

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. Mislavy (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 Mislavy 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” (Mislavy, 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” (Mislavy, 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 Mисlevy 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 Mисlevy 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.

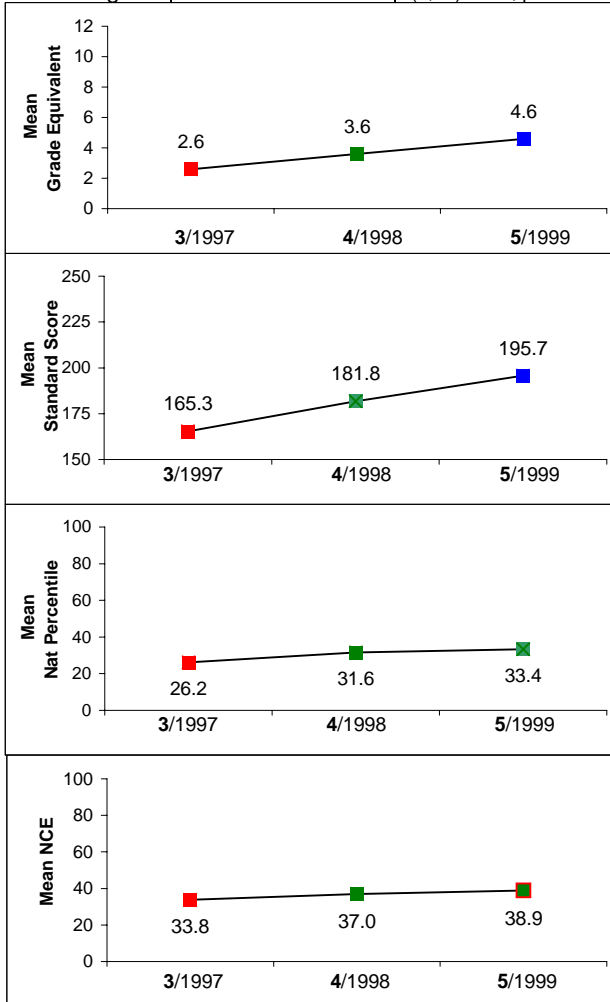
In terms of GE scores, the students are expected to progress at least one grade level per year. As seen in the results, over the two years the students progressed 2.3 grade levels in math, 2.0 grades in language, and 2.2 years on average in reading. In each subject analysis there was a statistically 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)

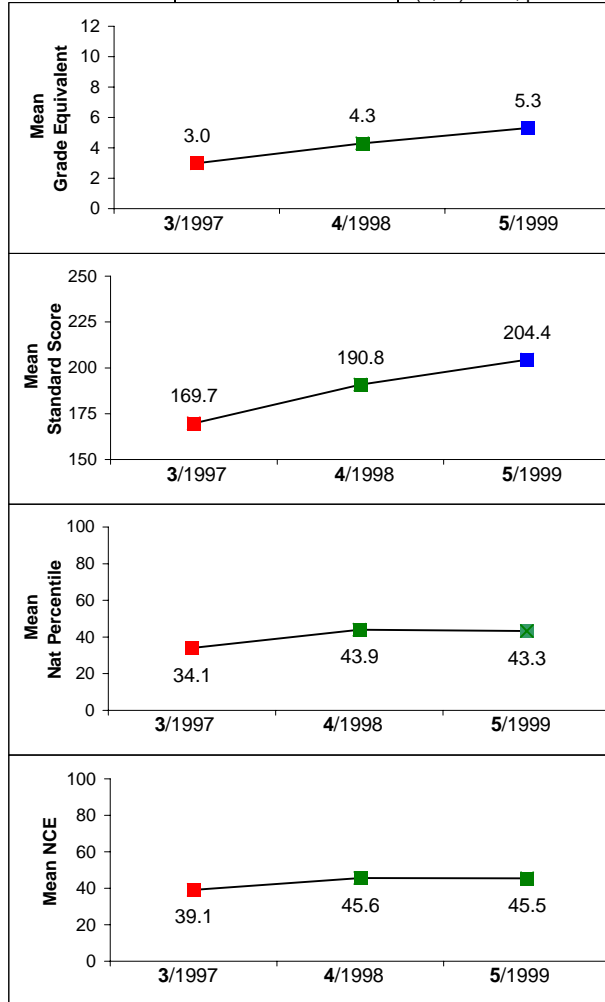
LANGUAGE

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



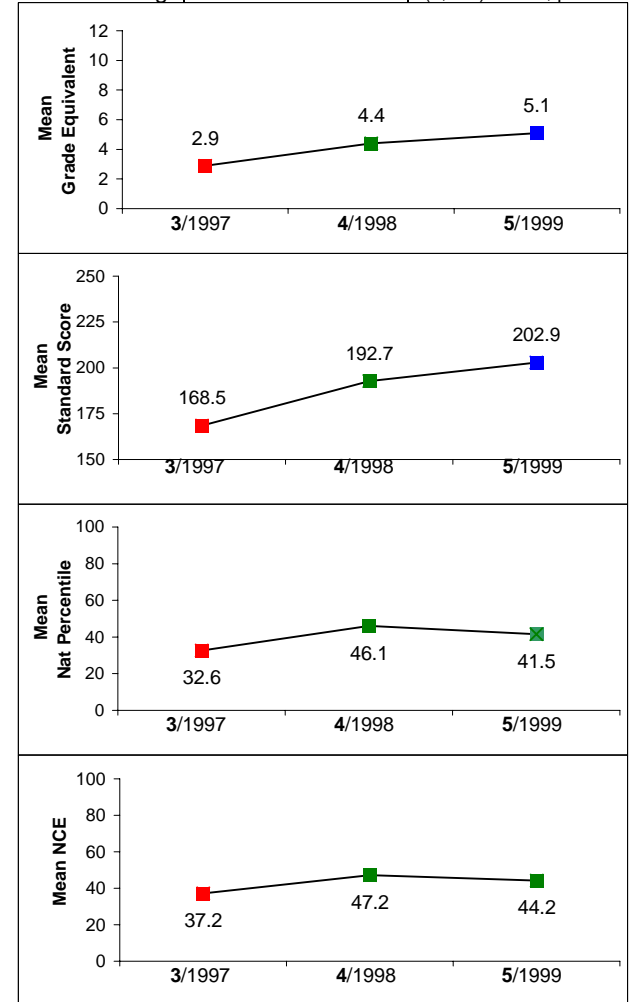
MATH

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.

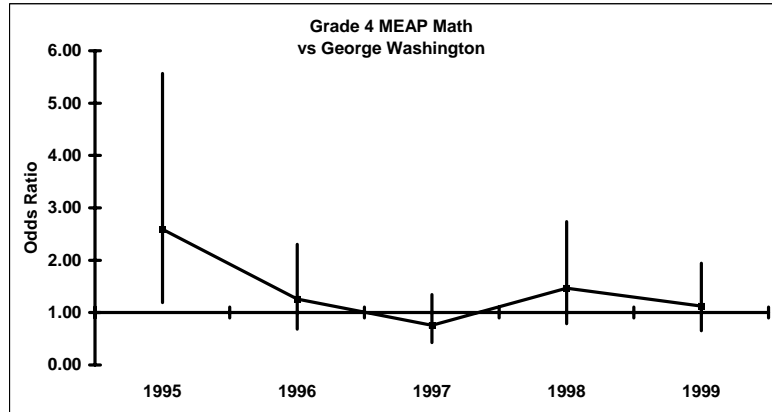
Exhibit 9:2 Grade 4 Odds Ratio Results When Compared with the District and State

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



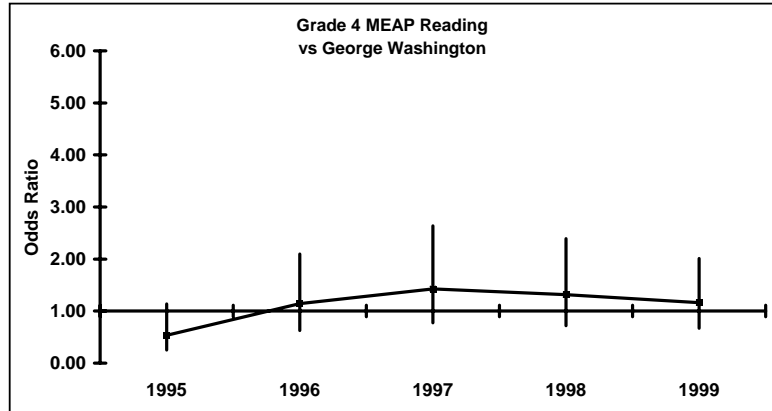
Note: George Washington Elementary is the only other Elementary School in the district with students at Grade 4 and 5.

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

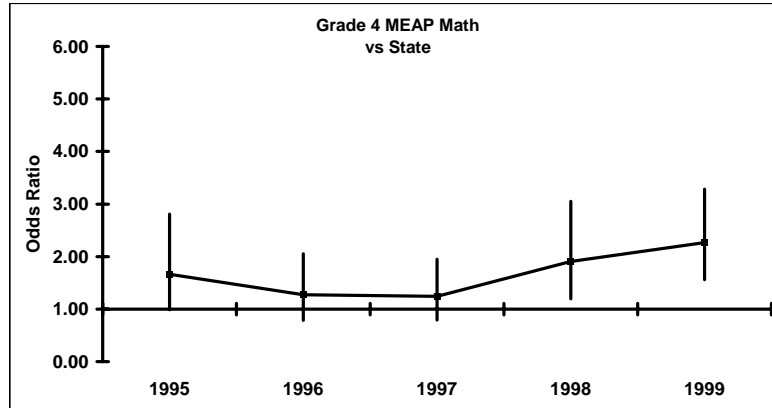


Grade 4 MEAP Math vs State

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

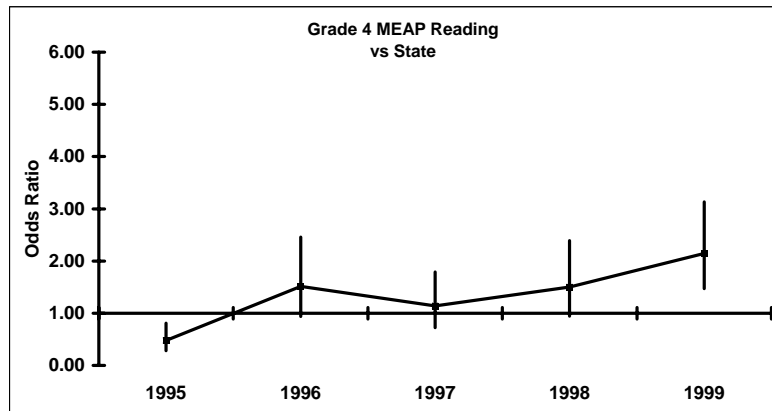
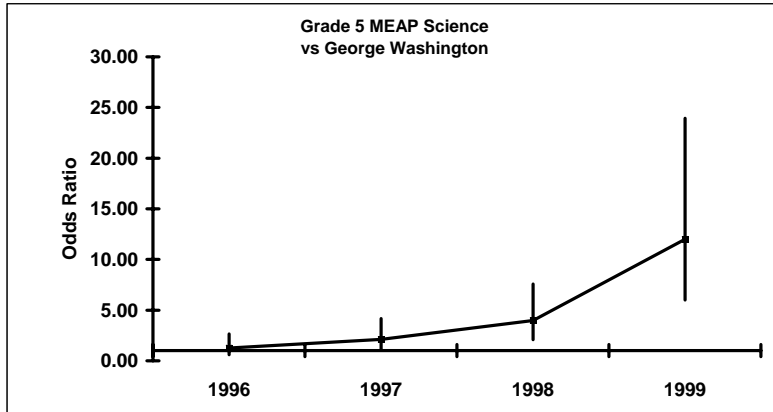


Exhibit 9:3 Grade 5 Odds Ratio Results When Compared with the District and State

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

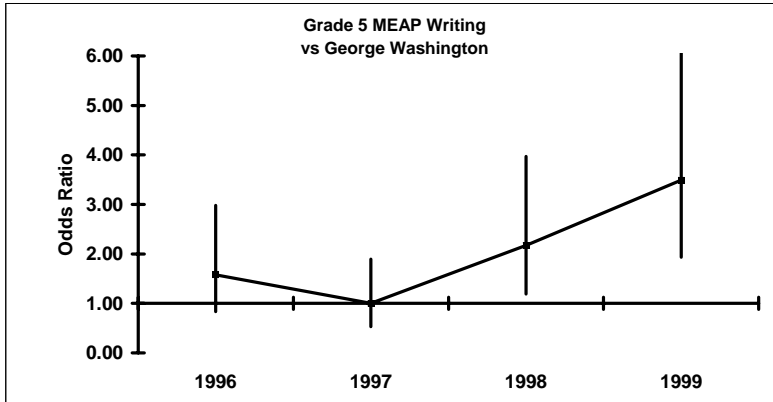


Note: George Washington Elementary is the only other Elementary School in the district with students at Grade 4 and 5.

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

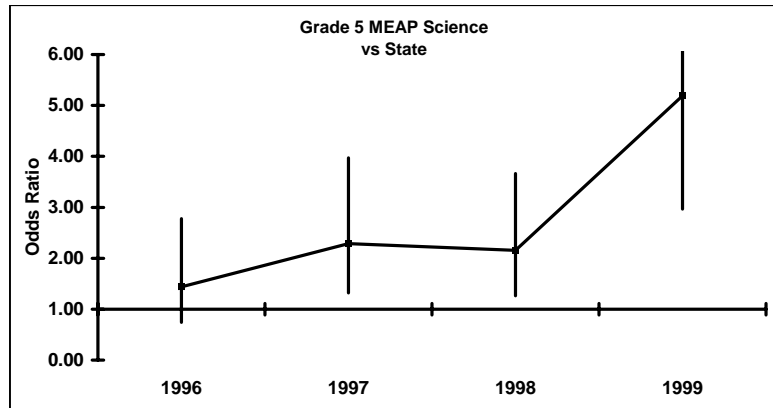
Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (3, N=771) = 8.434, p = .038



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

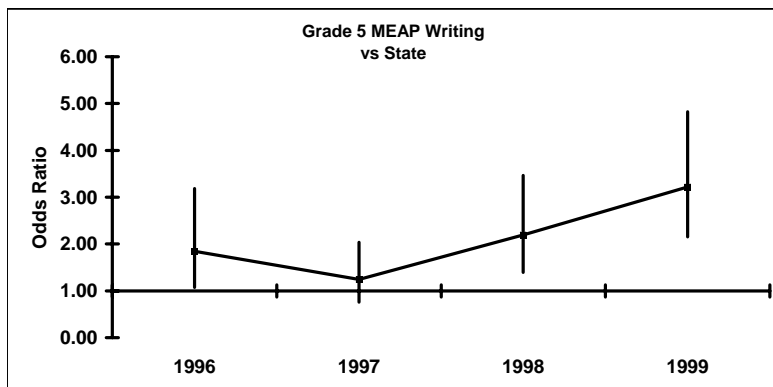
Breslow-Day for Homogeneity of Odd Ratio
Chi-Sq (3, N=458,782) = 8.775, p = .032



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

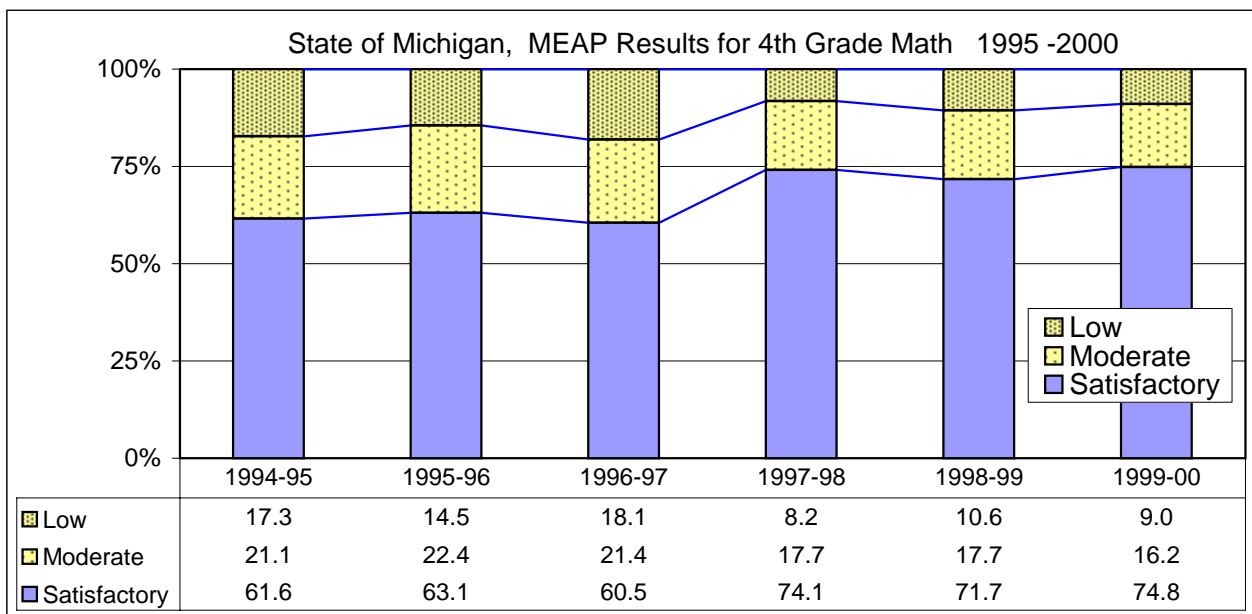
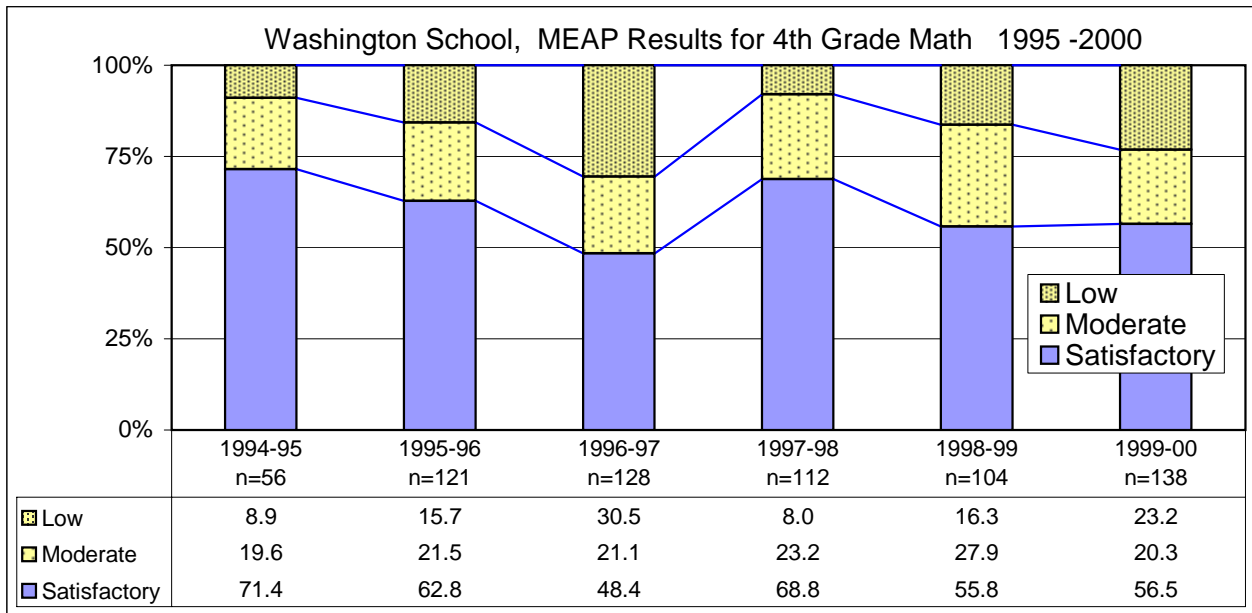
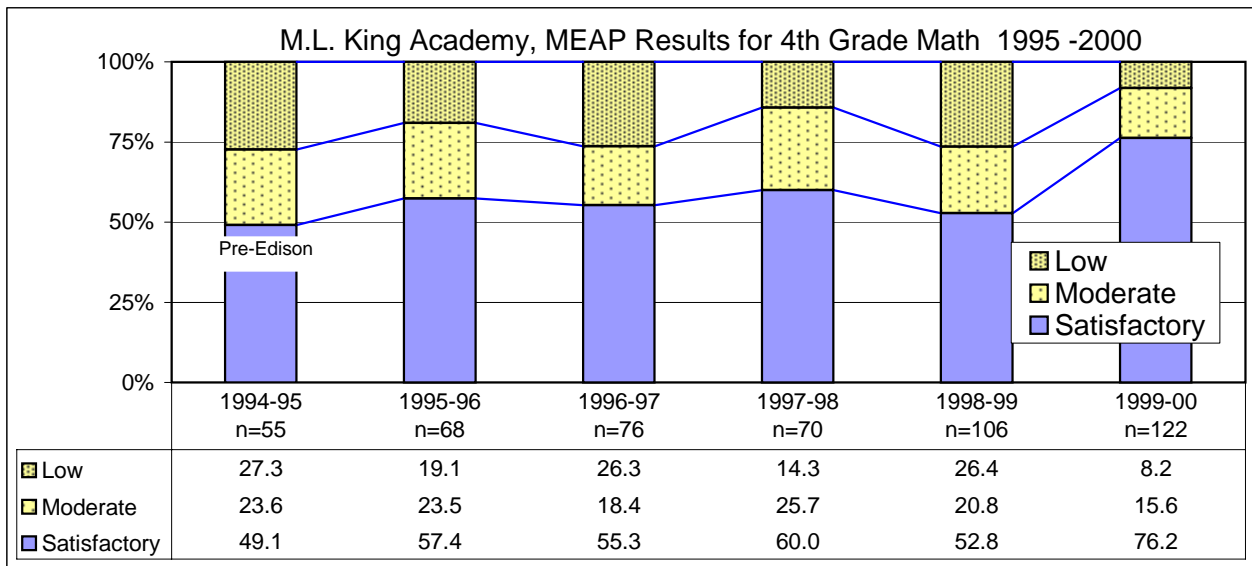


Exhibit 9:5 Performance on Grade 4 Reading for Edison, District, and State

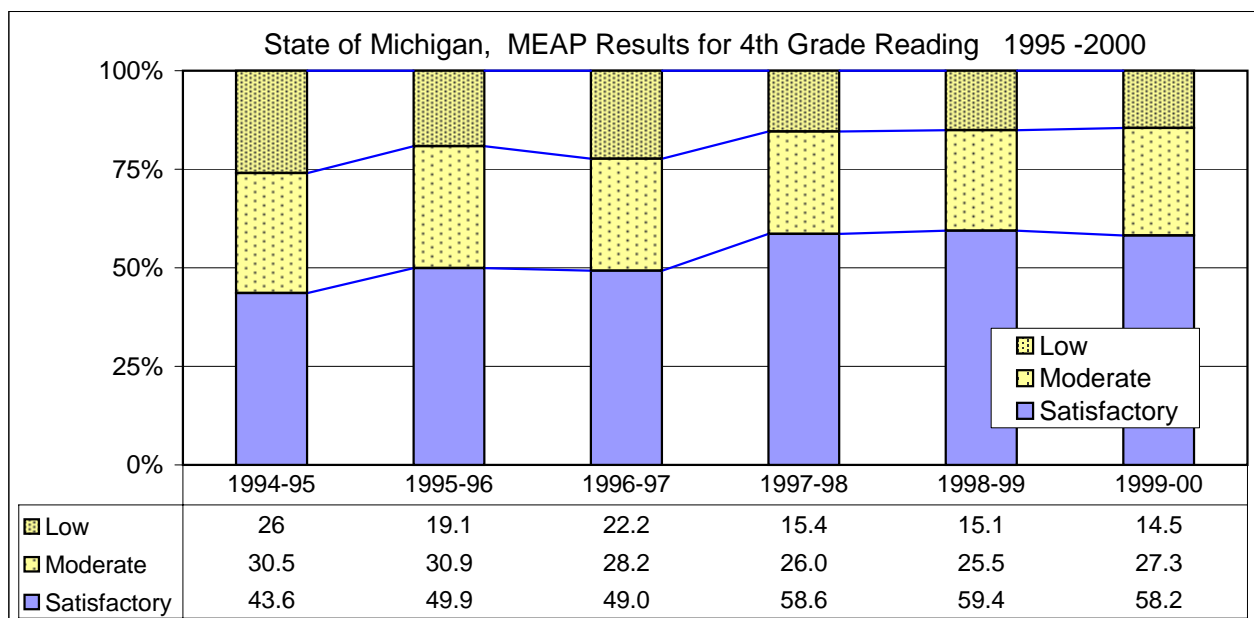
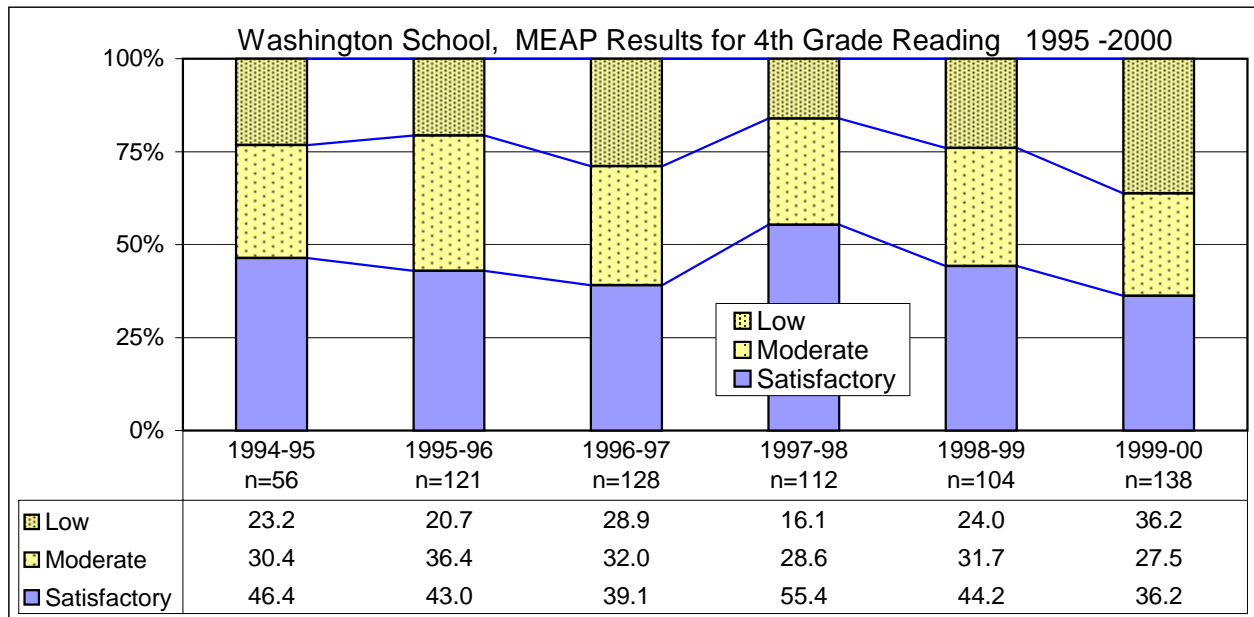
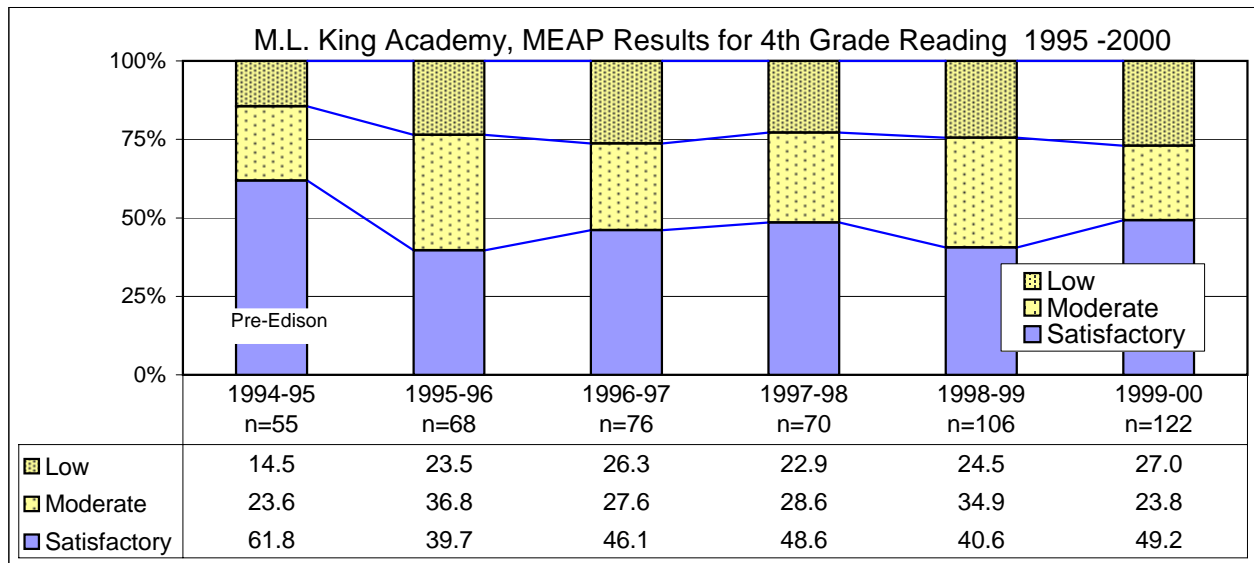


Exhibit 9:6 Performance on Grade 5 Science for Edison, District, and State

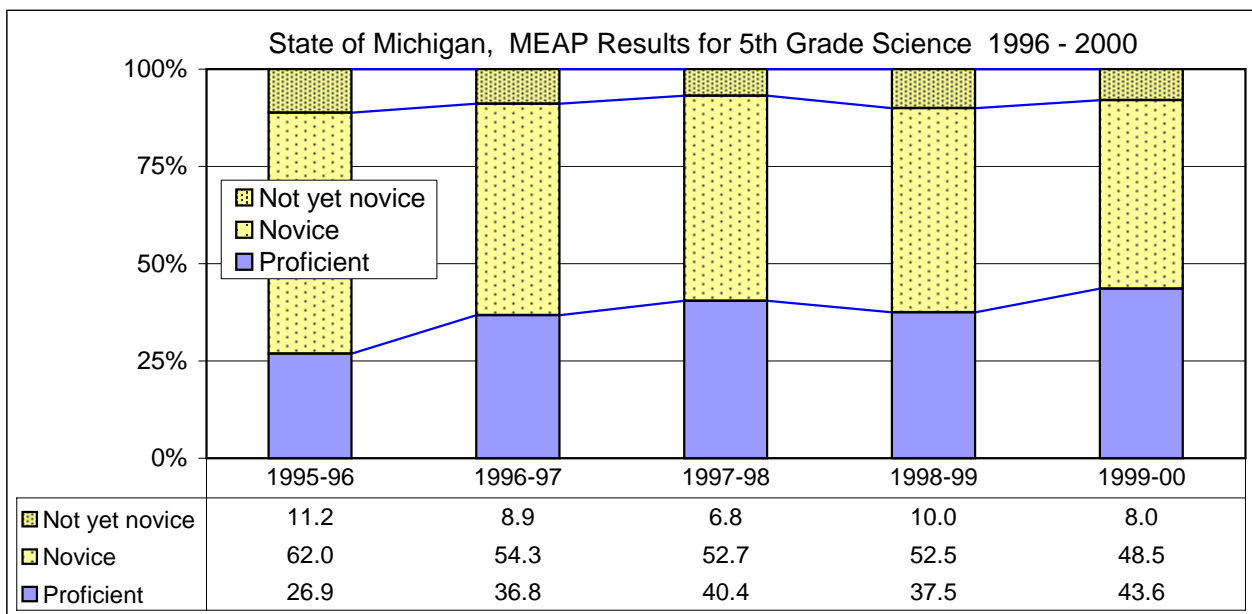
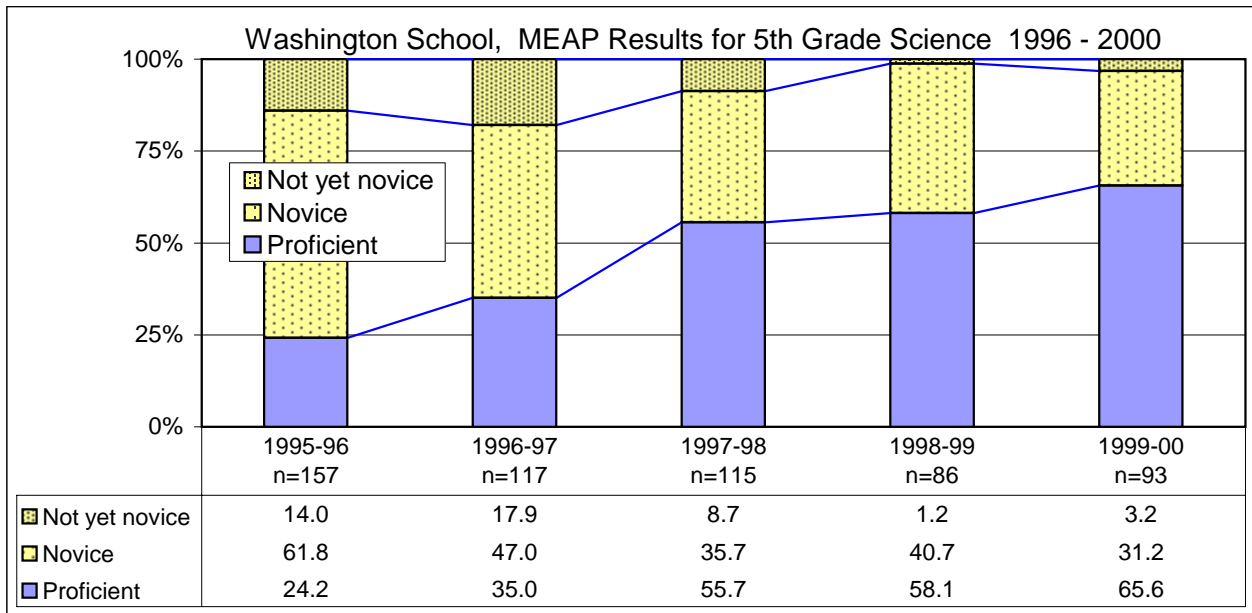
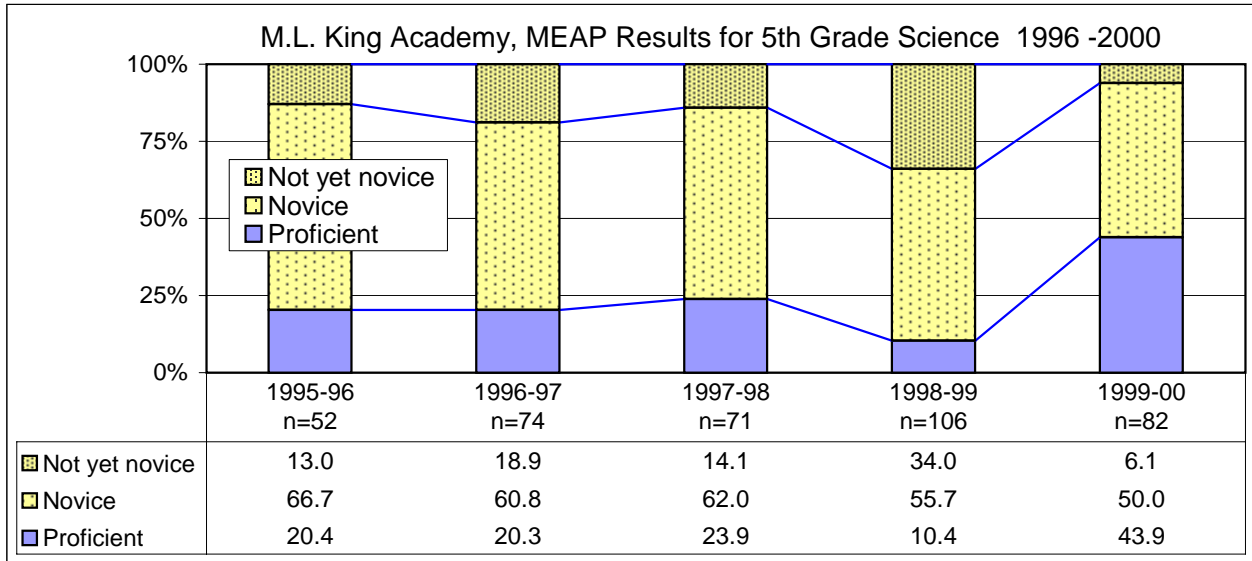


Exhibit 9:7 Performance on Grade 5 Writing for Edison, District, and State

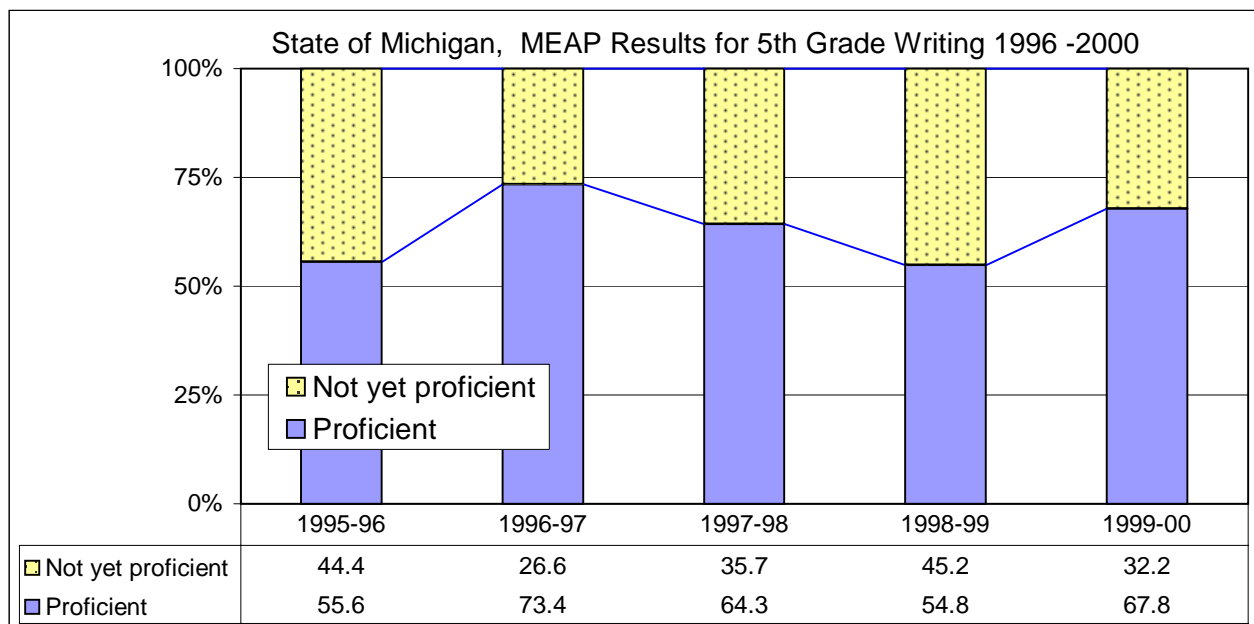
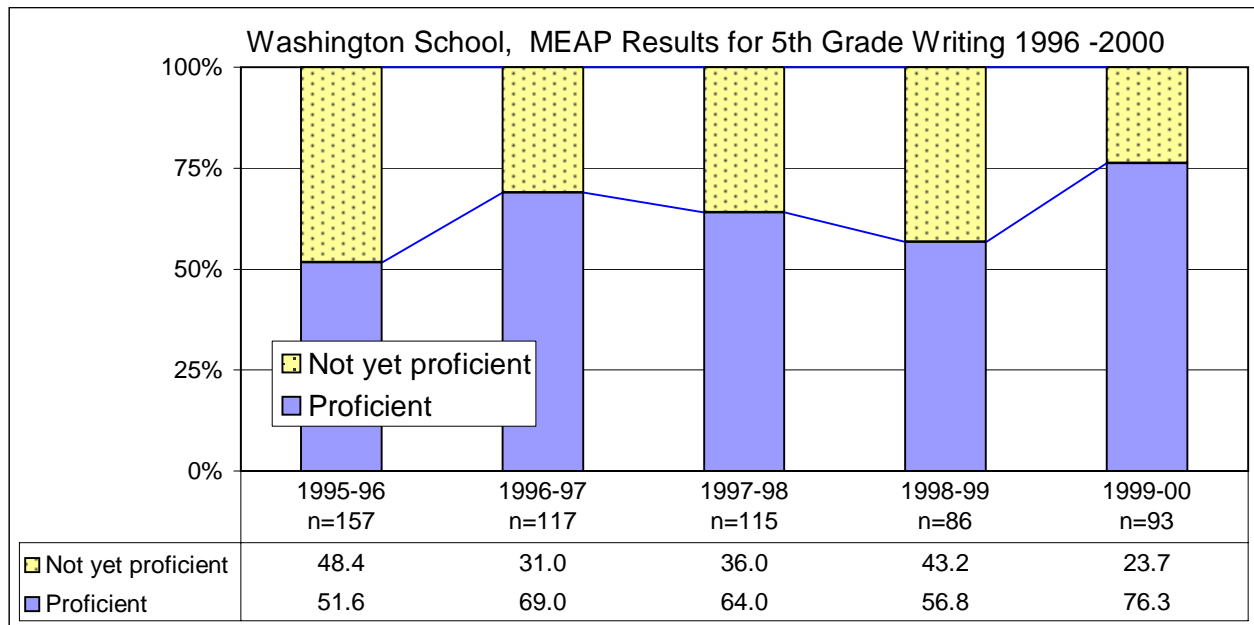
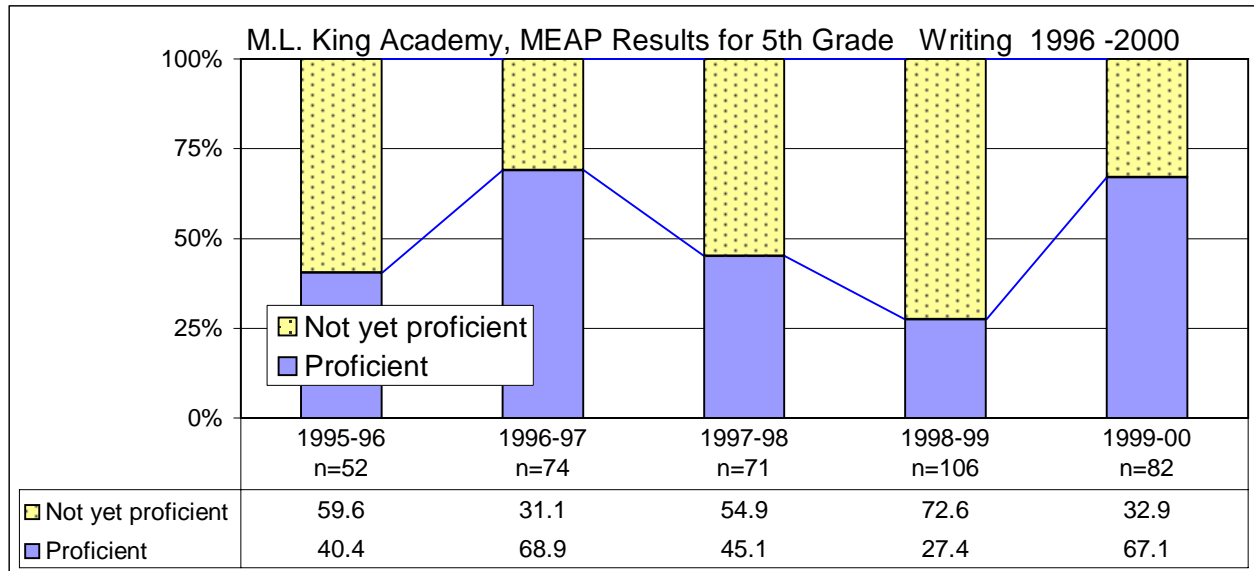
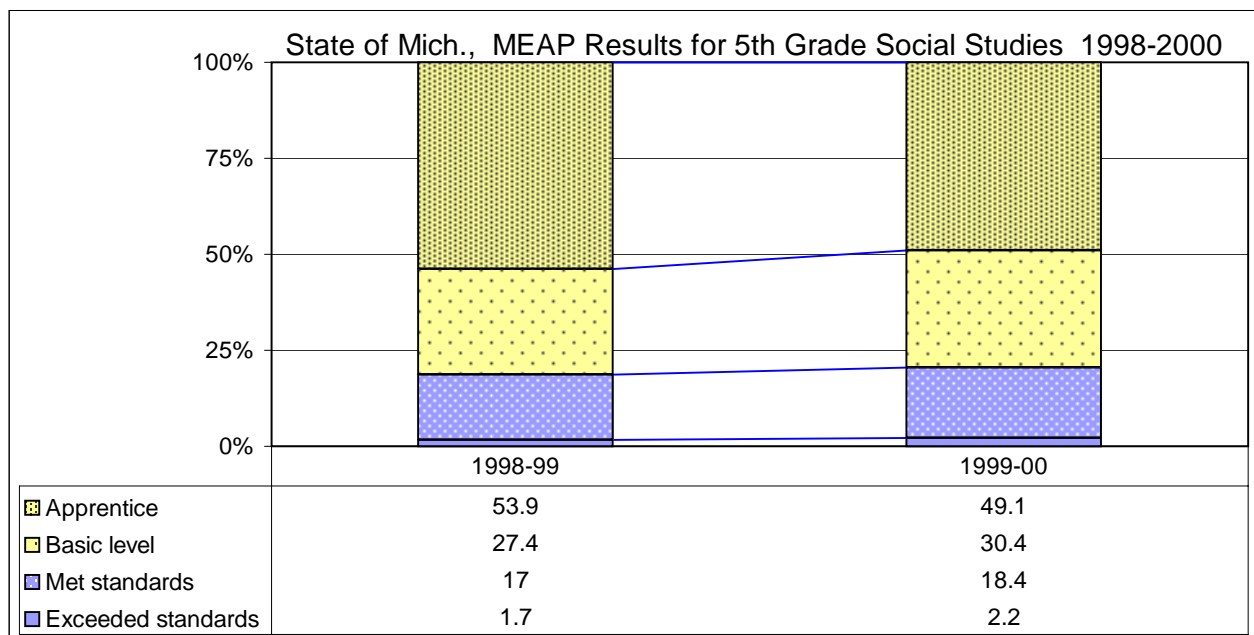
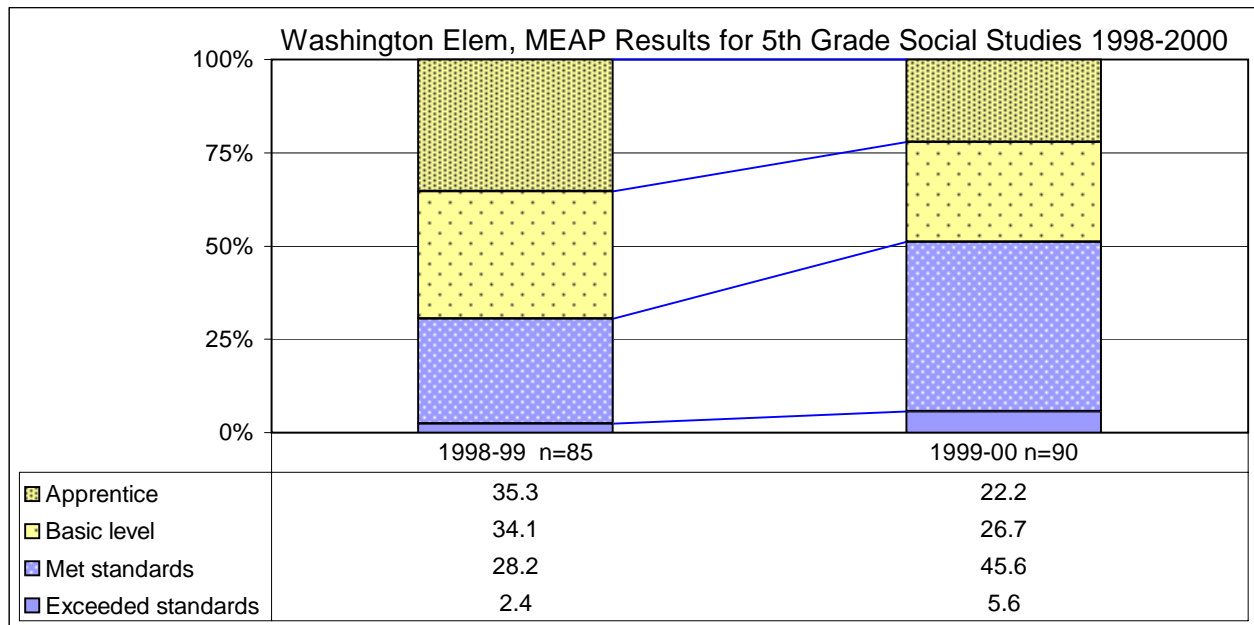
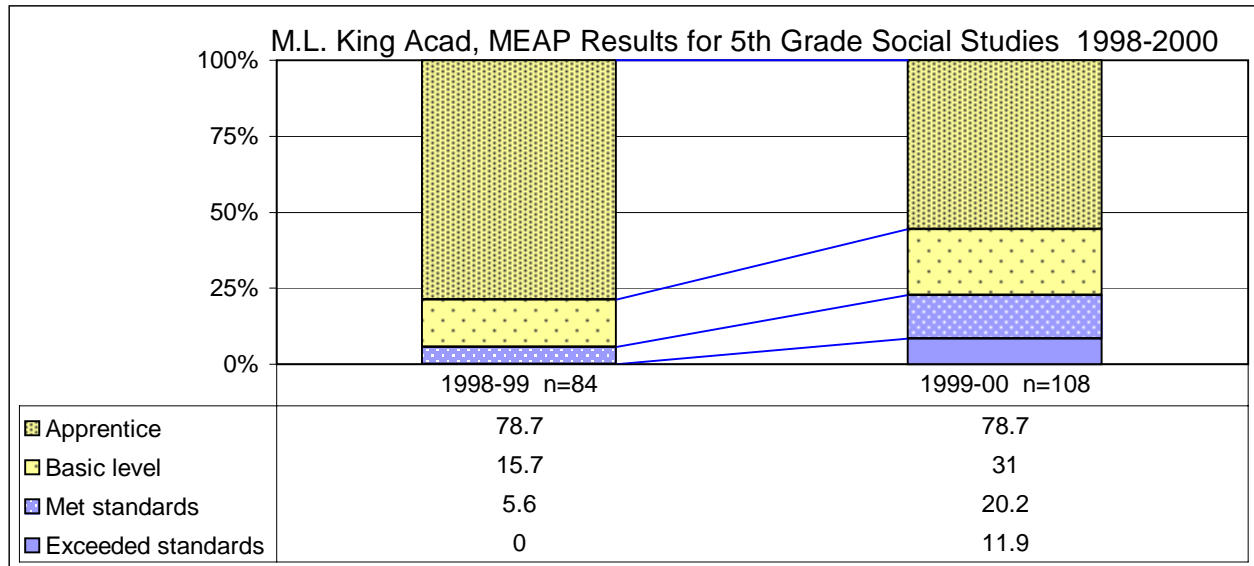


Exhibit 9:8 Performance on Grade 5 Social Studies for Edison, District, and State



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.