CHAPTER 6

HOW SUCCESSFUL HAS EDUCATION REFORM BEEN IN MASSACHUSETTS? WAS OPPORTUNITY MADE MORE EQUAL?

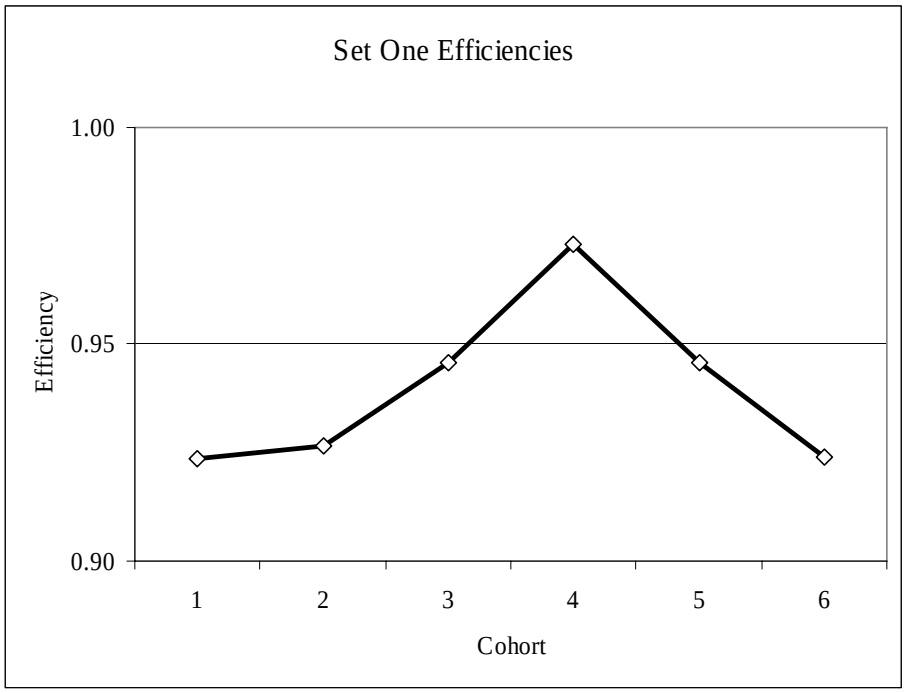
This chapter documents two sets of experiments. The first undertaken to test the proposition that "education standards have been raised and educational outcomes have improved as a consequence of MERA" – see Section 1 to 5 – and the second to test the proposition that opportunity was made more equal – see Sections 6 to 8.

A one-off improvement in test scores from a number of periods will be revealed from a CCR model as an increase in efficiency followed by a decrease, if the model uses earlier test scores from each wave or cohort of children as the inputs and later test scores from the same wave or cohort of children as outputs.

In order to be sure that the CCR model would behave in this way, two sets of synthetic data were generated to conform to two assumptions, the first of a one-off improvement and the second of a sustained improvement. The results from the synthetic data confirmed the proposition: – see Figure 6.01[24].

[24] Cohorts were constructed to replicate Grade 4 inputs to Grade 8 outputs as shown in Section 1 of this Chapter. Cohorts 1 and 2 assumed no improvement in education. Cohorts 3 and 4 assumed that the Grade 8 scores reflect an improvement. Cohorts 5 and 6 assume that Grade 4 scores also, now, reflect an improvement.

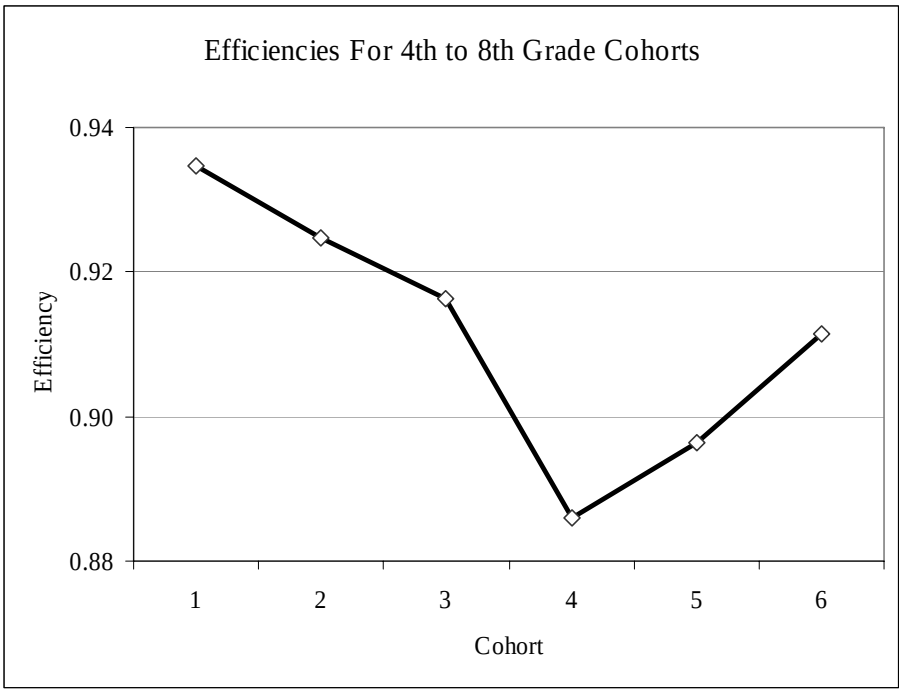
Figure 6.01 Average Efficiency Over Time From Test Data Set One.



Section 3 describes the actual data used. The results from actual data are given in Section 4. A decrease in efficiency after MERA is followed by an increase – see Figure 6.02[25] – which implies a one off deterioration in education. Section 5 looks to the percentage of all students in the grade who actually took the tests as a possible explanation for the shape of the graph

[25] Cohort 1 uses 1988 Grade 4 scores and 1992 Grade 8 scores. Cohort 2 uses 1990 Grade 4 and 1994 Grade 8 scores. Each subsequent Cohort uses scores from two years later than the Cohort before. If MERA prompted an improvement over the period 1994 to 1998 then Cohorts 3 and 4 would show increasing efficiency over Cohorts 1 and 2 with efficiency declining in Cohorts 5 and 6 as shown in Figure 6.01.

Figure 6.02 Average Efficiency Over Time 4th to 8th Grade Cohorts.



Sections 6, 7 and 8 look at whether educational opportunity was made more equal by MERA using per-pupil expenditures as input and test scores as outputs. Section 9 concludes that there is little in this analysis to suggest an improvement in educational outcomes and little to suggest otherwise. It also concludes that educational opportunity was made more equal.

6.1 Proposition Behind The First Experiments

Assume that educational test scores are a measure of education standards and that you just graduated from high school. Your younger sibling is four years younger and will enter the 9th grade next semester. Your other sibling is four years older. You all attended the same K-12 school district.

You scored 60% in your 4^{th} grade tests. As you entered the 5^{th} grade, the school district implemented an improvement plan and by the time you got to the 8^{th} grade you scored 70% in the 8^{th} grade tests.

Your younger sibling experienced the improvement on entering the first grade and scored 70% on the 4^{th} grade tests and also 70% on the 8^{th} grade tests. Your older sibling experienced no improvement until entering the ninth grade and had scored 60% on the 4^{th} grade tests and 60% on the 8^{th} grade tests.

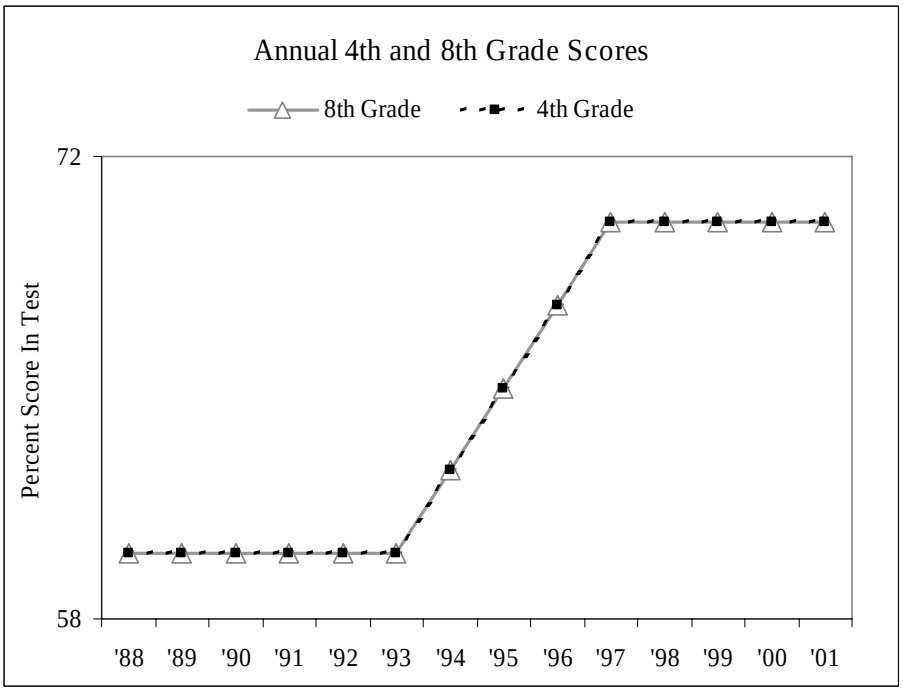
Treating your 4^{th} grade score as an input and your 8^{th} grade score as an output and applying a pricing to the input score that would make the ratio of your priced scores equal to one implies that the price, p, is equal to 7/6. ($70/60*p = 1 \Rightarrow p = 7/6$).

Applying the same pricing to your younger sibling's scores of 70 and 70 implies that the efficiency of your younger sibling was 85.7 percent. ($70/70*p = 70/70*(7/6) = 70/81.666 = 0.857$). Using your older sibling's scores of 60 and 60 estimates his efficiency as 85.7 percent too. ($60/60*p = 60/60*(7/6) = 60/70 = 0.857$.)

Due to the timing of the improvement in your school district the measurement of your efficiency in turning 4^{th} grade scores into 8^{th} grade scores shows you to have been more efficient than either your younger or older sibling.

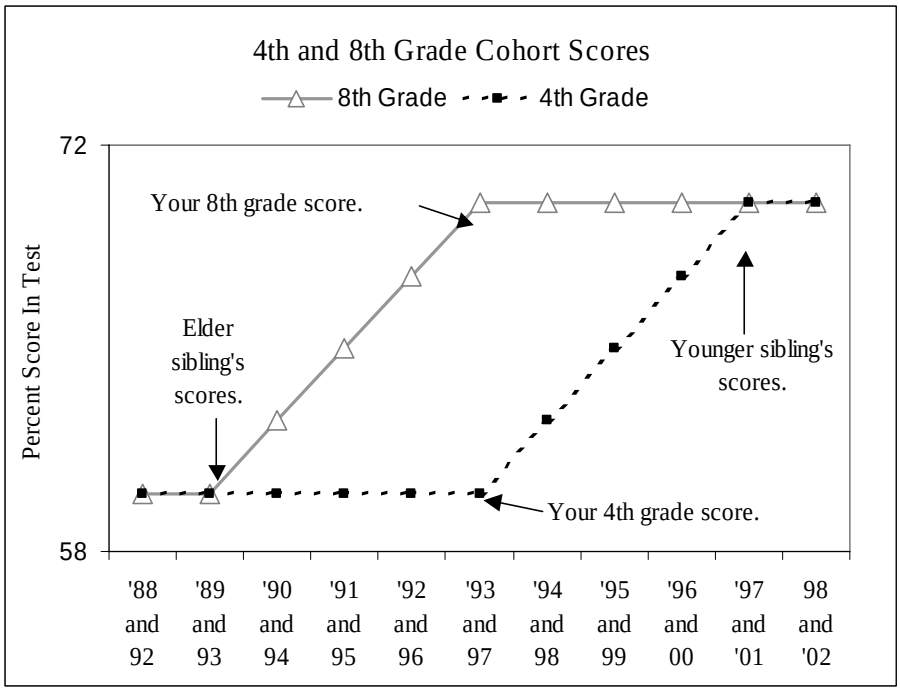
Figure 6.03 shows your test scores improving over three periods in a one off improvement starting in 1993 and ending in 1997.

Figure 6.03 – 4th and 8th Grade Test Scores Improving at One Time.



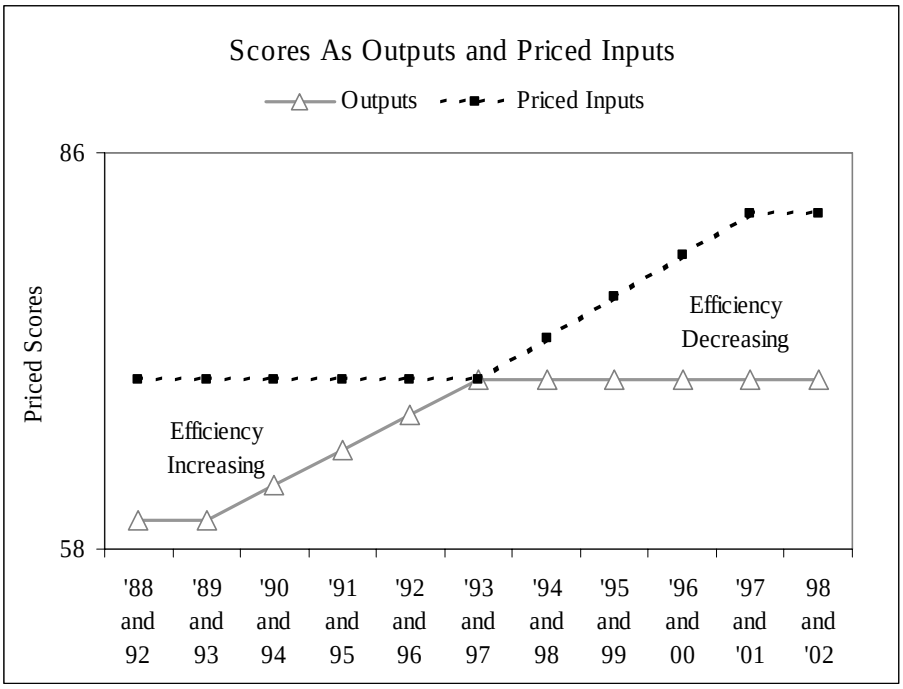
Imagine that your older sibling took tests in 1989 and 1993, you took tests in 1993 and 1997 and your younger sibling took tests in 1997 and 2001. Measuring your efficiency with your 4th grade score as an input and your 8th grade score as an output is easier if the two scores are lined up vertically. Figure 6.04 shows your 1997 8th grade score lined up vertically with your 1993 4th grade score.

Figure 6.04 – 4th and 8th Grade Test Scores of Cohorts.



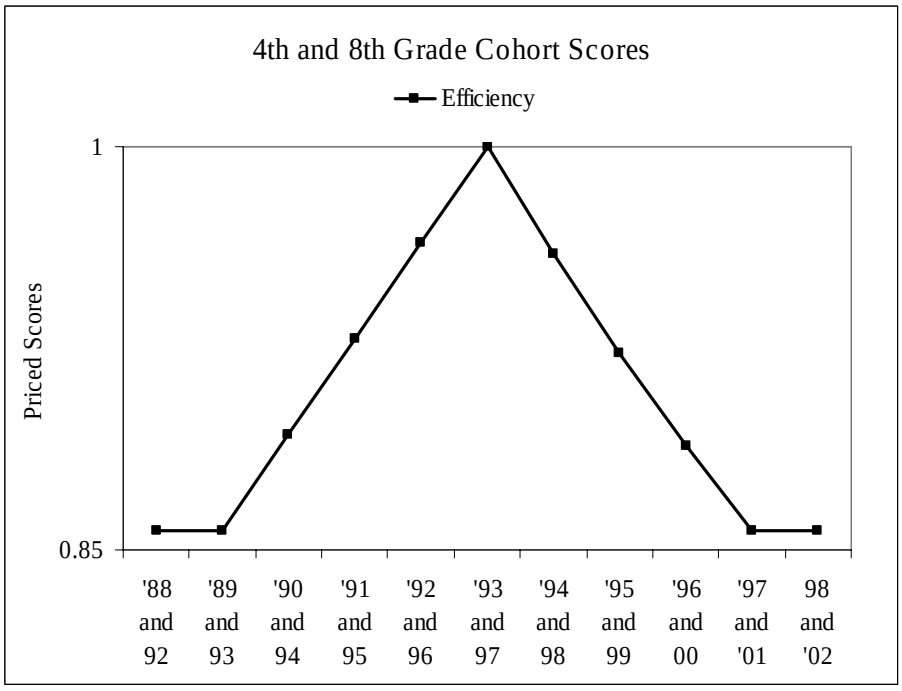
You have the minimum input to the maximum output, so when inputs are priced at 7/6 as calculated above, the inputs lie above the outputs for all cohorts except for the one that you are in. In other words all the others are inefficient – see Figure 6.05. Note that efficiency scores improve up to your efficiency and decline thereafter.

Figure 6.05 – 4th and 8th Grade Test Scores of Cohorts as Outputs and Priced Inputs.



When the efficiency scores are calculated as the ratio of the priced inputs to the outputs and graphed. The graph – Figure 6.06 – shows an increase in efficiency followed by a decrease in efficiency. So, if MERA gave rise to an improvement in education, then the CCR model should reveal increasing then decreasing average efficiency for school districts around the introduction of MERA.

Figure 6.06 – 4th and 8th Grade Cohorts Efficiency Scores.



6.2 Testing the Proposition

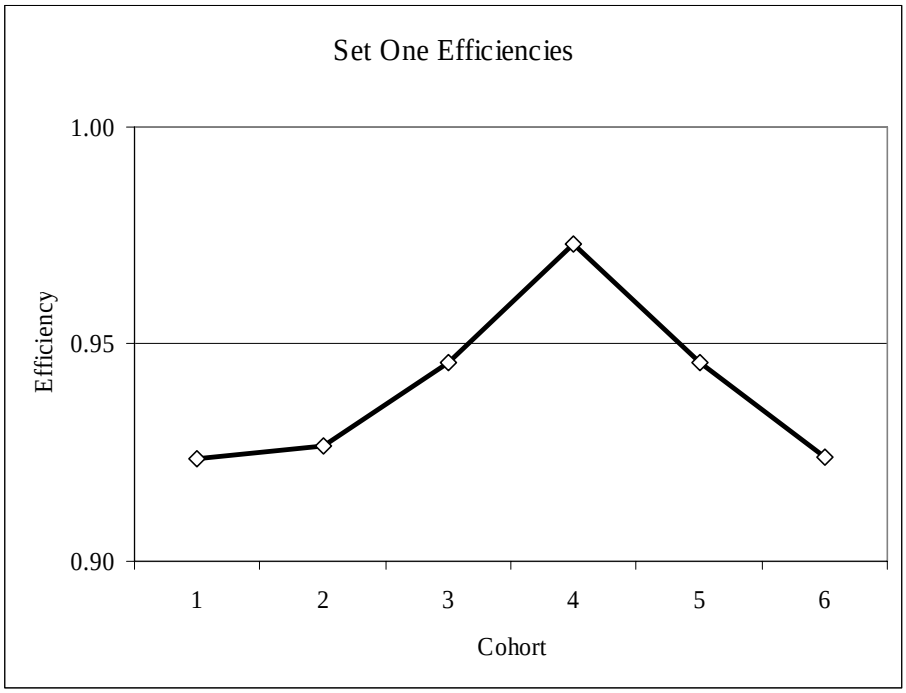
To test the proposition two sets of data were synthesized for even years from 1988 to 2002. Set One assumed a one off improvement evidenced over four years allowing for mean scores of 200 in each of 3 subjects for 1988, 1990, 1992 and 1994 in both the 4th and 8th grades. The mean increased to 205 for 1996 and increased again to 210 for 1998, leveling off at 210 for 2000 and 2002. Set Two assumed that improvement continued through to 2002, so Set Two was the same as Set One except for 2000 – a mean of 215 – and 2002 – a mean of 220.

Minitab was used to generate sets of randomized normally distributed data around these means with standard deviations of 5. The seed mean and standard deviations and the actual means and standard deviations of the generated data are given in Appendix J.

Grade 4 scores for 1988 formed inputs to grade 8 scores for 1992 as outputs and together they formed Cohort 1. Grade 4 scores for 1990 formed inputs for Cohort 2, which had 1994 grade 8 scores as outputs. And so on … through to Cohort 6, which used 1998 grade 4 scores as inputs and 2002 grade 8 scores as outputs. For each set of data, the data for the 6 cohorts was combined forming 1,080 "Decision Making Units".

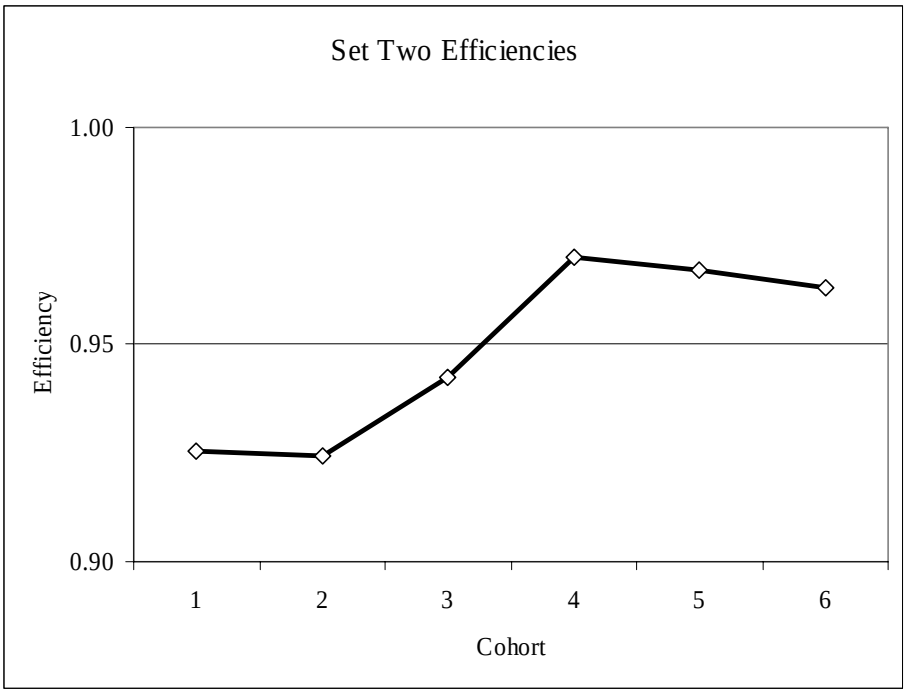
The results of these experiments confirmed the proposition that a one off improvement in Education should first result in increasing efficiency and then in decreasing efficiency – refer to Figure 6.07. This graph shows that the average efficiency first increased and then decreased as predicted.

Figure 6.07 – Average Efficiency Scores for Cohorts From Experimental Set One.



Set Two which assumed a continuous increase in test scores from 1994 onwards showed first increasing then decreasing efficiency although the rate of decrease was less dramatic than for Set One – refer to figure 6.08.

Figure 6.08 – Average Efficiency Scores for Cohorts From Experimental Set Two.



So the proposition that a one off improvement in education will lead to an increase in efficiency followed by a decrease in efficiency holds true.

6.3 The Actual Data

Forming cohorts for Grades 4 and 8 was relatively straightforward. 4 subjects tested in 1988, 1990, 1992, 1994, 1996 and (3 subjects in 1998) at grade 4 formed the inputs for each cohort. The average of the other three scores in 1998 was used to replace History for 1998.

The outputs at grade 8 were available for four subjects in 1992, 1994, 1996, (3 subjects in 1998), 2000 and (3 subjects in 2002). 1999's grade 8 History and Social

Studies score was appropriated to make the 1988 outputs up to four in total. 2000's Science and Technology/Engineering score and 2002's 7th grade English Language Arts score were appropriated to make the 2002 outputs up to four in total. The set of cohorts used is summarized in Table 6.01.

Table 6.01 – Sets of 4th to 8th Grade Cohorts.

Cohort	1	2	3	4	5	6
Inputs 4th Grade	1988 M,R,S,H	1990 M,R,S,H	1992 M,R,S,H	1994 M,R,S,H	1996 M,R,S,H	1998 M,R,S plus Average of M, R and S
Outputs 8th Grade	1992 M,R,S,H	1994 M,R,S,H	1996 M,R,S,H	1998 M,R,S plus 1999 H	2000 M,R,S,H	2002 M,H plus G7 R plus 2000 G8 S

Source: Massachusetts Department of Education

Notes:

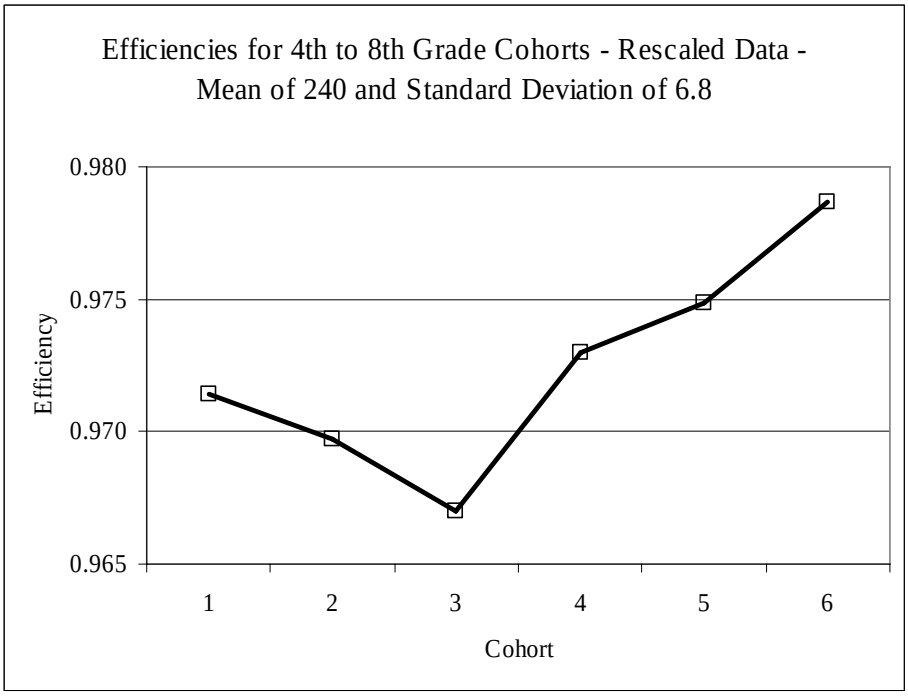
For MEAP Years (1988 to 1996), M=Mathematics; R=Reading; S=Science, and H=Social Studies.

For MCAS Years (1998 to 2002), M=Mathematics; E=English Language Arts; S=Science and Technology/Engineering, and H=History and Social Studies

6.4 The Results

The average of the efficiency scores derived using the CCR model was calculated for each cohort. The graph in Figure 6.09 shows the results for the 4th to 8th grade set of cohorts. Note that the range of scores is very small – 0.011681 (min 0.967016, max 0.978696)

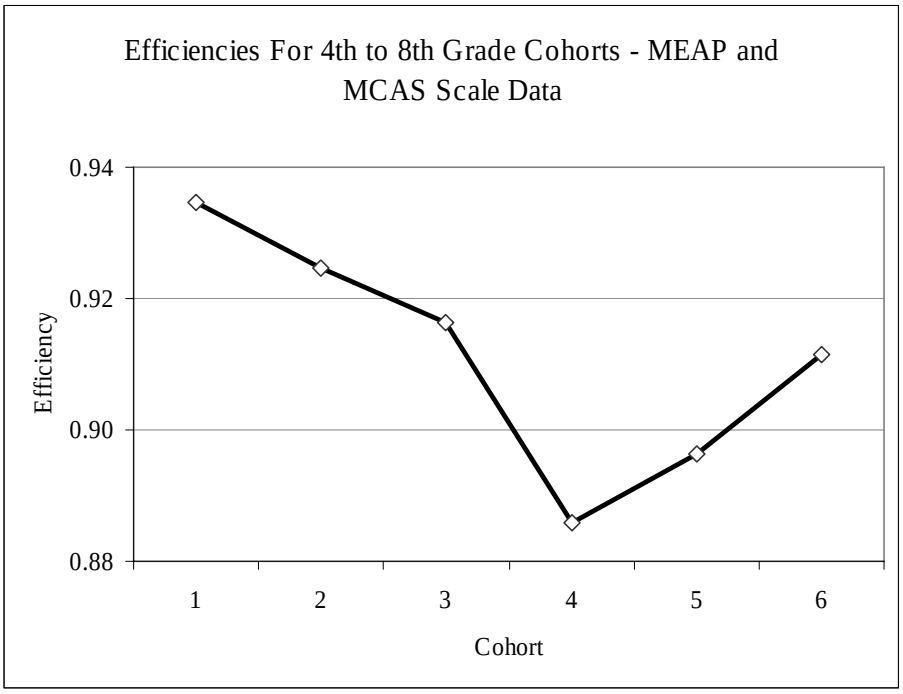
Figure 6.09 – Efficiencies for the 4th to 8th Grade Cohorts.



Cohort 2 includes the first 8th grade testing after MERA and cohort 4 includes the first 4th grade testing after MERA. If MERA had an immediate one off effect then the pattern of increasing and then decreasing efficiency should be seen, as in Figure 6.07, peaking at Cohort 4. Instead there is a steady decline up to Cohort 3 followed by an increase thereafter. The pattern seen is neither the pattern expected nor is it the opposite, since the reversal in trend occurs one cohort too early.

The Test Score data used was that scaled to a mean of 240 and a standard deviation of 6.8, so, in principle, the model should have revealed no changes at all over time. The model was re-run using the actual MCAS and MEAP scores and making an adjustment for the different scales of the scores as set out in Chapter 4, Section 6. The results are shown as Figure 6.10. Note that the range in scores is much higher than in the previous example: 0.048843 (min 0.885909, max 0.934752).

Figure 6.10 – Efficiencies for the 4th to 8th Grade Cohorts MCAS and MEAP Test Scores Adjusted for Scale Differences.



When the actual MEAP and MCAS scores are used the pattern revealed is exactly the opposite of the pattern that would be expected if there had been a one-off improvement as a consequence of MERA in other words the results suggest that MERA made things worse. In theory, individual student's MCAS scores are scaled to a common state mean and standard deviation. MEAP scores were also scaled to a common state mean and standard deviation. So, in theory, there should not have been an observable effect from models using actual MCAS and MEAP test scores. Recall, from Section 6 of Chapter 1, the quotations from Gipps (1988):

> Statistics of this kind are virtually meaningless because GCE grading is largely norm-referenced (when grades are awarded on the basis of how a student fares in comparison with other candidates) rather than criterion-referenced (where there is an attempt to compare a student's performance with some 'absolute' standard).

> The APU[26] has made little progress on its task of providing information on standards and how these are changing, because there is a major technical problem in measuring changes in performance on tests over time. That is, changes large enough to be meaningful will only be detected over a number of years, at least four or five, and any serious monitoring of performance would go on over a longer period than that.

In other words it is foolish to expect to derive meaningful comparisons of progress, statewide, from norm-referenced tests such as MEAP and MCAS.

This begs the question why is there an observable effect when using actual MCAS and MEAP test scores? The answer may lie in a comparison of the changes in the proportions of all students in a grade taking the tests with the changes in efficiency.

6.5 Participation in the Tests.

A table giving the percentages of enrolled students tested in each grade was given in Section 2 of Chapter 3 – Table 3.04. On the assumption that those who would avoid testing if possible would be the less able students, the impact of higher percentages taking a test should be to depress the average score. Taking the data from Table 3.04 for 4th and 8th graders and presenting it for the 6 cohorts allows the differences in the percentages tested at each grade within a cohort to be calculated as shown in Table 6.02.

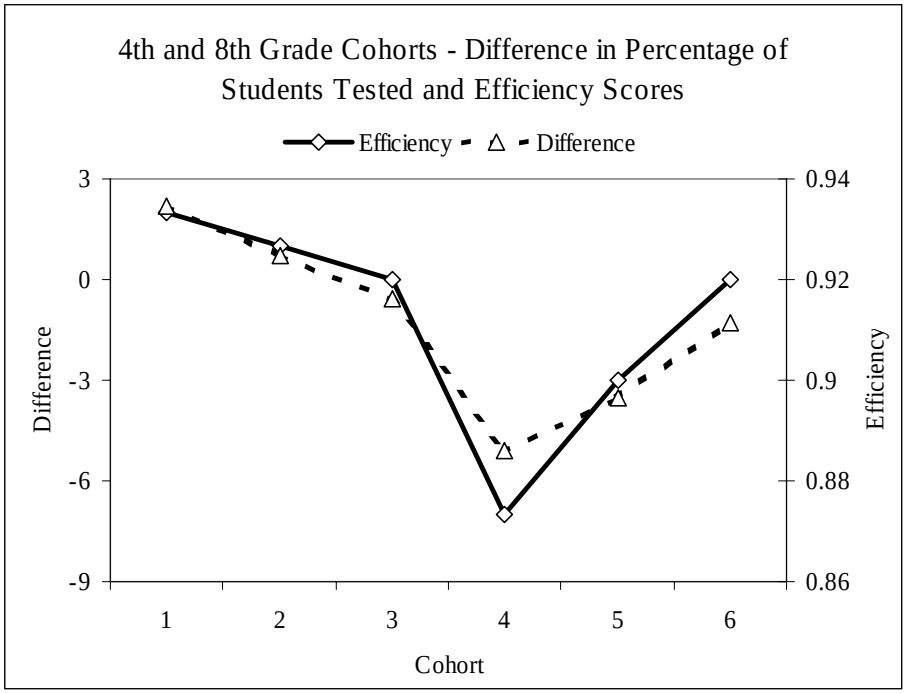
---

[26] APU is the Assessment and Performance Unit of the United Kingdom's Department of Education.

Table 6.02 – Percentages of Students Taking Exams Arranged By 4th to 8th Grade Cohort.						
	1	2	3	4	5	6
Cohort	1988 and 1992	1990 and 1994	1992 and 1996	1994 and 1998	1996 and 2000	1998 and 2002
Percent 4th Grade	90	90	89	90	90	97
Percent 8th Grade	88	89	89	97	93	97
4th less 8th	2	1	0	-7	-3	0
Source: Massachusetts Department of Education.						

When the difference is graphed on the same chart as the efficiency scores from Figure 6.10 – see Figure 6.11 – the result is two strikingly similar patterns, which suggests that the only information being captured by the analysis may be adjustments to the state mean made to reflect the changes in the percentages of students taking the examinations.

Figure 6.11 – 4th to 8th Grade Cohorts – Efficiency Scores and Differences in the Percentages Tested.



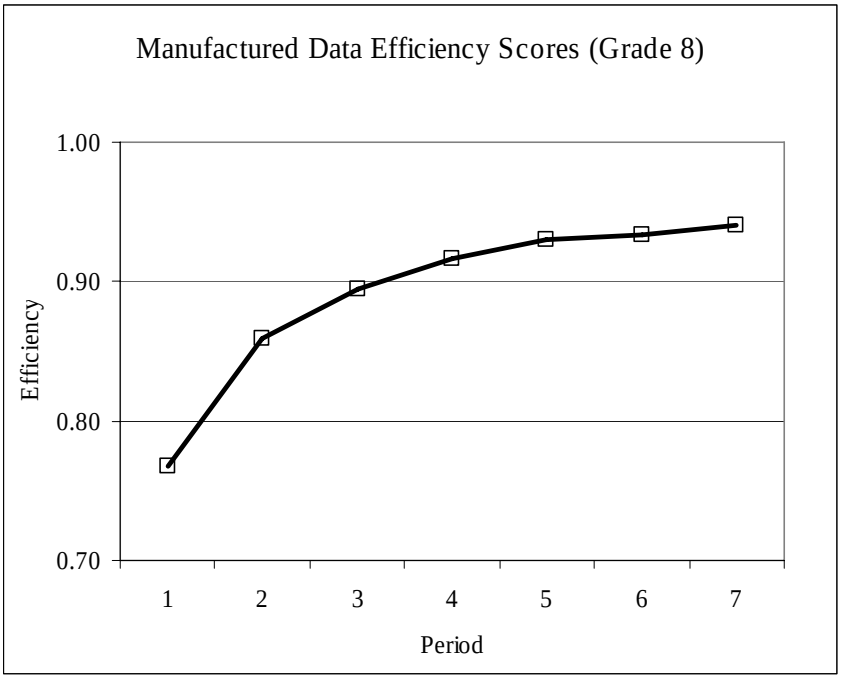
It appears that little can be said with any certainty about the impact of MERA on education in Massachusetts based on the analysis undertaken. The experiments described in Section 2 showed that if an even percentage of students had taken the tests over time and if an improvement had resulted from MERA then an improvement in efficiency scores would have been followed by a deterioration in efficiency scores. The pattern of changes in efficiency was precisely the opposite – a deterioration in efficiency scores followed by an improvement in efficiency scores.

6.6 Was Opportunity Made More Equal?

One of the key objectives of the Massachusetts Education Reform Act of 1993 ("MERA") was to ensure a basic minimum standard of educational opportunity. Raising the minimum, leaving other levels unchanged, would have the effect of making "educational opportunity more equal". The experiments undertaken in this chapter are intended to test whether educational opportunity was made more equal as a consequence of MERA.

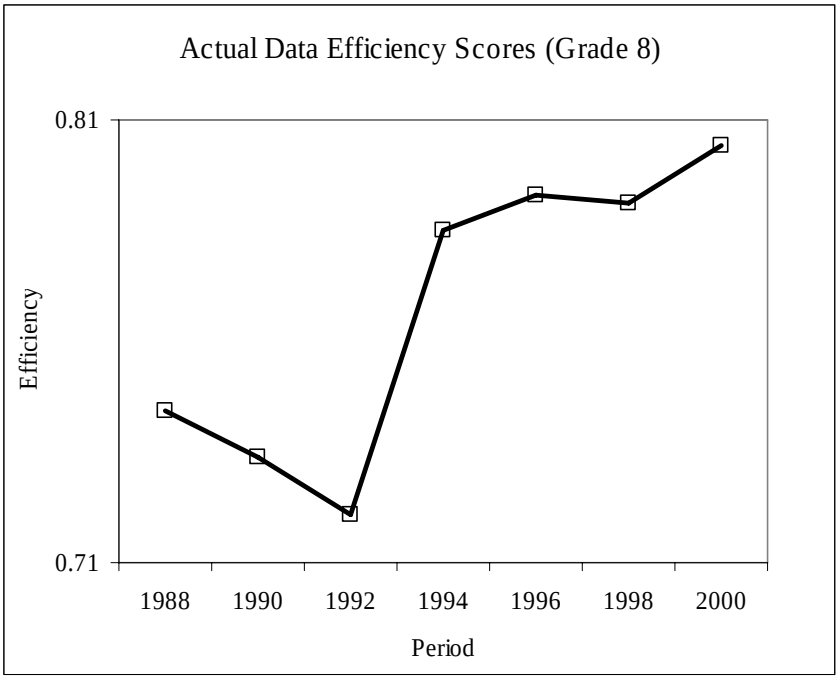
The proposition, set out in Section 7, is that if the base level of resources per pupil increases, then the efficiency with which poorer districts convert money into outcomes should decrease. To test this proposition data was synthesized to conform to the assumptions. The CCR model showed efficiency increasing as equity improved – see Figure 6.12.

Figure 6.12 – Manufactured Data Efficiency Scores – CCR Model.



The process was then applied to the actual data between 1988 and 2000. The results for Middle School expenditures and Grade 8 Test Scores are shown in Figure 6.13. Prior to MERA the average CCR efficiencies are lower than after MERA and the period from 1992 to 1994 shows a significant increase in average efficiency. The results for other grades, which are similar, are presented in Section 8.

Figure 6.13 – Efficiency – Actual Grade 8 Test Scores and Middle School Expenditures.



In conclusion it appears to be the case that MERA did coincide with an improvement in equity and given that MERA included significant increases in state aid for education targeted towards "poorer" towns it seems likely that MERA was a significant cause of the improvement in equity.

6.7 Measuring Variance in Base Resources – The Proposition

The basic assumption is that equality of opportunity in education can be measured by the size of the variance in the quantities of resources applied on behalf of each student. If MERA was successful in raising the base level of resources per pupil then the

efficiency with which poorer districts convert money into outcomes should decrease. It is assumed that better off districts continue to be as efficient (or inefficient) as before the reforms.

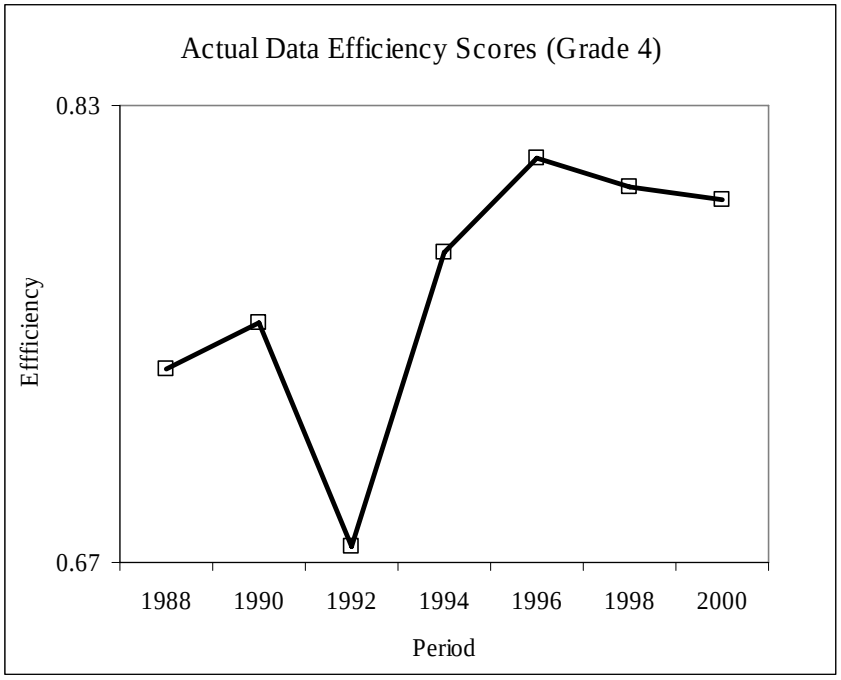
Consider two school districts, A and B, with the same tests scores. A spends twice as much as B. A is therefore 50% as efficient as B. Average efficiency is 0.75. Now increase B's budget to 80% of the amount spent by A. A is now 80% efficient against the benchmark set by B and the average efficiency has increased to 0.90. In this case the question is how, period by period, did the dispersion change. Analysis of the data using the first Panel Data approach – see Section 7 of Chapter 4 – putting all periods and DMUs into a single model will not reveal how dispersion changed period by period. So, models were run for each period separately.

If the assumptions are valid, then CCR models should show less variability in efficiency after MERA. Less variability in efficiency scores will be reflected in higher average efficiency scores. If equity improves over time, then so should average efficiency scores.

6.8 Results

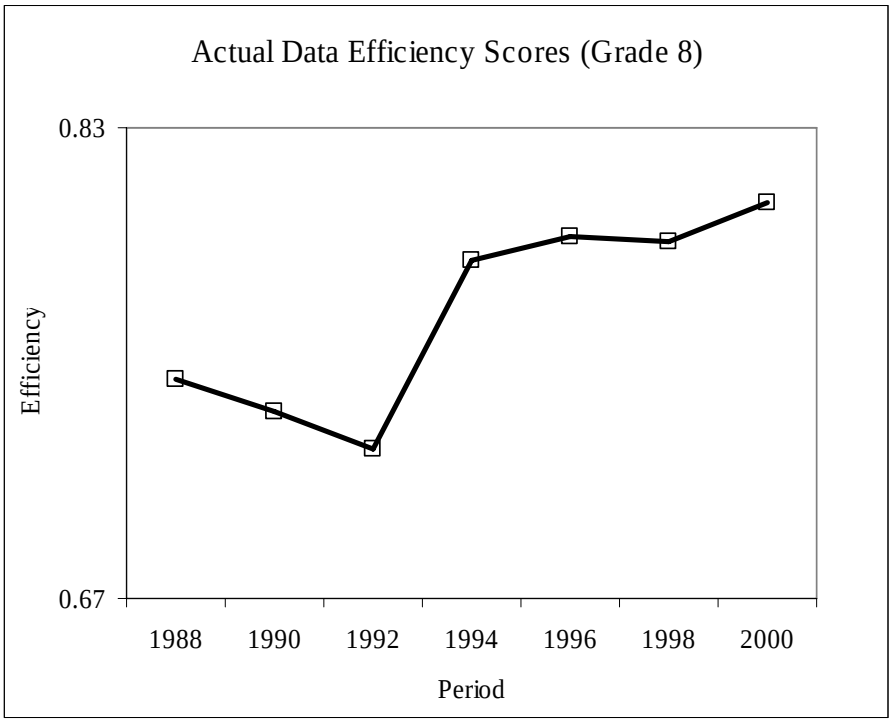
In the previous section it was established that the CCR model should show increasing efficiency accompanying greater equity when Per Pupil Expenditures are used as inputs to models that take test scores as outputs. The result derived from applying the process to actual data – see Figures 6.14, 6.15 and 6.16 – is precisely that efficiency improved from 1992/1994 onwards. This implies that MERA was effective in making opportunity more equal.

Figure 6.14 – Efficiency Scores from Actual Grade 4 Test Scores and Elementary School Expenditures.



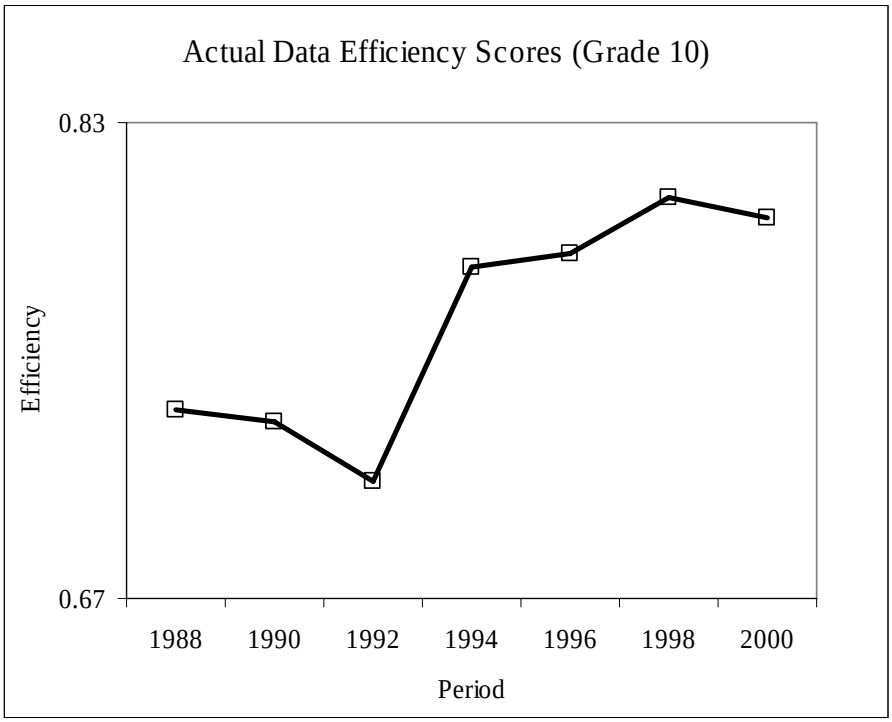
From the low point in 1992, before MERA, efficiency based on Elementary School Expenditures and Grade 4 Test Scores increased dramatically surpassing the levels seen in the years prior to MERA.

Figure 6.15 – Efficiency Scores from Actual Grade 8 Test Scores and Middle School Expenditures.



Efficiency scores based on Middle School Expenditures and Grade 8 Test Scores also increased from 1992 onwards and also exceed the levels seen prior to 1992. The "V" shape, down from 1990 and up to 1994, seen in the Grade 4 results is not as dramatic in the Grade 8 results. This suggests that the brunt of the budget crisis in the early 1990's was reflected in Elementary School Budgets.

Figure 6.16 – Efficiency Scores from Actual Grade 10 Test Scores and High School Expenditures.



The pattern of efficiency based on Grade 10 Test Scores and High School expenditures is broadly the same as that seen for middle school expenditures. There is little doubt that the base level of expenditure was increased for all school districts and this analysis suggests that equity also improved sharply between 1992 and 1994 and improved somewhat thereafter.

6.9 Conclusions

There is no evidence available from the analysis undertaken and presented in this Chapter to even suggest that education standards have been raised and educational outcomes have improved as a consequence of MERA. It is equally true to say that there is no evidence to the contrary.

There is evidence to support the conclusion that there was greater equity in funding in the years after MERA.