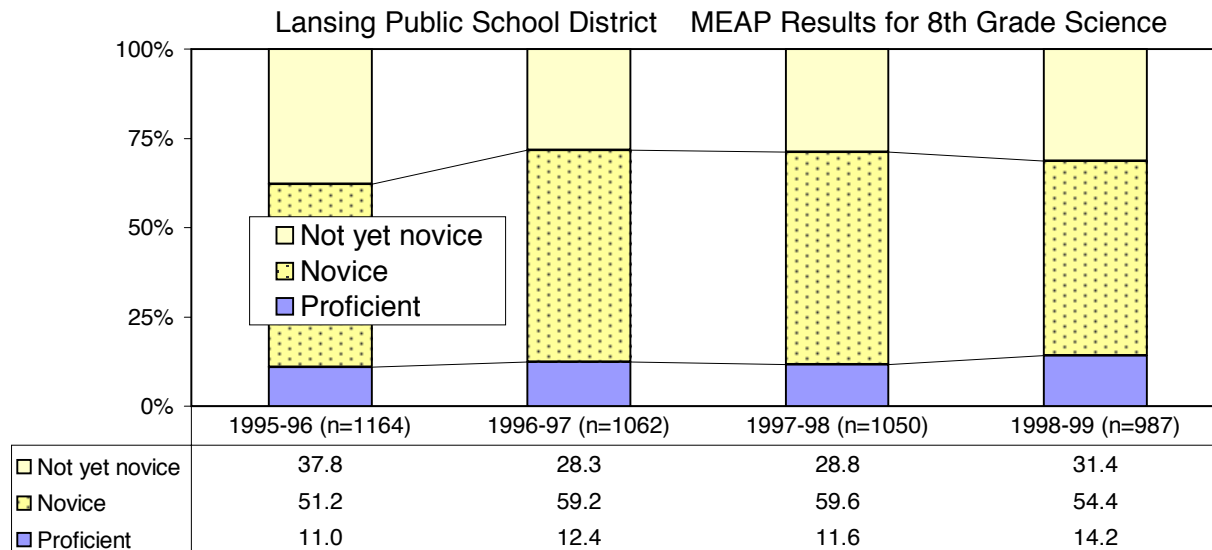
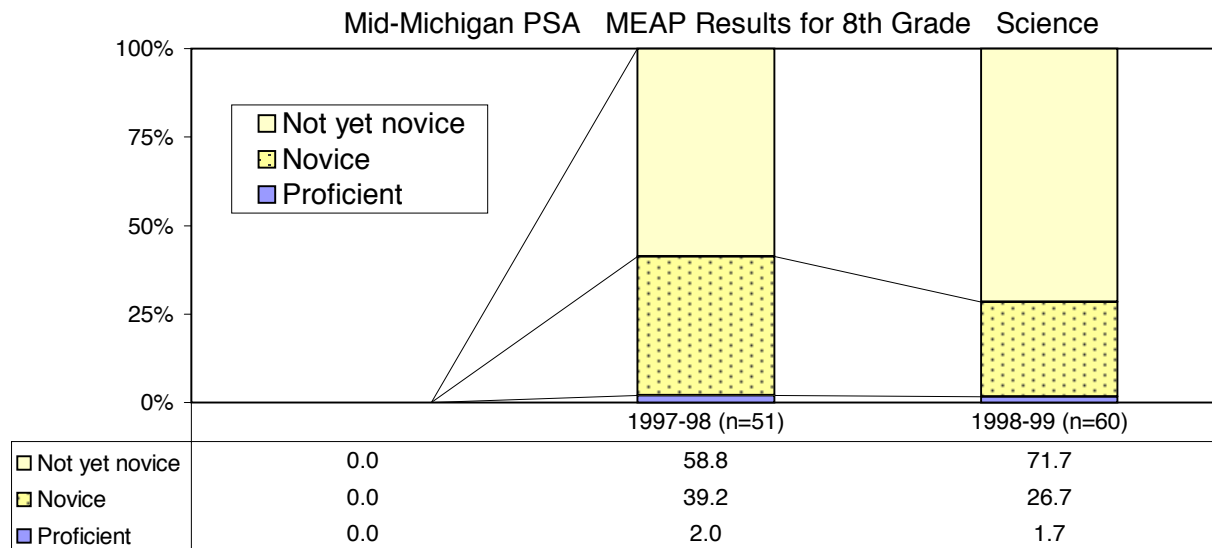


Exhibit 11:12 Grade 5 Writing, MEAP Results for Mid-Michigan, District, and State

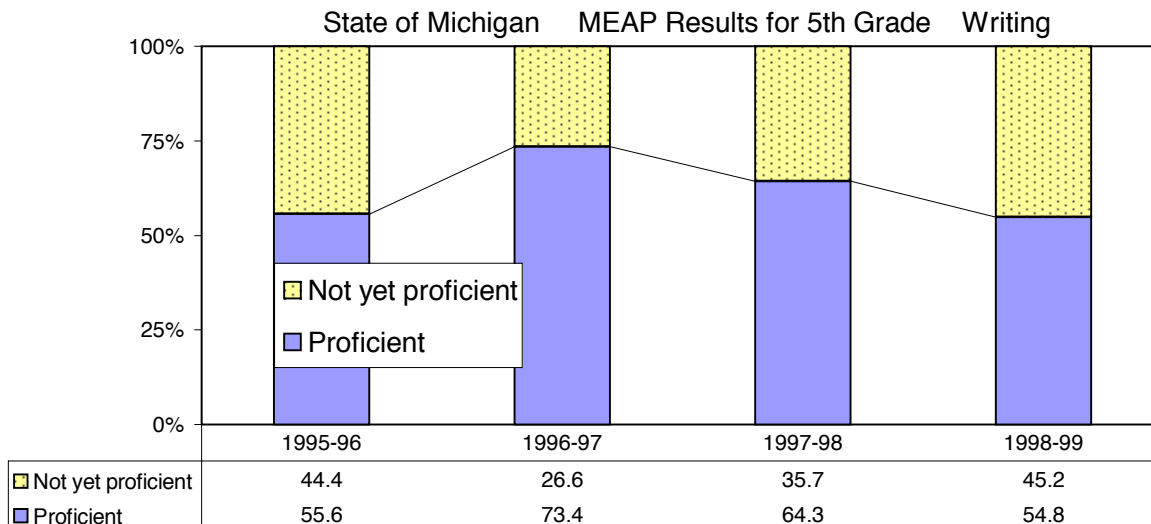
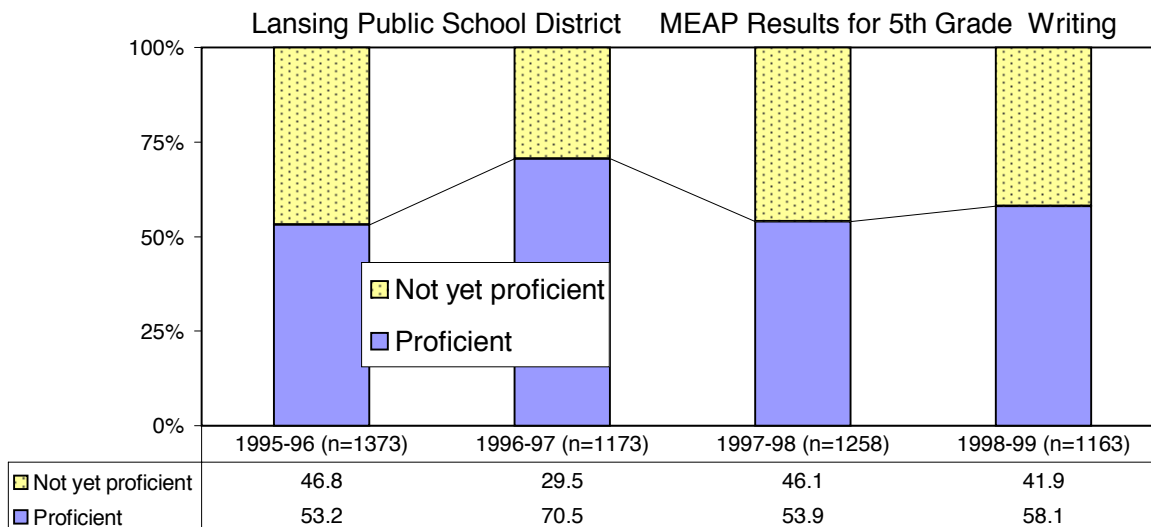
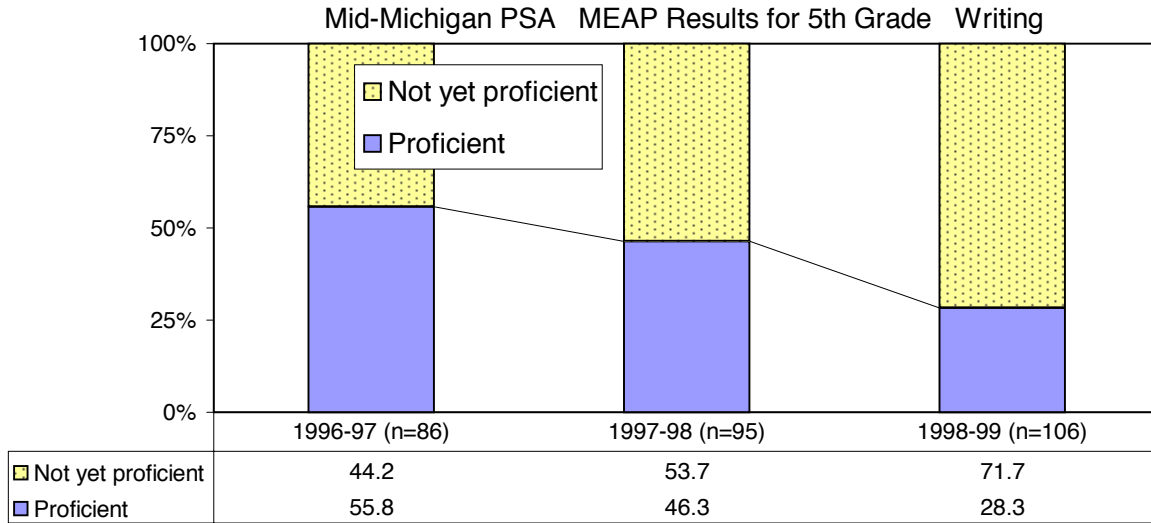


Exhibit 11:13 Grade 8 Writing, MEAP Results for Mid-Michigan, District, and State

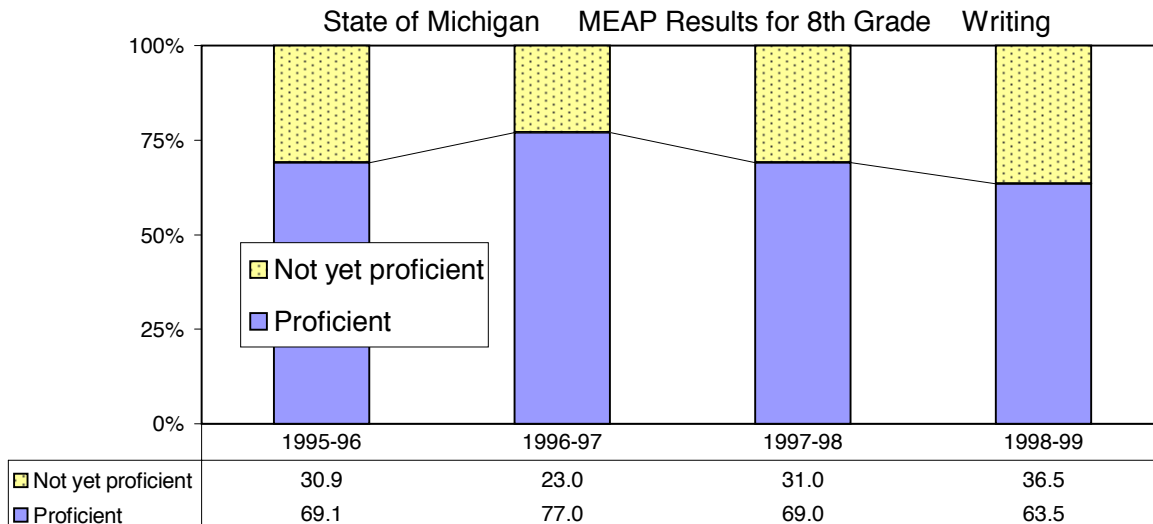
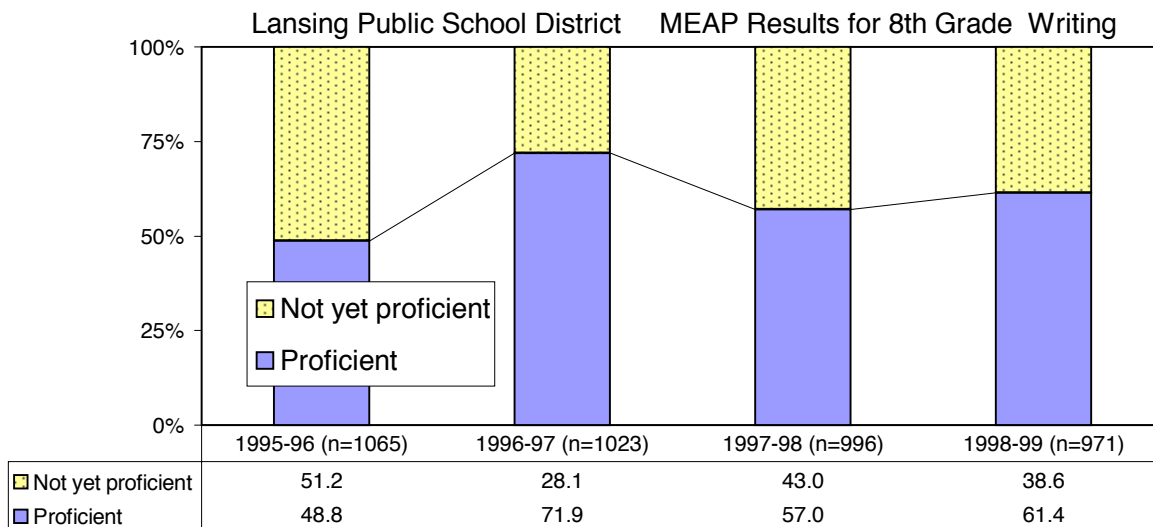
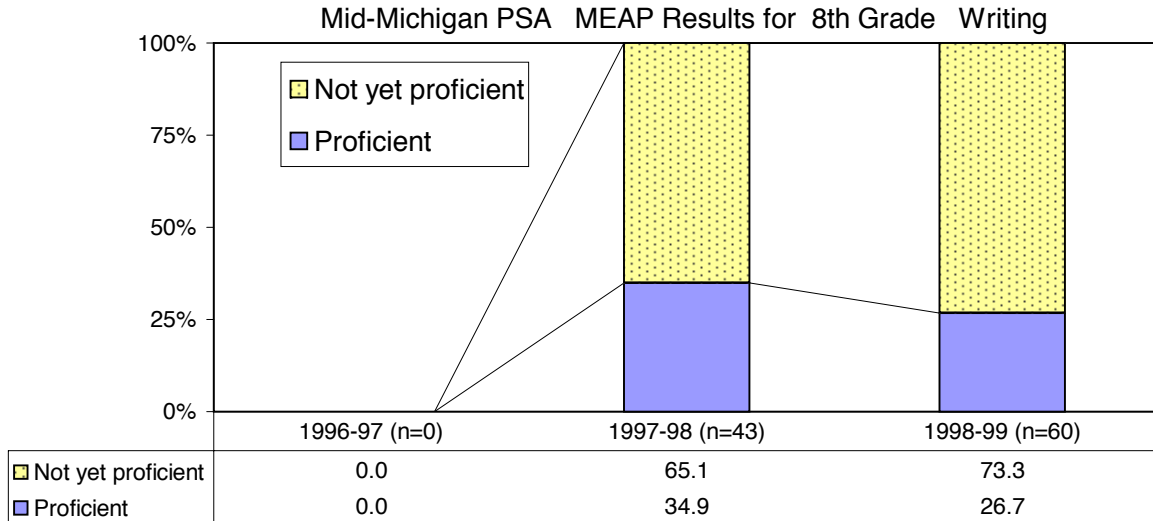
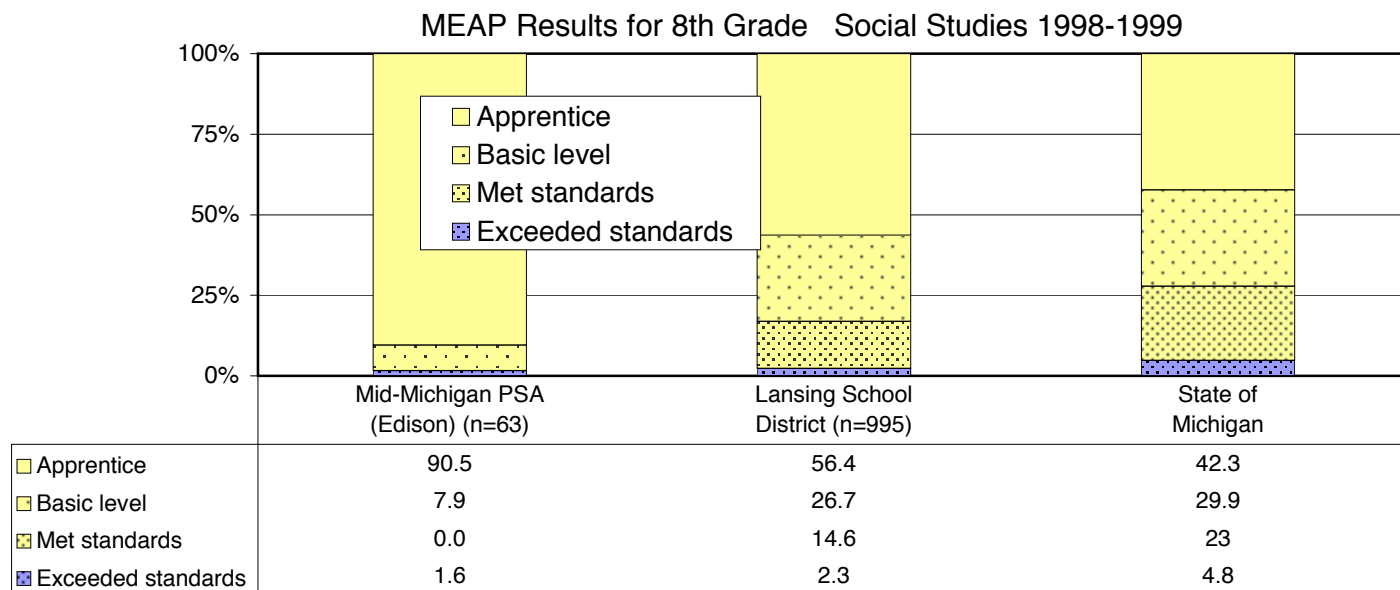
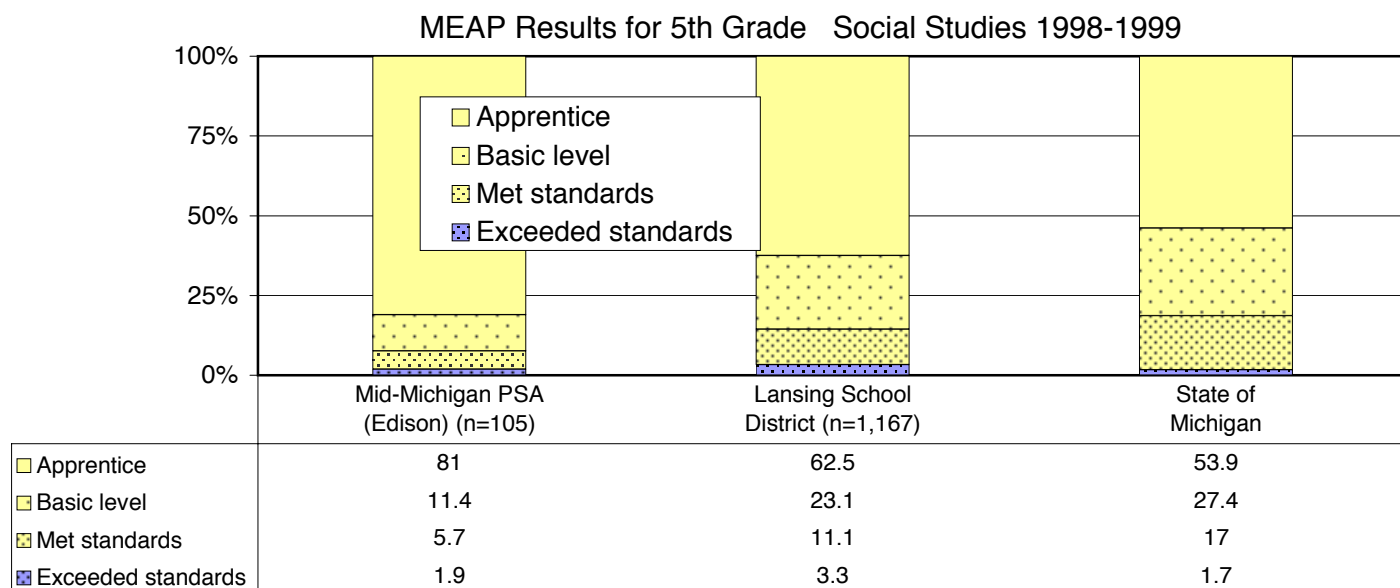


Exhibit 11:14 Grades 5 and 8 Social Studies: MEAP Results for Mid-Michigan, District, and State



11.7 Summary

Norm-referenced test findings

While Edison maintains that the gains in this school are “strongly positive,” based on our analysis this is a school that at best matches the student progress in other schools. In its second annual report Edison claimed that students were making great gains, which were measured by the Metropolitan Achievement Test. Our analysis of MAT-7 data do not corroborate this conclusion (Table 11:6). Our longitudinal analysis between 1997 and 1999 revealed that the Edison students went down slightly in most trends according to the national norm. In one subject, the decrease on the normal curve equivalent was significant over a two year period.

A summary score of -1 indicates a result that is unfavorable towards the sample school, a score of 0 indicates a neutral finding, and a score of 1 indicates a favorable result according to the criteria specified in Section 2.5. The effect size (ES) is the omega squared (ω^2) for a one way repeated measures ANOVA (Kepple, 1991).

Table 11:6 Summary of Results on Norm-Referenced Student Achievement Tests

Cohort A 3,4,5 grade (1997-99)	Standard Score		GE			NPR			NCE			Trend
	p-value	ES	p-value	Δ	ES	p-value	Δ	ES	p-value	Δ	ES	
Language	<.0001	.398	<.0001	2.8	.337	.1195	5.7	.010	.1365	5.0 (0)	.017	mixed (0)
Problem Solving(Math)	<.0001	.358	<.0001	2.0	.293	.4516	3.3	-.004	.2510	0.5 (0)	.006	mixed (0)
Reading Comprehension	<.0001	.258	<.0001	1.6	.222	.3653	-3.7	0.0	.2884	-3.1 (0)	.004	mixed (0)

Criterion-referenced test findings

Decisions regarding the OR were based on whether or not the $(1-\alpha)$ C.I. included 1.0. If the $(1-\alpha)$ C.I. fell completely below 1.0, this was interpreted as a protective odds ratio (1), thus favoring the Edison school. If the $(1-\alpha)$ C.I. included 1.0 (0) this was interpreted as an equal odds situation. If the $(1-\alpha)$ C.I. fell completely above 1.0 (-1), this was interpreted as an increase in odds for failing the state CRT relative to the comparison sample.

If the Breslow-Day statistic (B-D) is nonsignificant, one overall OR and $(1-\alpha)$ C.I. can be used to represent the odds for failing the CRT relative to the comparison group. Thus, there are no trends reported for each specific year in the tables, only a rating in the B-D column. If the B-D statistic is found to be statistically significant, then an overall common OR cannot be meaningfully interpreted; that is, there is a statistically significant change in the OR over years and thus yearly ORs are necessary. Thus, our summary ratings appear for each year of data and not in the B-D column (Table 11:7).

Table 11:7 Summary of Criterion-Referenced Test Results for Mid-Michigan

Mid-Michigan vs. District	1997	1998	1999	B-D	Trend
Grade 4 Reading				-1	negative (-1)
Grade 4 Math				-1	negative (-1)
Grade 5 Science				-1	negative (-1)
Grade 5 Writing	-1	0	-1		negative (-1)
Grade 7 Reading	no data			-1	negative (-1)
Grade 7 Math	no data			-1	negative (-1)
Grade 8 Science	no data			-1	negative (-1)
Grade 8 Writing	no data			-1	negative (-1)

Note: All state comparisons were negative.

Combined ratings

Given the total ratings for the trends that are highlighted in Table 11:8, we rate this school as Strongly Negative with a mean rating of -0.73. Our assessment of this school is similar to the assessment made by the school itself in its 1999 annual report (Mid-Michigan PSA, 1999). Edison's central office, however, reported a different picture. In its 1999 annual report, Edison rated this school as Mixed. In its 2000 annual report, it rated the 1999-2000 school year as Positive and the achievement gains since opening as Mixed.

Table 11:8 Combined Overall Trends for Mid-Michigan Public School Academy

	Positive	Mixed	Negative
Norm Referenced	0 of 3	3 of 3	0 of 3
Criterion Referenced	0 of 8	0 of 8	8 of 8
TOTALS	0 of 11	3 of 11	8 of 11

The design behind the trends in the norm-referenced results are based on tracing individual students over three years. While some would argue that this is a better design than was used with the criterion-referenced results (tracing consecutive cohorts of students), one also has to consider the sample sizes. The sample size for the norm-referenced test was very small. On the other hand, the test administration procedures are likely to be more regulated for the state-mandated MEAP, and the results we have for these trends include a much larger proportion of the total enrollment at the school.

Chapter Twelve

Washington Elementary School

Sherman, Texas

12.1 Descriptive Summary of School

Washington Elementary School is a district contract (partnership) school that Edison began to operate in 1995 to serve grades K-4. The picture of this school that is developed from available district and state data, as well as from the sketchy information contained in newspaper articles and in other reports, indicates that this is clearly among the poorest and most ethnically diverse schools in Sherman. It is no surprise that when a district contracts out one or more of its schools, it is most likely to select the schools with the poorest performance record and/or schools that require the most resources.

According to information obtained from the Campus AEIS Reports (TEA, 1995-2000), during the 1998/99 school year 445 students were enrolled at Washington Elementary. This is a rather small school for Edison. Of these, 23.8 percent were African Americans, 31.5 percent were Hispanic, 44.3 percent were white, and 0.4 percent were Native Americans. There were no students of Asian descent. From studying the descriptive data in Table 12:1, we can see that the enrollment at this school dropped by more than 10 percent over a 2-year period, while the total district enrollments dropped by a half percent.

The average number of pupils per teacher remained rather steady over the past 5 years. During the 1994/95 school year (pre-Edison) this figure was 14.6; since Edison took over the school, this figure has fluctuated between 13.0 and 16.5 pupils per teacher.

Since Edison took over the school, the enrollment of minority students has increased by around 10 percent, as compared with a 3 percent increase in the district. The campus comparison group, a set of 40 schools identified by the state that has similar demographics, showed a nearly 20 percent increase in minority students during the same period of time. There were 45 students (10.1 percent) enrolled in special education in 1998-99. In 1997-98 there were 47 (10.4 percent) enrolled in special education, and in 1996/97 there were 38 (7.6 percent).

While Edison reported that the student mobility in 1997-1998 was 7 percent (Edison, 1999), figures from the Academic Excellence Indicator System (AEIS) indicate the mobility¹ ran between 23.5 and 27.1 percent, which is approximately the same as in the district and in the campus comparison group. The difference in mobility figures is likely to be due to differences in the manner in which this indicator was calculated. The state average for mobility tends to run a little lower, at around 22 percent each year.

Table 12:1 Descriptive Information for Washington Elementary School (1994/95 - 1998/99)

	(Pre-Edison) 1994-95	1995-96	1996-97	1997-98	1998-99
<i>Enrollment</i>					
Washington Elem. (Edison)	369	492	499	454	445
Sherman ISD	5,777	5,964	6,149	6,049	6,109
Campus Group	50,912	18,948	19,362	18,835	21,632
State of Texas	3,670,196	3,740,260	3,828,975	3,891,877	3,945,367
<i>Number of students per teacher</i>					
Washington Elem. (Edison)	14.6	14.9	16.5	13.0	14.4
Sherman ISD	16.0	15.8	15.3	14.2	14.1
Campus Group	15.9	15.6	15.8	15.3	15.8
State of Texas	15.7	15.6	15.5	15.3	15.2
<i>Expenditures Per Pupil</i>					
Washington Elem. (Edison)	\$3,487	\$3,071	\$3,537	\$4,489	\$4,113
Sherman ISD	\$3,651	\$3,662	\$3,928	\$4,574	\$4,466
Campus Group	\$3,210	\$3,358	\$3,288	\$3,546	\$3,550
State of Texas	\$3,464	\$3,637	\$3,801	\$4,017	\$4,183
<i>Average Teacher Salary</i>					
Washington Elem. (Edison)	\$28,470	\$31,407	\$31,317	\$32,299	\$32,823
Sherman ISD	\$29,483	\$31,431	\$31,867	\$33,146	\$33,080
Campus Group	\$28,623	\$30,175	\$31,639	\$32,760	\$32,928
State of Texas	\$29,452	\$31,400	\$32,426	\$33,537	\$34,336
<i>Total Percent White</i>					
Washington Elem. (Edison)	52.0%	56.5%	54.9%	50.4%	44.3%
Sherman ISD	71.8%	71.4%	69.9%	68.5%	68.0%
Campus Group	36.2%	46.7%	43.2%	32.3%	22.6%
State of Texas	47.1%	46.4%	45.6%	45.0%	44.1%
<i>Special Education</i>					
Washington Elem. (Edison)	13.3%	9.3%	7.6%	10.4%	10.1%
Sherman ISD	12.6%	12.5%	12.7%	13.2%	13.8%
Campus Group	11.0%	11.9%	12.6%	12.1%	10.3%
State of Texas	11.1%	11.5%	11.6%	12.0%	12.1%

¹ A student is considered to be mobile if he or she has been at the school for less than 83 percent of the school year (i.e., has missed six or more weeks at a particular school) (AEIS Glossary, 1999).

	(Pre-Edison) 1994-95	1995-96	1996-97	1997-98	1998-99
<i>Economically Disadvantaged</i>					
Washington Elem. (Edison)	64.5%	68.7%	66.7%	75.8%	79.1%
Sherman ISD	35.6%	38.5%	38.2%	40.8%	41.4%
Campus Group	59.8%	68.8%	67.0%	75.7%	79.05
State of Texas	46.3%	46.9%	48.1%	48.5%	48.5%
<i>Percent Hispanic</i>					
Washington Elem. (Edison)	23.6%	22.0%	25.1%	28.6%	31.5%
Sherman ISD	7.5%	8.6%	9.6%	10.5%	11.2%
Campus Group	45.8%	28.5%	39.3%	50.9%	50.9%
State of Texas	36.1%	36.7%	37.4%	37.9%	38.6%
<i>Limited English Proficiency (LEP)</i>					
Washington Elem. (Edison)	20.6%	16.3%	18.8%	21.1%	24.3%
Sherman ISD	2.8%	3.1%	4.0%	4.4%	4.6%
Campus Group	14.6%	9.8%	15.6%	17.9%	26.0%
State of Texas	12.4%	12.8%	13.4%	13.4%	13.5%
<i>Average Attendance Rate</i>					
Washington Elem. (Edison)	95.6%	95.1%	95.2%	95.6%	
Sherman ISD	95.2%	95.0%	94.6%	94.9%	
Campus Group	96.5%	96.0%	96.0%	96.2%	
State of Texas	95.1%	95.1%	95.2%	95.3%	
<i>Mobility</i>					
Washington Elem. (Edison)	33.2%	23.5%	27.1%	25.9%	
Sherman ISD	23.4%	24.9%	26.8%	26.9%	
Campus Group	26.1%	23.2%	25.5%	28.4%	
State of Texas	21.5%	21.9%	22.1%	22.0%	

The school has had an increasingly high percentage of economically disadvantaged students who qualify for free or reduced lunch—79.1 percent in 1998/99, 75.8 percent in 1997/98, and 66.7 percent in 1996/97. Washington also has a high percentage of Spanish-speaking students. In 1998/99, 24.3 percent of the students had limited English proficiency (LEP). In 1997/98, 21.1 percent had limited proficiency and in 1996/97, 18.8 percent. Edison maintains that Washington's test scores improved despite this high percentage of economically disadvantaged and Spanish-speaking students. We found one year in which the school's students made large gains, but otherwise the performance of the school as measured by students' results on the TAAS has not been positive. In 1997 Washington was placed on a list of low-performing schools in Texas due to the low performance of Hispanic students in math on the TAAS. The school was later removed from the list after it was shown that the students were LEP.

In 1998/99 Washington had 39 total instructional staff, with 14.4 students per staff member. According to the AFT report (1998), although student enrollment increased between Edison's first and second year, the number of teachers and aides decreased, from 33 to 30 teachers and from 8 to 2 aides. Also, the percentage of teachers with less than 5 years' experience increased from 63 to 69 percent as compared with 30 percent in other schools in the district (TEA, 1997).

Total per pupil expenditures for 1998/99 were \$4,113, which is lower than the district figure of \$4,466. In 1997/98 total expenditures per pupil were \$4,489 and in 1996/97 were \$3,537. These funding levels are consistently lower than the average district expenditure levels, even though Washington Elementary has a higher level of low-income and LEP students. District officials reportedly did not agree that the AEIS figures were correct and calculated per pupil expenditures of around \$5,400 (Fowler, 1999). Average attendance rates for 1997/98 for Washington (95.6 percent) were nearly identical with the Sherman Intermediate School District (94.9 percent) and the state of Texas (95.3 percent).

Table 12:2 Trends in School Background Characteristics, Washington Elementary School

	(Pre-Edison) 1994/95*	1995/96	1996/97	1997/98	1998/99
Percent of Students Tested					
Washington Elem. (Edison)		96.1%	85.9%	80.9%	74.8%
Sherman ISD		94.7%	92.5%	92.3%	89.2%
Campus Group		89.7%	92.1%	89.7%	85.5%
State of Texas		89.6%	90.6%	91.1%	89.3%
Percent Exempted (LEP)					
Washington Elem. (Edison)	0.0%, 0.0%, 0.0%	0.0%	5.8%	12.6%	18.7%
Sherman ISD	0.2%, 0.1%, 0.2%	0.2%	1.3%	1.7%	2.2%
Campus Group	1.7%, 1.6%, 1.5%	0.7%	0.8%	1.5%	1.4%
State of Texas	5.3%, 5.2%, 5.3%	3.3%	2.4%	2.3%	2.2%
Percent Exempted (Special Ed. ARD)					
Washington Elem. (Edison)	1.8%, 8.5%, 1.8%	1.3%	7.3%	4.9%	4.4%
Sherman ISD	4.4%, 6.0%, 3.9%	3.6%	4.2%	4.5%	7.5%
Campus Group	9.1%, 9.5%, 8.5%	7.7%	4.3%	5.4%	9.1%
State of Texas	7.3%, 7.0%, 6.9%	5.5%	5.3%	5.2%	6.9%
Percent Exempted (Other)					
Washington Elem. (Edison)		1.9%	1.0%	1.6%	1.1%
Sherman ISD		0.9%	1.2%	0.9%	0.7%
Campus Group		0.6%	0.1%	0.2%	0.6%
State of Texas		0.8%	0.9%	0.7%	0.9%
Percent Exempted (Absent)					
Washington Elem. (Edison)		0.6%	0.0%	0.0%	0.0%
Sherman ISD		0.7%	0.9%	0.6%	0.4%
Campus Group		0.2%	0.0%	0.0%	0.0%
State of Texas		0.8%	0.8%	0.8%	0.7%

Notes: Students taking Spanish TAAS were LEP-exempt in 1995, but not in 1996.

The column figures for Washington Elementary, Sherman ISD, and the state consistently add up to 100 percent as expected, while figures for the campus comparison group fall short of 100 percent.

* Figures for the exemptions in 1994-95 could not be obtained across all subjects so they are reported separately. The three figures refer to exemption rates for reading, writing, and math, respectively.

Nonrenewal of contract

Edison and the district decided not to renew its contract with this school and the other school it operates in the Sherman district, Dillingham Intermediate School. This represents the first contract Edison has chosen not to renew. Both the Sherman School District and Edison are citing financial difficulties and a history of financial losses accrued by both parties since the management company took over Washington Elementary School in 1995. The Sherman School District says it spent \$4 million more on Edison schools than on its traditional public schools. Edison says it lost \$6.3 million (Doclar, 2000). The data from the state indicate that Edison was spending less per student than the district, even though the school apparently has students that are likely to require more resources. District officials did not agree with these figures, however. Further data on revenues as well as expenditures would help to sort this picture out.

Besides financial losses, Sherman board members are also concerned about test scores at the Edison schools that lag behind other Sherman elementary schools. Washington's scores on the Texas Assessment of Academic Skills (TAAS), while improving each year, are still the lowest of the district's six elementary schools, according to the Texas Education Agency. Edison cites the high percentage of Washington's students who are economically disadvantaged or have limited proficiency in English. Edison argues that to measure student achievement in an Edison school, an assessment test must be aligned with the objectives of its educational program. Edison personnel contend that traditional state assessment tests do not measure creativity or interdisciplinary problem solving, which is part of its educational program. Because this school has a large proportion of LEP students, many or most do not take the test, and exemptions from the state test have risen almost 25 percent at this school.

Sherman Intermediate School District, like the other districts that have contracted out one or more of their schools to Edison, chose its poor performing schools. Therefore, it is not surprising that Washington Elementary performs poorly in terms of absolute scores. Nonetheless, both Edison and the district expected that gains at the school would exceed those in the district. One last area of contention between Edison and the Sherman community has been the claims of district officials that Edison has not shared all test results with them. This includes test results from the Iowa Test of Basic Skills, as well as disaggregated data for Dillingham Intermediate School.

12.2 Past Studies and Evaluations and Available Test Data

In 1995-96 a reading study was conducted to compare Washington students (K-2) with a control group picked by Sherman district staff (Mislevy, 1996). The differences in gender, ethnicity, native language, and socioeconomic status between the Washington group and the control group were too great to provide reliable conclusions without making adjustments for these differences. Although unadjusted differences favored the control group, when variables were adjusted for language and demographic differences, there were no statistically significant differences between the groups (Mislevy, 1996). Because of excessive numbers of noncomparable students at Washington and the control group, this reading assessment study of Washington Elementary was terminated (Edison, 1999).

Tests administered at Washington Elementary include the Texas Assessment of Academic Skills (TAAS) in reading, math, and writing for Grades 3 and 4 ; the Mislevy Reading Study in 1995/1996

for Grade K-2; and the Iowa Test of Basic Skills (ITBS) in 1996 for Grade 1, in 1997 for Grade 2, and in 1998 for Grade 3.

Edison provided us no individual student data for norm-referenced tests at this school. Test data from the ITBS were not made available for this study. While it is understood that the ITBS is administered at this school, there is no reference to the results on the ITBS in either of Edison's two annual reports on student performance.

12.3 Analysis of Texas Assessment of Academic Skills Results

Like all Texas public schools, this school is required to take part in the state assessment program (TAAS). We secured TAAS test results for the 1995-96, 1996-97, 1997-98 and 1998-99 school years from the Texas Education Agency. This covers four of the five years that Washington Elementary was operated by Edison (1999-00 not available). Since the school provides only K-4 instruction, the only two grades from which we could obtain test data were grades 3 and 4.

The Texas state assessment program is well developed and in many respects has been a leader in terms of comprehensive reporting of school performance. The results available to the public include indicators that meet the state standards by subject, grade, and according to various background characteristics. A number of comparisons can be made with these data, including the following: (i) comparisons of consecutive class cohorts at a particular grade level; (ii) comparisons between the Edison school and the local school district average; (iii) comparisons between the Edison school and the campus group; and (iv) comparisons between the Edison school and the state average

Section 12.4 describes and discusses the results from the above-mentioned comparisons. Because there are so many years of data available, and because of the way the results are broken down by subject, grade, and demographic groups, the results from the comparison can yield extensive information about the progress of students in a given school. These comparisons are limited, however, in that they do not trace individual students or even groups of students. Rather, they compare one group of students with another succeeding group of students.

Nevertheless, the TAAS does allow for comparing groups of individual students by reporting on comparable improvements for each school. Comparable improvements compare a school with its control group of 40 demographically similar schools and, most importantly, are based on the growth in the students' Texas Learning Index ² scores on the TAAS reading and mathematics tests given from one year to the next. The comparable improvement is based only on students whose TAAS results can be matched from the current to prior year. Since only grades 3 and 4 are tested at Washington, the comparable improvement is based on the performance of fourth graders on the TAAS compared with how they performed in third grade (Texas Learning Index scores are used for comparing individual students from one year to the next). Section 12.5 describes and discusses the results on comparable improvement.

² The Texas Learning Index (TLI) is a score that describes a student's performance on the TAAS reading or mathematics test. It can be used to tell how far a student is above or below the passing standard.

Table 12:3 Washington Elementary Performance on TAAS by Grade and Subject, 1994-1999

	Grade	Year	State	District	Campus Group	Campus	African American	Hispanic	White	Native American	Asian/Pac.Is.	Male	Female	Econ. Disadv.	Special Educ.
Grade 3 TAAS % Passing															
Reading, including Spec.Ed.	3	1998-99	88.0%	86.5%	86.6%	75.8%	66.7%	87.5%	77.8%	-	-	83.3%	71.1%	65.9%	33.3%
Reading, including Spec.Ed.	3	1997-98	83.7%	85.5%	78.8%	74.2%	64.7%	62.5%	81.1%	-	-	75.9%	72.7%	66.7%	50.0%
Reading	3	1997-98	86.2%	88.6%	80.9%	76.8%	68.8%	57.1%	84.8%	-	-	81.5%	72.4%	68.3%	50.0%
Reading	3	1996-97	81.5%	81.5%	81.9%	69.6%	50.0%	50.0%	83.0%	-	-	63.2%	75.6%	65.3%	*
Reading	3	1995-96	80.5%	80.3%	75.0%	66.7%	63.6%	41.2%	78.0%	-	-	58.8%	74.3%	61.4%	33.3%
Reading	3	1994-95	79.5%	82.8%	78.8%	88.6%	100.0%	77.8%	88.0%	-	*	82.6%	95.2%	83.3%	*
Reading	3	1993-94	77.9%	77.4%	73.7%	75.0%	55.6%	*	79.3%	n/i	n/i	78.9%	71.4%	76.0%	*
Math, including Spec.Ed	3	1998-99	83.1%	81.9%	80.3%	73.1%	66.7%	77.8%	75.0%	-	-	79.3%	68.4%	64.6%	33.3%
Math, including Spec.Ed	3	1997-98	78.2%	78.8%	72.4%	72.1%	52.9%	50.0%	84.2%	-	-	72.4%	71.9%	66.7%	42.9%
Math	3	1997-98	81.0%	83.1%	74.1%	75.9%	56.3%	40.0%	90.9%	-	-	80.8%	71.4%	70.0%	42.9%
Math	3	1996-97	81.7%	76.9%	85.0%	53.8%	38.9%	14.3%	70.8%	-	-	53.8%	53.7%	55.1%	60.0%
Math	3	1995-96	76.7%	70.1%	75.3%	66.2%	45.5%	56.3%	75.6%	-	-	60.6%	71.4%	65.1%	50.0%
Math	3	1994-95	73.3%	74.2%	68.5%	68.2%	71.4%	66.7%	68.0%	-	*	66.7%	70.0%	62.5%	*
Math	3	1993-94	63.0%	59.7%	56.3%	58.5%	44.4%	*	62.1%	n/i	n/i	65.0%	52.4%	72.0%	20.0%
All Tests, including Spec.Ed	3	1998-99	78.9%	77.8%	74.7%	66.7%	57.9%	77.8%	68.3%	-	-	73.3%	61.5%	56.0%	27.3%
All Tests, including Spec.Ed	3	1997-98	73.3%	74.4%	64.1%	65.6%	44.4%	62.5%	76.3%	-	-	70.0%	61.8%	57.4%	37.5%
All Tests	3	1997-98	76.6%	79.3%	65.8%	69.6%	50.0%	57.1%	81.8%	-	-	77.8%	62.1%	61.0%	37.5%
All Tests	3	1996-97	74.2%	70.5%	75.0%	44.4%	26.3%	14.3%	60.4%	-	-	42.5%	46.3%	46.0%	60.0%
All Tests	3	1995-96	70.4%	65.6%	63.5%	53.6%	45.5%	29.4%	65.9%	-	-	47.1%	60.0%	52.3%	33.3%
All Tests	3	1994-95	67.4%	69.2%	64.2%	68.9%	71.4%	66.7%	69.2%	-	*	66.7%	71.4%	64.0%	*
All Tests			58.6%	54.7%	50.4%	51.2%	33.3%	*	55.2%	n/i	n/i	55.0%	47.6%	64.0%	20.0%
Grade 4 TAAS % Passing															
Reading, including Spec.Ed.			88.8%	89.2%	88.5%	74.1%	61.1%	50.0%	87.5%	-	-	76.0%	72.7%	66.7%	*
Reading, including Spec.Ed.			86.7%	87.9%	84.9%	81.7%	66.7%	81.8%	87.8%	*	-	78.4%	85.3%	77.6%	60.0%
Reading			89.7%	93.0%	90.6%	85.2%	80.0%	90.0%	86.1%	-	-	85.7%	84.8%	83.3%	60.0%
Reading			82.5%	83.0%	81.7%	64.5%	33.3%	58.8%	74.3%	-	*	65.5%	63.6%	65.9%	*
Reading	4	1995-96	78.3%	77.5%	69.4%	56.0%	55.6%	33.3%	58.6%	-	*	46.2%	66.7%	52.9%	*
Reading	4	1994-95	80.1%	78.6%	75.6%	50.0%	11.1%	50.0%	60.0%	-	*	52.4%	48.3%	50.0%	40.0%
Reading	4	1993-94	75.5%	81.5%	72.5%	71.1%	75.0%	66.7%	72.7%	n/i	n/i	70.6%	71.4%	64.3%	*
Writing, including Spec.Ed.	4	1998-99	88.4%	90.8%	87.4%	87.0%	88.9%	66.7%	90.0%	-	-	83.3%	90.0%	84.6%	*
Writing, including Spec.Ed.	4	1997-98	85.5%	75.4%	81.5%	64.3%	64.7%	54.5%	65.9%	*	-	58.3%	70.6%	56.3%	30.0%
Writing	4	1997-98	88.7%	80.7%	86.6%	70.0%	73.3%	66.7%	69.4%	-	-	67.9%	71.9%	65.9%	30.0%
Writing	4	1996-97	87.1%	81.8%	86.7%	53.2%	55.6%	58.8%	51.4%	-	*	41.4%	63.6%	46.5%	*
Writing	4	1995-96	86.3%	76.0%	80.5%	66.7%	60.0%	37.5%	76.7%	-	*	61.5%	72.0%	60.0%	*
Writing	4	1994-95	85.0%	86.1%	83.9%	69.4%	44.4%	70.0%	75.9%	-	*	65.0%	72.4%	63.3%	*
Writing	4	1993-94	85.5%	83.0%	85.9%	69.4%	62.5%	80.0%	66.7%	n/i	n/i	62.5%	75.0%	61.5%	*

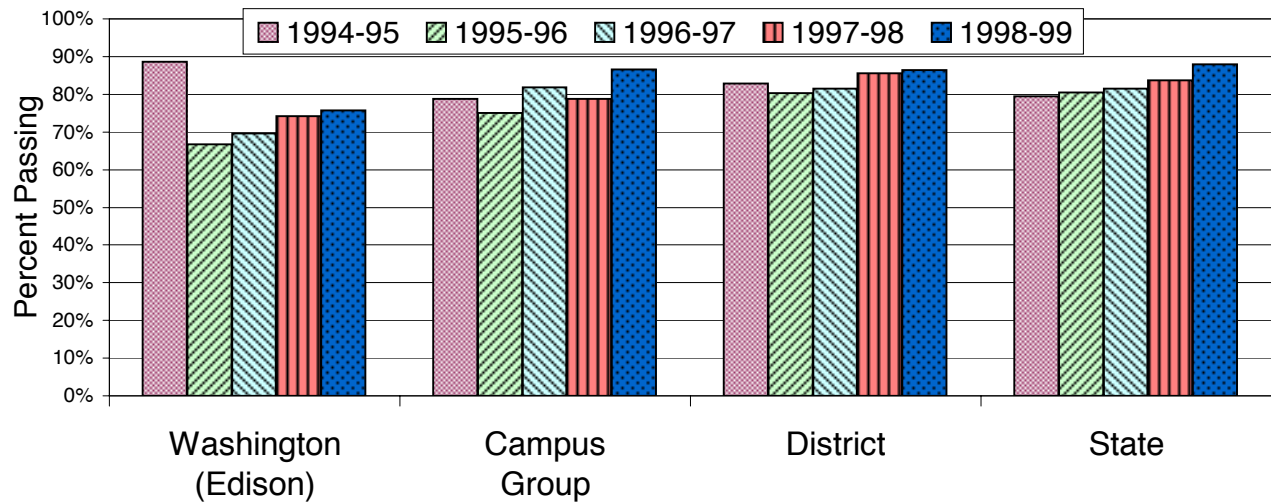
	Grade	Year	State	District	Campus Group	Campus	African American	Hispanic	White	Native American	Asian/Pac.Is.	Male	Female	Econ. Disadv.	Special Educ.
Grade 4 TAAS % Passing (Continued from previous page)															
Math, including Spec.Ed.	4	1998-99	87.6%	86.9%	86.6%	69.4%	57.9%	25.0%	85.7%	-	-	64.3%	73.5%	60.0%	57.1%
Math, including Spec.Ed.	4	1997-98	82.7%	77.6%	78.7%	82.4%	75.0%	90.0%	82.9%	*	-	82.9%	81.8%	80.0%	85.7%
Math			86.3%	82.0%	84.4%	82.0%	73.3%	88.9%	83.8%	-	-	82.8%	81.3%	80.5%	85.7%
Math			82.6%	74.3%	84.0%	48.4%	25.0%	41.2%	58.3%	-	*	48.3%	48.5%	48.8%	*
Math			78.5%	66.8%	71.7%	34.0%	22.2%	33.3%	35.7%	*	*	23.1%	45.8%	27.3%	*
Math			71.1%	68.1%	68.1%	52.0%	11.1%	50.0%	63.3%	-	*	57.1%	48.3%	56.7%	40.0%
Math			59.4%	54.9%	56.5%	31.6%	12.5%	33.3%	36.4%	n/i	n/i	23.5%	38.1%	32.1%	*
All Tests, including Spec.Ed.			78.4%	79.6%	76.4%	61.9%	50.0%	25.0%	77.1%	-	-	55.2%	67.6%	52.2%	50.0%
All Tests, including Spec.Ed.	4	1997-98	74.1%	66.0%	68.2%	59.5%	50.0%	58.3%	62.8%	*	-	55.3%	63.9%	51.0%	27.3%
All Tests	4	1997-98	78.6%	72.3%	73.4%	65.1%	60.0%	70.0%	65.8%	-	-	65.5%	64.7%	60.5%	27.3%
All Tests	4	1996-97	72.0%	65.5%	70.0%	38.1%	22.2%	41.2%	41.7%	-	*	31.0%	44.1%	38.6%	*
All Tests	4	1995-96	67.2%	55.4%	52.6%	30.2%	20.0%	22.2%	33.3%	*	*	18.5%	42.3%	25.0%	*
All Tests	4	1994-95	64.1%	61.8%	60.0%	41.2%	11.1%	45.5%	46.7%	-	*	42.9%	40.0%	45.2%	20.0%
All Tests	4	1993-94	54.8%	50.9%	50.0%	28.9%	12.5%	33.3%	36.4%	n/i	n/i	23.5%	33.3%	28.6%	*
1999 TAAS Participation															
Tested		1999-99	89.3%	89.2%	85.5%	75.8%	93.2%	31.5%	95.2%	-	-	65.9%	85.7%	70.7%	
Accountability Subset		1999-99	84.2%	84.8%	77.0%	72.5%	88.6%	31.5%	90.5%	-	-	64.8%	80.2%	68.6%	
Mobile Subset		1999-99	4.6%	4.3%	5.7%	3.3%	4.5%	0.0%	4.8%	-	-	1.1%	5.5%	2.1%	
Sci and/or Soc St Only		1999-99	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	-	-	0.0%	0.0%	0.0%	
Spanish (gr 4, 5, 6)		1999-99	0.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	-	-	0.0%	0.0%	0.0%	
Not Tested		1999-99	10.7%	10.8%	14.6%	24.2%	6.8%	68.5%	4.8%	-	-	34.1%	14.3%	29.3%	
Absent		1999-99	0.7%	0.4%	0.0%	0.0%	0.0%	0.0%	0.0%	-	-	0.0%	0.0%	0.0%	
ARD Exempt		1999-99	6.9%	7.5%	9.1%	4.4%	6.8%	1.9%	4.8%	-	-	7.7%	1.1%	5.0%	
LEP Exempt		1999-99	2.2%	2.2%	1.4%	18.7%	0.0%	63.0%	0.0%	-	-	24.2%	13.2%	22.9%	
Other		1999-99	0.9%	0.7%	0.6%	1.1%	0.0%	3.7%	0.0%	-	-	2.2%	0.0%	1.4%	
Total Answer Documents		1999-99	2,081,638	3,209	205	182	44	54	84	0	0	91	91	140	
1998 TAAS Participation															
Tested		1997-98	91.1%	92.3%	89.7%	80.9%	95.2%	43.5%	92.4%	100.0%	-	78.5%	83.3%	76.3%	
Accountability Subset		1997-98	76.0%	78.6%	70.4%	65.0%	73.8%	37.0%	77.2%	0.0%	-	60.2%	70.0%	60.4%	
Mobile Subset		1997-98	4.4%	4.9%	4.4%	4.9%	9.5%	0.0%	3.3%	50.0%	-	4.3%	5.6%	5.0%	
Special Education		1997-98	8.7%	8.9%	10.0%	10.9%	11.9%	6.5%	12.0%	50.0%	-	14.0%	7.8%	10.8%	
Spanish (gr 3-6)		1997-98	2.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	-	0.0%	0.0%	0.0%	
Not Tested		1997-98	8.9%	7.7%	10.4%	19.1%	4.8%	56.5%	7.6%	0.0%	-	21.5%	16.7%	23.7%	
Absent		1997-98	0.8%	0.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	-	0.0%	0.0%	0.0%	
ARD Exempt		1997-98	5.2%	4.5%	5.4%	4.9%	4.8%	0.0%	7.6%	0.0%	-	7.5%	2.2%	5.0%	
LEP Exempt		1997-98	2.3%	1.7%	1.5%	12.6%	0.0%	50.0%	0.0%	0.0%	-	11.8%	13.3%	16.5%	
Other		1997-98	0.7%	0.9%	0.2%	1.6%	0.0%	6.5%	0.0%	0.0%	-	2.2%	1.1%	2.2%	
Total Answer Documents		1997-98	2,052,472	3,177	211	183	42	46	92	2	0	93	90	139	

Source: Adapted from the Academic Excellence Indicator System, Campus Performance Report for Washington Elementary 1998-99 (Texas Education Agency, 2000)

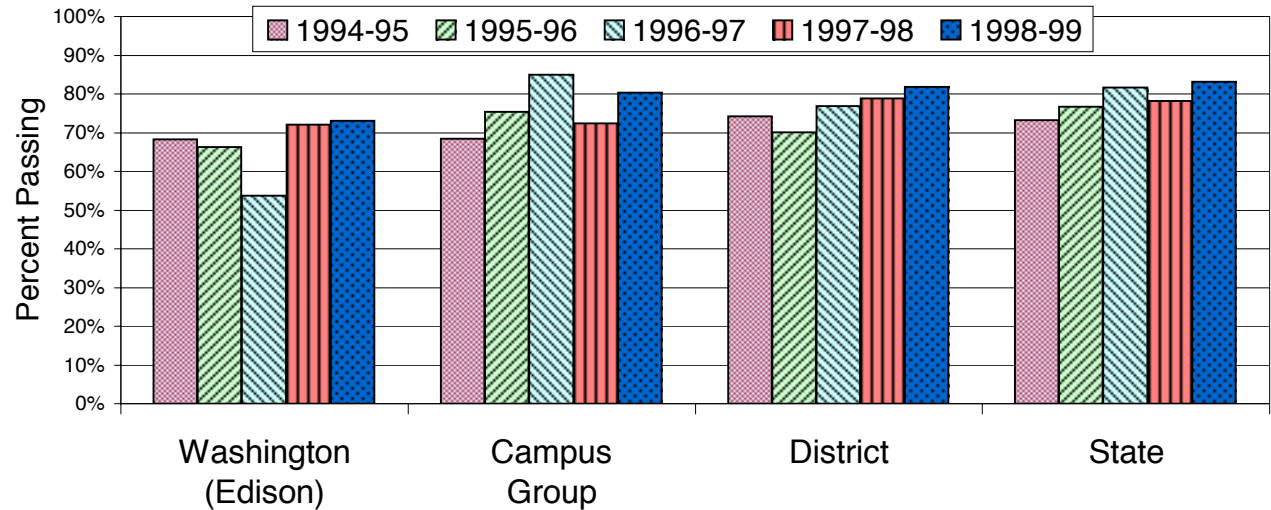
Note: The accountability rating was "Acceptable" for 1998-99. Enrollment was indicated as 445 for this school year.

Exhibit 12:1 Washington Elementary School, TAAS Results 1994-1999
Performance on Grade 3 Tests Compared with Campus Comparison Group,
Local School District, and the State

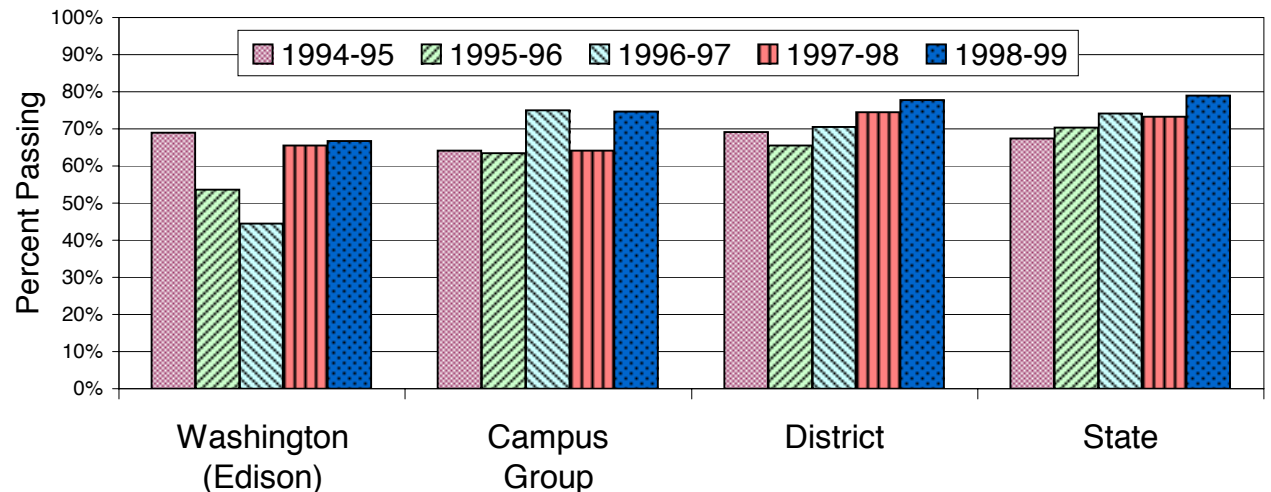
Grade 3 Reading



Grade 3 Math

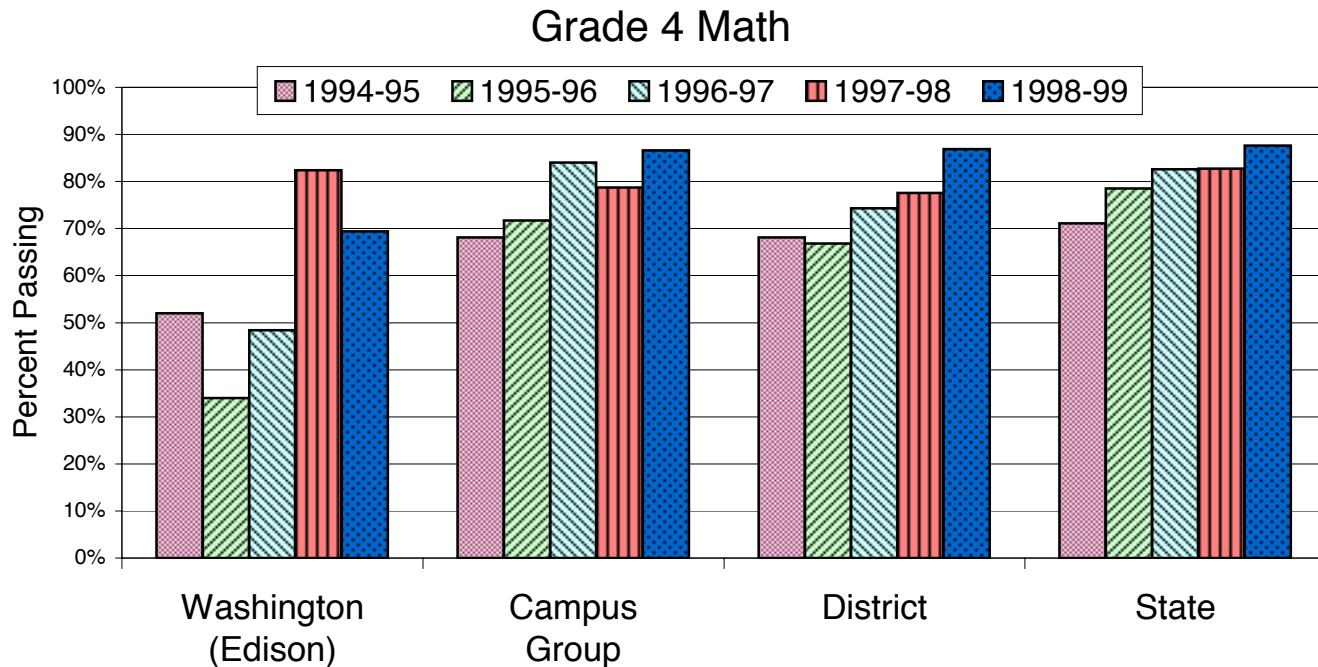
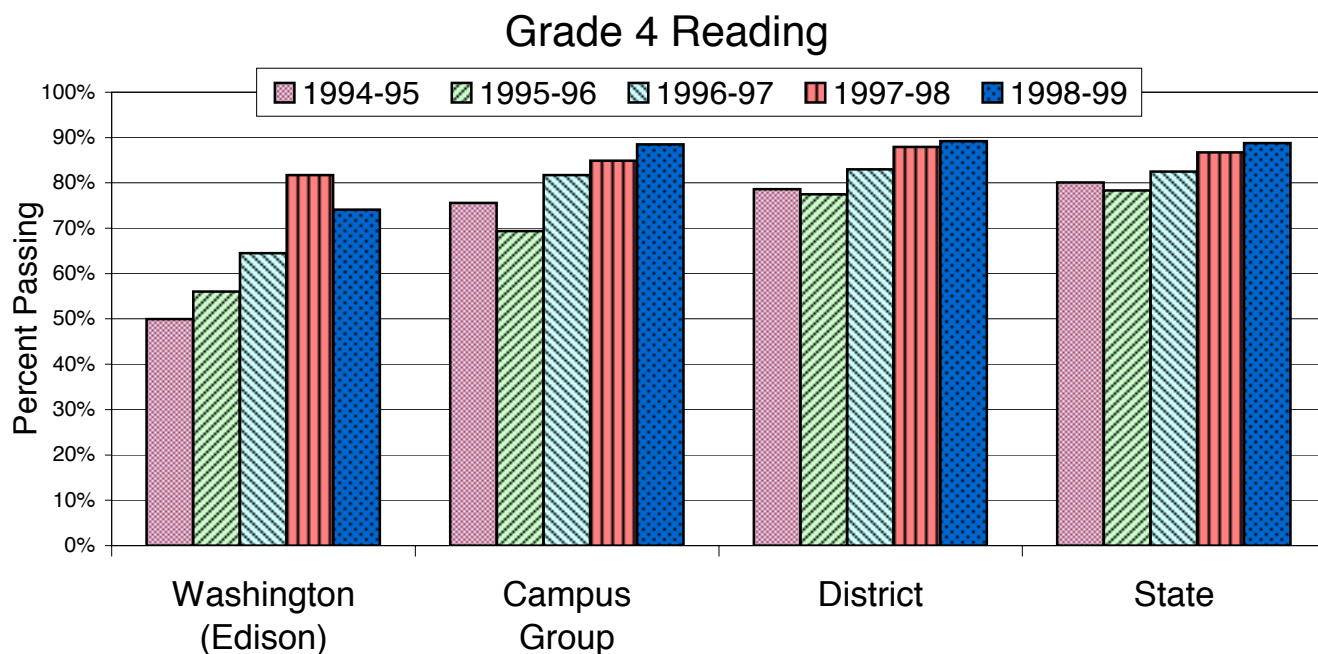


Grade 3 Reading and Math Combined



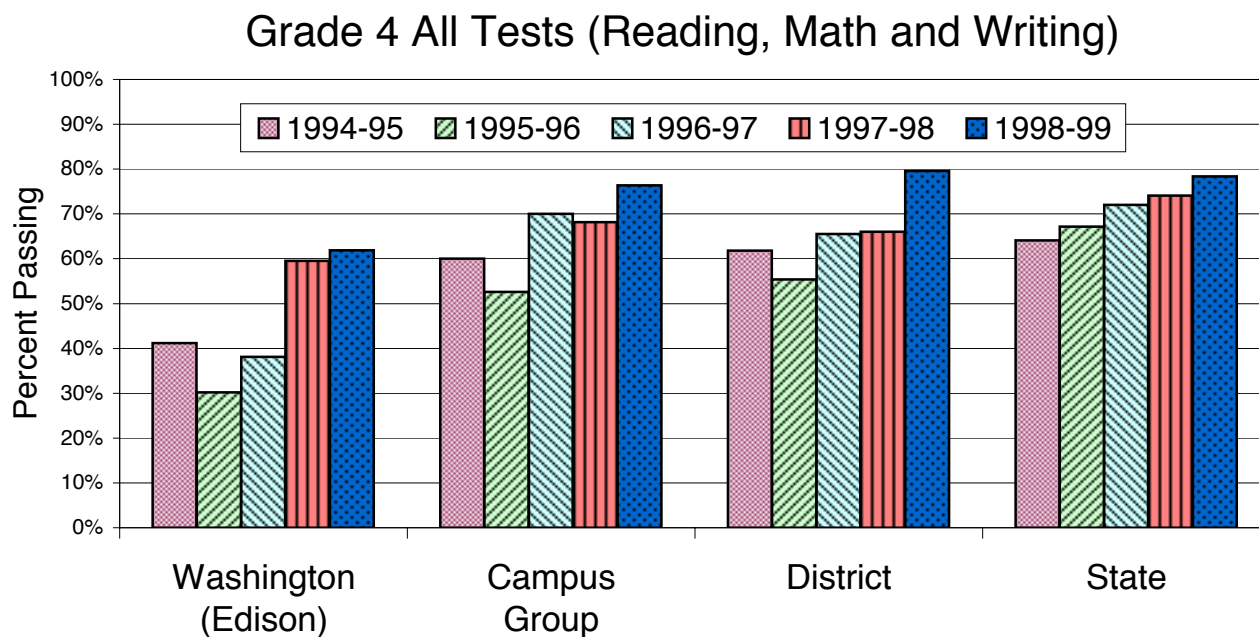
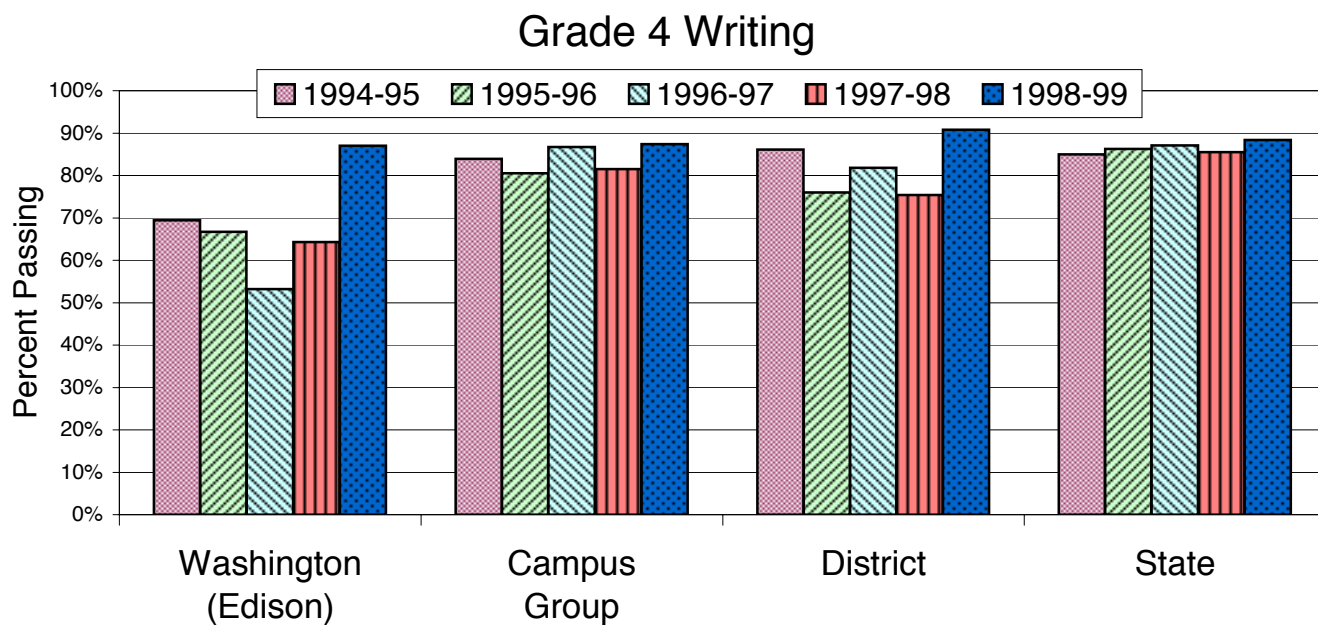
Note: The 1994-95 and 1995-96 school years are pre-Edison. All tests are administered in the spring. Figures for the 1997-98 and the 1998-99 school years include special education students.

Exhibit 12:2 Washington Elementary School, TAAS Results 1994-1999
Performance on Grade 4 Tests Compared with Campus Comparison Group,
Local School District, and the State



Note: The 1994-95 and 1995-96 school years are pre-Edison. All tests are administered in the spring. Figures for the 1997-98 and the 1998-99 school years include special education students.

Exhibit 12:3 Washington Elementary School, TAAS Results 1994-1999
Performance on Grade 4 Tests Compared with Campus Comparison Group,
Local School District, and the State



Note: The 1994-95 and 1995-96 school years are pre-Edison. All tests are administered in the spring. Figures for the 1997-98 and the 1998-99 school years include special education students.

12.4 Comparable Growth on the TAAS Based on Consecutive Class Cohorts

Table 12:3 summarizes the key results for Washington Elementary School and the three comparison groups (i.e., district, campus group, and the state) on the Texas Assessment of Academic Skills (TAAS). Results for Washington are also broken down by ethnic group and gender. Separate results are included for students from families that are economically disadvantaged and students who have special educational needs. The results are grouped by grade and subject. Each row in each section includes annual results, and these are listed chronologically from 1994 to 1999. Recall that Edison took over operation of the school during the 1995-96 school year so the results for 1994 and 1995 are pre-Edison results. The first results that can be attributed to Edison are the data for 1996, which are for the 1995-96 school year. TAAS test administration is always in the spring of each year. All results in Table 12:3 refer to the percent of students in each group that passed the TAAS.

Previous to 1998, the results were reported separately for students in the regular education program and students receiving special education services. After 1998, the school totals include both regular and special education students. For this reason the Texas Education Agency posted two sets of results for 1998, to facilitate comparisons for the year before and year after. Nevertheless, some students are still excluded from the TAAS; but this depends on a decision from a student's Admission, Review, and Dismissal (ARD) Committee. ARD committees are comprised of parent(s)/guardians, teachers, administrators, and other concerned parties. Table 12:2 includes information on the percent of students with disabilities who are still excluded from the TAAS (see rows designated as special education, ARD). During the first year Edison operated the school, 1.3 percent of the students with disabilities were exempted. This figure jumped to 7.3 percent the following year and then dropped below 5 percent during the latter years. In recent years, the comparison groups had a higher proportion of students with disabilities exempted from the TAAS due to decisions from the ARD.

A careful read of the results in the table will illustrate that students at Washington Elementary School have not matched the gains made in the campus comparison group, district, and state. Exhibits 12:1, 12:2, and 12:3 contain bar charts that illustrate the changes in the percent of students passing the TAAS each year. While the results included in Table 12:3 include two years of pre-Edison data, we have included only one year of pre-Edison results (1994-95) in the charts in Exhibits 12:1, 12:2, and 12:3. The first column for the Washington Edison bar charts are for the 1994-95 school year.

TAAS results for grade 3

Grade 3 results are reported for reading, math, and all tests (i.e., combination of reading and math). In all three series of bar charts for Grade 3, one can see that there were initial decreases in the percentage of students meeting the state standard during the first one or two years that the school was operated by Edison. Starting in the 1996-97 school year the reading scores begin to rise, and in the 1997-98 school year the math scores begin to increase. While the school did not return to the pre-

Edison scores in reading, the percentage of students passing the math component of TAAS eventually surpassed the pre-Edison score in math. In all three control groups, the annual gains were generally linear, with incremental gains each year in the percentage of students meeting the state standard. In terms of absolute passing rate, the score for grade 3 math and reading remain lower than the three control groups.

On the whole, white grade 3 students perform better on the TAAS than the African-American and Hispanic students. An interesting pattern among Hispanic students evolved over the years that this school has been operated by Edison. While the scores for white and African-American students generally decreased over time, the Hispanic students exhibited gains each year in grade 3 TAAS. In fact, there were between 10 and 15 percent gains each year in reading for this group; and while only 14 percent of the Hispanic students passed the math test in 1995-96, this figure increased to 77.8 percent in 1998-99. These gains are extremely large and stand out clearly among all the results for this school. Table 12:4 below includes these findings. In order to explain one likely reason for these gains, we also included data on the percent of students exempted from TAAS due to limited English proficiency. There is a clear correlation between increases in passing rates among Hispanic students and the percent of students exempted due to LEP. As can be seen in Table 12:2, exemptions for LEP among the district, campus comparison group, and the state as a whole remain largely the same from 1994-95 to 1998-99 and range from 0.3 percent to 3.3 percent. Likewise, the results for Hispanic students among the three control groups more or less parallel the growth among the other ethnic groups.

Table 12:4 Percent Hispanic Students Passing TAAS Compared with Exemptions Because of Limited English Proficiency

Percent of Students Passing TAAS at Washington	Pre-Edison 1994/95	1995/96	1996/97	1997/98	1998/99
Reading					
White	88.0%	78.0%	83.0%	81.1%	77.8%
African American	100.0%	63.6%	50.0%	64.7%	66.7%
Hispanic	77.8%	41.2%	50.0%	62.5%	87.5%
Math					
White	68.0%	75.6%	70.8%	84.2%	75.0%
African American	71.4%	45.5%	38.9%	52.9%	66.7%
Hispanic	56.3%	14.3%	40.0%	50.0%	77.8%
Percent Exempted (LEP)	0.0%	0.0%	5.8%	12.6%	18.7%

Note: Passing rates for 1997-98 and 1998-99 include special education students.

TAAS results for grade 4

Exhibits 12:2 and 12:3 contain the Washington-Edison results for grade 4. In grade 4, students are tested in reading, math, and writing. The results are reported separately for each subject as well as for all subjects combined. The percentage of grade 4 students passing the TAAS grade 4 increased gradually each year. In math and writing there were initial decreases in the percentage of students meeting the state standard during the first one or two years that the school was operated by Edison.

Table 12:5 Performance on TAAS for All Students Relative to Local District and State 1995-1999

	Pre-Edison					Washington (Edison)	District	State	Difference in Gains Compared with District	Difference in Gains Compared with State
	1995	1996	1997	1998	1999 *	CHG: 95-99	CHG: 95-99	CHG: 95-99	1995 to 1999	1995 to 1999
TAAS ALL TESTS TAKEN										
ALL STUDENTS	54.2%	43.4%	41.7%	67.2%	64.4%	10.2%	17.1%	17.6%	-6.9%	-7.4%
AFRICAN AMERICAN	37.5%	33.3%	25.0%	54.8%	53.8%	16.3%	22.5%	25.7%	-6.2%	-9.4%
HISPANIC	55.0%	26.9%	29.0%	64.7%	52.9%	-2.1%	35.6%	24.0%	-37.7%	-26.1%
WHITE	57.1%	52.1%	52.4%	73.2%	72.4%	15.3%	14.9%	13.1%	0.4%	2.2%
ECONOMICALLY DISADVANTAGED	53.6%	40.0%	42.6%	60.7%	54.2%	0.6%	19.7%	23.1%	-19.1%	-22.5%
TAAS READING										
ALL STUDENTS	68.1%	62.2%	67.4%	81.2%	75.0%	6.9%	8.8%	8.1%	-1.9%	-1.2%
AFRICAN AMERICAN	50.0%	60.0%	44.4%	74.2%	63.9%	13.9%	13.2%	15.2%	0.7%	-1.3%
HISPANIC	63.2%	38.5%	54.8%	76.5%	68.8%	5.6%	27.7%	11.6%	-22.1%	-6.0%
WHITE	72.7%	70.0%	79.3%	85.5%	82.4%	9.7%	6.3%	5.3%	3.4%	4.4%
ECONOMICALLY DISADVANTAGED	64.8%	57.7%	65.6%	75.9%	66.3%	1.5%	14.0%	12.1%	-12.5%	-10.6%
TAAS MATHEMATICS										
ALL STUDENTS	59.6%	52.5%	51.4%	79.1%	71.3%	11.7%	18.7%	19.8%	-7.0%	-8.1%
AFRICAN AMERICAN	37.5%	35.0%	34.6%	64.5%	62.2%	24.7%	25.9%	29.0%	-1.2%	-4.3%
HISPANIC	57.9%	48.0%	29.0%	71.4%	52.9%	-5.0%	35.5%	28.4%	-40.5%	-33.4%
WHITE	65.5%	59.4%	65.5%	87.1%	80.0%	14.5%	16.2%	13.3%	-1.7%	1.2%
ECONOMICALLY DISADVANTAGED	59.3%	48.7%	52.2%	75.3%	62.4%	3.1%	22.2%	27.3%	-19.1%	-24.2%
TAAS WRITING										
ALL STUDENTS	69.4%	66.7%	53.2%	70.0%	87.0%	17.6%	9.7%	6.2%	7.9%	11.4%
AFRICAN AMERICAN	44.4%	60.0%	55.6%	73.3%	88.9%	44.5%	21.8%	11.4%	22.7%	33.1%
HISPANIC			58.8%	66.7%	66.7%	-3.3%	28.9%	9.7%	-32.2%	-13.0%
WHITE			51.4%	69.4%	90.0%	14.1%	5.8%	3.4%	8.3%	10.7%
ECONOMICALLY DISADVANTAGED			46.5%	65.9%	84.6%	21.3%	19.7%	9.9%	1.6%	11.4%
ATTENDANCE RATE, GRADES 1-12										
ALL STUDENTS			95.1%	95.2%	95.6%	0.7%	-0.4%	0.2%	1.1%	0.5%

Source: Adapted from "Selected AEIS Campus Data: A Multi-year History" (Texas Education Agency, 1999)

* Beginning in 1999, results also include special education test takers and Spanish test takers for grades 3 and 4, reading and mathematics

Starting in the 1996-97 school year, the math scores began to rise, and the writing scores began to climb in the 1997-98 school year. The writing scores eventually increased to 87 percent passing, which is comparable to the campus group and state average and just under the passing rate of the Sherman School District, which was 90.9 percent. After the initial drop in scores when Edison began to operate the school, the school eventually surpassed the pre-Edison scores in math and reading. In all three control groups, the annual gains were generally linear, with incremental gains each year in the percentage of students meeting the state standard.

The pattern among grade 4 Hispanic students does not parallel the pattern that emerged in grade 3. While Hispanic students did make notable gains, they performed exceptionally poorly during the most recent test administration.

Changes in performance relative to district and state

Table 12:5 illustrates the relative growth on the TAAS at Washington Elementary School in comparison with district and state trends. The table combines results for grade 3 and 4 students, but presents the results by subject as well as for all subjects combined. The first set of data at the top of the table presents the combined results across all grades and subjects. One can see that the gains made in this school range from -2 percent to 15 percent between 1995 and 1999. The first year of data represents the school for the year before Edison began operating it. Ironically, the Hispanic students performed the worst between the 1994-95 and 1998-99 school years at Washington School. In terms of total differences, they showed a drop of 2 percent in terms of percent of students passing the test. If we look at the total difference among Hispanic students in Edison's first year of operation (1995-96) and the most recent test results (1998-99), we see that they are the group that made the largest gains (i.e., 35.6 percent). The explanation for this is that the Hispanic students performed extremely well in the year before Edison and performed extremely poorly in Edison's first year of operation, and then increased their scores over the next three years and approached the pre-Edison levels.

The pattern in terms of exemptions on the TAAS is of particular interest at this school since it differs greatly from the pattern of exemptions in the 3 control groups. More than half of all Hispanic students were excluded from the TAAS in 1997-98, most due to LEP exemption. During the 1998-99 TAAS, 68.5 percent of all Hispanic students were excluded from the TAAS at Washington Elementary (63 percent due to LEP and the remainder due to ARD or other exemptions). Exemptions for special education at this school are almost exclusively for white and black students, with 4.8 percent and 4.5 percent. Only 1.9 percent of the Hispanic students were exempted because of special educational needs.

The comparisons and findings based on the data in Table 12:5 can be helpful in examining the relative changes in performance levels at Washington Elementary. Nevertheless, the design for these comparisons is limited because the data focus on consecutive classes of students rather than on individual student gains. Additionally, the comparison groups (the district and state, in this case) combine results on the TAAS for grades 3-8 plus grade 10, while the performance data for Washington is based only on grades 3 and 4. The comparisons discussed earlier, based on the results

in Table 12:3, are not limited in this way since the data are presented and analyzed by grade as well as by subject. Therefore, the findings in Table 12:5 were included to provide further detail, but because of the limitation in making comparisons, we will not consider these findings in our summary.

12.5 Comparable Improvement Results Based on Individual Students' Gains on the Texas Learning Index

The Texas Education Agency provides extensive data on each public school in the state. Included in the publicly available data are indicators of growth based upon individual student results. Comparable improvements can be compared with the campus group of 40 demographically similar schools. Gains are based on the growth in the students' Texas Learning Index scores on the TAAS reading and mathematics tests given from one year to the next. The comparable improvement is based only on students whose TAAS results can be matched from the current to prior year.

Table 12:6 illustrates the annual gains/loses made by students at Washington Elementary on the Texas Learning Index. Gains/losses are calculated separately by year and subject. The table also includes information on the number of students that could be matched from grade 3 to grade 4 in consecutive years. If there were no attrition and if no students were exempted from the test, we would expect to have between 80 and 90 students matched each year. As one can see in the first data column in Table 12:6, the number of matched students ranges from 26 to 51. The number of matched students decreases over the most recent years, and this is likely due to the increase of students exempted from the TAAS.

Exhibit 12:4 includes bar charts that illustrate the relative gains and losses on the Texas Learning Index for matched groups of students at Washington Elementary and in the comparison group of elementary schools. The performance of the students at the Edison school was noticeably worse than in the matched students in the comparison group in both reading and math in all years except for 1997-98 when Washington students showed larger gains than the campus comparison group.

Because of the design of this analysis, this is likely to be the most reliable results we present in this chapter. Individual students' results are measured over two years; and gains, rather than absolute scores, are the focus of attention. Even while the Edison school excludes a far larger proportion of its students from the TAAS, the campus comparison group shows larger gains than the Edison group for 3 or 4 years. There are 40 schools in the campus group, and the quartile statistic in Table 12:6 indicates the relative status of the Edison group among the 40 schools in terms of annual gains on the TLI. The Edison school has a fourth quartile ranking in all subjects and years, except for math and reading in 1997-98, when they ranked in the top quartile of schools in the campus comparison group. The quartile position of a school indicates its comparable improvement from the previous year within its comparison group. For example, if a school is in the top quartile for reading, that means that the average growth in reading of the students tested at that school is better than that of at least three-quarters of the schools in its comparison group.

Table 12:6 Annual Gains/Loses on the Texas Learning Index Based on Matched Students at Washington Elementary School and its Comparison Group of Elementary Schools

1998 to 1999	Reading					Math				
	Number of matched students	1998 average TLI	1999 average TLI	Average growth on the TLI	Quartile	Number of matched students	1998 average TLI	1999 average TLI	Average growth on the TLI	Quartile
WASHINGTON EL GROUP AVERAGE	26 2156	68.96 71.43	67.85 76.4	-1.12 4.97	Q4	39 3001	70.38 71.58	72.41 77.91	2.03 6.34	Q4

1997 to 1998	Reading					Math				
	Number of matched students	1997 average TLI	1998 average TLI	Average growth on the TLI	Quartile	Number of matched students	1997 average TLI	1998 average TLI	Average growth on the TLI	Quartile
WASHINGTON EL GROUP AVERAGE	33 2661	66.06 69.44	75.58 76.89	9.52 7.45	Q1	46 3147	66.98 72.41	75.17 76.42	8.2 4.01	Q1

1996 to 1997	Reading					Math				
	Number of matched students	1996 average TLI	1997 average TLI	Average growth on the TLI	Quartile	Number of matched students	1996 average TLI	1997 average TLI	Average growth on the TLI	Quartile
WASHINGTON EL GROUP AVERAGE	51	67.67 70.8	68.22 75	0.55 4.2	Q4	49	65.82 72.1	63.73 76.4	-2.08 4.3	Q4

1995 to 1996	Reading					Math				
		1996 average TLI	Average growth on the TLI	Quartile		Number of matched students	1995 average TLI	1996 average TLI	Average growth on the TLI	Quartile
WASHINGTON EL GROUP AVERAGE		71.8 79.3	-8.91 0.9	Q4		44	74.5 72.7	64.5 76.1	-10.02 3.4	Q4

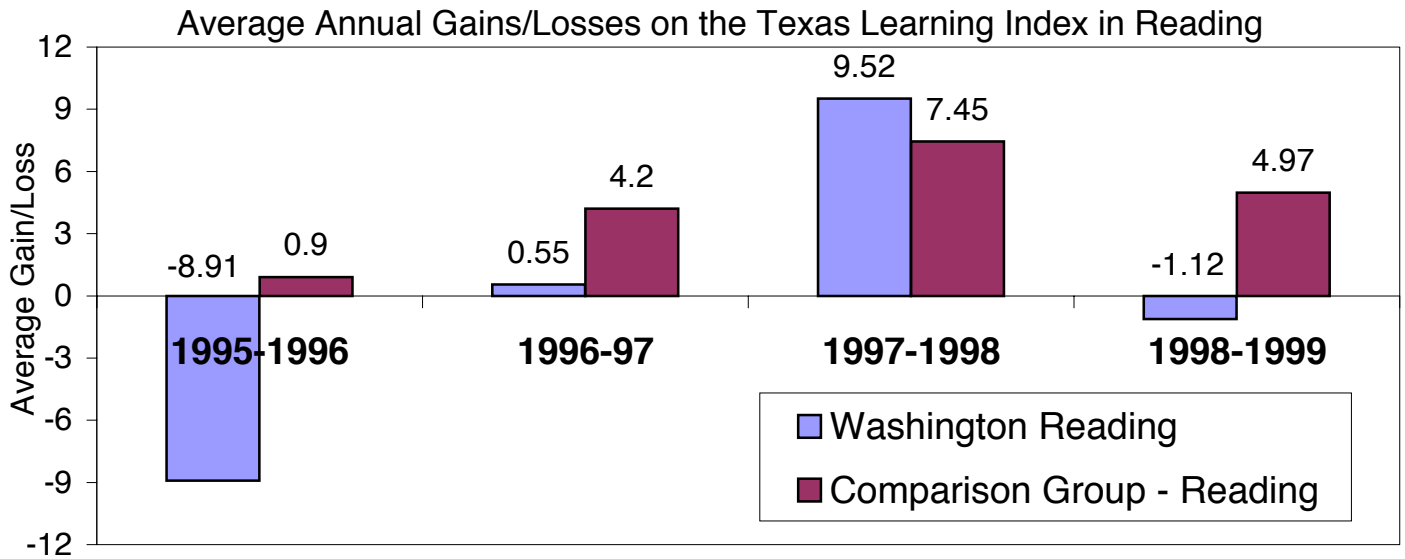
Source: Campus Reports of Academic Excellence Indicator System, 1995, 1996, 1997, 1998, 1999. Texas Education Agency.

Note: All values are based on students who can be matched from current to prior year on the TAAS.

Students move from grade 3 to grade 4 each year in both the Edison school and in the comparison group of schools.

CI is calculated separately for TAAS reading and TAAS mathematics. For each matched student and each subject, TLI growth is determined by subtracting the prior year TLI value from the current year TLI value. The student-level TLI growth values are then aggregated to the campus level to create a TLI Average Growth (TAG) for each campus.

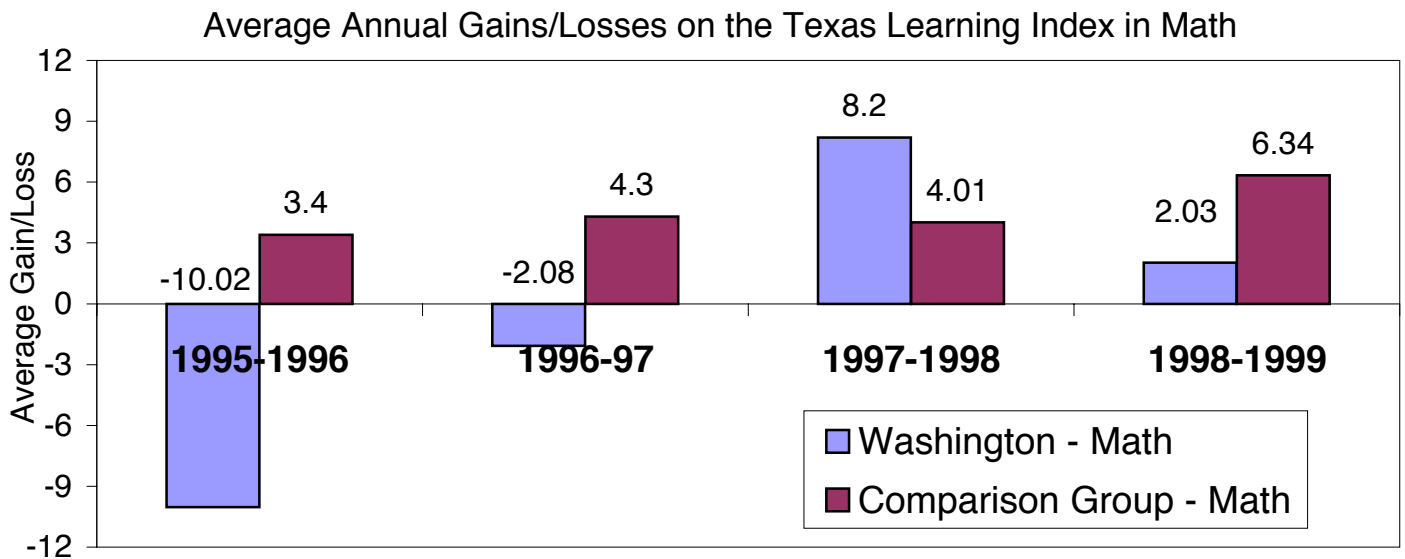
Exhibit 12:4 Average Gains/Losses on the Texas Learning Index, Washington Elementary School



Source: Campus Reports of Academic Excellence Indicator System, 1995-1999. Texas Education Agency.

Note: All values are based on students who can be matched from current to prior year on the TAAS.

Students move from grade 3 to grade 4 in both the Edison school and in the Comparison Group.



Source: Campus Reports of Academic Excellence Indicator System, 1995-1999. Texas Education Agency.

Note: All values are based on students who can be matched from current to prior year on the TAAS.

Students move from grade 3 to grade 4 in these comparisons.

12.6 Chi-Square Analysis of TAAS Data

A chi-square analysis was initiated on data obtained from the Texas Education Agency on the outcomes of the Texas Assessment of Academic Skills (TAAS) test, the state-mandated, criterion-referenced test. While the chi-square can help us distinguish the size and strength of the differences between Washington Elementary and the three control groups we compared it with, our main interest was in the change over time at this Edison school. The chi-square analysis was completed because it is one step in preparation for the odds ratio (OR) analysis, which allows us to examine relative changes over time in this school or in the control groups.

At Washington Elementary School, the TAAS is administered in grade 3 (reading and math) and grade 4 (reading, math, and writing). The test results indicate the percent of students meeting the state standard, or passing. Additional information on the TAAS is provided in Appendix A.

Three different comparison groups were used for the chi-square analyses. Our first comparison is with the local public school district in which Washington Elementary resides (i.e., Sherman Intermediate School District). The second comparison group we selected was the campus comparison group designated by the Texas Education Agency. This is a group of 40 elementary schools across the state with a similar demographic composition. The third comparison was made with the state totals. While the state demographics differ greatly from Washington Elementary, we believe that comparisons with state averages can yield further information regarding the relative gains of this Edison school. Also, since Edison claims that advances in other district schools are—in part—due to its presence, we use the campus group and the state as more distant points of comparison that cannot be easily influenced by the presence of Edison schools.

General procedure

Utilizing published data from the Texas Education Agency, we made comparisons over five years (1995 to 1999) for both grades 3 and 4. We did not convert percentage data (students in each scoring category) into raw frequency data prior to chi-square analysis because we were unable to obtain the specific number of students taking each component of the test for both the Edison school and control groups. Specifically, in the chi-square analyses, the p-values are not valid without appropriate sample sizes and thus are not presented. Moreover, the lack of appropriate sample sizes does not affect the OR estimate since it is based on proportions, but it does invalidate the 95 percent confidence interval (CI). If these data can be obtained, both the chi-square and odds-ratio analyses will be recalculated in order to obtain correct p-values for the chi-square analyses and valid 95 percent CI for the ORs.

Normally, we would evaluate three chi-square analyses for each subtest nested within year and grade level in order to compare passing rates at Washington with the district, campus comparison group, and the state, thus producing 2x2 contingency tables. However, in grade 3 we could not secure the TAAS writing test results; thus, we could only report on the reading and mathematics subtests. These chi-square analyses are testing the null hypothesis that the relative proportion of students in

the two scoring categories are the same for Washington Elementary and the comparison group (district, campus group, or the state).

Chi-square findings

Results of the chi-square analyses for grade 3 covered the TAAS administration for reading and math for the six years from 1994 to 1999. Individual contingency tables are not presented due to the lack of specific sample sizes. Should these sample sizes become available, we will add these data in tabulated form. The first set of comparisons was made against district data as the comparison group. Six separate 2x2 contingency tables were evaluated for each year. Results of the two subtests taken at grade 3 indicate that the proportions of students passing (or failing) more or less parallel each other, with comparable proportions passing each test. The only notable exception was on the 1997 mathematics test, where there was a higher proportion of Washington students in the failure category than reported by the district. Similar findings also were observed in the campus comparison group. Comparable proportions were observed passing each test with the only notable exceptions in the 1996 and 1997 mathematics test. A higher proportion of Washington students passed the test in 1996, but this flipped to a lower proportion in 1997 relative to the campus comparison group. The state comparison revealed similar findings. There were similar passing rates on the TAAS reading test in 1994, 1995, and 1998; but in 1996, 1997, and 1999, more Washington students scored in the failure category.

Results of the chi-square analyses for grade 4 covered the TAAS administration for reading, math, and writing for the six years from 1994 to 1999. Individual contingency tables are not presented due to the lack of specific sample sizes. Should these sample sizes become available, these data will be added in tabulated form. The first set of comparisons was made against district data as the comparison group. Six separate 2x2 contingency tables were evaluated for each year. Results at grade 4 indicate that the proportions of students failing all three tests tended to occur more often at Washington than in the any of the comparison groups (campus comparison group, district, or state). Notable exceptions were on the reading test in 1997, where comparable passing proportions were observed, and the 1998 and 1999 writing tests. Other than on these tests and years, students at Washington tended to fare worse on the TAAS than students in the comparison groups.

12.7 Odds Ratio Analysis of the TAAS Data

One of the many possible statistics that can be derived from a 2x2 contingency table is the odds ratio statistic (OR) and corresponding $1-\alpha$ confidence interval. As presented in Section 2.4 of this report, the 2x2 tables analyzed in the previous section can be thought of as representing consecutive class cohorts in a prospective design. From a classic epidemiological perspective, the students in the Edison school can be thought of as the “exposed” group, that is, exposed to the “Edison effect,” and students in the comparison group as the unexposed group. From this perspective, each yearly comparison is a new cohort, measured over a period of years. There is a minimal possibility for

cohort contamination if a number of students in one group are not promoted to the next grade level. However, we think this represents a very small number of possible cases and therefore has minimal impact on the validity of these analyses. Section 2.4 details the OR statistic and corresponding $1-\alpha$ confidence interval. We calculated and charted OR for each of the 2x2 tables constructed from the 2x2 analyses presented above. We have not, however, presented the 95 percent CI for these ORs because they are not valid due to the lack of specific sample sizes. Exhibits 12:5-12:10, as well as Tables 12:7 and 12:8 present these findings. The bar charts in Exhibit 12:2 illustrate the overall performance of Washington Elementary School on the TAAS in comparison with district, campus comparison group, and state performance levels. These charts complement odds ratio results reported in Tables 12:7 and 12:8 and more easily allow the reader to see the relative passing proportions for the Edison and control groups.

Odds Ratio Findings for Grade 3

Two OR analyses were evaluated, one for each subject test on the TAAS. In grade 3, the OR for a Washington student failing the 1995 reading TAAS relative to the students in the district as a whole was 0.619 and showed a substantial increase in the following year that remained more or less unchanged through 1999.

Table 12:7 Summary of Odds Ratio Findings for Washington Elementary School, Grade 3

	1995	1996	1997	1998	1999
<i>Odds of not meeting standard compared with district</i>					
Reading	0.619	2.035	1.924	2.050	2.046
Math	1.341	1.197	2.859	1.438	1.665
<i>Odds of not meeting standard compared with campus group</i>					
Reading	0.478	1.498	1.976	1.292	2.063
Math	1.014	0.167	4.866	1.015	1.500
<i>Odds of not meeting standard compared with state</i>					
Reading	0.499	2.061	1.924	1.785	2.341
Math	1.280	1.681	3.834	1.388	1.809

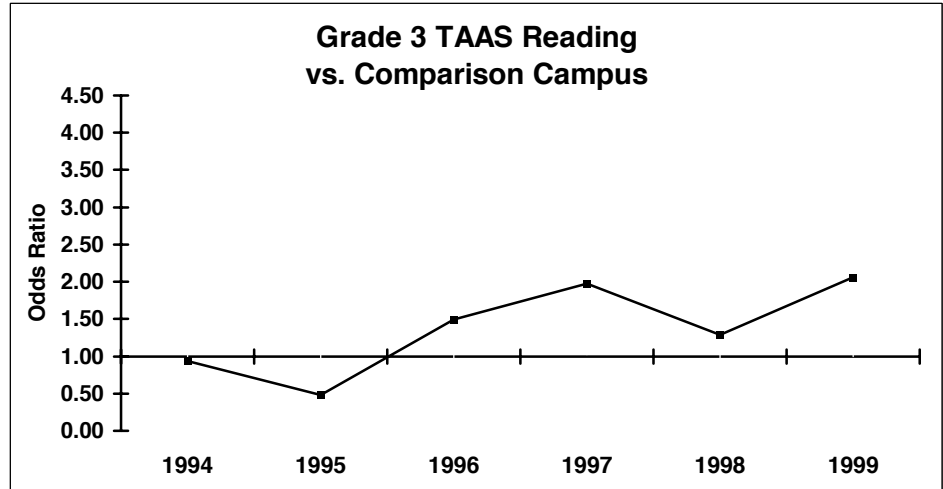
Table 12:8 Summary of Odds Ratio Findings for Washington Elementary School, Grade 4

	1995	1996	1997	1998	1999
<i>Odds of not meeting standard compared with district</i>					
Reading	3.673	2.706	2.687	1.627	2.887
Math	1.933	3.906	3.082	0.740	2.925
Writing	2.731	1.581	3.954	1.702	1.475
<i>Odds of not meeting standard compared with campus group</i>					
Reading	3.098	1.782	2.457	1.259	2.690
Math	1.971	4.918	5.597	0.789	2.850
Writing	2.298	2.061	5.735	2.446	1.036
<i>Odds of not meeting standard compared with state</i>					
Reading	4.025	2.835	2.595	1.460	2.771
Math	2.271	7.088	5.061	1.021	3.115
Writing	2.499	3.145	5.940	3.274	1.139

Exhibit 12:5 Results of the Odds Ratio Analysis for Washington Elementary (Grade 3) vs. Campus Comparison Group

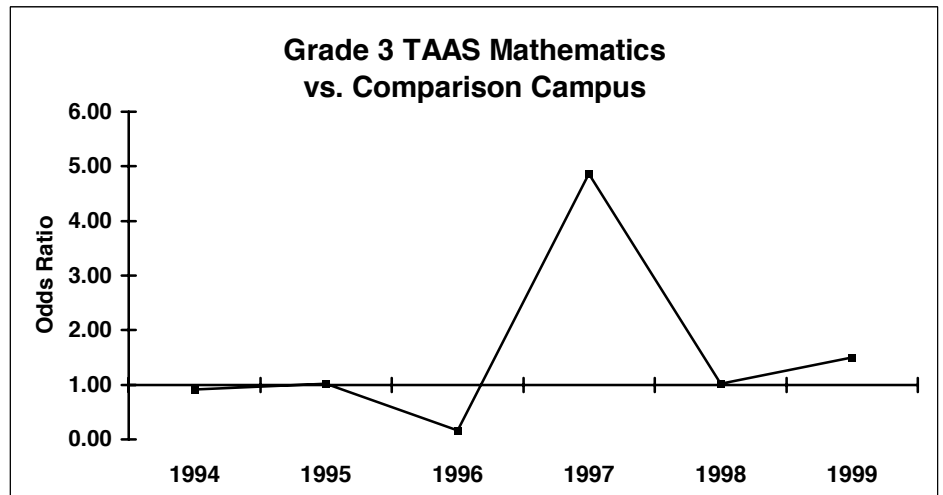
Grade 3 TAAS Reading

Year	U CI	L CI	OR
1994			0.934
1995			0.478
1996			1.498
1997			1.976
1998			1.292
1999			2.063



Grade 3 TAAS Math

Year	U CI	L CI	OR
1994			0.914
1995			1.014
1996			0.167
1997			4.866
1998			1.015
1999			1.500



Grade 3 TAAS All Tests

Year	U CI	L CI	OR
1994			0.968
1995			0.809
1996			1.506
1997			2.993
1998			0.936
1999			1.474

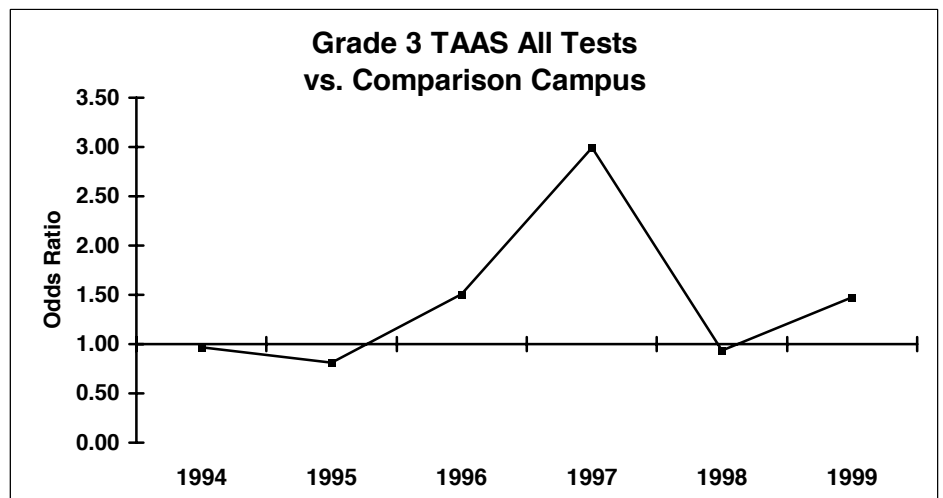
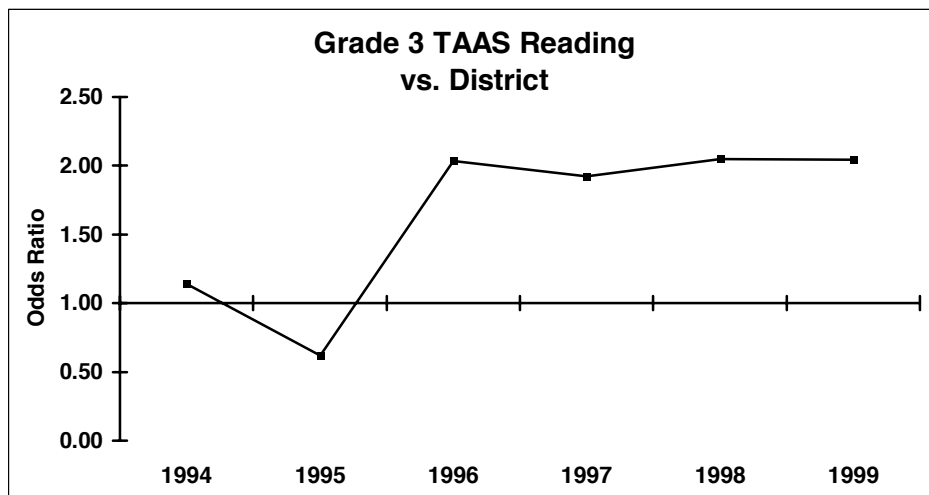


Exhibit 12:6 Results of the Odds Ratio Analysis for Washington Elementary (Grade 3) vs. District

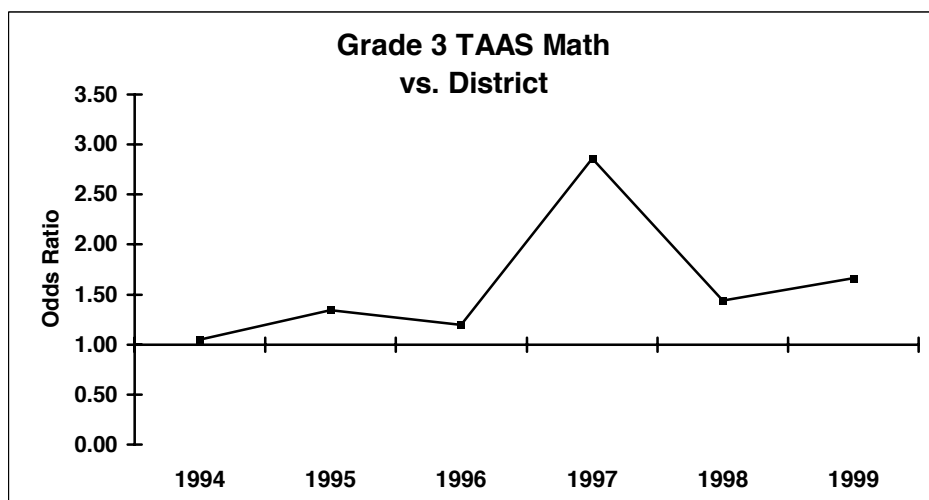
Grade 3 TAAS Reading

Year	U CI	L CI	OR
1994			1.142
1995			0.619
1996			2.035
1997			1.924
1998			2.050
1999			2.046



Grade 3 TAAS Math

Year	U CI	L CI	OR
1994			1.051
1995			1.341
1996			1.197
1997			2.859
1998			1.438
1999			1.665



Grade 3 TAAS All Tests

Year	U CI	L CI	OR
1994			1.151
1995			1.014
1996			1.651
1997			2.993
1998			1.524
1999			1.750

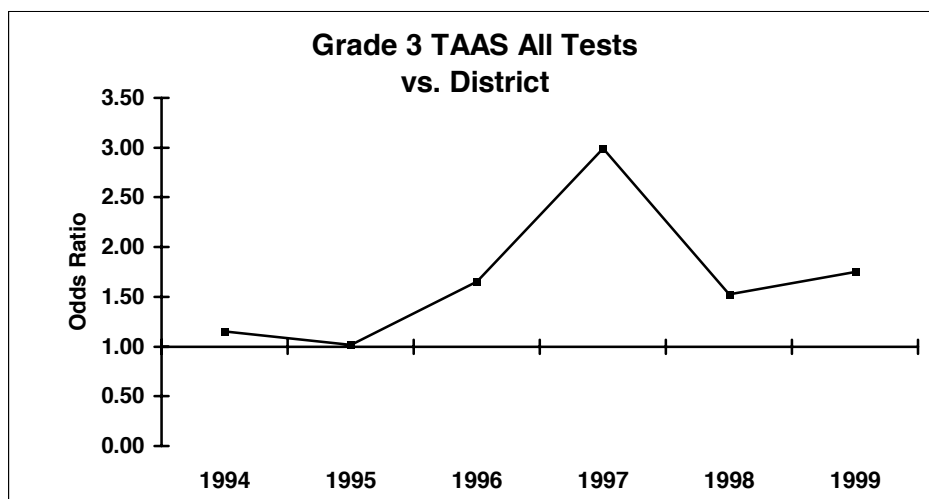
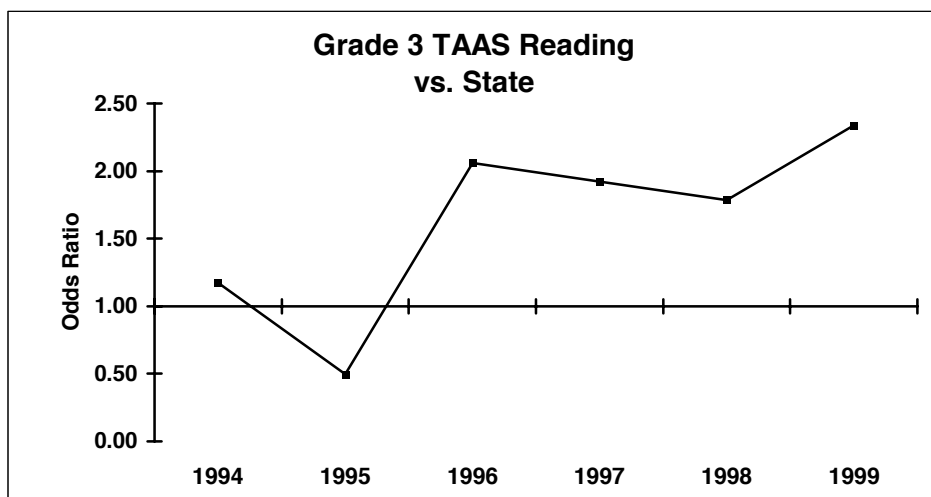


Exhibit 12:7 Results of the Odds Ratio Analysis for Washington Elementary (Grade 3) vs. State

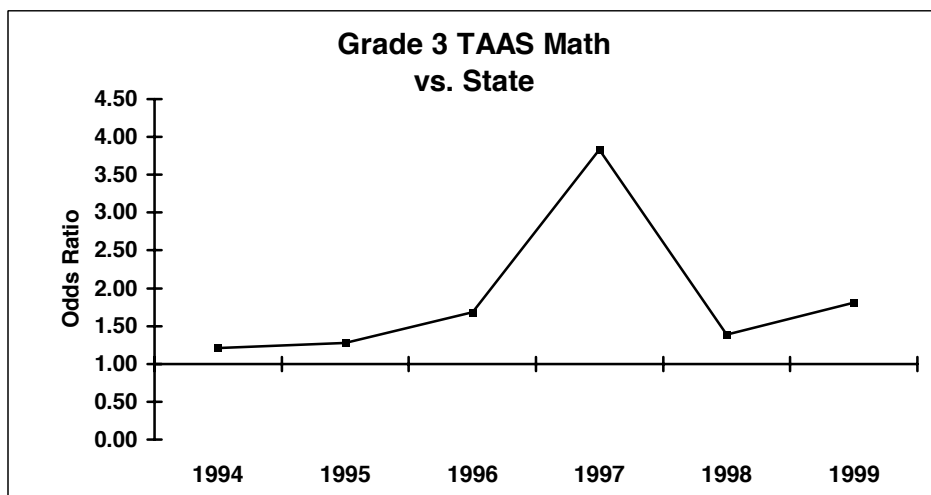
Grade 3 TAAS Reading

Year	U CI	L CI	OR
1994			1.175
1995			0.499
1996			2.061
1997			1.924
1998			1.785
1999			2.341



Grade 3 TAAS Math

Year	U CI	L CI	OR
1994			1.208
1995			1.280
1996			1.681
1997			3.834
1998			1.388
1999			1.809



Grade 3 TAAS All Tests

Year	U CI	L CI	OR
1994			1.349
1995			0.933
1996			2.059
1997			3.601
1998			1.440
1999			1.867

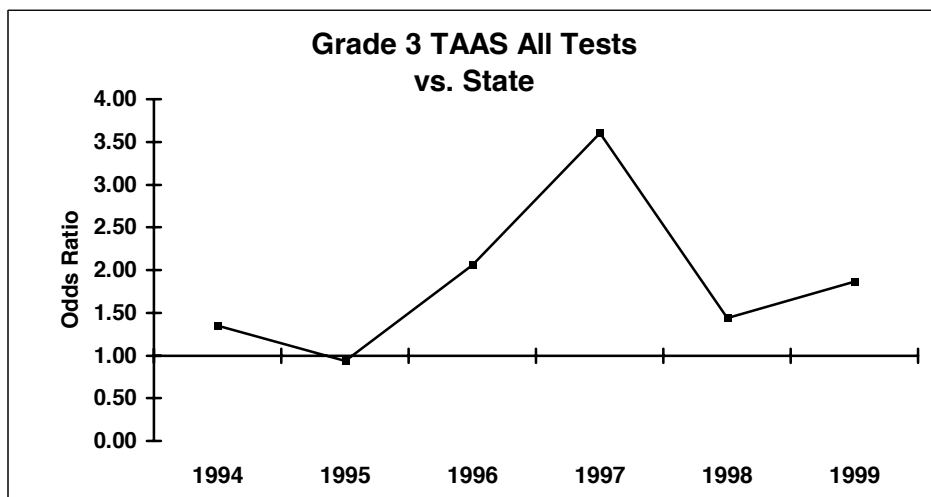
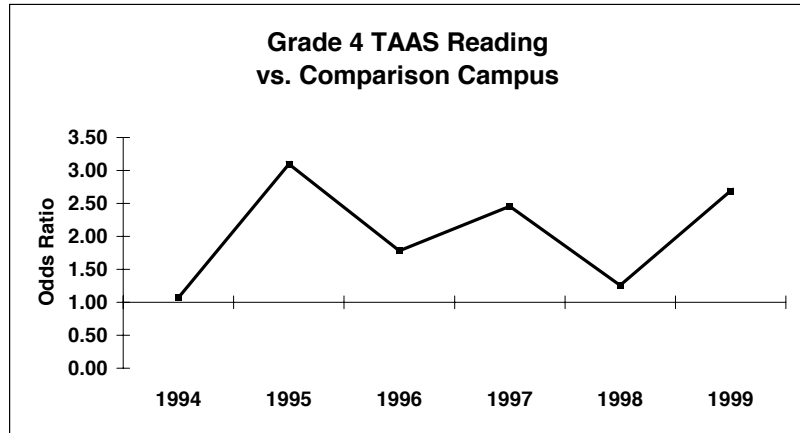


Exhibit 12:8 Results of the Odds Ratio Analysis for Washington Elementary (Grade 4 vs. Campus Comparison Group)

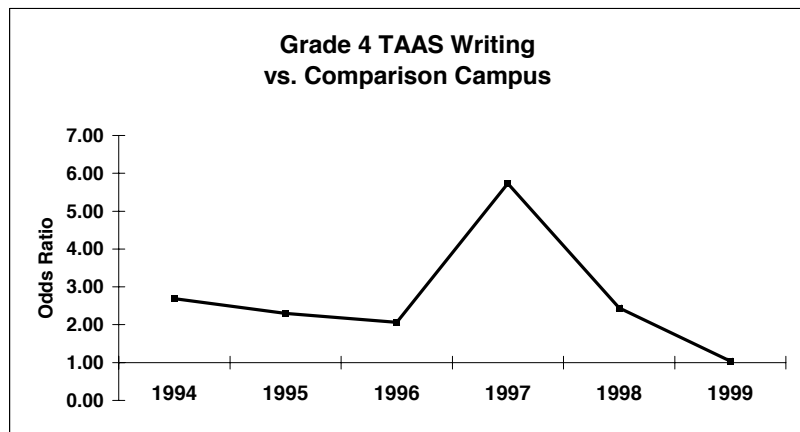
Grade 4 TAAS Reading

Year	U CI	L CI	OR
1994			1.072
1995			3.098
1996			1.782
1997			2.457
1998			1.259
1999			2.690



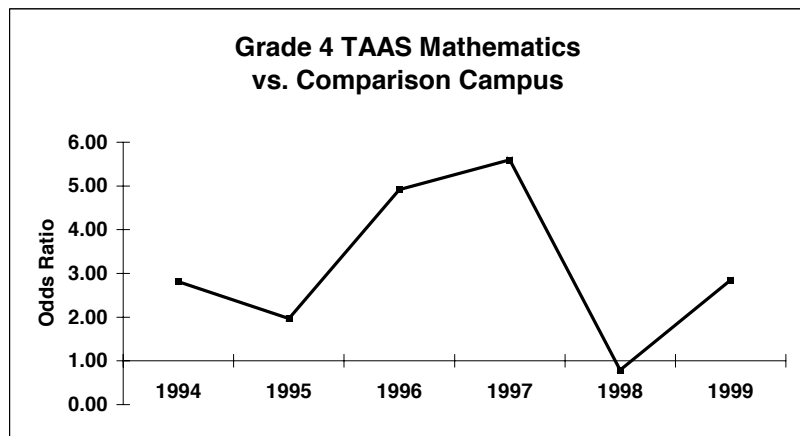
Grade 4 TAAS Writing

Year	U CI	L CI	OR
1994			2.686
1995			2.298
1996			2.061
1997			5.735
1998			2.446
1999			1.036



Grade 4 TAAS Math

Year	U CI	L CI	OR
1994			2.811
1995			1.971
1996			4.918
1997			5.597
1998			0.789
1999			2.850



Grade 4 TAAS All Tests

Year	U CI	L CI	OR
1994			2.460
1995			2.141
1996			2.565
1997			3.791
1998			1.460
1999			1.993

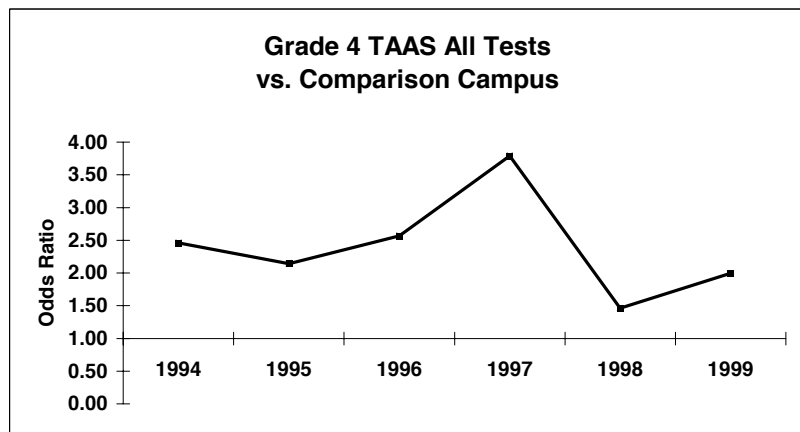
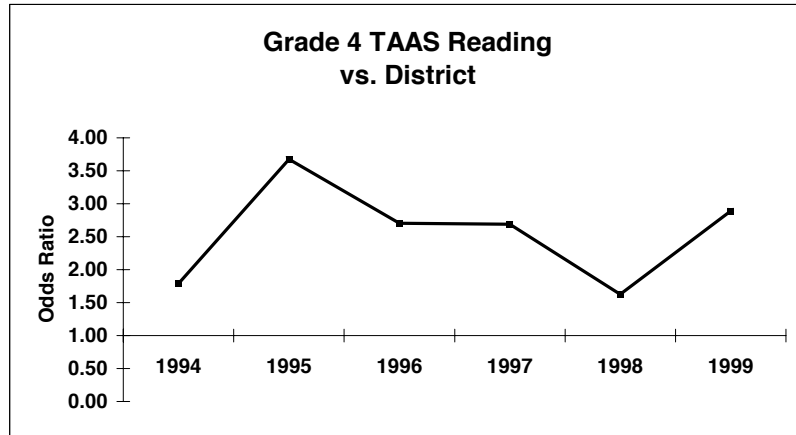


Exhibit 12:9 Results of the Odds Ratio Analysis for Washington Elementary (Grad vs. District)

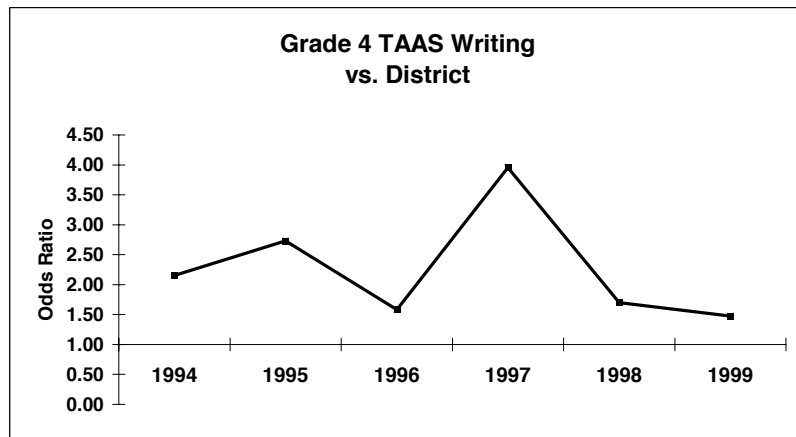
Grade 4 TAAS Reading

Year	U CI	L CI	OR
1994			1.791
1995			3.673
1996			2.706
1997			2.687
1998			1.627
1999			2.887



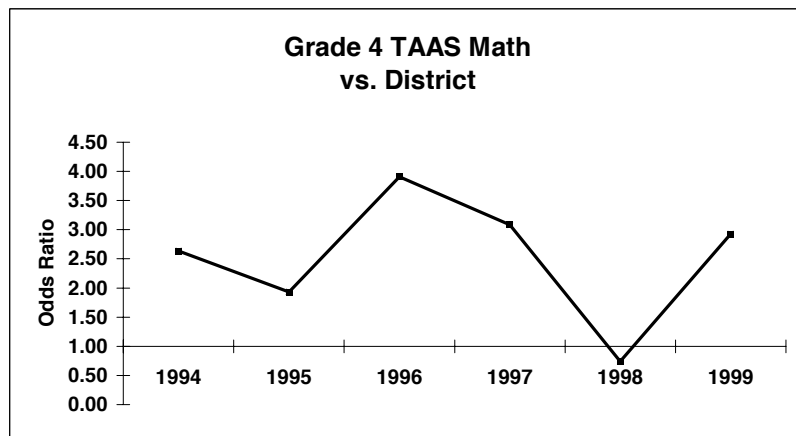
Grade 4 TAAS Writing

Year	U CI	L CI	OR
1994			2.153
1995			2.731
1996			1.581
1997			3.954
1998			1.702
1999			1.475



Grade 4 TAAS Math

Year	U CI	L CI	OR
1994			2.635
1995			1.933
1996			3.906
1997			3.082
1998			0.740
1999			2.925



Grade 4 TAAS All Tests

Year	U CI	L CI	OR
1994			2.550
1995			2.309
1996			2.871
1997			3.085
1998			1.321
1999			2.402

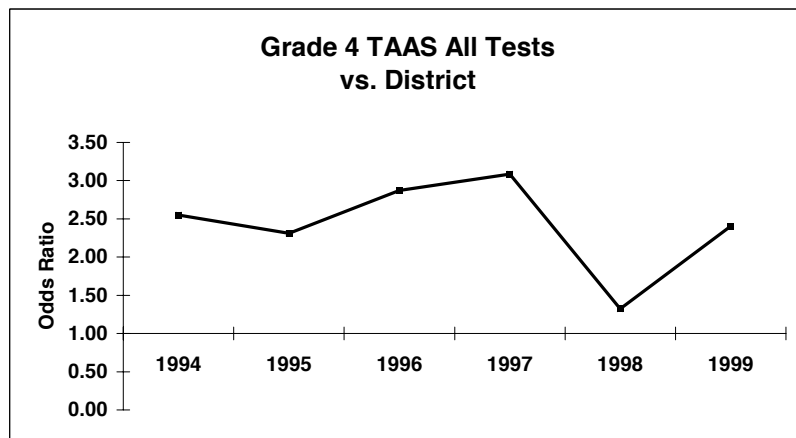
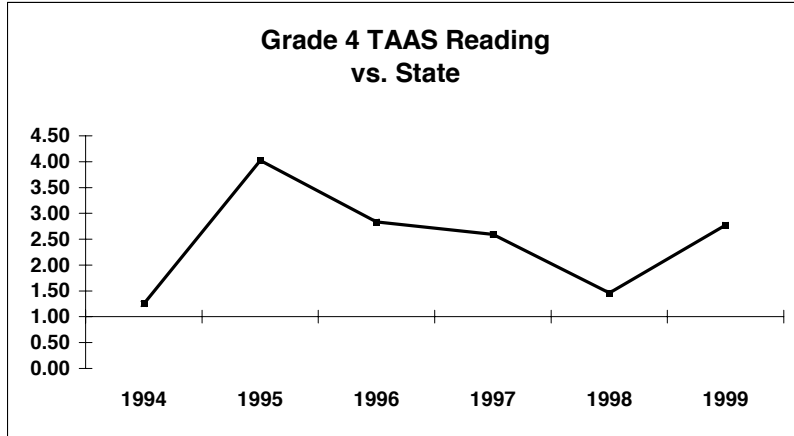


Exhibit 12:10 Results of the Odds Ratio Analysis for Washington Elementary (Grad vs. State)

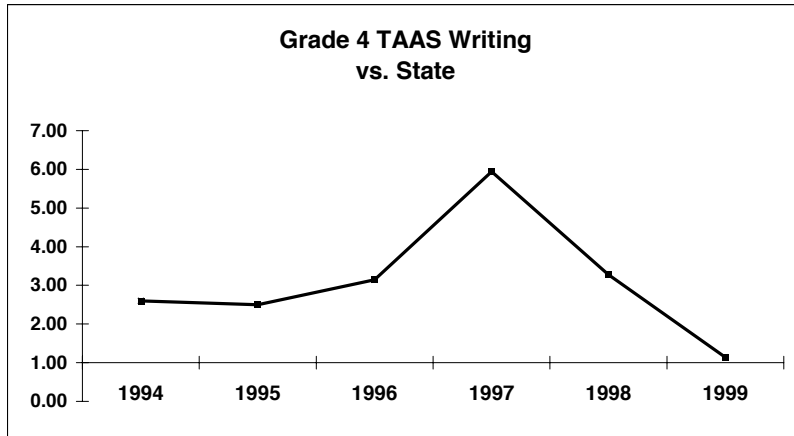
Grade 4 TAAS Reading

Year	U CI	L CI	OR
1994			1.253
1995			4.025
1996			2.835
1997			2.595
1998			1.460
1999			2.771



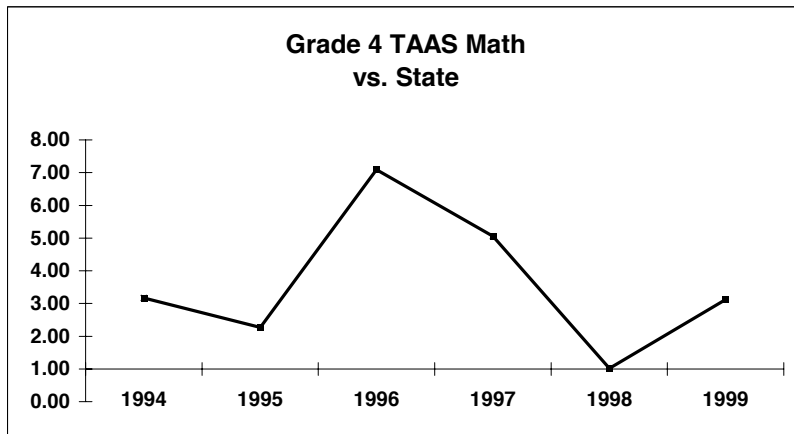
Grade 4 TAAS Writing

Year	U CI	L CI	OR
1994			2.600
1995			2.499
1996			3.145
1997			5.940
1998			3.274
1999			1.139



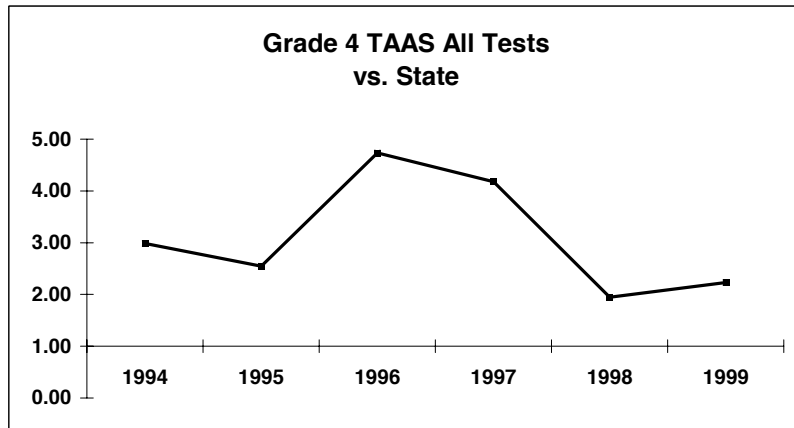
Grade 4 TAAS Math

Year	U CI	L CI	OR
1994			3.167
1995			2.271
1996			7.088
1997			5.061
1998			1.021
1999			3.115



Grade 4 TAAS All Tests

Year	U CI	L CI	OR
1994			2.983
1995			2.548
1996			4.735
1997			4.178
1998			1.947
1999			2.234



Over this four-year period, students attending Washington Elementary were about two times more likely to not meet state standards than students enrolled in the district. Since the odds ratio increased over the years the school was operated by Edison, this indicates that the performance on the TAAS at the Edison school is not as good as the performance of other grade 3 students in the district. Similar findings were observed relative to the state and campus comparison groups, although for this comparison the initial jump in odds of failing was not as dramatic as it was for the district or state comparisons. Results of the math analyses also showed relative stability over the five years except for a jump in odds in 1997 in all three comparisons. Overall, students at Washington Elementary were only a bit more likely to fail the TAAS math subtest.

Odds ratio findings for grade 4

Three OR analyses were evaluated in grade 4, one for each subject test on the TAAS. Washington student performance relative to students in all three comparison groups indicated that Washington students were uniformly more likely to fail the reading test. Similar results were observed on the math test except in 1998, where Washington students actually outperformed the other three groups. Results on the writing test seemed to parallel the findings on the reading test, with Washington students at greater odds for failure than students in the three comparison groups.

12.8 Summary

In Edison's first annual report on student performance, it was noted that "The state criterion referenced test, TAAS, has proved a bit of a struggle for third-grade students and teachers. . . . This is one of Edison's two cases of significant ineffectiveness" (Edison, 1997, p. 14). By the time Edison prepared its second annual report on student performance, the results for the 1997-98 school year were available. This was the single best year for this school during the five years it has been operated by Edison. With these results in hand, Edison characterized the trends from this school as "Strongly Positive." The test results for the following year, however, dropped again, and the size of the relative gains ranked this school in the bottom quartile for all subjects and grades on the Texas Learning Index, just as it had been ranked in the years previous to the successful 1997-98 school year.

The performance of the matched students at the Edison school that was measured with the Texas Learning Index was noticeably worse than that of the matched students in the comparison group in both reading and math in all years except for 1997-98 when Edison students showed larger gains than the campus comparison group.

Due to the unavailability of specific cell frequencies, the $(1-\alpha)$ C.I. around the OR could not be calculated and the Breslow-Day statistic could not be meaningfully calculated. Instead our "trend" is based upon a visual inspection of the consecutive cohorts' performance relative to the campus comparison group.

Table 12:10 Summary of Trends on the State Criterion-Referenced Test, TAAS

	Visual Trend in OR	Texas Learning Index*	Trend
Grade 3 Reading	negative (-1)	-4.38	negative (-1)
Grade 4 Reading	negative (-1)		negative (-1)
Grade 3 Math	mixed (0)	-4.98	mixed (0)
Grade 4 Math	negative (-1)		negative (-1)
Grade 4 Writing	positive (+1)		positive (+1)

* The four year net change in TLI relative to the campus comparison group. The TLI is based on the change in performance from a matched group of students from grade 3 to 4.

Table 12:11 Overall Achievement Trends Washington Elementary School

	Positive	Mixed	Negative
Norm Referenced	--	--	--
Criterion Referenced	1 of 5	1 of 5	3 of 5
TOTALS	1 of 5	1 of 5	3 of 5

We characterize the overall trends in this school based on student performance on the TAAS as negative. Except for one successful year, the school is largely unable to match even the gains made by the district when we consider matched students. In terms of consecutive class cohorts, the school has only been able to come close to matching the control groups on the grade 4 writing test. The results dropped considerably during Edison's first year of operation compared with the previous year when the school was run by the district.

On the third grade subject tests, the performance of students dropped and then eventually returned to pre-Edison performance levels. The control groups were making small but steady annual gains and thus retained higher absolute scores and overall higher gain scores.

The extensive background information on test takers and those exempted from test taking allowed us to piece together an interesting puzzle. A number of questions need to be raised regarding the pattern of exempting students from the test at Washington Elementary School, which clearly stands out in terms of the trend of exempting students. During the first year Edison operated the school, 3.9 percent of the students were exempted from the TAAS due to special educational needs, limited English proficiency, or other reasons. The proportion of students exempted increased annually to a total of 25.6 percent in the 1998-99 school year. At the same time, the rates of exemption at the three control groups largely remained unchanged. While a number of questions arise, we are not in a position to provide answers. More information is needed to explain these figures and to help determine if they had any impact on the overall performance of the school.

Given the total ratings for the trends that are highlighted in Table 12:11, we rate this school as Negative with a mean rating of -0.4. In its 1999 annual report, Edison rated this school as Strongly Positive. In its 2000 annual report, it rated the 1999-2000 school year as Positive and the achievement gains since opening as Positive.

Chapter Thirteen

Summary of the Findings

As previously noted, there has been a paucity of evaluations of Edison schools, and those conducted typically have been limited to one or two schools and limited in time. The most notable studies are the Miami Dade (Gomez & Shay, 1998, 1999, 2000), Minneapolis Public Schools (2000), Wichita State University (1996), AFT (1998, 2000), and the three annual reports published by Edison Inc. (1997, 1999, 2000). Of particular interest in this evaluation is the relationship of our results to the results previously published in Edison annual reports on student achievement. Thus, as part of this chapter, we will present a brief overview of Edison's three published annual reports (1997, 1999, 2000).

13.1 Edison's Findings

Edison prepared three annual reports on student achievement at its schools (Edison, 1997, 1999, 2000). The findings contained in the annual reports and the message Edison spreads in conference presentations and in the media indicate that the company is quite successful and that students enrolled in Edison schools are making large and substantial achievement gains.

Edison's First Annual Report on School Performance (Edison, 1997) was structured around 5 key performance areas: student achievement, customer satisfaction, the implementation of the school design, financial management, and systems growth. In this report, Edison summarized the effectiveness of its program across 7 different schools (all of which were considered in our study). A total of 40 trends apparently were traced, 28 of them positive. In other instances, there was only one year of data, and no rating could be given in such cases. In 2 trends (1 for M.L. King Academy and 1 for Washington Elementary) the evidence indicated "that students are not being more successful with The Edison Project than they would be elsewhere" (Edison, 1997, p. 7).

Edison published its second annual report on school performance in 1999, reporting achievement trends for 17 schools as well as baseline results for 8 additional schools. This report rated its school performance with a variety of assessment data including criterion-referenced and norm-referenced tests. Edison's evaluation derived student performance trends on a year-to-year basis. Year-to-year trends were inferred by examining gains and/or losses in scores for matched groups of students or consecutive class cohorts of students.

To establish achievement trends Edison reported that it used the following rules in its 1999 report: (i) whenever possible, measure trends using the same individual students; (ii) when the same individual students are not available, measure the same group of students; and (iii) when the same

individual students and the same group of students are not available, measure the same grade level. Edison calculated the net gain for every trend from the first observation to the final observation. Some of its reported trends were three years long, others two years, and some only one. Each net trend was then divided by its duration, yielding the average annual achievement gain or loss for each trend (Edison, 1999).

Edison claims that student achievement in its schools has been heading steadily upward. Edison indicated that 136 of 176 trends were positive, and that the average percentile gain on norm-referenced tests was 5 percent, and the average percentile gain on criterion-referenced tests was 6 percent. In its second annual report (Edison, 1999), the company listed the following summary statements about the achievement levels of its students/schools:

- The overwhelming majority of achievement trends—numbering nearly 200—were positive.
- On average, students were gaining more than 5 percentiles per year against state and national standards.
- Students in nearly every Edison school are achieving more today than when the school opened; in no Edison school are students achieving less.
- Fourteen of 17 Edison schools that now have established achievement trends have moved student achievement forward.
- The trends in 10 of the 17 schools were rated Strongly Positive, 4 were rated Positive, 3 were Mixed, and no schools were rated as Negative or Strongly Negative.

Edison's third annual report was released in September 2000 and noted that achievement gains were larger and more widespread than was the case in previous years: "For the 1999-2000 school year, the average gain of Edison students in the core areas of reading, language arts, spelling, writing, and mathematics was 5 percentiles on nationally normed tests and 7 percentage points on criterion-referenced tests, which also include science and social studies. These gains represent improvements of one point in each case over the gains reported for 1995-99, and are the highest gains reported by Edison to date" (Edison, 2000, p. 2).

The findings reported by Edison are very positive. Over recent years, the reporting procedures seem to have shifted. In some cases, the change in ratings made by Edison is due to the inclusion or exclusion of data. Edison also seems to have changed its priorities for data types and its willingness to apply a common set of criteria. These things can account for some of the changes between years. Later in this chapter we revisit this issue when we discuss the expectations that districts should have in terms of reporting results.

13.2 Evaluation Center Findings

The principal aim of this evaluation was to determine the effect of the Edison model on student performance on norm- or criterion-referenced standardized tests. This section presents a summary of the main findings from this evaluation. These findings are based on analyses conducted in each

of the cases presented in the preceding chapters. The 10 cases were selected because these schools had been operated by Edison for 4 or more contract years. The longer time period operating under an Edison contract should provide a more convincing picture of the impact the Edison model can have on student achievement. While some of these schools may have had more difficult start-ups than others, they have been in operation beyond what many consider the start-up phase. It was our intention to include all schools (11 total) that opened during the first years that Edison was operating schools (1995-96 and 1996-97); however, we were unable to secure any independently verifiable student achievement data for 1 school.

There is a possibility of a selection bias among the schools included in our sample. To guard against this possibility, we analyzed difference in mean trend ratings Edison gave to the 10 schools included in our study and the 33 nonsampled schools for which it reported trends in its 2000 annual report. Using Edison's own ratings we found no significant difference³ between the sampled and nonsampled schools. Thus, we believe that there is no strong indication that the 10 schools we evaluated are not an accurate representation of the schools for which Edison currently has trend data. This issue is discussed in further detail later in this chapter.

Methodology employed in the evaluation

The focus of the methodology employed in this evaluation was to capture the academic achievement gains of students enrolled in Edison schools. Since the focus of this evaluation was directed toward *achievement gain*, a comparison must be made. A multitude of possible comparisons could be made, and it is beyond the scope of this report to include them all. Rather, this evaluation focused on two primary types: (1) individual (within-subject) gains evidenced by longitudinal panel norm-referenced achievement test data and (2) cohort comparisons evidenced by longitudinal cohort groups on criterion-referenced test data.

Norm-referenced achievement test data. A variety of NRT data was provided to us by Edison Schools Inc. for a limited number of schools for a limited number of years of data collection. These data files did not contain data on a comparison sample and tended to be rather incomplete. Consequently, all NRT analyses examine only individual within-subject change and did not provide an external comparison, except what was imbedded in the score reporting scale, (e.g., NCR scale). Following a list-wise deletion strategy, repeated measures ANOVA was used to test for longitudinal trends over the available years. The list-wise deletion strategy utilized in these analyses resulted in analyzing only complete cases; however, this likely raised attrition rates. Specific attrition rates can be calculated for each case by cohort group, subject test, and reporting scale; and general attrition rates for the six cases are presented in Table 13:1. As can be seen from this table, there was a substantial amount of attrition in these analyses.

³ Independent t-test results of Edison's own mean trend rating in its 2000 annual report indicated that there was no statistically significant difference between Edison's ratings for the schools included in this evaluation and the schools not included [$t(41) = 0.213$, $p = .832$].

Table 13:1 Approximate Attrition Rates in the Norm-Referenced Test Analyses

Case	Cohort Groups/Time	Approximate Attrition Rate	Average Annual Attrition Rate
Roosevelt	2 Cohorts, 2 Years	36%	12%
Dodge-Edison	2 Cohorts, 3 Years	59%	20%
Boston Renaissance	5 Cohorts, 3 Years	44%	15%
	1 Cohort, 4 Years	59%	15%
Seven Hills	2 Cohorts, 3 Years	40%	13%
Dr. Martin Luther King Jr.	1 Cohort, 3 Years	35%	12%
Mid-Michigan	1 Cohort, 3 Years	58%	19%

Note: Approximate attrition rate was calculated for the only NCE score scale. For example, in the Roosevelt case there were 2 cohorts and 3 subject tests. This produced 6 different analyses per cohort. An approximate attrition rate was calculated by determining the specific attrition rate for each analyses and then averaging the rates together. In this case 12 rates were averaged, 6 for cohort A and 6 for cohort B. Since each panel represented 3 years, the approximate attrition rate was divided by 3 to get the average annual attrition rate.

There is clearly a need for a closer examination of the reasons and possible explanations for large groups of students not participating in the norm-referenced testing at these schools.

Criterion-referenced achievement test data. CRT data were culled from a variety of sources, but primarily from the Web. These data were typically reported as aggregates for schools (i.e., no individual performance data were made available to us; rather, the data reflect all students taking the test at each school, broken down by grade and subject test). Thus, the CRT data represented the performance of consecutive cohorts of students. Since these data are open to the public, we were able to define and construct comparison groups (detailed in each case study) for these analyses. The ability to define a comparison group allowed our analyses to test if the relative proportion of students in an Edison school scoring among the various levels coincided with either district or state scoring proportions. We examined these proportions via chi-square analysis and as a prospective cohort study by collapsing the ordinal responses on the various state tests into pass/fail categories.

General findings from the norm-referenced analyses

Table 13:2 presents an overall summary of the performance trends from the ten cases on norm-referenced achievement tests. Three norm-referenced tests were often used by the schools: the Iowa Test of Basic Skills (ITBS), the Metropolitan Achievement Test (MAT-7), and the Stanford Achievement Test (SAT-9). As noted earlier, Edison provided us with seven data sets on six cases of individual student data, which include ITBS, MAT-7, and SAT-9 results. Most often each achievement test could be analyzed over four different score scales (GE, SS, PR, NCE). When possible, we chose to consider only the NCE analyses in our summative ratings for each case (see Section 2.4), although we present analyses on each scale in the respective case study. Our ratings (positive, neutral, or negative) were based on a rating system developed and elaborated in Section 2.5. For example, if there was an average gain of 3.5 NCEs or more, we rated it as a positive gain.

If there was a average loss of greater than 3.5 NCEs, we rated it as negative. If the gain or loss was ± 3.50 NCEs, we rated it as mixed.

Table 13:2 Overall Norm-Referenced Trends

Case	Positive	Mixed	Negative
Roosevelt-Edison [†]	2 of 6	4 of 6	0 of 6
Henry E.S. Reeves [‡]	0 of 4	4 of 4	0 of 4
Dodge-Edison [†]	2 of 4	2 of 4	0 of 4
Jardine-Edison [£]	3 of 6	3 of 6	0 of 6
Boston Renaissance [†]	0 of 12	12 of 12	0 of 12
Seven Hills Charter School [†]	3 of 12	9 of 12	0 of 12
Dr. Martin Luther King Jr. [†]	0 of 3	3 of 3	0 of 3
Mt. Clemens Sec. Academies	—	—	—
Mid-Michigan PSA [†]	0 of 3	3 of 3	0 of 3
Washington Elementary	—	—	—
TOTALS	10 of 50	40 of 50	0 of 50

[†] Trends based on NCE scores.

[‡] Trends based on effect size calculation by Shay (2000).

[£] Trends based on national percentile rank scores.

The results from Table 13:2 can be summed up as follows:

- We charted 50 norm-referenced trends.
- Overall, the norm-referenced trends were either mixed or positive; none were negative.
- Students in Edison schools are generally showing academic achievement gains consistent with grade level advancement on norm-referenced tests. However, they do not consistently exceed grade level expectations on norm-referenced tests.

General findings from the criterion-referenced analyses

Table 13:3 presents an overall summary of the achievement trends from the ten sampled schools on criterion-referenced achievement tests. These criterion-referenced tests were different for each state since they were usually part of a mandated state assessment program. Our ratings (positive, neutral, or negative) were based on a rating system developed and elaborated in Section 2.5. We analyzed the CRT data a number of different ways; however, we based our rating on the consecutive cohort odds ratio analyses. A negative rating would be an OR greater than 1.0 with a CI whose lower bound was greater than 1.0. A mixed rating would be reflective of even odds in an OR analysis, where the CI eclipses 1.0. A positive rating was a protective OR, signified by an OR less than 1.0

with a CI whose upper bound was less than 1.0. To this general scoring system we tried to determine if a trend was present when there were two years of data present by examining the Breslow-Day statistic. In some cases we could not implement this criterion, as noted in the table.

Table 13:3 Overall Criterion-Referenced Trends

Case	Positive	Mixed	Negative
Roosevelt-Edison	0 of 3	2 of 3	1 of 3
Henry E.S. Reeves	0 of 3	3 of 3	0 of 3
Dodge-Edison	1 of 3	2 of 3	0 of 3
Jardine-Edison	1 of 3	1 of 3	1 of 3
Boston Renaissance	0 of 6	3 of 6	3 of 6
Seven Hills Charter School	0 of 6	1 of 6	5 of 6
Dr. Martin Luther King Jr.	0 of 4	2 of 4	2 of 4
Mt. Clemens Sec. Academies	2 of 8	6 of 8	0 of 8
Mid-Michigan PSA	0 of 8	0 of 8	8 of 8
Washington Elementary	1 of 5	1 of 5	3 of 5
TOTALS	5 of 49	21 of 49	23 of 49

Below, we have summarized the main findings from our analysis of the criterion-referenced tests (see Table 13:3):

- We charted 49 criterion-referenced trends, which is nearly equal to the number of norm-referenced trends.
- Student performance on criterion-referenced tests often lags behind district performance and almost always behind state performance levels.
- In nearly half the trends, we found that students enrolled in Edison schools were making smaller gains on the criterion-referenced tests than comparison groups were (i.e., 23 out of 49 trends).
- In 21 out of 48 trends, Edison students showed gains or changes in test results that were similar to the local districts and other comparison groups. In only 5 of the 49 trends did we find Edison students making larger gains than comparison groups.

Overall summary of findings

Given the varied designs and available data used in this and past evaluations, one can realize that we are far from having a true scientific experiment that can render a conclusive decision on whether or not students succeed in schools operated by Edison Schools Inc. Our results do not differ greatly from past evaluations from Miami-Dade, Minneapolis, or those studies conducted by Howard Nelson at AFT (1998, 2000). Statewide evaluations of charter schools that included schools operated by Edison also indicate that these schools are not better than other schools (Colorado Department of

Education, 2000; Horn & Miron, 1999, 2000; and Renewal Inspection Report, 1999). While there are differences in the quality, scope, and rigor of these studies, it is important to note that the most rigorous studies (Gomez & Shay, 2000; Shay, 2000; and Minneapolis Public Schools, 2000), although limited in scope, all indicate that Edison students are not doing better than control groups of students. Therefore, we are not alone in determining that students in schools operated by Edison—while they often start at levels below national norms and district averages—progress at rates comparable to students in other district schools. Unfortunately, this conclusion does not meet Edison’s goal, which is to have achievement performance levels that exceed the level at comparable schools.

Table 13:4 presents a list of all our trend ratings. Together we examined 99 trends: half from norm-referenced test results and half from criterion-referenced test results. A positive trend reflects students in the Edison school evidencing statistically significant and meaningful annual gains on norm-referenced tests or outperforming the comparison groups on state criterion-referenced tests. A mixed trend is indicative of Edison students gaining at grade-level expectation on norm-referenced tests or performing at the same levels as comparison group students on criterion-referenced tests. A negative trend is evidenced by lower than grade level gains on norm-referenced tests or a lower passing percentage than the comparison group on criterion-referenced tests.

Table 13:4 Combined Overall Trends Across All 10 Edison Schools in Our Study

	Positive	Mixed	Negative
Norm Referenced	10 of 50	40 of 50	0 of 50
Criterion Referenced	5 of 49	21 of 49	23 of 49
TOTALS	15 of 99	61 of 99	23 of 99

Our trends for both the norm-referenced tests and the criterion-referenced tests are based on gains in performance, and not absolute performance levels. To summarize the main findings in Table 13:4, we include the following points:

- The majority of the trends, both norm- and criterion-referenced, were mixed, indicating that students in Edison schools are achieving at levels comparable to students in the comparison groups.
- Edison students tend to show larger gains on norm-referenced tests than on state or district criterion-referenced tests. This is likely to be due to a combination of different factors, two of which are that (i) Edison’s curriculum might not be adequately oriented to state standards in the various states in which they work and (ii) the norm-referenced tests are administered to fewer students and selectivity may come into play.

13.3 Comparison Between Edison's Findings and The Evaluation Center's Findings

It is clear from our findings that across all schools in our sample, Edison students do not perform as well as Edison claims in its annual reports. Table 13:5 presents a summary of our numerical ratings (5-point scale from strongly positive to strongly negative, see Section 2.5) for the 10 cases. Table 13:5 also depicts the ratings Edison gave to these cases in its 2000 annual report (Edison, 2000). This table provides an overview of how our findings compare with the findings reported in Edison's annual performance reports.

Table 13:5 Comparison of Results by Edison Schools and The Evaluation Center

School	Edison Schools Inc.				Evaluation Center	
	Edison Rating from 1999 Annual Report	Edison Rating in 2000 for 1999-00 Alone	Edison Rating in Numerical 2000 for All Years Since Opening		Rating for All Years Since Opening	Numerical Rating
Roosevelt-Edison Charter School	Strongly Positive	Negative	Positive	4	Mixed	3
Henry E.S. Reeves Elementary School	Positive	Strongly Positive	Strongly Positive	5	Mixed	3
Dodge-Edison Elementary	Positive	Strongly Positive	Strongly Positive	5	Positive	4
Jardine-Edison Junior Academy	Positive	Negative	Strongly Positive	5	Positive	4
Boston Renaissance Charter School	Mixed	Strongly Positive	Positive	4	Mixed	3
Seven Hills Charter School	Strongly Positive	Mixed	Mixed	3	Mixed	3
Dr. Martin Luther King Jr. Academy	Strongly Positive	Strongly Positive	Strongly Positive	5	Negative	2
Mt. Clemens Secondary Academies	Strongly Positive	Strongly Positive	Strongly Positive	5	Positive*	4
Mid-Michigan Public School Academy	Mixed	Positive	Mixed	3	Strongly Negative	1
Washington Elementary School	Strongly Positive	Positive	Positive	4	Negative	2

* Indicates very weak basis for rating the school

Average performance rating **4.30**

2.90

In terms of “value-added” performance, over time we found that students at only three of the ten Edison schools were performing better than the comparison groups we examined (overall positive rating). On the other hand, there were three schools whose gains scores were less than those in the comparison groups (overall negative rating) and the remaining four schools showed mixed results.

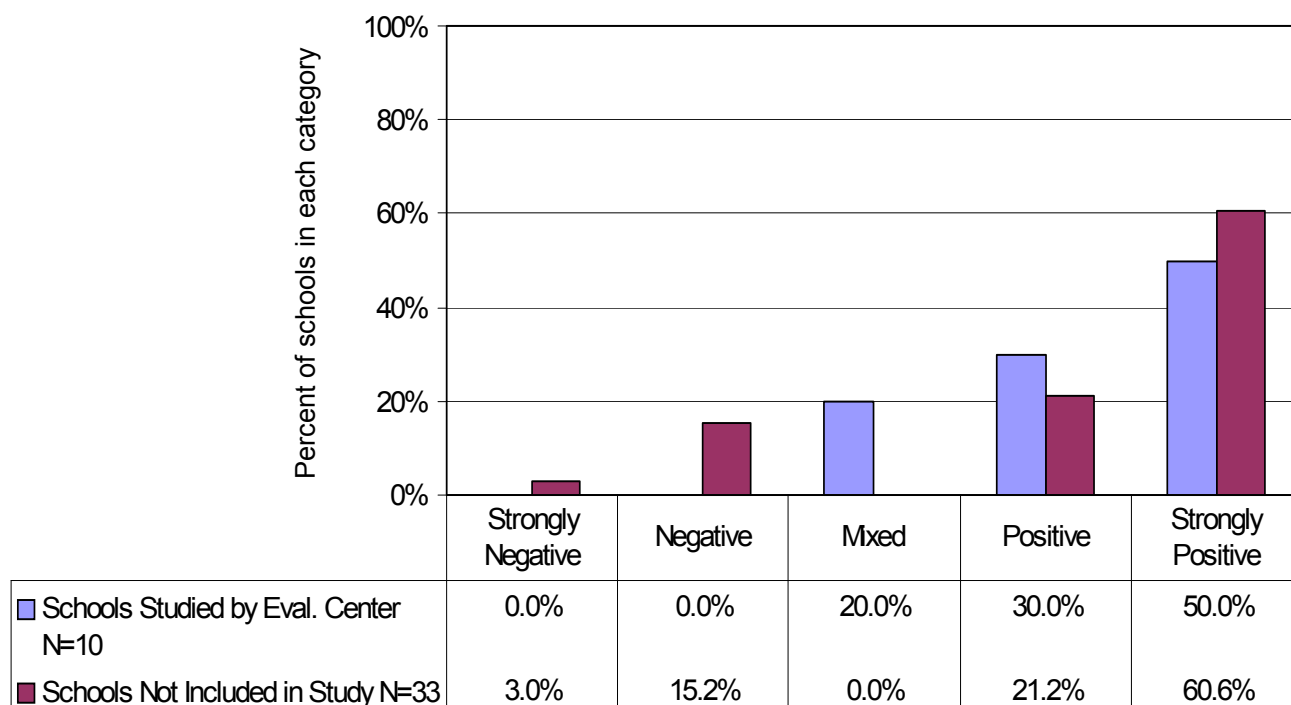


Exhibit 13:1 Edison’s Own Ratings of the Sampled Schools vs. Nonsampled Schools

There is a possibility of selection bias related to the schools selected for the evaluation. However, when we examined and analyzed the ratings Edison gave to the 10 sampled schools as compared with the 33 nonsampled schools, we found no significant difference. Edison rates each school on a 5-point scale, from Strongly Positive to Strongly Negative. Our analysis indicates there was no significant difference in the ratings Edison gave to the 10 schools included in our study and the 33 nonsampled schools for which it reported trends in its 2000 annual report.⁴ Exhibit 13:1 illustrates the difference in ratings Edison gave to the 10 sampled schools vs. the 33 nonsampled schools.

There was, however, a statistically significant⁵ difference in the mean ratings we gave the 10 schools based on the findings in this study as compared with Edison’s rating of these same 10 schools. Exhibit 13:2 illustrates the difference in the ratings based on the findings in this study as compared with the ratings Edison gave the schools in its 2000 annual report (Edison, 2000).

⁴ Wilcoxon rank-sum test, $p = .7709$

⁵ Wilcoxon signed ranks test, $p = .0039$

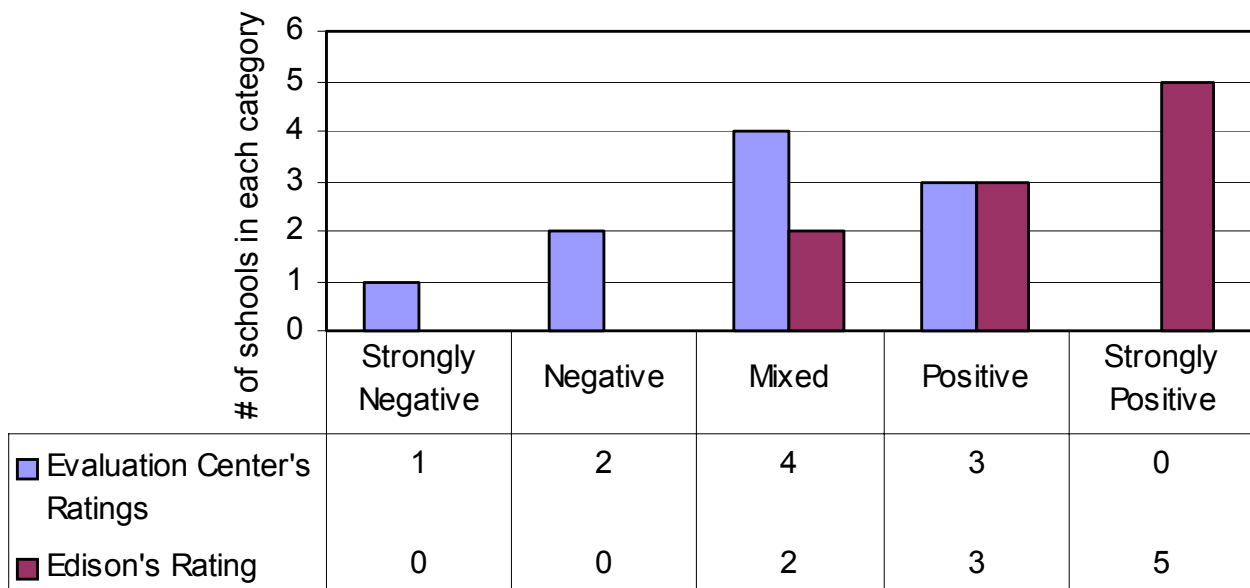


Exhibit 13:2 Edison's Ratings as Compared With The Evaluation Center Ratings
for the 10 Schools Considered in this Study

13.4 Recommendations Based on the Findings

Several different recommendations are presented below. The first are simply general recommendations aimed at improving the general nature of reporting student achievement data in evaluations such as this one. The second group of recommendations is more directly aimed at individuals and groups that are considering contracting out the operation of their schools or already have existing contracts with companies such as Edison Schools Inc.

Recommendations regarding reporting of student achievement

Listed below are some general recommendations for ways to more effectively report student achievement. There is obviously a large amount of data that a company such as Edison could share with participating schools, districts, and communities as well as with the research community. We offer the following suggestions for consideration:

- Prepare more comprehensive and complete annual reports.
- Define trends and how schools are rated more clearly, whether positive or negative.
- Follow the same trends from annual report to annual report.
- Report data across all years with consistent trends.
- Prepare academic reports as well as "lay" reports.
- Report all the sample size information with analyses.
- Cite sources for data on schools, particularly when these differ from district and state data.

Recommendations for Edison's partners or districts considering contracting out schools to a private company

It is especially important for district and charter school boards to ask questions and make requests in order to benefit from the collaboration with an outside company. Below, we have listed some relative suggestions to consider when contracting out schools to an outside company. These recommendations are based on the findings presented in the report or reflect lessons learned during the course of the study.

- Require that all student achievement data will be available to the public/researchers.
- Require both external and internal evaluations.
- Obtain and analyze both norm-referenced and criterion-referenced tests
- Ensure that all students are included in the test results and require evaluators to report the number of students included in test results.
- Ensure that all evaluation findings are made available to right-to-know audiences.

Additional recommendations for district or charter school boards when contracting out educational services to EMOs can be found in Lin & Hassel (1999) and Miron (2000).

When seeking new contracts, Edison promises districts and charter school groups that its model is a successful one. In this report, we examined the question of whether Edison did more in terms of student performance on standardized achievement tests. We selected schools that have a long record to trace, and we looked at a variety of test results with its first 10 schools. While our findings do not suggest that Edison did less, they do not suggest that the company did more with these schools in terms of gains on standardized tests.

The differences in Edison's and our own ratings for the 10 schools included in this study raise an important question. Since disparities exist between Edison's and our own results for these 10 schools, might they also exist if we were to analyze student achievement gains in its other schools?

Policymakers and investors want clear and unquestionable results whereas social scientists have a tendency to weigh their findings with cautious interpretations. We attempted to present our findings in a clear and cohesive manner and have tried to alert readers about the many limitations that need to be considered. Nevertheless, we believe that because of the many years of test results examined in our study and because of the multiple sources of student achievement data and multiple comparison groups, this is the most in-depth and extensive evaluation of student achievement in Edison schools to date. We hope that the extensive documentation of our methodology contained in this report will serve those individuals and groups who are interested in looking beyond summary tables and charts and who wish to pursue systematic evaluation of alternative schools.

References

- American Federation of Teachers. (1998). *Student achievement in Edison schools: Mixed results in an ongoing enterprise*. [On-line] Available: <http://www.aft.org/research/edisonproject/index.htm>.
- American Federation of Teachers. (2000). *Trends in student achievement for Edison Schools, Inc.: The emerging track record in an ongoing enterprise*. [On-line] Available: <http://www.aft.org/research/edisonproject/index.htm>.
- Balow, I.H., and Others. (1992). *Metropolitan achievement tests, seventh edition (MAT-7), form s, preprimer*. [On-line]. Available: <http://ericae.net/tc2/TC019276.htm>
- Chubb, J. E. (1997). Lessons in school reform from the Edison Project. In D. Ravitch & J. P. Viteritti (Eds.), *New schools for a new century* (pp. 86-122). New Haven: Yale University Press.
- Colorado Department of Education. (1997). *Colorado student assessment program: 1997 demonstration booklet, grade 4 reading and writing and Colorado student assessment program: 1997 administration manual, grade 4 reading and writing*. Denver, CO: Author. (ERIC Document Reproduction Service No. ED 409 529)
- Colorado Department of Education. (2000). *1998-99 Colorado charter schools evaluation study: The characteristics, status and performance record of Colorado charter schools*. Denver, CO: Author.
- Conoley, J. C., & Impara, J. C. (Eds.). (1995). *The twelfth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurement.
- Digest of Education Statistics. (1997). *Public elementary schools, by grade span and average school size, by state: 1995-96*. Table 97. [On-line] Available: <http://nces.ed.gov/pubs/digest97/d97t097.html>.
- Doclar, M. (2000, March 2). The Edison enigma: School management company praised, panned in Sherman. *Fort Worth Star-Telegraph*, p.1
- Edison. (1997). *Annual report on school performance*. New York: Author
- Edison. (1999). *Second annual report on school performance*. New York: Author
- Edison. (2000). *Third annual report on school performance*. New York: Author. [On-line] Available: <http://www.edisonschools.com/annualreport2000.pdf>.
- Farber, P. (1998). Boston: Renaissance charter school. *The American prospect*, 30.
- Florida Department of Education. (2000a). *Florida's comprehensive assessment test home page*. [On-line] Available: <http://www.firn.edu/doe/sas/fcat.htm>

- Florida Department of Education. (2000b). *Florida school indicators report*. [On-line] Available: <http://www.firn.edu/doe/doehome.htm>
- Fowler, D. (1999). Edison project's history outlined, Parts 1-4. (4 part article on Edison in the *Herald Democrat* (December 15-19, 1999).
- Glass, G. V, McGraw, B., & Smith, M. L. (1981). *Meta-analysis in the social research*. Beverly Hills, CA: Sage.
- Gomez J. J., & Shay, S. A. (1998). *Evaluation of the Edison project school first interim report: 1996-97 school year*. Miami: Miami-Dade County Public Schools.
- Gomez J. J., & Shay, S. A. (1999). *Evaluation of the Edison project school second interim report: 1997-98 school year*. Miami: Miami-Dade County Public Schools.
- Gomez J. J., & Shay, S. A. (2000). *Evaluation of the Edison project school third interim report: 1998-99 school year*. Miami: Miami-Dade County Public Schools.
- Hoover, H. D., and others, (1993). *Iowa Tests of Basic Skills, Forms K and L, Level 5*. [On-line] Available: <http://ericae.net/tc2/TC018925.htm>
- Horn, J., & Miron, G. (1999). *Evaluation of the Michigan charter school academy initiative*. Kalamazoo, MI: Western Michigan University Evaluation Center. [On-line] Available: <http://www.wmich.edu/evalctr/>
- Horn, J. & Miron, G. (2000). *An evaluation of the Michigan charter school academy initiative: Performance, accountability, and impact*. Kalamazoo, MI: Western Michigan University Evaluation Center.[On-line] Available: <http://www.wmich.edu/evalctr/>
- Impara, J. C., & Plake, B. S. (Eds.). (1998). *The thirteenth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurement.
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards*, 2nd edition. Thousand Oaks, CA: Sage.
- Kansas State Department of Education. (1999). *Kansas school building report card, school year 1998-1999*. [On-line]. Available: <http://www.ksbe.state.ks.us/rcard/D02591816.html> [2000, March 9].
- Kepple, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Lin, M., & Hassel, B. (1999). *Charting a clear course: A resource guide for building successful partnerships between charter schools and school management organizations*. Charter Friends National Network. [On-line]. Available: www.charterfriends.org/partnerships.html.

- Massachusetts Department of Education. (1998). *Massachusetts charter schools: Expanding the possibilities of public education*. [On-line] Available: <http://www.doe.mass.edu/cs.www/report98/default.htm>
- Massachusetts Department of Education. (1999). *School and district profiles*. [On-line] Available: <http://www.doe.mass.edu/pic.www/pic.html>
- Massachusetts Department of Education. (2000). *Massachusetts Department of Education homepage* [On-line] Available: <http://www.doe.mass.edu>
- McNeil, D. (1996). *Epidemiological research methods*. New York: Wiley.
- Michigan Department of Education. (1998). *Michigan educational assessment program*. Lansing, MI: Author.
- Michigan Department of Education. (2000). *1999 Michigan school report: District information*. [On line] Available: http://www.state.mis.us/mde/cfdata/msr99/msr_dist.cfm [March 12, 2000]
Building Information. [On line] Available: http://www.state.mis.us/mde/cfdata/msr99/msr_bldg.cfm [January 17, 2000]
- Mid-Michigan Public School Academy. (1999). *Annual education report 1998-1999*. Lansing, MI: Author.
- Mid-Michigan Public School Academy. (2000). *School profile report 1999-2000*. Lansing, MI: Author.
- Minneapolis Public Schools. (2000). *Edison project school information report*. Minneapolis: Author.
- Miron, G. (2000, April). "What's public about Michigan's charter schools: Lessons in school reform from statewide evaluations of charter schools." Paper presented at the AERA annual meeting, New Orleans.
- Mislevy⁶, R. J. (1996). *Reading achievement test-score analysis: Washington-Edison School, Sherman Independent School District* (Available from Edison Schools, 521 Fifth Avenue, 15th Floor, New York, NY 10175)
- Mislevy, R. J. (1996-1998). *Reading achievement test-score analysis: Boston Renaissance Charter School* (Available from Edison Schools, 521 Fifth Avenue, 15th Floor, New York, NY 10175)
- Mislevy, R. J. (1996-1998). *Reading achievement test-score analysis: Dodge-Edison vs. control schools, Wichita Unified School District #259* (Available from Edison Schools, 521 Fifth Avenue, 15th Floor, New York, NY 10175)

⁶ Note: Additional information regarding the Mislevy reading studies conducted for the Edison Schools is available on-line at <http://www.aft.org/research/edisonproject/sfa/index>.

- Mislevy, R. J. (1996-1998). *Reading achievement test-score analysis: King-Edison vs. control schools, Mount Clemens Community Schools* (Available from Edison Schools, 521 Fifth Avenue, 15th Floor, New York, NY 10175)
- Mislevy, R. J. (1997-1998). *Reading achievement test-score analysis: Roosevelt-Edison School, Colorado Springs, CO.* (Available from Edison Schools, 521 Fifth Avenue, 15th Floor, New York, NY 10175)
- Murphy, L. L., Impara, J. C., & Plake, B. S.(Eds.). (1999). *Tests in print V: An index to tests, test reviews, and the literature on specific tests.* Lincoln, NE: Buros Institute of Mental Measurements, University of Nebraska Press.
- Peterson, P. (1998). *A guide to recent studies of school effectiveness by the Edison Project and the American Federation of Teachers.* (Occasional Paper). Cambridge, MA: Program on Education Policy and Governance, Harvard University.
- Poggio, J. (1999). *An overview of the Kansas assessment programs.* [On-line] Available: <http://www.ksbe.state.ks.us/outcomes/ksassessments.html>.
- Psychological Corporation. (1999). *Stanford Achievement Test (SAT-9), primary 1, ninth edition.* [On-line] Available: <http://ericae.net/tc2/TC019831.htm>
- Renewal Inspection Report. (1999). *Renewal inspection report: Boston Renaissance Charter School.* Malden, MA: Massachusetts Department of Education.
On-line: http://www.doe.mass.edu/cs.www/reports/2000/inspection/boston_ren.pdf.
- School District Eleven. (1997). *Roosevelt-Edison achievement data 1996-1997.* (Available from School District Eleven, Colorado Springs, CO)
- Shay, S. A. (2000). *A longitudinal study of achievement outcomes in a privatized public school: A growth curve analysis.* Unpublished doctoral dissertation, University of Miami, Coral Gables, Florida.
- Texas Assessment of Academic Skills. (1998). 1997-98 report (Publication number 97.12.) [On-line] Available: <http://Erica.net/ericdc/ED4244260.htm>
- Texas Education Agency. (1999). *Student assessment division.* [On-line]. Available: <http://www.tea.state.tx.us/student.assessment/techdig/chapt1.htm>
- Wichita State University. (1996). *An independent program evaluation for the Dodge-Edison Partnership School: First year interim report.* Wichita, KS: Wichita State University, Department of Educational Administration.
- Worcester School Department. (1997). [Comparison: Iowa Test of Basic Skills Test, grade 3 reading test: Flagg Street and Seven Hills charter school.] Unpublished raw data.

Appendix A

Description of the Norm- and Criterion-Referenced Tests Considered in the Study

In this appendix, the various norm- and criterion-referenced tests that yielded results considered in this study are described. The norm-referenced tests include the Iowa Test of Basic Skills, the Metropolitan Achievement Test, and the Stanford Achievement Test. The criterion-referenced tests are all state-mandated tests, except for one district-mandated test. These tests are part of state or district assessment programs. Below the various tests are listed and briefly described. For ease of reference, we have listed them in alphabetical order.

Colorado Student Assessment Program

The Colorado Student Assessment Program (CSAP) was mandated by the state legislature in 1993. The first statewide assessments of reading and writing were implemented in April 1997. These criterion-referenced tests were developed to assess achievement in relation to state content area standards.

As of Fall 1999, CSAP is assessing reading and writing at grades 3, 4, and 7 and math at grade 5. Assessments in math and science were scheduled to begin at grade 8 in 2000; and reading, writing, and math assessments are scheduled for grade 10 beginning in 2001. State reports of results provide information regarding the percentage of students achieving at each of four performance levels in each test area. In addition, breakdowns by gender, race/ethnicity, disability, and district are available (Colorado Department of Education, 1997).

Five Edison schools administer the CSAP as a tool to measure student achievement, and one of these is included in our study: Roosevelt-Edison Charter School.

District Achievement Levels Tests

The District Achievement Levels Tests (DALT) are a series of tests designed to assess student achievement in specific areas of a district's curriculum. Curriculum areas covered by these tests include reading, language, and mathematics. The tests are designed to determine students' mastery at specific achievement levels and are used to monitor student progress in learning the established curriculum. These tests are administered at the beginning and end of each school year. The DALT's validity is enhanced by the use of Rasch scaling, which is a statistical method that employs test items representing narrow bands of increasing difficulty. Progress is monitored through the use of growth scores referred to as Rasch units or, more simply, RIT scores. The DALT is required for all schools in the Colorado Springs school district. Therefore, Roosevelt-Edison Charter School participates in this criterion-referenced test.

Florida Comprehensive Assessment Test

Florida Comprehensive Assessment Test (FCAT) is a statewide assessment program designed to measure challenging learning milestones for elementary, middle, and high school levels in Florida. FCAT measures the first four standards of Goal 3 of Florida's System of School Improvement and Accountability, with an emphasis on reading and mathematics as defined by the Sunshine State Standards.

FCAT was administered for the first time in January and February 1998 in order to establish baseline information on the achievement of Florida students and schools. The second FCAT administration took place during January and February 1999. The FCAT is administered in grade 4 (reading), grade 5 (mathematics), and grades 8 and 10 (reading and mathematics). Scores on the FCAT can range from 100 to 500. Five student achievement levels categorize students based on the scale score ranges. Level 5 is the highest level and Level 1 is the lowest (Florida Department of Education, 2000a). State, district, and school results are available on-line at <<http://www.firn.edu/doe/sas/fcat.htm>>.

Henry E.S. Reeves Elementary School, a school included in our study, is currently the only Edison school that has taken part in the FCAT.

Florida Writing Assessment

The Florida Writing Assessment, also known as Florida Writes, was mandated by the state legislature in 1990 and was first implemented in the spring of 1995. The test is administered to students in grades 4, 8, and 10. Scores of students with identified disabilities are not included in the test reports. This criterion-referenced test is a performance assessment with levels determined in relationship to proficiency in four areas: focus, organization, support, and conventions. According to the Florida Department of Education, the results of the Florida Writing Assessment are best used to identify strengths and weaknesses in writing instruction programs, rather than individual student achievement (Florida Department of Education, 2000a). State, district, and school results are available on-line at <<http://www.firn.edu/doe/sas/fwaphome.htm>>.

Students are required to read, plan, and respond independently to a specified topic within a 45-minute time frame. Fourth grade students receive a prompt directing them to Write to Tell a Story or to Write to Explain, while eighth and tenth grade students are prompted to Write to Explain or Write to Convince. Student responses are scored by readers who are trained to use Florida's holistic scoring method to evaluate a piece of writing for its overall quality while considering four elements: focus, organization, support (i.e., quality of details), and conventions (i.e., punctuation, capitalization, spelling, etc.). The writing is scored on a 6-point scale, with the best writing samples receiving a 6 and those writing samples not meeting expectations in the four areas of consideration receiving a 1.

One Edison school in our study, Henry E.S. Reeves Elementary School, participates in the Florida Writing Assessment.

Iowa Test of Basic Skills

The Iowa Test of Basic Skills (ITBS) has been available since 1955 (Murphy, Impara, & Plake, 1999). The ITBS, a norm-referenced test, is designed to assess student achievement in basic skills, covering grades K-9. In our analyses, we considered the following results from this test: grade equivalents, standard scores, national percentiles, and normal curve equivalents. The ITBS is available in three forms: Complete Battery, Core Battery, and Survey Battery. The Complete Battery assesses major objectives expected at each grade level. The Core Battery consists of the entire tests in reading, language, and mathematics found in the Complete Battery. The Survey Battery contains select questions from the Complete Battery and provides scores in reading, language, and mathematics (Hoover et al., 1993). Reviews of ITBS can be located in *The Thirteenth Mental Measurements Yearbook* (Impara & Plake, 1998).

At least four of the ten schools included in our study use the ITBS. For two of these schools, Roosevelt-Edison Charter School and Dr. Martin Luther King Jr. Academy, we received data sets from Edison that contained individual student results on the ITBS so that we could analyze individual student gains.

Kansas Assessment Program

The Kansas Assessment Program includes criterion-referenced tests in math, reading, writing, science, and social studies. These tests are based on state curriculum standards. These tests are reviewed and redesigned on a yearly basis by a committee of educators, representing specific content areas, selected from local nominations. The only three subject area tests considered in our analyses are the reading, math, and writing assessments. The reading test is administered at grades 3, 7, and 10. Math assessments are administered at grades 4, 7, and 10; and the writing assessments are administered to grades 5, 8, and 11. Scores are reported for each student and summarized by grade for each building and district (Poggio, 1999).

Four Edison schools currently list the Kansas Reading and Math Assessments as a tool used to measure student achievement. Two of those schools are included in our study: Dodge-Edison Elementary School and Jardine-Edison Junior Academy.

Massachusetts Comprehensive Assessment System

The Massachusetts Comprehensive Assessment System (MCAS) was mandated by the Education Reform Law of 1993. MCAS tests in English language arts, mathematics, and science and technology were implemented at grades 4, 8, and 10 in May 1998. MCAS is a criterion-referenced test based on the Massachusetts Curriculum Frameworks for each content area covered. The Education Reform Law of 1993 requires all students to participate in the MCAS testing, although the results for students with special needs and/or students who have limited English proficiency are reported separately from the totals for any given school. The Education Reform Law requires that students pass the grade 10 tests in order to receive a high school diploma. This requirement will be implemented beginning with the class of 2003 (Massachusetts Department of Education, 1999).

Student results on the MCAS tests are reported according to four performance levels: Advanced, Proficient, Needs Improvement, and Failing. The Massachusetts Department of Education (2000) has elaborated descriptions of what is expected for each performance level by grade and subject. The performance level descriptions were used as a basis for determining the minimum score for each performance level on each MCAS test. The descriptions are meant to help teachers, students, parents, and others understand the meaning of the MCAS results. The general performance level definitions are as follows:

Advanced. Students at this level demonstrate a comprehensive and in-depth understanding of rigorous subject matter and provide sophisticated solutions to complex problems.

Proficient. Students at this level demonstrate a solid understanding of challenging subject matter and solve a wide variety of problems.

Needs improvement. Students at this level demonstrate a partial understanding of subject matter and solve some simple problems.

Failing. Students at this level demonstrate a minimal understanding of subject matter and do not solve even simple problems.

Two Edison schools currently list the MCAS as a tool used to measure student achievement. Both schools are included in our study: Boston Renaissance Charter School and Seven Hills Charter School.

Metropolitan Achievement Tests, 7th Edition

The Metropolitan Achievement Tests (MAT) series has been in publication since 1931 (Murphy, Impara, and Plake, 1999). The seventh edition (MAT-7) is a norm-referenced series of tests designed to provide achievement data on students in grades K-12. Skill and content areas covered by MAT-7 include reading, mathematics, language, science, and social studies. The MAT-7 emphasizes critical thinking skills for solving problems in realistic contexts. Guidelines from the National Council of Teachers of Mathematics Curriculum and Evaluation Standards for School Mathematics were used to develop subtests in mathematics relevant to either traditional or progressive instructional objectives (Balow, 1992). Reviews of MAT-7 are contained in *The Twelfth Mental Measurements Yearbook* (Conoley & Impara, 1995).

Among the many Edison schools that list the MAT-7 as a tool to measure student achievement, six are included in our study. We received data sets from Edison containing the individual student results for three of these schools: Dodge Edison Elementary, Seven-Hills Charter School, and Mid-Michigan Public School Academy. For Jardine-Edison, Boston Renaissance, and Martin Luther King Academy, the other three schools in our study that administer the MAT-7, we included grade level averages on these tests if the results were available in a verifying form.

Michigan Educational Assessment Program

The Michigan Educational Assessment Program (MEAP) originated in 1969. From 1969 to 1973 the program used commercially produced norm-referenced tests to measure student achievement at the fourth and eighth grade levels. In 1973, the MEAP began using criterion-referenced tests, developed by teachers and curriculum specialists, in the areas of reading and mathematics for grades 4 and 7; grade 10 was added in 1979. The first objective-referenced test in science was developed in 1986. These objective-referenced tests were designed to measure student achievement of basic skills. Beginning in 1989, new tests designed to measure skills in context were implemented. The Essential Skills Reading Test for grades 4, 7, and 10 was implemented in 1989. The Essential Skills Mathematics Test was implemented for grades 4, 7, and 10 in 1991. High School Proficiency Tests in mathematics, science, and communication arts (reading and writing) were implemented in the spring of 1996. In the spring of 1998, the communication arts endorsement was divided into separate reading and writing components. Students graduating from Michigan public schools can earn endorsements in mathematics, science, reading, and/or writing by scoring at any of the top three levels of the individual proficiency examinations. These endorsements are recorded on the students' permanent transcripts. (Michigan Department of Education, 1998).

MEAP is a criterion-referenced testing program based on specific criteria set by Michigan educators. Criteria for each test are based on what students should have learned up to the grade that the test is administered.

The *MEAP Handbook* (MEAP, 1999) suggests that while the results should not be compared across grade levels, comparing MEAP trends over time within a grade is an excellent use of the data. The fact that subject tests are not administered in consecutive grades is one of the main constraints for making such comparisons across grades. In our own analysis, we used consecutive class cohorts as suggested by the Michigan Educational Assessment Program.

Fourteen Edison schools currently list the MEAP as a tool used to measure student achievement. Three of those schools are included in our study: (1) Martin Luther King Jr. Academy, (2) Mount Clemens Secondary Academy, and (3) Mid-Michigan Public School Academy.

Stanford Achievement Test Series, 9th Edition

The Stanford Achievement Test series (SAT) has been in publication since 1923 (Murphy, Impara, & Plake, 1999). The 9th edition (SAT-9) is a norm-referenced series of tests designed to provide achievement data on students in grades K-13. The series includes both multiple-choice and open-ended subtests that can be used independently or in combinations. Areas covered by the multiple-choice subtests include reading, mathematics, language, spelling, study skills, listening, science, and social science. Areas covered by the open-ended subtests include reading, mathematics, science, social science, and language. The open-ended subtests are designed for those instructional objectives that are more accurately measured using performance-based or student-generated responses (Psychological Corp., 1999). Available: <http://ericae.net/tc2/TC019831.htm>. Reviews of SAT-9

are contained in *The Thirteenth Mental Measurements Yearbook* (Impara & Plake, 1998; see test 292).

Among the many Edison schools that list SAT-9 as a tool used to measure student achievement, three are included in our study: (1) Henry E. S. Reeves Elementary School, (2) Boston Renaissance Charter School, and (3) Seven Hills Charter School.

Texas Assessment of Academic Skills (TAAS)

The Texas Assessment of Academic Skills (TAAS) was first administered during the 1990-91 school year. The TAAS is a criterion-referenced test designed to measure student mastery of the statewide curriculum in reading, writing, and mathematics. Students are assessed in reading and mathematics in grades 3 through 8, writing in grades 4 and 8, and all three areas at the exit level (Texas Assessment, 1998). Available: <http://Erica.net/ericdc/ED4244260.htm>). In 1994, science and social studies assessments were added to the TAAS series. These tests are normally administered in grades 4, 8, and exit level. Tests for Biology I and Algebra I were “benchmarked” (a standard-setting process) in 1994. The Algebra I test was rebenchmarked in the spring of 1995, and both the Biology I and Algebra I end-of-course tests were given to eligible students in the fall of 1995. In the spring of 1998 the U.S. History and English II end-of-course tests were benchmarked. In 1996, the Texas Education Agency contracted with Harcourt Brace Educational Measurement to obtain national comparative data in the subject areas assessed. The Metropolitan Achievement Tests, Seventh Edition were used to collect data using a stratified classroom sampling design. Results of this study can be obtained from the Texas Education Agency.¹ According to the Texas Education Agency, results from the TAAS are “useful for providing a snapshot of individual student performance, an indicator of areas in which further diagnosis is warranted, and a mechanism for providing a ‘level playing field’ for comparing the performance of campuses and districts” (Texas Education Agency, 1999).²

Seven Edison schools currently list TAAS as a tool used to measure student achievement. One of those schools was included in the ten under study: Washington Elementary.

¹ The URL is < <http://www.tea.state.tx.us/student.assessment/results/studies/ncds.htm>>

² The URL is <<http://www.tea.stte.tx.us/student.assessment/techdig/chap1.pdf>>

Appendix B

Normal Curve Equivalents Based on National Percentile Ranks

Percentile Rank	NCE	Percentile Rank	NCE	Percentile Rank	NCE	Percentile Rank	NCE
99	99	74	64	49	49	24	35
98	93	73	63	48	49	23	34
97	90	72	62	47	48	22	34
96	87	71	62	46	48	21	33
95	85	70	61	45	47	20	32
94	83	69	60	44	47	19	32
93	81	68	60	43	46	18	31
92	80	67	59	42	46	17	30
91	78	66	59	41	45	16	29
90	77	65	58	40	45	15	28
89	76	64	58	39	44	14	27
88	75	63	57	38	44	13	26
87	74	62	56	37	43	12	25
86	73	61	56	36	42	11	24
85	72	60	55	35	42	10	23
84	71	59	55	34	41	9	22
83	70	58	54	33	41	8	20
82	69	57	54	32	40	7	19
81	68	56	53	31	40	6	17
80	68	55	53	30	39	5	15
79	67	54	52	29	38	4	13
78	66	53	52	28	38	3	10
77	66	52	51	27	37	2	7
76	65	51	51	26	36	1	1
75	64	50	50	25	36		

Source: U.S. Department of Education (1999)
<http://www.ed.gov/offices/OERI/BlueRibbonSchools/nce.html>)

Appendix C

Criteria for Rating Student Achievement Trends Presented by Edison and AFT

Peterson (1998) was asked by The Edison Project to compare the results presented by AFT and Edison. In his analysis, he used the following categories to rate the designs behind each trend presented by Edison and AFT (Peterson, 1998, pp. 4-8):

1. *Gold Medal: Randomized Experiments.* Studies designed as randomized experiments win the Gold Medal. Subjects are randomly assigned to test and control groups. The two groups are followed over time to see if differences appear. Any differences may reasonably be attributed to the test, because the two groups can be assumed to be similar in all other respects. Unfortunately, the Gold Standard, though often achieved in medical research, is seldom realized in the field of education.
2. *Silver Medal: Matched Comparison Group.* The Silver Medal goes to designs that compare treatment groups with a comparison group that initially is similar to the treatment group in all identifiable respects. In education, baseline data are needed to determine average initial test scores, ethnicity, family education, and other family characteristics. Statistical adjustments can be made for minor differences in initial conditions. Still, this design does not win the Gold Medal, because the two groups may differ in some unknown respect.
3. *Bronze Medal: Similar Groups, Statistical Controls.* The Bronze Medal is given to studies designed to compare a treatment group with a comparison group that is broadly similar (though not matched on key characteristics). Although the two groups do not match, statistical analyses may be able to adjust for differences in initial conditions, if differences between the two groups are moderate. However, to the extent that the two groups differ in both known and unknown characteristics, comparisons may be biased.
4. *Changes by individuals over time, comparisons with a similar group but with no statistical control.* Fourth place goes to designs that compare changes in the performances of the same individuals to changes over the same time period of similar individuals, but with no statistical controls for such differences. Such a study assumes that the individuals' initial characteristics do not affect the changes over time. This is a problematic assumption. This type of study also assumes that attrition from the sample does not systematically differ between test and control group.

5. *Changes by individuals over time, comparisons with a dissimilar group without appropriate statistical control.* The quality of studies with design 4, as with designs 6, 7, and 8, depends upon the groups of individuals being compared. If the group is different, the comparison is less likely to be valid. Edison has often been asked to administer schools that are in trouble. If Edison schools are compared with state or national norms or with schools in other parts of the country, any differences in gains may underestimate Edison school effects.
6. *Changes of cohorts over time, no statistical control.* In sixth place, one finds studies that assess the gains in the performance of a cohort of students (as compared with students in another location or to national or state norms) from one year to the next. This design provides information on many of the same students from one time period to the next, but it also includes newcomers to the school and excludes those who have left the school. To the extent that the students are not the same, this study design has less validity than design 5 studies. The rate of turnover is especially high in low-income neighborhoods.
7. *Changes in successive cohorts, similar comparison group.* Next are designs that compare the gains made by a succession of cohorts in a treatment place to the gains made by a succession of students in a similar comparison place, with no statistical adjustments for differences in the initial characteristics of the two groups. Although this analytical technique is widely practiced in estimating the effects of a school, it has many defects. Most seriously, the gains or losses of successive cohorts in either the treatment or comparison places are not gains or losses realized by the same individuals. To show that successive cohorts of students have gained or lost may only show that the composition of that cohort has changed. As economist Eric Hanushek has pointed out, “approaches that do not concentrate on individual gains give very . . . inferior estimates of school performance.”
8. *Changes in successive cohorts, comparison group not similar.* In eighth place, one finds studies that compare a cross-section of individuals in a treatment place to a prior cross-section of students in that same place. The only comparison groups are district, state, or national norms. These are highly problematic studies. As Robert Meyer has pointed out, “indicators other than the value-added performance indicator [gain scores] convey potentially inaccurate information about school quality and therefore could severely harm the policymaking process and distort the school choices of students and families.”
9. *Performance levels, no change information.* Finally, we find studies that compare the level of performance of a test group of students with those of some other group or with state or national standards at any point in time. The test is deemed successful if the standard is matched. This widely used method of estimating school effects is profoundly misleading, because the characteristics of the test group may be dramatically different from those of the comparison group.

Appendix D

Results Provided by Edison for Dillingham Intermediate School, Sherman, TX

Edison Students at Dillingham Intermediate School

Texas Assessment of Academic Skills				
<i>Percentage of Students Met Minimum Expectations (1998-1999)</i>				
Grade \ Year	Reading		Mathematics	
	Spring 98	Spring 99	Spring 98	Spring 99
5	84.3%	83.0%	73.9%	84.4%
6	76.1%	84.7%	57.0%	73.0%

Texas Assessment of Academic Skills				
<i>Percentage of Students Met Minimum Expectations (1997-1998)</i>				
Grade \ Year	Reading		Mathematics	
	1996-97	1997-98	1996-97	1997-98
5	78 %	85 %	71 %	53 %
6	82 %	82 %	63 %	77 %

All Students at Dillingham Intermediate School

Texas Assessment of Academic Skills				
<i>Percentage of Students Met Minimum Expectations (1998-1999)</i>				
Grade \ Year	Reading		Mathematics	
	Spring 98	Spring 99	Spring 98	Spring 99
5	83.2%	76.0%	70.8%	71.8%
6	81.2%	87.1%	71.4%	82.4%

Appendix E Chi Square Tables

Roosevelt Charter School (Colorado Springs,CO) Chi square results from CSAP

2x4	Level				Row total	Grade 3 Reading 1998
	Unsatisfactory	Partial	Proficient	Advanced		
Roosevelt	34	27	48	3	112	
%	30.36	24.11	42.86	2.68		
District	310	529	1,462	182	2483	
%	12.48	21.3	58.88	7.33		
State	6,251	10,448	30,328	4,187	51,214	
%	12.21	20.4	59.22	8.18		

$\chi^2 (3, N=2,595) = 34.196, p < .001$

$\chi^2 (3, N=51,326) = 40.013, p < .001$

2x2	Level		Row total	Grade 3 Reading 1998
	Failed	Passed		
Roosevelt	61	51	112	
%	54.46	45.54		
District	839	1644	2,483	
%	33.79	66.21		
State	16,699	34,515	51,214	
%	32.61	67.39		

$\chi^2 (1, N=2595) = 20.220, p < .001$

$\chi^2 (1, N=51326) = 24.280, p < .001$

2x4	Level				Row total	Grade 3 Reading 1999
	Unsatisfactory	Partial	Proficient	Advanced		
Roosevelt	22	32	50	2	106	
%	20.75	30.19	47.17	1.89		
District	449	1067	787	76	2379	
%	18.87	44.85	33.08	3.19		
State	5919	10769	31274	4319	52,281	
%	11.32	20.6	59.82	8.26		

$\chi^2 (3, N=2485) = 11.655, p = .009$

$\chi^2 (3, N=52387) = 21.062, p < .001$

2x2	Level		Row total	Grade 3 Reading 1999
	Failed	Passed		
Roosevelt	54	52	106	
%	50.94	49.06		
District	1516	863	2,379	
%	63.72	36.28		
State	16,688	35,593	52,281	
%	31.92	68.08		

$\chi^2 (1, N=2485) = 7.126, p = .008$

$\chi^2 (1, N=52387) = 17.606, p < .001$

2x4	Level				Row total	Grade 4 Reading 1997
	Unsatisfactory	Partial	Proficient	Advanced		
Roosevelt	20	33	40	2	95	
%	21.05	34.74	42.11	2.11		
State	5516	14561	24619	1024	45720	
%	11.32	20.6	59.82	8.26		

$\chi^2 (3, N=45815) = 9.022, p = .029$

2x2	Level		Row total	
	Failed	Passed		
Roosevelt	53	42	95	
%	55.79	44.21		
State	20077	25643	45,720	$\chi^2 (1, N=45815) = 5.429, p = .020$
%	43.91	56.09		

**Grade 4 Reading
1997**

2x4	Level				Row total	
	Unsatisfactory	Partial	Proficient	Advanced		
Roosevelt	25	55	33	1	114	
%	21.93	48.25	28.95	0.88		
District	227	700	1300	150	2377	$\chi^2 (3, N=2491) = 47.931, p < .001$
%	9.55	29.45	54.69	6.31		
State	5182	15565	26520	3123	50,390	$\chi^2 (3, N=50504) = 43.830, p < .001$
%	10.28	30.89	52.63	6.2		

**Grade 4 Reading
1998**

2x2	Level		Row total	
	Failed	Passed		
Roosevelt	80	34	114	
%	70.18	29.82		
District	927	1450	2,377	$\chi^2 (1, N=2491) = 43.904, p < .001$
%	39	61		
State	20,747	29,643	50,390	$\chi^2 (1, N=50504) = 39.482, p < .001$
%	41.17	58.83		

**Grade 4 Reading
1998**

2x4	Level				Row total	
	Unsatisfactory	Partial	Proficient	Advanced		
Roosevelt	25	46	40	3	114	
%	21.93	40.35	35.09	2.63		
District	205	746	1289	150	2390	$\chi^2 (3, N=2504) = 33.639, p < .001$
%	8.58	31.21	53.93	6.28		
State	5314	15436	27721	3734	52,205	$\chi^2 (3, N=52319) = 30.072, p < .001$
%	10.18	29.57	53.1	7.15		

**Grade 4 Reading
1999**

2x2	Level		Row total	
	Failed	Passed		
Roosevelt	71	43	114	
%	62.28	37.72		
District	951	1439	2,390	$\chi^2 (1, N=2504) = 22.783, p < .001$
%	39.79	60.21		
State	20,750	31,455	52,205	$\chi^2 (1, N=52319) = 24.107, p < .001$
%	39.75	60.25		

**Grade 4 Reading
1999**

2x4	Level				Row total	
	Unsatisfactory	Partial	Proficient	Advanced		
Roosevelt	41	34	18	0	93	
%	44.09	36.56	19.35	0		
District						
%						
State	10896	21343	13902	1491	47,632	$\chi^2 (3, N=47725) = 25.619, p < .001$
%	22.88	45.81	29	3		

**Grade 4 Writing
1997**

2x2	Level		Row total	
	Failed	Passed		
Roosevelt	75	18	93	Grade 4 Writing 1997
%	80.65	19.35		
District				
%				
State	32,239	15,393	47,632	$\chi^2 (1, N=47725) = 7.132, p = .008$
%	67.68	32.32		

2x4	Level				Row total	
	Unsatisfactory	Partial	Proficient	Advanced		
Roosevelt	35	54	18	3	110	Grade 4 Writing 1998
%	31.82	49.09	16.36	2.73		
District	417	1000	735	148	2300	$\chi^2 (3, N=2410) = 21.715, p < .001$
%	10.18	29.57	53.1	7.15		
State	10890	21275	15589	3118	50,872	$\chi^2 (3, N=50982) = 16.326, p < .001$
%	18.13	43.48	31.96	6.43		

2x2	Level		Row total	
	Failed	Passed		
Roosevelt	89	21	110	Grade 4 Writing 1998
%	80.91	19.09		
District	1417	883	2,300	$\chi^2 (1, N=2410) = 16.683, p < .001$
%	61.61	38.39		
State	32,165	18,707	50,872	$\chi^2 (1, N=50982) = 14.766, p < .001$
%	63.23	36.77		

2x4	Level				Row total	
	Unsatisfactory	Partial	Proficient	Advanced		
Roosevelt	26	36	20	1	83	Grade 4 Writing 1999
%	31.33	43.37	24.1	1.2		
District	383	1138	451	50	2022	$\chi^2 (3, N=2105) = 9.318, p = .025$
%	18.94	56.28	22.3	2.47		
State	8516	23427	16530	1601	50,074	$\chi^2 (3, N=50157) = 13.207, p = .004$
%	17.01	46.78	33.01	3.2		

2x2	Level		Row total	
	Failed	Passed		
Roosevelt	62	21	83	Grade 4 Writing 1999
%	74.7	25.3		
District	1521	501	2,022	$\chi^2 (1, N=2105) = 0.012, p = .914$
%	75.22	24.78		
State	31,943	18,131	50,074	$\chi^2 (1, N=50157) = 4.269, p = .039$
%	63.79	36.21		

2x4	Level				Row total	
	Unsatisfactory	Partial	Proficient	Advanced		
Roosevelt	34	58	23	5	120	Grade 5 Math 1999
%	28.33	48.33	19.17	4.17		
District	302	1,057	856	252	2,467	$\chi^2 (3, N=2,587) = 35.952, p < .001$
%	12.24	42.85	34.7	10.21		
State	6,892	21,049	18,709	7,051	53,701	$\chi^2 (3, N=53,821) = 40.688, p < .001$
%	12.83	39.2	34.84	13.13		

2x2	Level		Row total	
	Failed	Passed		
Roosevelt	92	58	120	
%	76.67	23.33		
District	1,359	1,108	2,467	$\chi^2 (1, N=2,587) = 21.636, p < .001$
%	55.09	44.91		
State	27,941	25,760	53,701	$\chi^2 (1, N=53,821) = 29.118, p < .001$
%	47.97	7		

Grade 5 Math
1999

2x4	Level				Row total	
	Unsatisfactory	Partial	Proficient	Advanced		
Roosevelt	54	49	30	0	133	
%	40.6	36.84	22.56	0		
District	321	676	1221	50	2268	$\chi^2 (3, N=2401) = 84.850, p < .001$
%	14.15	29.81	53.84	2.2		
State	6975	14550	28086	2163	51,774	$\chi^2 (3, N=51907) = 105.859, p < .001$
%	13.47	28.1	54.25	4.18		

Grade 7 Reading
1998

2x2	Level		Row total	
	Failed	Passed		
Roosevelt	103	30	133	
%	77.44	22.56		
District	997	1271	2,268	$\chi^2 (1, N=2401) = 56.741, p < .001$
%	43.96	56.04		
State	21,525	30,249	51,774	$\chi^2 (1, N=51907) = 70.220, p < .001$
%	41.57	58.43		

Grade 7 Reading
1998

2x4	Level				Row total	
	Unsatisfactory	Partial	Proficient	Advanced		
Roosevelt	7	89	17	0	113	
%	6.19	78.76	15.04	0		
District	43	1162	883	0	2088	$\chi^2 (2, N=2201) = 37.602, p < .001$
%	2.06	55.65	42.29	0		
State	1074	26405	21611	541	49,631	$\chi^2 (3, N=49744) = 44.549, p < .001$
%	2.16	53.2	43.54	1.09		

Grade 7 Reading
1999

2x2	Level		Row total	
	Failed	Passed		
Roosevelt	96	17	113	
%	84.96	15.04		
District	1205	883	2,088	$\chi^2 (1, N=2201) = 32.922, p < .001$
%	57.71	42.29		
State	27,479	22,152	49,631	$\chi^2 (1, N=49744) = 39.955, p < .001$
%	55.37	44.63		

Grade 7 Reading
1999

Henry E.S. Reeves Elementary School (Miami-Dade, FL)
Chi square results from FCAT

2x4	Level				Level 5	Row total	
	Level 1	Level 2	Level 3	Level 4			
Reeves	103	22	30	14	0	169	
%	60.95	13.02	17.75	8.28	0		
District	9,887	4,160	5,778	2,774	232	22,831	$\chi^2 (4, N=23,000) = 22.097, p < .001$
%	43.31	18.22	25.31	12.15	1.02		
State	47,903	26,304	47,976	27,860	4,646	154,689	$\chi^2 (4, N=154,858) = 74.050, p < .001$
%	30.97	17	31.01	18.01	3		

**Grade 4 Reading
1999**

2x2	Level		Row total	
	Failed	Passed		
Reeves	103	66	169	
%	60.95	39.05		
District	9,887	12,944	22,831	$\chi^2 (1, N=23,000) = 21.251, p < .001$
%	43.31	56.69		
State	47,903	106,786	154,689	$\chi^2 (1, N=154,858) = 70.933, p < .001$
%	30.97	69.03		

**Grade 4 Reading
1999**

2x4	Level				Level 5	Row total	
	Level 1	Level 2	Level 3	Level 4			
Reeves	100	44	17	5	2	168	
%	59.52	26.19	10.12	2.98	1.19		
District	9,031	7,083	3,769	1,771	221	21,881	$\chi^2 (4, N=22,049) = 25.796, p < .001$
%	41.27	32.37	17.22	8.12	1.01		
State	42,030	49,609	34,590	21,060	3,007	150,296	$\chi^2 (4, N=150,464) = 89.338, p < .001$
%	27.96	33.01	23.01	14.01	2.00		

**Grade 5 Math
1999**

2x2	Level		Row total	
	Failed	Passed		
Reeves	100	68	168	
%	59.52	40.48		
District	9,031	12,850	21,881	$\chi^2 (1, N=22,049) = 22.888, p < .001$
%	41.27	58.73		
State	42,030	108,266	150,296	$\chi^2 (1, N=150,464) = 82.905, p < .001$
%	27.96	72.04		

**Grade 5 Math
1999**

2x2	Level		Row total	
	Failed	Passed		
Reeves	73	105	178	
%	41.01	58.99		
District	6,944	15,515	22,459	$\chi^2 (1, N=22,637) = 8.410, p = .004$
%	30.92	69.08		
State	44,978	105,014	149,992	$\chi^2 (1, N=150,170) = 10.289, p < .001$
%	29.99	70.01		

**Grade 4 Florida Writes
1998**

2x2	Level		Row total	
	Failed	Passed		
Reeves	55	113	168	
%	32.74	67.26		
District	6,699	16,422	23,121	$\chi^2 (1, N=23,289) = 1.148, p = .284$
%	28.97	71.03		
State	41,703	112,787	154,490	$\chi^2 (1, N=154,658) = 2.809, p < .094$
%	26.99	73.01		

**Grade 4 Florida Writes
1999**

Boston Renaissance (Boston, MA)
Chi Square Results from MCAS

2x4	Level				Row total	Grade 4 English 1998 $\chi^2 (2, N=4,392) = 2.198, p = .333$ $\chi^2 (3, N=75,197) = 50.704, p < .001$
	Fail	Need Improvement	Proficient	Advanced		
Boston Renaissance	39	67	2	0	108	
%	36.11	62.04	1.85	0.00		
District	1,692	2,417	175	0	4,284	
%	39.50	56.42	4.08	0.00		
State	11,129	49,071	14,144	745	75,089	
%	14.82	65.35	18.84	0.99		

2x2	Level		Row total	Grade 4 English 1998 $\chi^2 (1, N=4,392) = 1.358, p = .244$ $\chi^2 (1, N=75,197) = 21.945, p < .001$
	Failed	Passed		
Boston Renaissance	106	2	108	
%	98.15	1.85		
District	4,109	175	4,284	
%	95.92	4.08		
State	60,200	14,889	75,089	
%	80.17	19.83		

2x4	Level				Row total	Grade 4 English 1999 $\chi^2 (2, N=4,612) = 1.662, p = .436$ $\chi^2 (2, N=76,137) = 70.050, p < .001$
	Fail	Need Improvement	Proficient	Advanced		
Boston Renaissance	39	66	3	0	108	
%	36.11	61.11	2.78	0.00		
District	1,452	2,822	230	0	4,504	
%	32.24	62.66	5.11	0.00		
State	9,098	50,945	15,986	0	76,029	
%	11.97	67.01	21.03	0.00		

2x2	Level		Row total	Grade 4 English 1999 $\chi^2 (1, N=4,612) = 1.192, p = .275$ $\chi^2 (1, N=76,137) = 21.647, p < .001$
	Failed	Passed		
Boston Renaissance	105	3	108	
%	97.22	2.78		
District	4,274	230	4,504	
%	94.89	5.11		
State	60,043	15,986	76,029	
%	78.97	21.03		

2x4	Level				Row total	Grade 4 Math 1998 $\chi^2 (3, N=4,645) = 2.784, p = .426$ $\chi^2 (3, N=75,988) = 94.320, p < .001$
	Fail	Need Improvement	Proficient	Advanced		
Boston Renaissance	64	39	3	1	107	
%	59.81	36.45	2.8	0.93		
District	2,610	1,557	278	93	4,538	
%	57.51	34.45	6.13	2.05		
State	17,240	33,065	17,301	8,275	75,881	
%	22.72	43.57	22.80	10.91		

2x2	Level		Row total	Grade 4 Math 1998 $\chi^2 (1, N=4,645) = 2.773, p = .096$ $\chi^2 (1, N=75,988) = 42.969, p < .001$
	Failed	Passed		
Boston Renaissance	103	4	107	
%	96.26	3.74		
District	4,167	371	4,538	
%	91.82	8.18		
State	50,305	25,576	75,881	
%	66.29	33.71		

2x4	Level				Row total	Grade 4 Math 1999
	Fail	Need Improvement	Proficient	Advanced		
Boston Renaissance	59	40	9	0	108	
%	54.63	37.04	8.33	0.00		
District	2,063	1,986	522	193	4,764	
%	43.30	41.69	10.96	4.05		
State	14,572	33,843	18,473	9,241	76,129	
%	19.14	44.45	24.27	12.14		

$\chi^2 (3, N=4,872) = 8.702, p = .034$

$\chi^2 (3, N=76,237) = 96.511, p < .001$

2x2	Level		Row total	Grade 4 Math 1999
	Failed	Passed		
Boston Renaissance	99	9	108	
%	91.67	8.33		
District	4,049	715	4,764	
%	84.99	15.01		
State	48,415	27,714	76,129	
%	63.6	36.4		

$\chi^2 (1, N=4,872) = 3.719, p = .054$

$\chi^2 (1, N=76,237) = 36.723, p < .001$

2x4	Level				Row total	Grade 4 Science & Tech 1998
	Fail	Need Improvement	Proficient	Advanced		
Boston Renaissance	42	52	13	0	107	
%	39.25	48.60	12.15	0.00		
District	2,022	2,012	503	47	4,584	
%	44.11	43.89	10.97	1.03		
State	8,986	30,040	31,584	4,514	75,124	
%	11.96	39.99	42.04	6.01		

$\chi^2 (3, N=4,691) = 2.316, p = .510$

$\chi^2 (3, N=75,231) = 97.457, p < .001$

2x2	Level		Row total	Grade 4 Science & Tech 1998
	Failed	Passed		
Boston Renaissance	94	13	107	
%	87.85	12.15		
District	4,034	550	4,584	
%	88	12		
State	39,026	36,098	75,124	
%	51.95	48.05		

$\chi^2 (1, N=4,691) = .002, p = .962$

$\chi^2 (1, N=75,231) = 55.176, p < .001$

2x4	Level				Row total	Grade 4 Science & Tech 1999
	Fail	Need Improvement	Proficient	Advanced		
Boston Renaissance	27	64	15	0	106	
%	25.47	60.38	14.15	0.00		
District	1,417	2,341	851	96	4,705	
%	30.12	49.76	18.09	2.04		
State	6,134	27,657	35,657	7,700	76,898	
%	7.98	35.97	46.04	10.01		

$\chi^2 (3, N=4,811) = 6.140, p = .105$

$\chi^2 (3, N=77,004) = 92.038, p < .001$

2x2	Level		Row total	Grade 4 Science & Tech 1999
	Failed	Passed		
Boston Renaissance	91	15	106	
%	85.85	14.15		
District	3,758	947	4,705	
%	79.87	20.13		
State	33,791	43,107	76,898	
%	43.94	56.06		

$\chi^2 (1, N=4,811) = 2.315, p = .128$

$\chi^2 (1, N=77,004) = 75.445, p < 0.001$

2x4	Level				Row total	Grade 8 English 1998
	Fail	Need Improvement	Proficient	Advanced		
Boston Renaissance	38	50	24	0	112	
%	33.93	44.64	21.43	0.00		
District	1,232	1,419	1,128	40	3,819	
%	32.26	37.16	29.54	1.05		
State	9,550	21,180	35,589	2,055	68,374	
%	13.97	30.98	52.05	3.01		

$\chi^2 (3, N=3,931) = 5.340, p = .149$

$\chi^2 (3, N=68,486) = 62.091, p < .001$

2x2	Level		Row total	Grade 8 English 1998
	Failed	Passed		
Boston Renaissance	88	24	112	
%	78.57	21.43		
District	2,651	1168	3,819	
%	69.42	30.58		
State	30,730	37,644	68,374	
%	44.94	55.06		

$\chi^2 (1, N=3,931) = 4.317, p = .038$

$\chi^2 (1, N=68,486) = 51.088, p < .001$

2x4	Level				Row total	Grade 8 English 1999
	Fail	Need Improvement	Proficient	Advanced		
Boston Renaissance	30	60	40	1	131	
%	22.90	45.80	30.53	0.76		
District	1,093	1,504	1,283	39	3,919	
%	27.89	38.38	32.74	1.00		
State	8,398	21,713	37,185	2,106	69,402	
%	12.10	31.29	53.58	3.03		

$\chi^2 (3, N=4,050) = 3.205, p = .361$

$\chi^2 (3, N=69,533) = 36.578, p < .001$

2x2	Level		Row total	Grade 8 English 1999
	Failed	Passed		
Boston Renaissance	90	41	131	
%	68.7	31.3		
District	2,597	1322	3,919	
%	66.27	33.73		
State	30,111	39,291	69,402	
%	43.39	56.61		

$\chi^2 (1, N=4,050) = .337, p = .562$

$\chi^2 (1, N=69,533) = 34.108, p < .001$

2x4	Level				Row total	Grade 8 Math 1998
	Fail	Need Improvement	Proficient	Advanced		
Boston Renaissance	104	10	0	0	114	
%	91.23	8.77	0.00	0.00		
District	2,808	605	451	123	3,987	
%	70.43	15.17	11.31	3.09		
State	28,882	17,933	15,783	5,521	68,119	
%	42.40	26.33	23.33	8.10		

$\chi^2 (3, N=4,101) = 26.194, p < .001$

$\chi^2 (3, N=68,233) = 112.865, p < .001$

2x2	Level		Row total	Grade 8 Math 1998
	Failed	Passed		
Boston Renaissance	114	0	114	
%	100	0		
District	3,413	574	3,987	
%	85.6	14.4		
State	46,815	21,304	68,119	
%	68.73	31.27		

$\chi^2 (1, N=4,101) = 19.083, p < .001$

$\chi^2 (1, N=68,233) = 51.838, p < .001$

2x4	Level				Row total	Grade 8 Math 1999
	Fail	Need Improvement	Proficient	Advanced		
Boston Renaissance	88	35	8	0	131	
%	67.18	26.72	6.11	0.00		
District	2,562	848	539	168	4,117	
%	62.23	20.6	13.09	4.08		
State	28,443	22,076	15,684	4,280	70,483	$\chi^2 (3, N=4,248) = 12.942, p = .005$
%	40.35	31.32	22.25	6.07		

2x2	Level		Row total	Grade 8 Math 1999
	Failed	Passed		
Boston Renaissance	123	8	131	
%	93.89	6.11		
District	3,410	707	4,117	
%	82.83	17.17		
State	50,519	19,964	70,483	$\chi^2 (1, N=4,248) = 11.106, p < .001$
%	71.68	28.32		

2x4	Level				Row total	Grade 8 Science & Tech 1998
	Fail	Need Improvement	Proficient	Advanced		
Boston Renaissance	104	9	1	0	114	
%	91.23	7.89	0.88	0.00		
District	2,972	647	368	0	3,987	
%	74.54	16.23	9.23	0.00		
State	28,184	21,379	17,938	1,380	68,881	$\chi^2 (3, N=68,995) = 119.931, p < .001$
%	40.92	31.04	26.04	2.00		

2x2	Level		Row total	Grade 8 Science & Tech 1998
	Failed	Passed		
Boston Renaissance	113	1	114	
%	99.12	0.88		
District	3,619	368	3,987	
%	90.77	9.23		
State	49,563	19,318	68,881	$\chi^2 (1, N=4,101) = 9.444, p = .002$
%	71.95	28.05		

2x4	Level				Row total	Grade 8 Science & Tech 1999
	Fail	Need Improvement	Proficient	Advanced		
Boston Renaissance	113	16	3	0	132	
%	85.61	12.12	2.27	0.00		
District	3,052	617	335	42	4,046	
%	75.43	15.25	8.28	1.04		
State	31,977	19,238	16,399	3,566	71,180	$\chi^2 (3, N=4,178) = 9.644, p = .022$
%	44.92	27.03	23.03	5.01		

2x2	Level		Row total	Grade 8 Science & Tech 1999
	Failed	Passed		
Boston Renaissance	129	3	132	
%	97.73	2.27		
District	3,669	377	4,046	
%	90.68	9.32		
State	51,215	19,965	71,180	$\chi^2 (1, N=4,178) = 7.674, p = .006$
%	71.95	28.05		

Seven Hills Charter School (Worcester, MA)

Chi Square Results from MCAS

2x4	Level				Row total	Grade 4 English 1998
	Fail	Need Improvement	Proficient	Advanced		
Seven Hills	34	45	4	0	83	
%	41	54	5	0		
District	395	1241	210	0	1846	
%	21	67	11	0		
State	11134	49094	14142	745	75,115	
%	15	65	19	1		

2x2	Level		Row total	Grade 4 English 1998
	Failed	Passed		
Seven Hills	79	4	83	
%	95.18	4.82		
District	1636	210	1,846	
%	88.62	11.38		
State	60,228	14,887	75,115	
%	80.18	19.82		

$\chi^2 (1, N=1929) = 3.462, p = 0.063$

$\chi^2 (1, N=75,198) = 11.746, p < 0.001$

2x4	Level				Row total	Grade 4 English 1999
	Fail	Need Improvement	Proficient	Advanced		
Seven Hills	38	41	4	0	83	
%	46	50	5	0		
District	319	1269	313	0	1901	
%	17	67	17	0		
State	9098	50971	15985	0	76,054	
%	12	67	21	0		

2x2	Level		Row total	Grade 4 English 1999
	Failed	Passed		
Seven Hills	79	4	83	
%	95.18	4.82		
District	1588	313	1,901	
%	83.53	16.47		
State	60,069	15,985	76,054	
%	78.98	21		

$\chi^2 (1, N=1984) = 8.034, p = 0.005$

$\chi^2 (1, N=76,137) = 13.113, p < 0.001$

2x4	Level				Row total	Grade 4 Math 1998
	Fail	Need Improvement	Proficient	Advanced		
Seven Hills	44	31	7	1	83	
%	53	37	8	1		
District	456	950	374	139	1919	
%	24	50	19	7		
State	17260	33073	17297	8275	75,905	
%	22.7	43.5	22.7	10.9		

2x2	Level		Row total	Grade 4 Math 1998
	Failed	Passed		
Seven Hills	75	8	83	
%	90	10		
District	1406	513	1,919	
%	73	27		
State	50,333	25,572	75,905	
%	66.31	33.69		

$\chi^2 (1, N=2002) = 12.076, p < 0.001$

$\chi^2 (1, N=75,988) = 21.476, p < 0.001$

2x4	Level				Row total	Grade 4 Math 1999
	Fail	Need Improvement	Proficient	Advanced		
Seven Hills	34	31	15	3	83	
%	41	37	18	4		
District	494	903	371	179	1947	
%	25.37	46.38	19.01	9.19		
State	14597	33852	18467	9238	76,154	$\chi^2 (3, N=2030) = 11.655, p = 0.009$
%	19	44	24	12		$\chi^2 (3, N=76,237) = 27.730, p < 0.001$

2x2	Level		Row total	
	Failed	Passed		
Seven Hills	65	18	83	
%	78	22		
District	1397	550	1,947	
%	71.75	28.25		
State	48,449	27,705	76,154	$\chi^2 (1, N=2030) = 1.701, p = 0.192$
%	64	36		$\chi^2 (1, N=76,237) = 7.735, p = 0.005$

2x4	Level				Row total	Grade 4 Science & Tech. 1998
	Fail	Need Improvement	Proficient	Advanced		
Seven Hills	20	46	16	1	83	
%	24	55	19	1		
District	260	855	725	79	1919	
%	14	45	38	4		
State	9008	30046	31581	4513	75,148	$\chi^2 (3, N=2002) = 17.463, p < 0.001$
%	12	40	42	6		$\chi^2 (3, N=75,231) = 28.471, p < 0.001$

2x2	Level		Row total	Grade 4 Science & Tech. 1998
	Failed	Passed		
Seven Hills	66	17	83	
%	80	20		
District	1115	804	1,919	
%	58	42		
State	39,054	36,094	75,148	$\chi^2 (1, N=2002) = 15.082, p < 0.001$
%	52	48		$\chi^2 (1, N=75,231) = 25.209, p < 0.001$

2x4	Level				Row total	Grade 4 Science & Tech. 1999
	Fail	Need Improvement	Proficient	Advanced		
Seven Hills	23	45	13	2	83	
%	28	54	16	2		
District	220	869	718	140	1947	
%	11	45	37	7		
State	6137	27676	35408	7699	76,920	$\chi^2 (3, N=2030) = 32.086, p < 0.001$
%	8	36	46	10		$\chi^2 (3, N=77,003) = 69.426, p < 0.001$

2x2	Level		Row total	Grade 4 Science & Tech. 1999
	Failed	Passed		
Seven Hills	68	15	83	
%	82	18		
District	1089	858	1,947	
%	56	44		
State	33,813	43,107	76,920	$\chi^2 (1, N=2030) = 21.948, p < 0.001$
%	44	56		$\chi^2 (1, N=77,003) = 48.510, p < 0.001$

2x4	Level				Row total	Grade 8 English 1998
	Fail	Need Improvement	Proficient	Advanced		
Seven Hills	12	20	20	1	53	
%	23	38	38	2		
District	368	535	598	15	1516	
%	24	35	40	1		
State	9576	21211	35593	2054	68,434	
%	14	31	52	3		

$\chi^2 (3, N=1569) = 0.585, p = 0.900$

$\chi^2 (3, N=68,487) = 5.900, p = 0.117$

2x2	Level		Row total	Grade 8 English 1998
	Failed	Passed		
Seven Hills	32	21	53	
%	60	40		
District	903	613	1,516	
%	60	40		
State	30,787	37,647	68,434	
%				

$\chi^2 (1, N=1569) = 0.014, p = 0.906$

$\chi^2 (1, N=68,487) = 5.068, p = 0.024$

2x4	Level				Row total	Grade 8 English 1999
	Fail	Need Improvement	Proficient	Advanced		
Seven Hills	26	37	17	1	81	
%	32	46	21	1		
District	422	587	511	16	1536	
%	27	38	33	1		
State	8402	21736	37208	2107	69,453	
%	12	31	54	3		

$\chi^2 (3, N=1617) = 5.285, p = 0.152$

$\chi^2 (3, N=69,534) = 48.959, p < 0.001$

2x2	Level		Row total	Grade 8 English 1999
	Failed	Passed		
Seven Hills	63	18	81	
%	78	22		
District	1009	527	1,536	
%	66	34		
State	30,138	39,315	69,453	
%	43	57		

$\chi^2 (1, N=1617) = 5.031, p = 0.025$

$\chi^2 (1, N=69,534) = 38.933, p < 0.001$

2x4	Level				Row total	Grade 8 Math 1998
	Fail	Need Improvement	Proficient	Advanced		
Seven Hills	33	9	9	2	53	
%	62	17	17	4		
District	964	329	216	46	1555	
%	62	21	14	3		
State	28953	17935	15864	5519	68,271	
%	42.41	26.27	23	8		

$\chi^2 (3, N=1608) = 0.890, p = 0.828$

$\chi^2 (3, N=68,324) = 8.771, p = 0.032$

2x2	Level		Row total	Grade 8 Math 1998
	Failed	Passed		
Seven Hills	42	11	53	
%	79.25	20.75		
District	1293	262	1,555	
%	83	17		
State	46,888	21,383	68,271	
%	69	31		

$\chi^2 (1, N=1608) = 0.555, p = 0.456$

$\chi^2 (1, N=68,324) = 2.749, p = 0.097$

2x4	Level				Row total	Grade 8 Math 1999
	Fail	Need Improvement	Proficient	Advanced		
Seven Hills	64	12	4	0	80	
%	80	15	5	0		
District	936	405	213	33	1587	
%	59	25.52	13.42	2		
State	28466	22099	15688	4280	70,533	$\chi^2 (3, N=1667) = 14.792, p = .002$
%	40	31	22	6		$\chi^2 (3, N=70,613) = 53.431, p < 0.001$

2x2	Level		Row total	Grade 8 Math 1999
	Failed	Passed		
Seven Hills	76	4	80	
%	95	5		
District	1341	246	1,587	
%	84.5	15.5		
State	50,565	19,968	70,533	$\chi^2 (1, N=1667) = 6.588, p = 0.010$
%	72	28		$\chi^2 (1, N=70,613) = 21.406, p < 0.001$

2x4	Level				Row total	Grade 8 Science & Tech. 1998
	Fail	Need Improvement	Proficient	Advanced		
Seven Hills	29	16	8	0	53	
%	55	30	15	0		
District	1000	338	201	0	1539	
%	65	22	13	0		
State	28259	21373	17931	1380	68,943	$\chi^2 (2, N=1592) = 2.555, p = 0.279$
%	41	31	26	2		$\chi^2 (3, N=68,996) = 5.933, p = 0.115$

2x2	Level		Row total	Grade 8 Science & Tech. 1998
	Failed	Passed		
Seven Hills	45	8	53	
%	85	15		
District	1338	201	1,539	
%	87	13		
State	49,632	19,311	68,943	$\chi^2 (1, N=1592) = 0.186, p = 0.666$
%	72	28		$\chi^2 (1, N=68,996) = 4.382, p = 0.036$

2x4	Level				Row total	Grade 8 Science & Tech. 1999
	Fail	Need Improvement	Proficient	Advanced		
Seven Hills	62	13	3	2	80	
%	77.5	16	4	2.5		
District	1037	282	210	30	1559	
%	66.52	18	13	2		
State	32028	19241	16398	3563	71,230	$\chi^2 (3, N=1639) = 7.173, p = 0.067$
%	45	27	23	5		$\chi^2 (3, N=71,310) = 36.130, p < 0.001$

2x2	Level		Row total	Grade 8 Science & Tech. 1999
	Failed	Passed		
Seven Hills	75	5	80	
%	94	6		
District	1319	240	1,559	
%	84.61	15.39		
State	51,269	19,961	71,230	$\chi^2 (1, N=1639) = 5.005, p = 0.025$
%	72	28		$\chi^2 (3, N=71,310) = 18.792, p < 0.001$

Martin Luther King Jr. Academy (Mt. Clemens, MI)
Chi Square Results from MEAP

2x3	Level			Row total	Grade 4 Math 1995
	Satisfactory	Moderate	Low		
Martin Luther King Acad.	27	13	15	55	
%	49.09	23.64	27.27		
George Washington Sch.	40	11	5	56	
%	71.43	19.64	8.93		
State	72,250	24,744	20,284	117,278	
%	61.61	21.1	17.3		
					$\chi^2 (2, N=111) = 7.681, p = .021$
					$\chi^2 (2, N=117,333) = 4.729, p = .094$

2x2	Level		Row total	Grade 4 Math 1995
	Failed	Passed		
Martin Luther King Acad.	28	27	55	
%	50.91	49.09		
George Washington Sch.	16	40	56	
%	28.57	71.43		
State	45,028	72,250	117,278	
%	38.39	61.61		
				$\chi^2 (1, N=111) = 5.787, p = .016$
				$\chi^2 (1, N=117,333) = 3.640, p = .056$

2x3	Level			Row Total	Grade 4 Math 1996
	Satisfactory	Moderate	Low		
Martin Luther King Acad.	39	16	13	68	
%	57.35	23.53	19.12		
George Washington Sch.	76	26	19	121	
%	62.81	21.49	15.7		
State	72,485	25,729	16,652	114,866	
%	63.1	22.4	14.5		
					$\chi^2 (2, N=189) = 0.595, p = .743$
					$\chi^2 (2, N=114,934) = 1.396, p = .498$

2x2	Level		Row total	Grade 4 Math 1996
	Failed	Passed		
Martin Luther King Acad.	29	39	68	
%	42.65	57.35		
George Washington Sch.	45	76	121	
%	37.19	62.81		
State	42,381	72,485	114,866	
%	36.9	63.1		
				$\chi^2 (1, N=189) = 0.544, p = .461$
				$\chi^2 (1, N=114,934) = 0.965, p = .326$

2x3	Level			Row total	Grade 4 Math 1997
	Satisfactory	Moderate	Low		
Martin Luther King Acad.	42	14	20	76	
%	55.26	18.42	26.32		
George Washington Sch.	62	27	39	128	
%	48.44	21.09	30.47		
State	69,462	24,571	20,774	114,807	
%	60.5	21.4	18.09		
					$\chi^2 (2, N=204) = 0.890, p = .641$
					$\chi^2 (2, N=114,883) = 3.496, p = .174$

2x2	Level		Row total	
	Failed	Passed		
Martin Luther King Acad.	34	42	76	
%	44.74	55.26		
George Washington Sch.	66	62	128	$\chi^2 (1, N=204) = 0.889, p = .346$
%	51.56	48.44		
State	45,345	69,462	114,807	$\chi^2 (1, N=114,883) = 0.873, p = .350$
%	39.5	60.5		

**Grade 4 Math
1997**

2x3	Level			Row	
	Satisfactory	Moderate	Low		
Martin Luther King Acad.	42	18	10	70	
%	60	25.71	14.29		
George Washington Sch.	77	26	9	112	$\chi^2 (2, N=182) = 2.228, p = .328$
%	68.75	23.21	8.04		
State	84,314	20,139	9,328	113,781	$\chi^2 (2, N=113,851) = 7.756, p = .023$
%	74.1	17.7	8.2		

**Grade 4 Math
1998**

2x2	Level		Row total	
	Failed	Passed		
Martin Luther King Acad.	28	42	70	
%	40	60		
George Washington Sch.	35	77	112	$\chi^2 (1, N=182) = 1.457, p = .227$
%	31.25	68.75		
State	29,467	84,314	113,781	$\chi^2 (1, N=113,851) = 7.248, p = .007$
%	25.9	74.1		

**Grade 4 Math
1998**

2x3	Level			Row	
	Satisfactory	Moderate	Low		
Martin Luther King Acad.	56	22	28	106	
%	52.83	20.75	26.42		
George Washington Sch.	58	29	17	104	$\chi^2 (2, N=210) = 3.666, p = .160$
%	55.77	27.88	16.35		
State	84,420	20,832	12,461	117,713	$\chi^2 (2, N=117,819) = 30.861, p < .001$
%	71.72	17.7	10.59		

**Grade 4 Math
1999**

2x2	Level		Row total	
	Failed	Passed		
Martin Luther King Acad.	50	56	106	
%	44.74	55.26		
George Washington Sch.	46	58	104	$\chi^2 (1, N=210) = 0.183, p = .669$
%	44.23	55.77		
State	33,293	84,420	117,713	$\chi^2 (1, N=117,819) = 18.617, p < .001$
%	28.28	71.72		

**Grade 4 Math
1999**

2x3	Level			Row total	
	Satisfactory	Moderate	Low		
Martin Luther King Acad.	34	13	8	55	
%	61.82	23.64	14.55		
George Washington Sch.	26	17	13	56	$\chi^2 (2, N=111) = 2.782, p = .249$
%	46.43	30.36	23.21		
State	51,062	35,614	30,462	117,138	$\chi^2 (2, N=117,193) = 7.794, p = .020$
%	43.59	30.4	26.1		

**Grade 4 Reading
1995**

2x2	Level		Row total	
	Failed	Passed		
Martin Luther King Acad.	21	34	55	
%	38.18	61.82		
George Washington Sch.	30	26	56	$\chi^2 (1, N=111) = 2.646, p = .104$
%	53.57	46.43		
State	66,076	51,062	117,138	$\chi^2 (1, N=117,193) = 7.427, p = .006$
%	56.41	43.59		

**Grade 4 Reading
1995**

2x3	Level			Row total	
	Satisfactory	Moderate	Low		
Martin Luther King Acad.	27	25	16	68	
%	39.71	36.76	23.53		
George Washington Sch.	52	44	25	121	$\chi^2 (2, N=189) = 0.278, p = .870$
%	42.98	36.36	20.66		
State	57,266	35,453	21,914	114,633	$\chi^2 (2, N=114,701) = 2.870, p = .238$
%	49.96	30.93	19.12		

**Grade 4 Reading
1996**

2x2	Level		Row total	
	Failed	Passed		
Martin Luther King Acad.	41	27	68	
%	60.29	39.71		
George Washington Sch.	69	52	121	$\chi^2 (1, N=189) = 0.191, p = .662$
%	57.02	42.98		
State	57,367	57,266	114,633	$\chi^2 (1, N=114,701) = 2.856, p = .091$
%	50.04	49.96		

**Grade 4 Reading
1996**

2x3	Level			Row total	
	Satisfactory	Moderate	Low		
Martin Luther King Acad.	35	21	20	76	
%	46.05	27.63	26.23		
George Washington Sch.	50	4	37	91	$\chi^2 (2, N=167) = 18.076, p < .001$
%	54.95	4.4	40.66		
State	56,229	32,360	25,471	114,060	$\chi^2 (2, N=114,136) = 0.717, p = .699$
%	49.3	28.37	22.33		

**Grade 4 Reading
1997**

2x2	Level		Row total	
	Failed	Passed		
Martin Luther King Acad.	41	35	76	
%	53.95	46.05		
George Washington Sch.	41	50	91	$\chi^2 (1, N=167) = 1.310, p = .252$
%	45.05	54.95		
State	57,831	56,229	114,060	$\chi^2 (1, N=114,136) = 0.320, p = .572$
%	50.7	49.3		

**Grade 4 Reading
1997**

2x3	Level			Row total	
	Satisfactory	Moderate	Low		
Martin Luther King Acad.	34	20	16	70	
%	48.57	28.57	22.86		
George Washington Sch.	62	32	18	112	$\chi^2 (2, N=182) = 1.438, p = .487$
%	55.36	28.57	16.07		
State	66,578	29,534	17,489	113,601	$\chi^2 (2, N=113,671) = 3.910, p = .142$
%	58.61	26	15.4		

**Grade 4 Reading
1998**

2x2	Level		Row total	
	Failed	Passed		
Martin Luther King Acad.	36	34	70	
%	51.43	48.57		
George Washington Sch.	50	62	112	$\chi^2 (1, N=182) = 0.796, p = .372$
%	44.64	55.36		
State	47,023	66,578	113,601	$\chi^2 (1, N=113,671) = 2.904, p = .088$
%	41.39	58.61		

**Grade 4 Reading
1998**

2x3	Level			Row total	
	Satisfactory	Moderate	Low		
Martin Luther King Acad.	43	37	26	106	
%	40.57	34.91	24.53		
George Washington Sch.	46	33	25	104	$\chi^2 (2, N=210) = 0.330, p = .848$
%	44.23	31.73	24.04		
State	69,936	30,004	17,763	117,703	$\chi^2 (2, N=117,809) = 16.263, p < .001$
%	59.42	25.49	15.09		

**Grade 4 Reading
1999**

2x2	Level		Row total	
	Failed	Passed		
Martin Luther King Acad.	63	43	106	
%	59.43	40.57		
George Washington Sch.	58	46	104	$\chi^2 (1, N=210) = 0.289, p = .591$
%	55.77	44.23		
State	47,767	69,936	117,703	$\chi^2 (1, N=117,809) = 15.606, p < .001$
%	40.58	59.42		

**Grade 4 Reading
1999**

2x3	Level			Row total	
	Satisfactory	Moderate	Low		
Martin Luther King Acad.	11	36	7	54	
%	20.37	66.67	12.96		
George Washington Sch.	38	97	22	157	$\chi^2 (2, N=211) = 0.438, p = .803$
%	24.2	61.78	14.01		
State	31,565	72,795	13,106	117,466	$\chi^2 (2, N=117,520) = 1.199, p = .549$
%	27	62	11		

**Grade 5 Science
1996**

2x2	Level		Row total	
	Failed	Passed		
Martin Luther King Acad.	43	11	54	
%	79.63	20.37		
George Washington Sch.	119	38	157	$\chi^2 (1, N=211) = 0.331, p = .565$
%	75.8	24.2		
State	85,901	31,565	117,466	$\chi^2 (1, N=117,520) = 1.161, p = .281$
%	73.13	26.87		

**Grade 5 Science
1996**

2x3	Level			Row total	
	Satisfactory	Moderate	Low		
Martin Luther King Acad.	15	45	14	74	
%	20.27	60.81	18.92		
George Washington Sch.	41	55	21	117	$\chi^2 (2, N=211) = 5.047, p = .080$
%	35.04	47.01	17.95		
State	41,841	61,784	10,175	113,800	$\chi^2 (2, N=113,874) = 13.651, p < .001$
%	36.77	54.29	8.94		

**Grade 5 Science
1997**

2x2	Level		Row total	
	Failed	Passed		
Martin Luther King Acad.	59	15	74	
%	79.73	20.27		
George Washington Sch.	76	41	117	$\chi^2 (1, N=191) = 4.773, p = .029$
%	64.96	35.04		
State	71,959	41,841	113,800	$\chi^2 (1, N=113,874) = 8.658, p = .003$
%	63.23	36.77		

**Grade 5 Science
1997**

2x3	Level			Row total	
	Satisfactory	Moderate	Low		
Martin Luther King Acad.	17	44	10	71	
%	23.94	61.97	14.08		
George Washington Sch.	64	41	10	115	$\chi^2 (2, N=186) = 17.975, p < .001$
%	55.65	35.65	8.7		
State	46,016	60,121	7,816	113,953	$\chi^2 (2, N=114,024) = 11.034, p < .001$
%	40.38	52.76	6.86		

**Grade 5 Science
1998**

2x2	Level		Row total	
	Failed	Passed		
Martin Luther King Acad.	54	17	71	
%	76.06	23.94		
George Washington Sch.	51	64	115	$\chi^2 (1, N=186) = 17.953, p < .001$
%	44.35	55.65		
State	67,937	46,016	113,953	$\chi^2 (1, N=114,024) = 7.964, p = .005$
%	59.62	40.38		

**Grade 5 Science
1998**

2x3	Level			Row total	
	Satisfactory	Moderate	Low		
Martin Luther King Acad.	11	59	36	106	
%	10.38	55.66	33.96		
George Washington Sch.	50	35	1	86	$\chi^2 (2, N=192) = 62.768, p < .001$
%	58.14	40.7	1.16		
State	42,501	59,457	11,300	113,258	$\chi^2 (2, N=113,364) = 81.941, p < .001$
%	37.53	52.5	9.98		

**Grade 5 Science
1999**

2x2	Level		Row total	
	Failed	Passed		
Martin Luther King Acad.	95	11	106	
%	89.62	10.38		
George Washington Sch.	36	50	86	$\chi^2 (1, N=192) = 49.966, p < .001$
%	41.86	58.14		
State	70757	42501	113,258	$\chi^2 (1, N=113,364) = 33.303, p < .001$
%	62.47	37.53		

**Grade 5 Science
1999**

2x2	Level		Row total	
	Failed	Passed		
Martin Luther King Acad.	31	21	52	
%	59.62	40.38		
George Washington Sch.	74	79	153	$\chi^2 (1, N=205) = 1.966, p = .161$
%	48.37	51.63		
State	51846	64842	116,688	$\chi^2 (1, N=116,740) = 4.853, p = .028$
%	44.43	55.57		

**Grade 5 Writing
1996**

2x2	Level		Row total	
	Failed	Passed		
Martin Luther King Acad.	23	51	74	
%	31.08	68.92		
George Washington Sch.	35	78	113	$\chi^2 (1, N=187) = 0.000, p = .998$
%	30.97	69.03		
State	30104	83082	113,186	$\chi^2 (1, N=113,260) = 0.762, p = .383$
%	26.6	73.4		

**Grade 5 Writing
1997**

2x2	Level		Row total	
	Failed	Passed		
Martin Luther King Acad.	39	32	71	
%	54.93	45.07		
George Washington Sch.	41	73	114	$\chi^2 (1, N=185) = 6.411, p = .011$
%	35.96	64.04		
State	40389	72786	113,175	$\chi^2 (1, N=113,246) = 11.445, p < .001$
%	35.69	64.31		

**Grade 5 Writing
1998**

2x2	Level		Row total	
	Failed	Passed		
Martin Luther King Acad.	77	29	106	
%	72.64	27.36		
George Washington Sch.	38	50	88	$\chi^2 (1, N=194) = 17.287, p < .001$
%	43.18	56.82		
State	50948	61833	112,781	$\chi^2 (1, N=112,887) = 32.256, p < .001$
%	45.17	54.83		

**Grade 5 Writing
1999**

2x4	Level				Row total	
	Level1	Level2	Level3	Level4		
Martin Luther King Acad.	0	6	17	85	108	
%	0	5.56	15.74	78.7		
George Washington Sch.	2	24	29	30	85	$\chi^2 (3, N=193) = 40.063, p < .001$
%	2.35	28.24	34.12	35.29		
State	191	19,133	30,830	60,596	110,750	$\chi^2 (3, N=110,858) = 25.789, p < .001$
%	0.17	17.28	27.84	54.71		

**Grade 5 Social Studies
1999**

2x2	Level		Row total	
	Failed	Passed		
Martin Luther King Acad.	108	0	108	
%	100	0		
George Washington Sch.	83	2	85	$\chi^2 (1, N=193) = 2.568, p = .109$
%	97.65	2.35		
State	110559	191	110,750	$\chi^2 (1, N=110,858) = 0.187, p = .666$
%	99.83	0.17		

**Grade 5 Social Studies
1999**

Mt. Clemens Secondary Academies (Mt. Clemens, MI)
Chi Square Results from HSPT MEAP

2x4	Level				Row total	Grade 11 Math 1999
	Endorsed	Met	Basic	Not Endorsed		
Mt. Clemens Academy	3	7	5	7	22	
%	13.6	31.8	22.7	31.8		
District	13	38	20	28	99	$\chi^2 (3, N=121) = 0.345, p = .951$
%	13.1	38.4	20.2	28.3		
State	18,391	33,362	14,157	15,376	81,286	$\chi^2 (3, N=81,308) = 3.533, p = .317$
%	22.6	41	17.4	18.9		

2x2	Level		Row total	Grade 11 Math 1999
	Endorsed	Not Endorsed		
Mt. Clemens Academy	15	7	22	
%	71.7	28.3		
District	71	28	99	$\chi^2 (1, N=121) = 0.109, p = .741$
%	68.2	31.8		
State	65,910	15,376	81,286	$\chi^2 (1, N=81,330) = 2.387, p = .122$
%	81.1	18.9		

2x4	Level				Row total	Grade 11 Reading 1999
	Endorsed	Met	Basic	Not Endorsed		
Mt. Clemens Academy	2	13	4	4	23	
%	8.7	56.5	17.4	17.4		
District	8	52	26	25	111	$\chi^2 (3, N=134) = 0.966, p = .810$
%	7.2	46.6	23.4	22.5		
State	13,972	41,009	12,172	14,623	81,776	$\chi^2 (3, N=81,799) = 1.234, p = .745$
%	22.6	41	17.4	18.9		

2x2	Level		Row total	Grade 11 Reading 1999
	Endorsed	Not Endorsed		
Mt. Clemens Academy	19	4	23	
%	82.6	17.4		
District	86	25	111	$\chi^2 (1, N=134) = 0.296, p = .587$
%	77.5	22.5		
State	67,153	14,623	81,776	$\chi^2 (1, N=81,799) = 0.004, p = .951$
%	82.1	17.9		

2x4	Level				Row total	Grade 11 Science 1999
	Endorsed	Met	Basic	Not Endorsed		
Mt. Clemens Academy	1	8	10	4	23	
%	4.4	34.8	43.5	17.4		
District	2	38	33	38	111	$\chi^2 (3, N=134) = 3.400, p = .334$
%	1.8	34.2	29.7	34.2		
State	5,819	35,397	23,674	15,920	80,810	$\chi^2 (3, N=80,333) = 2.328, p = .507$
%	7.2	43.8	29.3	19.7		

2x2	Level		Row total	Grade 11 Science 1999
	Endorsed	Not Endorsed		
Mt. Clemens Academy	19	4	23	
%	82.6	17.4		
District	73	38	111	$\chi^2 (1, N=134) = 2.512, p = .113$
%	65.8	34.2		
State	64,890	15,920	80,810	$\chi^2 (1, N=80,333) = 0.078, p = .781$
%	82.1	17.9		

2x4	Level				Row total	Grade 11 Writing 1999
	Endorsed	Met	Basic	Not Endorsed		
Mt. Clemens Academy	0	9	7	2	18	
%	0	50	38.9	11.1		
District	1	55	39	13	108	$\chi^2 (3, N=126) = 0.213, p = .975$
%	1	50.9	36.1	12		
State	5,337	35,855	25,812	11,534	78,538	$\chi^2 (3, N=78,556) = 1.653, p = .647$
%	7.2	43.8	29.3	19.7		

2x2	Level		Row total	Grade 11 Writing 1999
	Endorsed	Not Endorsed		
Mt. Clemens Academy	16	2	18	
%	88.9	11.1		
District	95	13	108	$\chi^2 (1, N=126) = 0.013, p = .911$
%	88	12		
State	67,004	11,534	78,538	$\chi^2 (1, N=78,556) = 0.184, p = .668$
%	82.1	17.9		

2x3	Level			Row total	Grade 7 Reading 1999
	Satisfactory	Moderate	Low		
Mt. Clemens Academy	42	31	43	116	
%	36.21	26.72	37.07		
District	28	25	45	98	$\chi^2 (2, N=214) = 1.988, p = .370$
%	28.57	25.51	45.92		
State	60,573	30,391	23,288	114,252	$\chi^2 (2, N=114,368) = 21.995, p < .001$
%	53.02	26.6	20.38		

2x2	Level		Row total	Grade 7 Reading 1999
	Failed	Passed		
Mt. Clemens Academy	74	42	116	
%	63.79	36.21		
District	70	28	98	$\chi^2 (1, N=214) = 1.407, p = .236$
%	71.43	28.57		
State	53,697	60,573	114,288	$\chi^2 (1, N=114,252) = 13.146, p < .001$
%	46.98	53.02		

2x3	Level			Row total	Grade 7 Math 1999
	Satisfactory	Moderate	Low		
Mt. Clemens Academy	59	36	21	116	
%	50.86	31.03	18.1		
District	35	28	31	94	$\chi^2 (2, N=210) = 6.821, p = .033$
%	37.23	29.79	32.98		
State	72,211	26,265	15,713	114,189	$\chi^2 (2, N=114,305) = 7.645, p < .022$
%	63.24	23	13.77		

2x2	Level		Row total	Grade 7 Math 1999
	Failed	Passed		
Mt. Clemens Academy	57	59	116	
%	49.14	50.86		
District	59	35	94	$\chi^2 (1, N=210) = 3.900, p = .048$
%	62.77	37.23		
State	41,978	72,211	114,189	$\chi^2 (1, N=114,305) = 7.634, p = .006$
%	36.76	63.24		

2x2	Level		Row total	
	Failed	Passed		
Mt. Clemens Academy	74	43	117	
%	63.25	36.75		
District	53	34	87	$\chi^2 (1, N=204) = 0.115, p = .734$
%	60.92	39.08		
State	41,018	71,445	112,463	$\chi^2 (1, N=112580) = 36.153, p < .001$
%	36.47	63.53		

**Grade 8 Writing
1998**

2x3	Level			Row total	
	Satisfactory	Moderate	Low		
Mt. Clemens Academy	9	75	30	114	
%	7.89	65.79	26.32		
District	13	44	29	86	$\chi^2 (2, N=200) = 4.998, p = .082$
%	15.12	51.16	33.72		
State	26,146	64,862	22,486	113,494	$\chi^2 (2, N=113608) = 15.260, p < .001$
%	23.04	57.15	19.81		

**Grade 8 Science
1998**

2x2	Level		Row total	
	Failed	Passed		
Mt. Clemens Academy	105	9	114	
%	92.11	7.89		
District	73	13	86	$\chi^2 (1, N=200) = 2.611, p = .106$
%	84.88	15.12		
State	87,348	26,146	113,494	$\chi^2 (1, N=113,608) = 14.735, p < .001$
%	76.96	23.04		

**Grade 8 Science
1998**

2x4	Level				Row total	
	Level1	Level2	Level3	Level4		
Mt. Clemens Academy	1	15	32	56	104	
%	0.96	14.42	30.77	53.85		
District	1	5	17	63	86	$\chi^2 (3, N=190) = 8.373, p = .039$
%	1.16	5.81	19.77	73.26		
State	5,446	26,087	33,900	47,949	113,382	$\chi^2 (3, N=113,486) = 9.831, p = .020$
%	4.8	23.01	29.9	42.29		

**Grade 8 Social Studies
1999**

2x2	Level		Row total	
	Failed	Passed		
Mt. Clemens Academy	103	1	104	
%	99.04	0.96		
District	85	1	86	$\chi^2 (1, N=190) = 0.018, p = .892$
%	98.84	1.16		
State	107,936	5,446	113,382	$\chi^2 (1, N=113,486) = 3.356, p = .067$
%	95.2	4.8		

**Grade 8 Social Studies
1999**

Mid-Michigan Public School Academy (Lansing, MI)
Chi Square Results from MEAP

2x3	Level			Row Total	Grade 4 Reading 1997
	Satisfactory	Moderate	Low		
Mid-Michigan	25	23	55	103	
%	24.27	22.33	53.4		
State	56,239	32,358	25,436	114,033	
%	22.6	41	17.4		

$\chi^2 (2, N=114,136) = 58.965, p < .001$

2x2	Level		Row total	Grade 4 Reading 1997
	Failed	Passed		
Mid-Michigan	78	25	103	
%	79.63	30.37		
District	854	382	1,236	
%	69.09	30.91		

$\chi^2 (1, N=1,339) = 1.978, p = .160$

$\chi^2 (1, N=114,136) = 25.827, p < .001$

2x3	Level			Row Total	Grade 4 Reading 1998
	Satisfactory	Moderate	Low		
Mid-Michigan	33	30	28	91	
%	36.26	32.97	30.77		
State	66,579	29,524	17,477	113,580	
%	58.62	25.99	15.39		

$\chi^2 (2, N=113,671) = 23.425, p < .001$

2x2	Level		Row total	Grade 4 Reading 1998
	Failed	Passed		
Mid-Michigan	58	33	91	
%	63.74	36.26		
District	624	523	1,147	
%	54.4	45.6		

$\chi^2 (1, N=1,238) = 2.696, p = .085$

$\chi^2 (1, N=113,671) = 18.730, p < .001$

2x3	Level			Row total	Grade 4 Reading 1999
	Satisfactory	Moderate	Low		
Mid-Michigan	23	32	41	96	
%	23.96	33.33	42.71		
State	69,956	30,009	17,748	117,713	
%	59.43	25.49	15.08		

$\chi^2 (2, N=117,809) = 71.129, p < .001$

2x2	Level		Row total	Grade 4 Reading 1999
	Failed	Passed		
Mid-Michigan	73	23	96	
%	76.04	23.96		
District	680	492	1,172	
%	58.02	41.98		

$\chi^2 (1, N=1,268) = 11.948, p = .001$

$\chi^2 (1, N=117,809) = 50.044, p < .001$

2x3	Level			Row Total
	Satisfactory	Moderate	Low	
Mid-Michigan	27	22	57	106
%	36.26	32.97	30.77	
State	69,477	24,563	20,737	114,777
%	60.53	21.4	18.07	

$\chi^2 (2, N=114,883) = 96.134, p < .001$

**Grade 4 Math
1997**

2x2	Level		Row total
	Failed	Passed	
Mid-Michigan	79	27	106
%	74.53	25.47	
District	731	507	1,238
%	59.05	40.95	
State	45,300	69,477	114,777
%	39.47	60.53	

$\chi^2 (1, N=1,344) = 9.773, p = .002$
 $\chi^2 (1, N=114,883) = 54.474, p < .001$

**Grade 4 Math
1997**

2x3	Level			Row Total
	Satisfactory	Moderate	Low	
Mid-Michigan	38	31	24	93
%	40.86	33.33	25.81	
State	84,318	20,126	9,314	113,758
%	74.12	17.69	8.19	

$\chi^2 (2, N=113,851) = 61.884, p < .001$

**Grade 4 Math
1998**

2x2	Level		Row total
	Failed	Passed	
Mid-Michigan	55	38	93
%	59.14	40.86	
District	447	706	1,153
%	38.77	61.23	
State	29,440	84,318	113,758
%	25.88	74.12	

$\chi^2 (1, N=1,246) = 14.845, p < .001$
 $\chi^2 (1, N=113,851) = 53.554, p < .001$

**Grade 4 Math
1998**

2x3	Level			Row total
	Satisfactory	Moderate	Low	
Mid-Michigan	32	20	44	96
%	33.33	20.83	45.83	
State	84,444	20,834	12,445	117,723
%	71.73	17.7	10.57	

$\chi^2 (2, N=117,819) = 132.774, p < .001$

**Grade 4 Math
1999**

2x2	Level		Row total
	Failed	Passed	
Mid-Michigan	64	32	96
%	66.67	33.33	
District	528	642	1,170
%	45.13	54.87	
State	33,279	84,444	117,723
%	28.27	71.73	

$\chi^2 (1, N=1,266) = 16.533, p < .001$
 $\chi^2 (1, N=117,819) = 69.698, p < .001$

**Grade 4 Math
1999**

2x4	Level				Row total	Grade 5 Social Studies 1999
	Level1	Level2	Level3	Level4		
Mid-Michigan	2	6	12	85	105	
%	1.9	5.71	11.43	80.95		
District	37	124	285	644	1,090	
%	3.39	11.38	26.15	59.08		
State	189	19,133	30,835	60,596	110,753	
%	0.17	17.28	27.84	54.71		

$\chi^2 (3, N=1,195) = 19.330, p < .001$

$\chi^2 (3, N=110,858) = 49.780, p < .001$

2x2	Level		Row total	Grade 5 Social Studies 1999
	Failed	Passed		
Mid-Michigan	103	2	105	
%	66.67	33.33		
District	1,053	37	1,090	
%	96.61	3.39		
State	110,564	189	110,753	
%	99.83	0.17		

$\chi^2 (1, N=1,195) = 0.673, p = .412$

$\chi^2 (1, N=110,858) = 18.341, p < .001$

2x2	Level		Row total	Grade 5 Writing 1997
	Failed	Passed		
Mid-Michigan	38	48	86	
%	44.19	55.81		
District	308	779	1,087	
%	45.13	54.87		
State	30,089	83,085	113,174	
%	26.59	73.41		

$\chi^2 (1, N=1,173) = 9.629, p = .002$

$\chi^2 (1, N=113,260) = 13.633, p < .001$

2x2	Level		Row total	Grade 5 Writing 1998
	Failed	Passed		
Mid-Michigan	51	44	95	
%	53.68	46.32		
District	529	634	1,163	
%	45.49	54.51		
State	40,377	72,774	113,151	
%	35.68	64.32		

$\chi^2 (1, N=1,258) = 2.376, p = .123$

$\chi^2 (1, N=113,246) = 13.398, p < .001$

2x2	Level		Row total	Grade 5 Writing 1999
	Failed	Passed		
Mid-Michigan	76	30	106	
%	71.7	28.3		
District	411	646	1,057	
%	38.88	61.12		
State	50,949	61,832	112,781	
%	45.18	54.82		

$\chi^2 (1, N=1,163) = 42.620, p < .001$

$\chi^2 (1, N=112,887) = 30.076, p < .001$

2x3	Level			Row total	
	Satisfactory	Moderate	Low		
Mid-Michigan	10	48	30	88	
%	11.36	54.55	34.09		
District	236	652	225	1,113	$\chi^2 (2, N=1,201) = 11.478, p = .003$
%	32	32	36		
State	41,831	61,781	10,145	113,757	$\chi^2 (2, N=113,845) = 77.790, p < .001$
%	36.77	54.31	8.92		

**Grade 5 Science
1997**

2x2	Level		Row total	
	Failed	Passed		
Mid-Michigan	78	10	88	
%	88.64	11.36		
District	887	236	1,113	$\chi^2 (1, N=1,201) = 4.848, p = .028$
%	78.8	21.2		
State	71,926	41,831	113,757	$\chi^2 (1, N=113,845) = 24.422, p < .001$
%	63.23	36.77		

**Grade 5 Science
1997**

2x3	Level			Row total	
	Satisfactory	Moderate	Low		
Mid-Michigan	15	57	22	94	
%	15.96	60.64	23.4		
District	351	709	117	1,117	$\chi^2 (2, N=1,271) = 20.241, p < .001$
%	29.82	60.24	9.94		
State	46,018	60,108	7,804	113,930	$\chi^2 (2, N=114,024) = 52.497, p < .001$
%	40.39	52.76	6.85		

**Grade 5 Science
1998**

2x2	Level		Row total	
	Failed	Passed		
Mid-Michigan	79	15	94	
%	84.04	15.96		
District	826	351	1,117	$\chi^2 (1, N=1,271) = 8.160, p = .004$
%	70.18	29.82		
State	67,912	46,018	113,930	$\chi^2 (1, N=114,024) = 23.293, p < .001$
%	59.61	40.39		

**Grade 5 Science
1998**

2x3	Level			Row total	
	Satisfactory	Moderate	Low		
Mid-Michigan	12	53	42	107	
%	11.21	49.53	39.25		
District	288	622	162	1,072	$\chi^2 (2, N=1,179) = 43.357, p < .001$
%	26.87	58.02	15.11		
State	42,500	59,463	11,294	113,257	$\chi^2 (2, N=113,364) = 111.564, p < .001$
%	37.53	52.5	9.97		

**Grade 5 Science
1999**

2x2	Level		Row total	
	Failed	Passed		
Mid-Michigan	95	12	107	
%	88.79	11.21		
District	784	288	1,072	$\chi^2 (1, N=1,179) = 12.562, p < .001$
%	73.13	26.87		
State	70,757	42,500	113,257	$\chi^2 (1, N=113,364) = 31.573, p < .001$
%	62.47	37.53		

**Grade 5 Science
1999**

2x3	Level			Row total	
	Satisfactory	Moderate	Low		
Mid-Michigan	26	25	35	86	
%	30.23	29.07	40.7		
State	56,596	32,347	27,113	116,056	$\chi^2 (2, N=116,142) = 17.148, p < .001$
%	48.77	27.87	23.36		

**Grade 7 Reading
1998**

2x2	Level		Row total	
	Failed	Passed		
Mid-Michigan	60	26	86	
%	69.77	30.23		
District	657	334	991	$\chi^2 (1, N=1,077) = 0.428, p = .513$
%	66.3	33.7		
State	59,460	56,596	116,056	$\chi^2 (1, N=116,142) = 11.815, p < .001$
%	51.23	48.77		

**Grade 7 Reading
1998**

2x3	Level			Row total	
	Satisfactory	Moderate	Low		
Mid-Michigan	17	25	38	80	
%	21.25	31.25	47.5		
State	60,598	30,397	23,293	114,288	$\chi^2 (2, N=114,368) = 44.698, p < .001$
%	53.02	26.6	20.38		

**Grade 7 Reading
1999**

2x2	Level		Row total	
	Failed	Passed		
Mid-Michigan	63	17	80	
%	78.75	21.25		
District	546	349	895	$\chi^2 (1, N=975) = 9.861, p = .002$
%	61.01	38.99		
State	53,690	60,598	114,288	$\chi^2 (1, N=114,368) = 32.397, p < .001$
%	46.98	53.02		

**Grade 7 Reading
1999**

2x3	Level			Row total	
	Satisfactory	Moderate	Low		
Mid-Michigan	18	25	41	84	
%	21.43	29.76	48.81		
State	71,345	27,985	16,812	116,142	$\chi^2 (2, N=116,226) = 91.233, p < .001$
%	61.43	24.1	14.48		

**Grade 7 Math
1998**

2x2	Level		Row total	
	Failed	Passed		
Mid-Michigan	66	18	84	
%	78.57	21.43		
District	571	434	1,005	$\chi^2 (1, N=1,089) = 15.112, p < .001$
%	56.82	43.18		
State	44,797	71,345	116,142	$\chi^2 (1, N=116,226) = 56.669, p < .001$
%	38.57	61.43		

**Grade 7 Math
1998**

2x3	Level			Row total	
	Satisfactory	Moderate	Low		
Mid-Michigan	18	19	47	84	
%	21.43	22.62	55.95		
State	72,252	26,282	15,734	114,268	$\chi^2 (2, N=114,352) = 131.442, p < .001$
%	63.23	23	13.77		

**Grade 7 Math
1999**

2x2	Level		Row total	
	Failed	Passed		
Mid-Michigan	66	18	84	
%	78.57	21.43		
District	534	347	881	$\chi^2 (1, N=965) = 10.517, p < .001$
%	60.61	39.39		
State	42,016	72,252	114,268	$\chi^2 (1, N=114,352) = 63.064, p < .001$
%	36.77	63.23		

**Grade 7 Math
1999**

2x2	Level		Row total	
	Failed	Passed		
Mid-Michigan	28	15	43	
%	65.12	34.88		
District	400	553	953	$\chi^2 (1, N=996) = 8.993, p = .003$
%	41.97	58.03		
State	33,822	75,338	109,160	$\chi^2 (1, N=109,203) = 23.412, p < .001$
%	30.98	69.02		

**Grade 8 Writing
1998**

2x2	Level		Row total	
	Failed	Passed		
Mid-Michigan	44	16	60	
%	73.33	26.67		
District	331	580	911	$\chi^2 (1, N=971) = 32.509, p < .001$
%	36.33	63.67		
State	41,048	71,472	112,520	$\chi^2 (1, N=112,580) = 35.139, p < .001$
%	36.48	63.52		

**Grade 8 Writing
1999**

2x3	Level			Row total	
	Satisfactory	Moderate	Low		
Mid-Michigan	1	20	30	51	
%	1.96	39.22	58.82		
District	121	606	272	999	$\chi^2 (2, N=1,050) = 24.889, p < .001$
%	12.11	60.66	27.23		
State	24,407	66,119	20,220	110,746	$\chi^2 (2, N=110,797) = 58.810, p < .001$
%	22.04	59.7	18.26		

**Grade 8 Science
1998**

2x2	Level		Row total	
	Failed	Passed		
Mid-Michigan	50	1	51	
%	98.04	1.96		
District	878	121	999	$\chi^2 (1, N=1,050) = 4.869, p = .027$
%	87.89	12.11		
State	86,339	24,407	110,746	$\chi^2 (1, N=110,797) = 11.964, p < .001$
%	77.96	22.04		

**Grade 8 Science
1998**

2x3	Level			Row total	
	Satisfactory	Moderate	Low		
Mid-Michigan	1	16	43	60	
%	1.67	26.67	71.67		
District	139	521	267	927	$\chi^2 (2, N=987) = 49.058, p < .001$
%	14.99	56.2	28.8		
State	26,154	64,916	22,473	113,543	$\chi^2 (2, N=113,603) = 103.075, p < .001$
%	23.03	57.17	19.79		

**Grade 8 Science
1999**

2x2	Level		Row total	
	Failed	Passed		
Mid-Michigan	59	1	60	
%	98.33	1.67		
District	788	139	927	$\chi^2 (1, N=987) = 8.224, p = .004$
%	85.01	14.99		
State	87,389	26,154	113,543	$\chi^2 (1, N=113,603) = 15.450, p < .001$
%	76.97	23.03		

**Grade 8 Science
1999**

2x4	Level				Row total	
	Level1	Level2	Level3	Level4		
Mid-Michigan	1	0	5	57	63	
%	1.59	0	7.94	90.48		
District	22	145	261	502	930	$\chi^2 (3, N=993) = 32.862, p < .001$
%	2.37	15.59	28.06	53.98		
State	5,446	26,120	33,927	47,946	113,421	$\chi^2 (3, N=113,484) = 60.612, p < .001$
%	4.8	23.01	29.91	42.27		

**Grade 8 Social Studies
1999**

2x2	Level		Row total	
	Failed	Passed		
Mid-Michigan	62	1	63	
%	98.41	1.59		
District	908	22	930	$\chi^2 (1, N=993) = 0.158, p = .691$
%	97.63	2.37		
State	107,975	5,446	113,421	$\chi^2 (1, N=113,484) = 1.424, p = .233$
%	95.2	4.8		

**Grade 8 Social Studies
1999**