THE APPLICATION OF DATA ENVELOPMENT ANALYSIS TO PUBLICLY FUNDED K-12 EDUCATION IN MASSACHUSETTS IN ORDER TO EVALUATE THE EFFECTIVENESS OF THE MASSACHUSETTS EDUCATION REFORM ACT OF 1993 IN IMPROVING EDUCATIONAL OUTCOMES.

A Dissertation Presented

by

ANDREW D. J. HALL

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2005

Management

UMI Number: 3193907

THE APPLICATION OF DATA ENVELOPMENT ANALYSIS TO PUBLICLY FUNDED K-12 EDUCATION IN MASSACHUSETTS IN ORDER TO EVALUATE THE EFFECTIVENESS OF THE MASSACHUSETTS EDUCATION REFORM ACT OF 1993 IN IMPROVING EDUCATIONAL OUTCOMES.

A Dissertation Presented

by

ANDREW D. J. HALL

Approved as to style and content by:

Robert Nakosteen, Chair

Iqbal Agha, Member

Sara McComb, Member

Gary D. Malaney, Member

D. Anthony Butterfield
Ph.D. Program Director
Management

DEDICATION

This dissertation is dedicated to good teachers everywhere because they make a positive difference to all our lives.

ACKNOWLEDGMENTS

Without the invaluable and unfailing support and inspiration of my advisor, Robert Nakosteen, none of this would have materialized. His advice has always been insightful, considered, appropriate to the circumstances and sound both at the time and with hindsight. He has been a superb mentor and a good friend. I can think of no greater compliment than to say that he is a gifted teacher.

Any true understanding that I may have of Data Envelopment Analysis is the result of active discussion with Iqbal Agha. He, literally, taught me most of the things that I know about Mathematical Modeling and Linear Programming. He is an excellent teacher and gifted researcher. He has generously given his time and his work. In particular his software, IDEAS, was used to solve DEA models.

It has been good to have the opportunity to work with the other members of my dissertation committee: Sara McComb and Gary Malaney. They both brought a fresh perspective and constructively critical eyes to this dissertation.

The Department of Finance and Operations Management in the Isenberg School of Management has been generous in its financial support. My thanks are due to the Dean, Tom O'Brien; to the directors of the Ph.D program, Tony Butterfield and Hossein Kazemi; to the Department Chair, Nelson Lacey, and to Professor Anna Nagurney who taught me everything I know about Math Programming.

The Massachusetts Department of Education ("DOE") has been the most significant provider of data either from their Web Site or from published materials. Thanks are due to Roger Hatch, Robert Curtin and, in particular, Judy Chan Li for their assistance in furnishing specially requested data.

ABSTRACT

THE APPLICATION OF DATA ENVELOPMENT ANALYSIS TO PUBLICLY FUNDED K-12 EDUCATION IN MASSACHUSETTS IN ORDER TO EVALUATE THE EFFECTIVENESS OF THE MASSACHUSETTS EDUCATION REFORM ACT OF 1993 IN IMPROVING EDUCATIONAL OUTCOMES.

SEPTEMBER 2005

ANDREW D. J. HALL, B.A., UNIVERSITY COLLEGE LONDON

M.B.A., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Robert Nakosteen

The Charnes Cooper Rhodes ratio DEA model ("CCR") is used, with panel data from a large sample of Massachusetts' school districts, to test three propositions concerning the Massachusetts Education Reform Act of 1993 ("MERA"). First, did the degree of positive correlation between Socio-Economic Status ("SES") and educational outcomes decrease, secondly did educational opportunity become more equal among towns in Massachusetts, and finally were education standards raised overall?

The CCR model is a Linear Programming method that estimates a convex production function using Koopmans' (1951) definition of technical efficiency and the radial measurements of efficiency proposed by Farrell (1957). It has been widely used in Education Production Function research.

The pursuit, through state and federal courts, of equitable funding, allied to the belief that smaller class sizes improve outcomes, has made K-12 education expensive. The belief that outcomes are in constant decline has led to calls for "Accountability" and to "Standards" reform.

Standards reform was combined, in MERA, with reform of state aid formulas and additional state funding, to ensure a minimum basic level of education pursuant to the decision of the Massachusetts Supreme Court in McDuffy v. Robertson.

The one certain relationship revealed by decades of research is a strong positive correlation between SES and outcomes. If MERA ensured a higher basic level of education, then the correlation between SES and outcomes should have weakened as the education of less well SES-endowed children improved. The CCR model was used first to measure "correlation" between multiple input and multiple output variables. Strong positive correlation was shown to exist and it appeared to strengthen rather than weaken. Next the CCR model was used to determine if there were changes in the distribution of per pupil expenditures and, lastly to determine whether outcomes improved between after MERA. The analysis suggested that the distribution of expenditures improved but that outcomes deteriorated. This deterioration seems to be closely related to the changes in the proportion of all students, in a grade, actually taking the tests.

There is little evidence that MERA achieved anything and no basis upon which to argue that it achieved nothing.

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The thesis of this dissertation is that the success of the Massachusetts Education Reform Act of 1993 ("MERA") can be evaluated by testing the following propositions using data for the years from 1988 to 2002 and for a representative sample of Massachusetts' school districts and using a set of linear programming models collectively known as Data Envelopment Analysis ("DEA"):

1. Educational opportunity became more equal as a consequence of MERA.
2. The degree to which socio-economic status is a determinant of educational outcomes in Massachusetts has decreased as a consequence of MERA.
3. Education standards have been raised and educational outcomes have improved as a consequence of MERA.

The dissertation is arranged as follows: Chapter 1 looks at the context of education reform in the United States. Chapter 2 looks at the background to MERA, its provisions and their implementation. Chapter 3 documents the selection of the sample of 180 school districts and describes the data – expenditure, socio economics and test score – used in the analysis.

Chapter 4 looks at the selection of the DEA model type used in the analysis and justifies the selection of the CCR model. The nature of efficiency measured by the CCR model is explored in some detail and 8 factors which impact on the efficiency results from the CCR model are identified and discussed.

Chapter 5 looks at the high level of positive correlation between Socio-Economic Status and Educational Outcomes. It identifies the fact that the two phenomena have become more highly positively correlated since MERA rather than less highly positively correlated. This implies that MERA has not been successful in improving outcomes.

Chapter 6 reports the finding that the base level of expenditure has increased, but the distribution of expenditures is unchanged. Poor schools got richer, but rich schools got richer to the same degree. It also reports that test scores show no improvement, rather they show education getting worse. When the proportion of students taking the tests is taken into account: there is no clear evidence of improvement or otherwise.

A summary of the results and the conclusions is presented as Chapter 7. Briefly, Education finds itself caught between the movement for Standards, Taxpayer Revolt and Litigation. MERA was a good faith attempt to improve education in Massachusetts. The extra cash was made available but with little evidence that it made any difference to the outcomes. The approach of the Massachusetts Department of Education to Standards reform was painstaking, but, to date, there is no evidence that the Standards reform components of MERA have made or will make any difference.

CHAPTER 1

GENERAL TOPICS IN EDUCATION

Three major political trends, one article of faith and one over-arching assumption define education in the United States.

The first trend is the pursuit of equity in outcomes. Outcomes are assumed to be dependent on equity in funding and equity in funding has been pursued in the Courts. Scholars refer to three waves of "School Finance Litigation" which are described in Section 1. School finance lawsuits were argued in Massachusetts in all three waves and the reforms that started in the early 1990s were in part court ordered reforms.

The article of faith is that smaller class sizes improve outcomes. Since teachers are the main education cost this belief leads to higher costs per pupil, which lead to higher property taxes and result in the second political trend: taxpayer revolt manifested in the adoption of caps on property taxes. Section 2 presents evidence.

The third trend is the demand for "results" and "accountability". Outcomes are widely believed to have worsened, although the evidence, given in Section 3, is not at all clear-cut. "Accountability" movements seek to use testing and test results as a means of holding children, parents, teachers and administrators responsible for educational outcomes. Section 4 describes the failure of "Accountability" reforms in the 1970's. "Accountability" reform, upgraded and re-branded as "Standards" reform, was a component of the reforms in Massachusetts.

Educational Production Function Research, considered in Section 5, has assumed, from the Coleman Report onwards, that educational outcomes are strongly related to Socio-Economic Status ("SES").

Educational Production Function Research has produced little else in the way of statistically significant relationships between outcomes ("outputs") and "inputs". Where a relationship is alleged to exist little is known and even less is proven about the direction of causation.

A key component of both Educational Production Function Research and Standards based reforms is measurement of educational outcomes. Measurement is very problematic for a number of reasons. Section 6 looks briefly at the problems and issues in measurement and the use of test scores in research.

1.1 School Finance Litigation

Implicit in School Finance Litigation is the idea that money begets results. Thus parents in poorer towns sued their states to make the inputs to education more equitable between school districts.

Commentators divide school finance litigation into three waves. Readers are referred to Thro (1994), Brown (1994) and, for a more extensive review, to Jordan and Lyons (1992).

The first wave, starting in the late sixties, was characterized by the use of the U.S. Constitution's Equal Protection Clause and ended with the 1973 United States Supreme Court decision, in San Antonio Independent School District v. Rodriguez[1], that the Texas's funding system was acceptable under the Equal Protection Clause.

In the second wave, litigants turned to the individual state constitutions' Equal Protection Clauses. It began with the New Jersey Supreme Court's decision, in Robinson

---

[1] San Antonio Independent School District v. Rodriguez. No. 71-1332 Supreme Court of the United States 411 US 1; 93 S. Ct. 1278; 1973 US.

v. Cahill[2] that New Jersey's funding system unconstitutionally discriminated against students in low-wealth areas. Plaintiffs also prevailed in Arkansas, California, Connecticut, New Jersey, Washington, West Virginia and Wyoming, but the overwhelming majority of the cases resulted in victories for the state.

The third wave relied upon state constitutions' Education Clauses. Instead of placing emphasis on equality of expenditures the third wave cases emphasized the quality of the education being delivered, arguing that all children were entitled to a minimum level of education. It started, in 1989, with the decision by Kentucky's highest court, in the case known as Rose v. Council for Better Education Inc.[3], in which the court invalidated, not only the finance system, but also every statute relating to public schools and ordered the legislature to design a new system.

The plaintiffs have been more successful in the third wave cases than in either of the other two waves. Supreme Courts in Alabama, Kentucky, Massachusetts, Montana, New Jersey, Tennessee and Texas have all struck down their respective finance systems.

1.2 Instructor Pupil Ratios, Costs and Taxpayer Revolt

It is implicit in all of the school finance cases and resulting educational reforms that smaller classes result in better educational outcomes, but the evidence suggests little effect if any. See, in particular, Woodhall and Blaug (1968), Levin et al. (1976), and West (1983).

---

[2] Robinson v. Cahill, 118 N.J.Super. 223, 287 A.2d 187 (Law Div.1972), Robinson v. Cahill, 62 N.J. 473, 303 A.2d 273 (1973), Robinson v. Cahill, 63 N.J. 196, 306 A.2d 65, cert. denied, 414 U.S. 976, 94 S.Ct. 292, 38 L. Ed.2d 219 (1973), Robinson v. Cahill, 67 N.J. 35, 335 A.2d 6 (1975), and Robinson v. Cahill, 69 N.J. 133, 351 A.2d 713, cert. denied 423 U.S. 913, 96 S.Ct. 217, 46 L. Ed.2d 141 (1975)

[3] Rose v. Council for Better Education, 790 SW.2d 186, 60 Ed. Law Rep. 1289 (1989).

According to the U.S. Department of Education, National Center for Education Statistics, there were 24.8 students per instructor or four instructors for every hundred pupils in 1960. By 1999 eight instructors were available for every hundred K-12 students.

Fuelled by the imperative towards smaller classes, by the tendency for fiscal equality to be achieved by equalizing expenditures upward, and driven by demographics: public school budgets ballooned both in nominal dollars and in real terms. Education's share of U.S. Gross Domestic Product rose from 1.85 percent in 1949-50 to 3.75 percent in 1969-70.[4]

There was bound to be a backlash. Proposition 13 was adopted by California voters on June 6th, 1978 with an almost two-thirds majority Sexton et al. (1999). Within 2 years, according to Hatward (1998), 43 states implemented some kind of property tax relief. Massachusetts voted, in 1980, in favor of Proposition 2½, which, like the California initiative, amended the state constitution to limit property taxes.

The effect was to make it even harder for poor school districts to afford the same levels of expenditure as richer towns, making state aid formulas more important and leading to more school finance litigation.

1.3 Educational Outcomes

Implicit in every call for reform of K-12 education is the assumption that the education system is failing. See West (1983), Hanushek (1994), and many others.

Some evidence supported this idea. In 1964, the federal government appropriated funds the National Assessment of Education Program, a national three-year cycle of

---

[4] Using the ratio of Revenue Applied to Public Education, from the U.S. Department of Education, National Center for Education Statistics and the U.S. Gross Domestic Product from the U.S. Department of Commerce, Bureau of Economic Analysis, National Accounts Data.

testing in each of four age groups 9, 13, 17 and 26-35. According to Guthrie (1980), in 1977 NAEP estimated that 13 percent of 17 year-old high school students were functionally illiterate. Wirtz (1977) tracked a steady decline in SAT scores between 1963 and 1977 in an investigation that looked at factors such as the more than tripling in the numbers taking the test each year and concluded that:

> there has been a lowering of educational standards and that this is a factor in the decline in SAT scores. We conclude at the same time that the correction of the various elements in this situation requires the collaboration of teachers, students, parents, and the broader community in the establishment of standards that can be truly considered higher only as they recognize youths' essential diversity.

Other evidence did not support the idea that the education system was failing. The number and percentage of persons aged 25 years or older with four or more years of high school education increased steadily from 34 percent in 1950 to 77.6 percent in 1990 and those with four or more years of college increased from 6 percent to 21.4 percent over the same time period, see Behrman and Stacey (1997) and Bracey (1997). Wirtz (1977) made the similar point that between 1960 and 1970 the number of students graduating from high school increased by 153 percent and the number taking the SAT increased by 300 percent. In the 35 years to 1985, according to Bracey (1997), school and college curricula broadened and children spent more time studying.

More children and a greater percentage of children are getting a broader education with a wider curriculum. If accurate, this is not a description of educational failure.

1.4 Accountability

1.4.1 Accountability in the 1970's

Whatever the truth, the public perception was one of clear educational failure. Money, which had often been given as the cause of the schools' problems, had been provided: with no apparent improvement. Congressional Quarterly Inc (1981) describe the reaction:

> One reaction to this perceived decline has been the movement toward returning to traditional methods of teaching. And one offshoot of the trend back to basics is the adoption by many states of standardized, mandatory minimal competency tests, especially as requirements for high school graduation.

The movement was referred to by some – Baron and Sergi (1979) – as "Minimum Competency Testing", and by others as the "Scientific Accountability Movement" – Levinson (1999) – and by yet others as the "Technocratic Accountability Movement" – Guthrie (1980).

Reform was to have consequences, so, a cut-off score was defined, below which some penalty was imposed on the student who had failed. Baron and Sergi (1979) point out the political problem with this approach:

> In practice this tends to become a question of 'How many children can we afford to fail?' in terms financial, remedial education etc and in terms of public relations.

In Florida, the tests were struck down by court action – Futrell and Brown (2000). The Fifth Circuit Court ruled that the test, curriculum, and teaching are inextricably linked – Debra P. v. Turlington[5]. In other states the tests simply became so unpopular that expenditure on them would be one of the first cuts at the next budget crisis in the state.

Teachers did not escape. They were to be trained into competency in spite of the fact that, according to Guthrie (1980):

> Teaching continued to be far more of an art than a science. ... the scientific base of pedagogy was simply too thin to justify competency-based teacher education, and the idea generally was short lived.

Murphy and Cohen (1974) concluded that:

> It may be that with time, research, and modest field trials, things can be improved. Certainly an effort ought to be made. But if Michigan is any guide, at this point scientific accountability hardly merits full-scale implementation.

And Guthrie (1980) concluded that:

> For all the publicity, money, and effort, the technocratic accountability movement appeared by the early 1970s to have produced little by way of results.

1.4.2 The Standards Movement

In the mid 1980's; as state budgets took on a greater role in education finance, after taxpayer revolts, and as a result of school finance litigation; education became more centralized in the states. Greater centralization raised the need for more mechanical control over education contributing to the need for "accountability". Accountability movements morphed into Standards movements.

---

[5] Debra P. v. Turlington, 474 F.Supp. 244 (MD FL 1979) and Debra P. v. Turlington, 633 F. 2nd 397 (5th Cir. 1981).

According to Hanushek and Raymond (2001); setting standards; measuring outcomes and holding students, teachers, schools and parents responsible for the outcomes, would lead to better performance.

Altering the incentives would change the behavior of students, teachers, administrators, and parents in a way that improved learning – Betts and Costrell (2001) and Finn and Kanstoroom (2001). The results of standards assessments would be powerful tools for local change, according to Smith et al. (1998). Clear information about performance would enable communities to invest their time, money, and energy in schools more effectively. For this to have a chance of working the following three conditions must be met:

1. All of the different levels of authority in the process of education from the political masters and paymasters (legislatures and town governments), the Departments of Education, the town and city governments, the school district superintendents, the school boards, the teachers, the parents, the school councils and the pupils need to be encouraged to work, in a coherent manner, towards the educational goals set by society through the political process.

2. A fair system of assessment that properly measures progress towards the educational goals is needed to provide a basis for awarding the incentives and disincentives, and

3. Everyone involved must refrain from altering the system too frequently, gaming the system or outright cheating.

In other words educational "Standards Based" reforms defy the usual laws of political gravity.

1.5 Educational Production Function Research

Education Production Function Research assumes a strong relationship between Socio-Economic Status and academic achievement. Education Production Function Research has produced no clear results. It began with the Coleman Report.

1.5.1 The Coleman Report

Section 402 of the Civil Rights Act of 1964 called on the Department of Health, Education and Welfare to undertake a survey of educational opportunity in the United States. The resulting report, which was published in July 1966, was titled "Equality of Educational Opportunity" – Coleman (1966). James Samuel Coleman was credited as the author and so the report came to be known as "The Coleman Report".[6]

A key finding of the Coleman Report was that schools and their characteristics have very little impact on student achievement.

> "It is known that socio-economic factors bear a strong relationship to academic achievement. When these factors are statistically controlled, however, it appears that the differences between schools account for only a fraction of differences in pupil achievement."

This finding found its way into public consciousness as meaning that schools have no effect on learning. In 1986, William J. Bennet, Reagan's Secretary of Education wrote:

> Its conclusion that unequal achievement could *not* be ascribed to unequal school resources so offended the conventional wisdom of the time that the next 20 years of educational research have been dominated by the quest for contrary evidence. – United States Department of Education (1986)

---

[6] The full list of authors follows: James Samuel Coleman, John Hopkins University; Ernest Q. Campbell, Vanderbilt University; Carol J. Hobson; James McPartland; Alexander M. Mood, Frederic D. Weinfeld and Robert L. York of the U.S. Office of Education.

Following the lead given in the Coleman Report, researchers hypothesized an "Educational Production Function", to which Kain and Hanushek gave the following form in Chapter 3 of Mosteller and Moynihan (1972):

$A_{it} = g\left(F_i^t, P_i^t, I_i, S_i^t\right)$ where

A_{it} is educational achievement of the ith student at time t

F_i^t is the cumulative individual and family characteristics for the ith student at time t

P_i^t peer influences - student body characteristics of the other students to time t

I_i individual endowments of the ith student

S_i^t school inputs relevant to the ith student cumulative to time t

Keisling (1967), Raymond (1968), Cohn (1968), Hanushek in Michelson and Levin (Eds.) (1970) Chapter 4, Bowles (1974), Levin et al (1976), Link and Ratledge (1979), and, Sebold and Dato (1981) use a similar framework. Hanushek (1996) tabulated the results of 377 Educational Production Function studies and showed that they were rarely statistically significant and often contradictory.

Regression Analysis was yielding little by way of useful results, so researchers turned to more sophisticated techniques: canonical regression developed in Aigner and Chu (1968) and used by Chizmar and Zak (1983); or Nested-Error Components analysis used by Montmarquette and Mahseredjian (1989), and various DEA models – refer to Appendix B. Unfortunately these studies tend to say more about the models that they use than they do about education. Reliance on test scores as inputs or outputs to the models also brings with it its own problems.

1.6 Assessment, Testing and Test Scores – Some General Issues

People cheat on exams because a positive outcome is important to them. One impact of accountability systems based upon the results of assessments is to align the students' interest in cheating with that of the teachers and schools. According to Breckheimer et al. (2001), such cheating led North Carolina to a complex set of procedures to maintain test security and to prevent "administrative irregularities" by teachers who administer the tests.

The average results for a school will be sensitive to the distribution of those taking the tests, so Schools can also cheat by holding back students, classifying them into categories that are not tested or by encouraging them to drop out. Departments of Education, bowing to political realities, can also cheat by creating special "re-test" exams.

Breckheimer et al. (2001), say that using test results to evaluate educational outcomes over time requires that the tests be valid and of consistent level of difficulty over time i.e. "Criterion Referenced". Criterion referenced assessments are those which measure performance against a "fixed" criterion. Much as stability is desirable, time changes things, so, curricula will change. The questions have to change, and experts must be brought in to "normalize" the criterion referenced assessments each time assessment is undertaken. This adds subjectivity and cost to the process.

Criterion referenced assessments whose results are scaled to a mean and standard deviation lose information critical to the evaluation of education over time. Norm referencing takes the results of tests and scales them to a norm of some kind, which is unchanging. They do not allow any meaningful inter-temporal comparison of one student to another as Gipps (1988) explains:

> 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).

Assuming that criterion referenced scores are available then the issue becomes one of the size and temporal scope of a meaningful change in scores. Since K-12 education takes place over a number of years and measurement is subject to a good deal of "statistical noise"; trends can only be identified and verified over an extended period of time. Gipps (1988) again:

> The APU[7] 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.

1.7 Conclusions

Education reform in Massachusetts in the 1990s did not arise in a vacuum. As we shall see, litigation resulted in the Massachusetts Supreme Court ordering the Executive

---

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

and Legislative Branches to ensure a basic level of education. In common with other states, Massachusetts' taxpayer revolt had capped increases in Property Taxes in Proposition 2½. As a consequence, State Aid for education had to be increased. The Executive wanted "accountability" for the extra State Aid and supported Standards reform.

These elements all came together in the Massachusetts Education Reform Act of 1993 ("MERA"), which is the subject of the next chapter.

CHAPTER 2

EDUCATION IN MASSACHUSETTS

2.1 Education in Massachusetts

Massachusetts has a long and admirable history of concern for the quality of education. The Massachusetts law of 1647 made each town responsible for educating all children to enable them to understand the religion and laws of the society and to provide them with the rudiments of a vocation. In McDuffy v. Robertson[8] the Massachusetts Supreme Court described this law as the beginning the history of public education in America.

When the Constitution of the Commonwealth, which included a separate chapter devoted to education, was adopted in 1780, a system of public schools had been in place for over 130 years and a college had been in existence for over 140 years.

In 1846, in Cushing v. Newburyport[9], the question was whether the power of towns to tax themselves to support schools was limited to the power to raise money to support only the number and type of schools that they were required by State Statute to maintain. The court held that the Statute set a minimum and that towns could tax themselves to support more schools than they were required to maintain. This prevents the Commonwealth from capping the level of expenditure on education to achieve greater equity.

In 1978, a school finance suit was filed by 16 students attending schools in Brockton, Belchertown, Berkley, Carver, Hanson, Holyoke, Lawrence, Leicester, Lowell,

---

[8] 415 Mass. 545; 615 N.E.2d 516; 1993 Mass.
[9] 10 Met 508 (1846)

Lynn, Rockland, Rowley, Salisbury, Springfield, Whitman and Winchendon; with the Massachusetts Supreme Judicial Court, under the caption Webby v. Dukakis[10]. Shortly thereafter, the Legislature enacted "School Funds and State Aid for Public Schools", St. 1978, c. 367 Section 70 C; known to some as the Collins Boverini legislation and, following this legislation, the case was suspended for five years: see Gaudet (1994).

In 1980, voters passed Proposition 2½, which put town budgets, particularly in poorer towns under stress. The plaintiffs re-opened the case by initiating discovery in 1983. In July 1985, as the court prepared to hear the evidence, the Legislature passed "An Act improving the public schools of the Commonwealth", St. 1985. c.188 Section 12 and proceedings in the case were suspended once more.

One part of the reforms was the creation of Equal Opportunity Grants under Massachusetts General Laws Chapter 70A Section 1, which deals with State financial aid for K-12 education. The second part of the reforms were two uniform statewide testing programs:

> The Massachusetts Educational Assessment Program is designed to improve curriculum and instruction in the public schools. The Basic Skills Testing Program aims to identify and assist students who are deficient in mastery of basic skills in reading, writing and mathematics. – Massachusetts Department of Education (1987)

When the state budget went into crisis, in 1989, education was one of the first items on the budget to suffer. The Basic Skills Testing Program was eliminated in 1991 as part of that year's budget cuts.

---

[10] The name then given to the case that in time would be known as McDuffy v. Robertson.

One of the provisions of Proposition 2½ had been to transfer ultimate responsibility for the education budget from the School Committee to the Town government. The town of Lawrence appeared to have taken advantage of this fact and the absence of a clear classification of the amounts of State Aid given, into amounts for education and for other expenditures, to divert approximately $33 million from education to the general town budget, between financial years 1985 and 1992. Not surprisingly, the plaintiffs re-opened the case once again, in 1990. The question in the early nineties was who would act first, the Massachusetts Supreme Court or the Legislature.

By May 2, 1993, there were two versions of an education reform bill in the State House. Both drew heavily on the Massachusetts Business Alliance's report entitled "Every Child a Winner" – Massachusetts Department of Education and Massachusetts Business for Education (1991).

On June 15, 1993, the Massachusetts Supreme Court published its decision in McDuffy v. Robertson[11]. It decided that the Massachusetts' Constitution made the education of all the school children in Massachusetts a duty of the Commonwealth, rather than, as the defendants had argued, an aspiration to which the State should work.

The plaintiffs, perhaps having learned the lessons of Serrano v. Priest[12], had constructed their case cleverly so as not to insist on absolute equality of expenditures; instead they sought:

---

[11] 415 Mass. 545; 615 N.E.2d 516; 1993 Mass.

[12] Serrano v. Priest, 5 Cal.3d 584 (1971) (Serrano I); Serrano v. Priest, 18 Cal.3d 728 (1976) (Serrano II); Serrano v. Priest, 20 Cal.3d 25 (1977) (Serrano III).

> … a declaratory judgment that these constitutional provisions require the State to provide every young person in the Commonwealth with an 'adequate' education. The plaintiffs argue that Part II, c. 5, Section 2, and arts. 1 and 10, each require 'equal access to an adequate education, not absolute equality.'

The standard that the Court adopted for a minimum education followed the guidelines set out in the first major case of the "third wave" of School Finance Equity Litigation, Rose v. Council for Better Education Inc.[13]

> The guidelines set forth by the Supreme Court of Kentucky fairly reflect our view of the matter and are consistent with the judicial pronouncements found in other decisions. An educated child must possess 'at least the seven following capabilities: (i) sufficient oral and written communication skills to enable students to function in a complex and rapidly changing civilization; (ii) sufficient knowledge of economic, social, and political systems to enable students to make informed choices; (iii) sufficient understanding of governmental processes to enable the student to understand the issues that affect his or her community, state, and nation; (iv) sufficient self-knowledge and knowledge of his or her mental and physical wellness; (v) sufficient grounding in the arts to enable each student to appreciate his or her cultural and historical heritage; (vi) sufficient training or preparation for advanced training in either academic or vocational fields so as to enable each child to choose and pursue life work intelligently; and (vii) sufficient level of academic or vocational skills to enable public school students to compete favorably with their counterparts in surrounding states, in academics or in the job market.'

The Court agreed that an adequate level of education was not being reached in towns in Massachusetts and passed the problem to the Legislature with the threat of further action if the Legislature failed to act in an appropriate way within a reasonable

[13] Rose v. Council for Better Educ., Inc., 790 S.W.2d 186, 212 (Ky. 1989).

time. By June 17 1993 the Massachusetts House and Senate bills were reconciled and passed. On June 18 1993 the Massachusetts Education Reform Act of 1993 ("MERA") was signed into law by Governor Weld.

2.2 The Massachusetts Education Reform Act of 1993

In assembling this précis of the elements of the Massachusetts Education Reform Act, I relied on the Act itself,[14] on Rossman and Anthony (1994), and on Parker (1997).

The Act encompassed a 7 to 10 year plan for standards reform and adequate minimum per student funding of education. The key elements were:

- 4 New Goals for Education in Massachusetts.
- Common Core, Curricula and Standards.
  - A high level common core was to be developed with extensive public participation and consultation. Academic standards, that would lend themselves to objective measurement, were to be developed in seven disciplines. Curriculum Frameworks were to be drawn up for the seven core disciplines.
  - Testing: a comprehensive diagnostic assessment of individual students was to be conducted at least in the fourth, eighth and tenth grades.
- The Act mandated that all students must pass the tenth grade test in order to graduate.
- Using a Foundation Budget and State Aid formulas, the Commonwealth would assure fair and adequate minimum per student funding for public schools.
  - The Foundation Budget for each school district was to be calculated by multiplying the enrollments in different student categories by their

[14] Chapter 71 of the Acts of 1993.

"standard" expense, adjusted by the application of a Wage Adjustment Factor. By the Year 2000, all districts in the state would be at their foundation level. The average foundation budget, in 1993, was to be $5,500 per pupil.

  - The State Aid formulas were based on the proposal in "Every Child a Winner" that state aid should be set so that a town's minimum contribution to meeting Foundation Budget levels of expenditure would be no more than $10 per thousand dollars of residential property value.

- Organizational / Procedural Elements.
  - Professional Development.

    Beginning in 1998, all new teachers were required to pass two tests to become certified to teach in public schools in Massachusetts.

  - School profiles allowing comparison of schools and school districts were to be published to "empower" parents.[15]
  - Better-defined roles for school committees.
  - Site Based Management – School Improvement Plans and School Councils.
  - Line management was given greater control over hiring and firing.
  - Principals were given broader powers to expel students.
  - School receiverships.

2.3 Implementation – Curricula

Between September 1993 and June 1994 a 40 member Commission on the Common Core of Learning, labored to produce the Common Core. It was adopted by the State Board of Education in July 1994 – Massachusetts Department of Education (1994).

[15] "Profiles Of School Districts Show Disparities", Laura Pappano, The Boston Globe, July 31, 1994, Third Edition, West Weekly Section, Page 1.

Curriculum Frameworks are the foundation upon which "Standards" are built. They are the standards. In all it took fully ten years to produce a full set of Curriculum Frameworks.

Debra v. Turlington – see Section 4 in Chapter 1 – meant that testing in a discipline had to be dependent on the existence of a Curriculum Framework, for the discipline. Assessment under MERA, the MCAS began late: in 1998. Only Mathematics and English Language Arts have been tested in all years and grades since then.

In the context of a 7 to 10 year plan, finishing the first step after 10 years is clearly failure. This failure was both political and predictable. Curricula are not easy to define. Should contraception be taught as part of human reproductive health?

The Development of Massachusetts Curriculum Frameworks can be split into two distinct, political, phases. The first, teacher centered phase, lasted from 1993 to January 1996 when the Board of Education Chairman, Martin Kaplan, resigned in the light of criticism of the "slow pace" of curriculum development and in the knowledge that Governor Weld intended to name John Silber, Dean of Boston University, as his replacement – see EyeOnEducation (2002).

Curriculum Frameworks were published in January 1996 for Arts; Health; Foreign Language; Mathematics; and Science and Technology / Engineering.

For the second phase, from 1996 onward, Silber appointed small teams of Board of Education members and "curriculum experts" to whom he gave the responsibility of revising and finalizing the frameworks. Revised and, as of January 2005, current Curriculum Frameworks, were released in the order given in Table 2.01:

Table 2.01 - Curriculum Frameworks Release Dates.	
Date	Curriculum Framework
August 1999	Foreign Languages
October 1999	Arts; Comprehensive Health
November 2000	Mathematics
May 2001	Science and Technology / Engineering
June 2001	English Language Arts
October 2003	History and Social Sciences
May 2004	Revised English Language Arts and Mathematics
Source: Massachusetts Department of Education.	

2.4 Implementation – Organization and Procedures

Investment in Professional Development appears to be the singular success of the Organizational / Procedural reforms.

> In February 1997, Governor William Weld executed Executive Order 393 establishing the Educational Management Accountability Board (EMAB). EMAB's mission was to review, investigate and report on the expenditure of funds by school districts, consistent with the goals of improving student achievement. The Order also directed the Board to verify the accuracy of reports submitted by school districts to DOE, and review progress under Education Reform. – Massachusetts Department of Revenue (2002).

Between November 1997 and February 2000, EMAB undertook 19 audits[16] summarized in Massachusetts Department of Revenue (2002). On the face of it the schools themselves were not doing very much to implement the reforms. Although nearly

[16] Agawam, Auburn, Braintree, Brockton, Cambridge, East Longmeadow, Everett, Gardner, Lexington, Lowell, Malden, Milton, New Bedford, North Attleborough, North Reading, Salem Triton, Worcester and Woburn.

all schools had created school improvement plans, many plans did not address student performance or test results, and while most school districts had met the minimum time & learning requirement, few districts had extended the school year. Investment in professional development is described as the one singular success of MERA, but EMAB had concerns over the effectiveness of the spending.

MERA removed principals from collective bargaining units, and they were supposed to be held accountable through performance-based contracts. EMAB found, however that most districts had not implemented the changes.

EMAB found that necessary data was not coherently collected and that limited student improvement data was available, because the MCAS started behind schedule. Districts are required to submit biannual reports to the state, but it was clear to the Board that district data in these reports was often incorrect, inconsistent and too late, and that the information was never reconciled by the Department of Education.

2.5 Implementation – Competency Determination

MERA mandated that students should earn a competency determination on assessments as a condition of receiving a high school diploma. In January 2000, the Board of Education established regulations that require students to attain a performance level of "Needs Improvement" or higher on both the grade 10 English Language Arts and Mathematics tests of the MCAS to earn a competency determination. Students have multiple opportunities, prior to their scheduled graduation date, to retake the test(s) that they did not pass.

The graduating class of 2003 which first took the 10^{th} Grade MCAS in 2001, was the first required to earn a competency determination and to have had the opportunity to take up to three retests. By the end of the third retest a significant improvement in the overall percentage achieving-competency is evidenced – see Table 2.02.

Table 2.02 - Class of 2003 Competency Determination April 2003.				
	Cumulative Percent Passed			
	Grade 10 Test	Retest 1	Retest 2	Retest 3
White	77	82	87	94
Asian	68	75	83	91
Native American	48	73	83	92
African American/Black	37	48	56	76
Hispanic	29	41	50	71
Source: Massachusetts Department of Education				

Students re-testing take special re-tests, which have a maximum scaled score of 23, are clearly constructed differently to the regular tests which have a maximum scaled score of 280. This at least begs the question, is the MCAS being adjusted in the light of the political realities and has the question, once more, become:

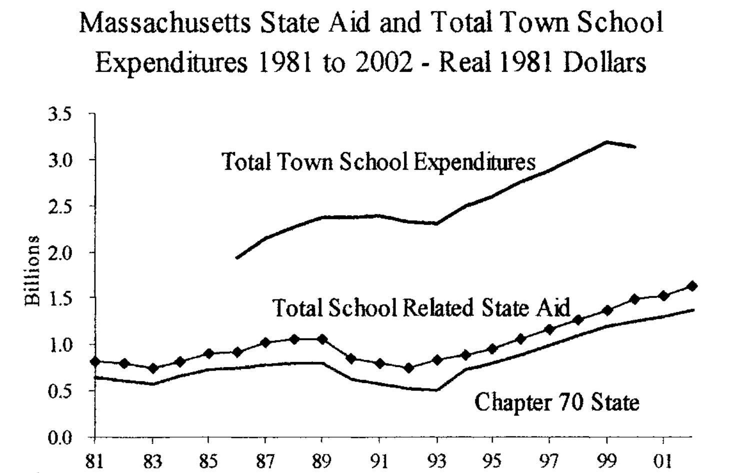
> 'How many children can we afford to fail?' in terms financial, remedial education etc and in terms of public relations.' – Baron and Sergi (1979).

If MCAS has succumbed to politics it will provide a flimsy foundation for any research into the changing state of education in Massachusetts.

2.6 Implementation – State Aid

There is little doubt that State Aid under Chapter 70 and total State Aid to education increased in every year after MERA. Chapter 70 aid increased: by 46.9 percent between 1993 and 1994, by between 11.9 and 14.3 percent per annum to 1999 and then by 9.2 percent between 1999 and 2000, the beginning of the current budget crisis. Figure 2.01 shows the real increase in both Total Town School Expenditures and State aid.

Figure 2.01 – Inflation Adjusted Total Town School Expenditures and State Aid For Education in Massachusetts 1981-2001.



Source: Massachusetts Department of Revenue, Cherry Sheets
Using: U.S. Dept. of Commerce, Bureau of Economic Analysis, State and Local Government Implicit Price Deflator.

Statewide, enrollment grew by only between 1.84 and 2.85 percent per annum in the same period, according to Department of Education analysis of Foundation Enrollment, so the increase in expenditures resulted in a per pupil increase in expenditure.

MERA proposed that a town's minimum contribution to meeting Foundation Budget levels of expenditure would be no more than $10 per thousand dollars of residential property value. By 2000, State Aid made up the difference between property taxes of, on average, $8.95 per thousand of residential property value and a town's Foundation Budget. State Aid was redistributive towards income poorer towns. For the poorest quartile of towns by average incomes, meeting the Foundation Budget without State Aid would have required property taxes of, on average, $22.30 per thousand dollars of residential property value. With State Aid the "Minimum Contribution" was, on average, $8.20 per thousand dollars of residential property value – see Table 2.03.

Table 2.03 – The Effect of State Aid and Minimum Contributions on Town Budgets - 2000. Levy in dollars per thousand dollars of property valuation.

Quartile by Average Income	Average of Average Income	Levy to Meet Foundation Budget **With** State Aid	Levy to Meet Foundation Budget **Without** State Aid
Lowest	37,635	8.2	22.3
2nd	44,906	9.4	19.0
3rd	51,727	9.2	14.8
Highest	73,640	8.0	9.0

Source: Massachusetts Department of Revenue and Massachusetts Department of Education and Author's Calculations.

The Foundation Budget for K-12 Education and State Aid Formulae placed towns, in Massachusetts, on a more or less equal footing. As we have seen, Foundation Budgets were used in State Aid calculations, but EMAB concluded that the Foundation Budget was not used to develop the school district budget.

2.7 Did Per Pupil Expenditure Become More Equal?

So the question then becomes, did this increase in expenditure overall result in the leveling up of the poorer towns as the Act intended?

The Integrated Total Cost of Schools[17] for each of town or city was divided by the Enrollment for each municipality, for the school years 1994 to 2000. No adjustment is made for inflation. Two small towns, Gosnold (1994 enrollment of 3 students) and Rowe (1994 enrollment of 31 students) were excluded. The statistics suggest that nothing

[17] See Section 3 of Chapter 4 for a definition of "Integrated Total Cost".

changed in the distribution of expenditures other than an increase in the average – refer to Table 2.04.

Table 2.04 - Massachusetts' Statewide Per Pupil Expenditure, 1994 to 2000. Excluding Gosnold and Rowe.

	1994	1995	1996	1997	1998	1999	2000
Minimum	3,934	4,246	4,442	4,673	4,900	3,818	5,873
Maximum	10,008	12,411	12,616	11,474	13,094	14,199	16,547
Median	5,274	5,657	5,934	6,238	6,574	7,098	7,550
Average	5,494	5,908	6,187	6,529	6,853	7,384	7,869
Standard Deviation	966	1,046	1,010	1,074	1,084	1,194	1,394
Std. Dev. /Average	18.0%	18.0%	16.0%	16.0%	16.0%	16.0%	18.0%
(Ave-Median) /Ave	4.0%	4.3%	4.1%	4.5%	4.1%	3.9%	4.1%

Source: Massachusetts Department of Education and Author's calculations.

However, looking at the towns and cities individually produces a slightly different picture, one in which there is movement in the rankings within an unchanged, on average, distribution. The towns were categorized according to whether they increased their expenditure between 1994 and 2000, when the increase was measured as a percentage of the average for all the towns and cities. Those towns that did were labelled "Advancers" and those that did not "Retreaters". For example if Northampton spent 120 percent of Massachusetts' towns average per pupil spending in 1994 and 80 percent of Massachusetts' towns average per pupil spending in 2000 it would be a "Retreater".

The towns were split into four sub-groups: those that were above the average in 1994 and continued above the average in 2000; those that were above the average in 1994 and fell to below the average in 2000; those that were below the average in 1994 and continued below the average in 2000; and those that were below the average in 1994 and rose to be above the average in 2000. Given the nature of the first split Advancers / Retreaters this gave six categories.

2.8 Enrollment Growth

The advance of "Advancers" and the retreat of "Retreaters" are related to trends in enrollment. Average enrollment growth for all the advancers between 1994 and 2000 was 5.65 percent as against 14.96 percent for all of the retreaters. Those that did best, the advancers above average in 1994 and above in 2000, saw enrollments shrinking on average by 3.84 percent between 1994 and 2000 and those who did next best saw positive growth of only 3.70 percent, compared with 16.64 percent growth in enrollments for the worst performers – see Table 2.05.

Table 2.05 - Enrollment Growth by Change in Town Per Pupil Expenditure Ranking 1994-2000.

Schools Status	Percent Average Enrollment Growth
All "Advancers"	5.65
Above in 1994 & above in 2000	-3.84
Below in 1994 & above in 2000	3.70
Below in 1994 & below in 2000	10.79
All "Retreaters"	14.96
Above in 1994 & above in 2000	16.80
Above in 1994 & below in 2000	8.78
Below in 1994 & below in 2000	16.64

Source: Massachusetts Department of Education and Author's calculations.

This analysis suggests that the pressure of rising enrollments is the main reason for a town's per pupil expenditure to retreat with respect to the average of all towns and cities in the Commonwealth, rather than the impacts of Proposition 2½ or indeed of the Foundation Budget. It makes sense in terms of human nature that it would be harder to get a budget increase when enrollment is increasing than it would be to keep a stable budget even though enrollments were decreasing.

The increase in the average expenditure from $5,494 in 1994 to $7,869 in 2000 owes more to the improving and then booming economy, than to any acts of the Legislature.

2.9 Conclusions

MERA evolved from a political process that married Equity to Standards reform. Implementation of the Equity components, increased State Aid and a maximum Minimum Contribution, together with a buoyant economy led to a higher basic level of funding per pupil. The rankings by expenditure per pupil appear to be more sensitive to growth in enrollment than to the impacts of Foundation Budgets or State Aid formulae.

Implementation of Standards has yet to really begin. Ten years of Curriculum development has meant that only two of the seven disciplines has been tested each year and also resulted in testing starting late. Competency Determination, as a condition for graduation, also began late and there is evidence that the tests and re-tests were adjusted to make failure less easy. Apart from Professional Development the Procedural / Administrative elements of reform have been largely ignored by school districts.

The next chapter looks at the data available and at the selection of a sample of 180 School Districts used in the analysis.

CHAPTER 3

THE DATA AND SAMPLE USED

The objective of this dissertation is to test the following propositions:

1. Educational opportunity became more equal as a consequence of MERA.
2. The degree to which socio-economic status is a determinant of educational outcomes in Massachusetts has decreased as a consequence of MERA.
3. Education standards have been raised and educational outcomes have improved as a consequence of MERA.

Educational opportunity is measured by expenditure per pupil. Socio-Economic Status is assessed using Education, Median Income, Poverty and a composite index known as the TSEI2. Education standards are measured by the results of two sets of standardized tests: scores from MEAP and scores from MCAS. The largest possible sample of school districts and towns was selected to address Factor 3 in DEA efficiency evaluation – refer to Section 8 of Chapter 4.

3.1 The Sample School Districts and Towns

In Massachusetts, in theory, each of the 351 Cities and Towns is a school district. In practice, there are School Districts that are too small to actually be operative, and there are Regional School Districts in which Towns pool their students either at all levels or for particular grades such as High School, or High School and Middle School. Adding to the complexity are Regional Vocational and Agricultural School Districts, School Choice, Private Schools and the emergence since the mid 1990's of a growing number of Charter Schools.

In the period from 1987 to 2000 approximately 10 percent of the K-12 population in Massachusetts were in Private Schools. About 75 percent were in local schools, around 10 percent were in Academic Regional School districts and about 2 percent were in Vocational Regional Schools – see Table 3.01.

Table 3.01 - K-12 Enrollment in Massachusetts 1988 to 2001.

Year	Total	Local		Academic Regional		Vocational Regional		Private	
1987	944,644	711,006	75.3%	76,792	8.1%	23,330	2.5%	119,825	12.7%
1988	938,016	713,509	76.1%	72,538	7.7%	21,598	2.3%	115,461	12.3%
1989	935,561	712,926	76.2%	74,151	7.9%	21,914	2.3%	113,794	12.2%
1990	941,234	721,941	76.7%	73,676	7.8%	21,359	2.3%	111,570	11.9%
1991	944,926	723,714	76.6%	78,197	8.3%	20,819	2.2%	109,518	11.6%
1992	952,040	726,282	76.3%	81,586	8.6%	20,405	2.1%	109,237	11.5%
1993	966,854	742,052	76.7%	80,365	8.3%	20,154	2.1%	109,078	11.3%
1994	980,767	742,612	75.7%	95,947	9.8%	20,925	2.1%	105,922	10.8%
1995	1,003,501	761,428	75.9%	94,251	9.4%	22,102	2.2%	108,543	10.8%
1996	1,017,891	771,064	75.8%	97,970	9.6%	23,007	2.3%	107,474	10.6%
1997	1,034,073	781,479	75.6%	101,796	9.8%	21,859	2.1%	108,249	10.5%
1998	1,039,701	796,248	76.6%	94,009	9.0%	22,460	2.2%	102,972	9.9%
1999	1,038,030	787,252	75.8%	92,533	8.9%	22,793	2.2%	108,336	10.4%
2000	1,061,008	799,133	75.3%	104,624	9.9%	22,119	2.1%	104,751	9.9%

Source: The Massachusetts Department of Education.

Some Towns form Academic Regional Districts for high school students only and some for junior and middle. Vocational Regional districts take high school students and tend to cover many more towns than do the Academic Regions. Given that about 95 percent of students attending public schools are covered by Academic Regions and Local Districts: it made sense to combine towns into their Academic Regions and to aggregate data to achieve this.

For example, Up-Island is a K - 8 region bringing together students from Aquinnah, Chilmark and West Tisbury. Together with Edgartown, Oak Bluffs and Tisbury, these three towns form Marthas Vineyard Regional. For the purposes of the analysis it made sense to treat these six towns as a single region[18]. So for 8th Grade MCAS scores, for example, the Up-Island average MCAS score multiplied by the number of 8th Grade students at Up-Island would be added to Aquinnah's average MCAS score multiplied by the number of 8th Grade students in Aquinnah. Add in the multiple for Chilmark and for West Tisbury and divide by the total number of 8th Grade students at Up-Island, Aquinnah, Chilmark and West Tisbury and the result is an average 8th Grade MCAS score for Marthas Vineyard.

Adjustments were made for vocational regions – see Appendix K. Models were run using data including and excluding vocational regions and the results were similar as to trend so Vocational Regions are not be considered further.

Approximately 3 percent of K-12 students either attend collaborative schools (0.4%) or exercise school choice (2.6% in 2000) – see Table 3.02. At its simplest, school choice involves students from one town opting to go to another. Data for school choice is sparse, so no attempt was made to adjust for it.

From 1995 onwards the data for school choice include Charter Schools, which took 1.3 percent of enrollment in 2000. Charter Schools are concentrated heavily in Boston, Worcester and Springfield. Boston was excluded from the sample partly for this reason. The other adjustments made to test scores to account for Charter Schools are detailed in Appendix D.

---

[18] Because data for the number of students in these towns and for MCAS 4th Grade English Language Arts tests are missing for the years 1999 and 2000 these towns were, in fact, dropped from the analysis.

Table 3.02 - Massachusetts K-12 Enrollment Other Public Number and Percent of Total.							
				School Choice			
Year	Total	Colaboratives		Total		Charter	
1995	1,003,501	3,123	0.3%	14,054	1.4%	2,396	0.2%
1996	1,017,891	3,171	0.3%	15,205	1.5%	5,195	0.5%
1997	1,034,073	3,207	0.3%	17,483	1.7%	6,572	0.6%
1998	1,039,701	3,516	0.3%	20,496	2.0%	9,797	0.9%
1999	1,038,030	3,788	0.4%	23,328	2.2%	12,518	1.2%
2000	1,061,008	3,992	0.4%	27,373	2.6%	13,799	1.3%
Source: The Massachusetts Department of Education.							

The sample is listed in Appendix C. It is indistinguishable from the State when compared on a number of factors – see Appendix M.

3.2 MEAP and MCAS – Reporting and Scores 1988 to 2002

In the late 1980's and early 1990's the Massachusetts Department of Education ("DOE") published "District Data Books" giving MEAP scores at school district level for the State. MCAS scores are published on the DOE Web Pages.

MEAP and MCAS scores are reported on hybrid norm-referenced and criterion-referenced scales: refer to Section 6 of Chapter 1. In each case proficiency levels were set with criteria against which the difficulty of the assessments could be measured and the raw scores would then be scaled onto a curve that followed a normal distribution. MEAP scores were scaled to a State mean of 1300 and a standard deviation of 100.

> After a scale has been established at a given grade level, it is maintained across subsequent MEAP administrations to permit comparisons of school and district performance over time. That is, scaled scores below 1300 indicate a decline in performance from the initial year of testing, and scaled scores above 1300 indicate an improvement in performance. Real educational changes are detected when scaled scores rise or fall at least 50 points. – Massachusetts Department of Education (1996 October).

Scores were not produced for schools with fewer than 20 students tested in a subject and score for schools with fewer than 60 students should be viewed with caution.

MEAP Proficiency Levels were more criterion-referenced in nature and described students' performance in five different bands on a scale that was related to, but not identical to the scaled scores. Criterion-referenced scores are not available at the level of granularity that would make them usable in the analysis undertaken in this research.

For the 2001 MCAS results, the reporting scale was adjusted – refer to Table 3.03. Average scaled scores in 2001 are not directly comparable to scores from previous test administrations.

From the raw to scaled score conversions, given in Table 3.03, it seems that, in 2001, all the tests were deemed to be have a much tougher baseline than in previous years, and thus the bottom end raw scores translated onto higher scaled scores than in the previous three years. Perhaps the large number of raw scores translating onto 200 in the first three years was the problem? If, on the other hand, the scaling changed without a change in the nature of the assessments, then the effect is to increase the average scaled scores, without an underlying improvement in the children assessed.

Table 3.03 - Scaled and Raw MCAS Scores - Grade 10 - English Language Arts - 2001.

Scaled Score	1998	1999	2000	2002	2001
280	84.	72.	72.	72, 71, 70.	71, 70, 69, 68, 67.
278	82, 81.	71, 70.	71, 70, 69.	69.	None.
276	None.	None.	None.	None.	66.
274	79.	69.	68.	68.	65.
272	None.	None.	None.	None.	64.
270	75.	67.	67.	67.	None.
268	74.	66.	66.	66.	63.
266	72.	65.	65.	65.	62.
264	71.	64.	64.	64.	61.
262	69.	63.	63.	63.	60.
260	68.	62.	62.	62.	59.
258	66.	61.	61.	61.	58.
256	65, 64.	60, 59.	60, 59.	60.	57.
254	None.	None.	None.	None.	None.
252	61.	57.	58.	59.	56.
250	60.	56.	57.	58.	55.
248	59.	55.	56.	57.	54.
246	57.	54.	55.	56.	53.
244	56, 55.	53, 52.	54, 53.	55.	52.
242	None.	None.	None.	54, 53.	51.
240	52.	50.	52.	52.	50.
238	51.	49.	51.	None.	49.
236	50, 49.	48, 47.	50, 49.	51.	48.
234	None.	None.	None.	50.	47.
232	None.	None.	None.	49.	None.
230	47, 46.	46, 45.	48, 47.	48.	46.
228	43.	43.	46.	47.	45.
226	42.	42.	45.	46.	44.
224	41, 40.	41, 40.	44, 43.	45.	43.
222	39.	39.	42.	44.	42.
220	37.	38.	41.	43, 42, 41.	41, 40, 39.
218	36.	37.	40.	40, 39, 38, 37, 36.	38, 37, 36, 35.
216	35.	36.	39.	35, 34, 33, 32, 31.	34, 33, 32, 31, 30.
214	34.	35.	38.	30, 29, 28, 27, 26.	29, 28, 27, 26.
212	32.	34.	37.	25, 24, 23.	25, 24, 23, 22.
210	30.	32.	36.	22, 21, 20, 19.	21, 20, 19.
208	29.	31.	35.	18, 17.	18, 17, 16.
206	28.	30.	34.	16, 15, 14.	15, 14.
204	26.	29.	33.	13, 12, 11, 10.	13, 12, 11, 10.
202	25, 24.	28, 27.	32, 31.	9, 8, 7, 6.	9, 8, 7, 6.
200	23, 21, 20, 19, 18, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 3, 2, 1, 0.	26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 5, 4, 3, 2, 0.	30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1,0	5, 4, 3, 2, 1, 0.	5, 4, 3, 2, 1, 0.

Source: Massachusetts Department of Education.

The average score of those scoring below 220 is estimated to be likely to have been increased by up to 10 points. It looks fortuitous, to say the least, that this change should have occurred in the year that Graduation should become dependent on achieving a scaled score of at least 220.

Raw scores are not published for either MEAP or MCAS. Conversion tables exist for the MCAS as we have seen above, but they are many (raw) to one (scaled score) and thus unsuitable for backwards conversion. Raw scores would anyway be dependent on the difficulty of the assessment. Criterion-referenced scores that were consistently scaled and consistently referenced to solid criteria would be ideal for comparisons across time, but MEAP is not scaled consistently with the MCAS and the criteria were different (some of the assessments were of different subjects). The MCAS itself is not scaled consistently and scaled scores would not appear to have been consistently referenced to the criteria.

One way to consider the data is to think of it as coming from at least three different systems. MEAP, MCAS to 2000 and MCAS 2001 and later.

A number of other issues and problems are associated with using the results of the assessments to measure progress in education over time. Looking first at who is being tested. If we make the assumption that the less good students avoid being tested, if they can, then in general MEAP average scores should be higher since Special Education and Other Needs students were exempted from testing and as a consequence a lower percentage of students took MEAP tests than take the MCAS – see Table 3.04.

Table 3.04 - Percentage of Enrolled Students Tested - 1988-2002.										
	MEAP						MCAS			
Grade	1988	1990	1992	1994	1996	1998	1999	2000	2001	2002
3	-	-	-	-	-				97	97
4	90	90	89	90	90	97	96	95	94	96
6	-	-	-	-	-	-	-	-	98	98
7	-	-	-	-	-	-	-	-	94	95
8	90	89	88	89	89	97	96	93	94	97
10	-	-	-	86	85	96	95	93	92	94
Source: Massachusetts Department of Education.										

Grade retention is a way in which schools can avoid their worst students being tested. A record number of 2000-2001 grade nine students were retained (8.4 percent[19]). If we assume that the reason for this is that students who would not do well on the Grade 10 tests being held back, then the effect of retention should be to increase the average scaled scores for 2001.

3.2.1 Sample School Districts Test Scores

It should come as no surprise that average test scores for school districts are highly correlated at a high level of significance (all p-values were less than 0.000)– see Table 3.05.

[19] According to Statistics from the Massachusetts Department of Education.

Table 3.05 - Pearson Correlations between the Rankings of 2000 Test Scores.										
	Grade 4			Grade 8				Grade 10		
Grade 4	ELA	Math	Sci.	ELA	Math	Sci.	Soc.	ELA	Math	Sci.
ELA	1.00									
Math	0.94	1.00								
Science	0.84	0.85	1.00							
Grade 8										
ELA	1.00	0.94	0.84	1.00						
Math	0.82	0.84	0.94	0.82	1.00					
Science	1.00	0.94	0.84	1.00	0.82	1.00				
Social	0.77	0.79	0.89	0.77	0.90	0.77	1.00			
Grade 10										
ELA	0.75	0.75	0.79	0.75	0.81	0.75	0.77	1.00		
Math	0.76	0.78	0.81	0.76	0.85	0.76	0.78	0.92	1.00	
Science	0.71	0.74	0.79	0.71	0.82	0.71	0.78	0.93	0.93	1.00

3.2.2 Normality of Test Scores Distributions

Six sets of test scores for the Sample of 180 School Districts were tested for normality in Minitab. The results are summarized in Table 3.06. Normality plots are given in Appendix E. The results for five of the six sets of scores support the hypothesis that the scores come from Normal distributions at the 0.10, 0.05 and 0.02 significance levels. 1988 12^{th} Grade Mathematics is problematic in that the hypothesis that the data comes from a normal distribution is only supported at a 0.20 significance level.

Table 3.06 - Anderson-Darling Normality Test Results for Six Sets of Sample School Districts' Average Test Scores.		
Test Year, Grade and Subject	A-Squared	P-Value
1988 4th Grade Reading	1.138	0.005
1988 8th Grade Science	0.981	0.013
1988 12th Grade Mathematics	0.513	0.191
2002 4th Grade English Language Arts	1.007	0.012
2002 7th Grade English Language Arts	0.947	0.016
2002 10th Grade Mathematics	0.923	0.019
Note: Anderson-Darling Adjusted Scores From Minitab.		

Notwithstanding the P-Value in the Anderson-Darling Normality test for 1988 8th Grade Mathematics, the Sample School Districts' Average Test scores were accepted as being from normal distributions.

MEAP and MCAS Scale scores were converted to z-scores and then rescaled to a mean of 240 and a standard deviation of 6.8 being approximately the scale of MCAS scores.

3.3 Per Pupil Expenditures

Per pupil expenditures were published by DOE under a number of headings from 1988 until 2000. Consideration of expenditures was therefore limited to that period. Per Pupil Integrated Cost, Per Pupil Special Education Expenditures, Per Pupil Regular Day Expenditures and Per Pupil Expenditures by Grade Level (Elementary, Middle and High School) are the measures used. Each is described in turn in the following sub-sections.

3.3.1 Per Pupil Integrated Cost

The Integrated Cost Per Pupil counts all of the resident public school-children in a city or town, regardless of where they are enrolled. Tuition and other expenditures associated with those educated outside the district are factored in. If a community belongs to one or more regional school districts, those districts' expenditures are apportioned back to the member town in accordance with its share of enrollment. A city or town's integrated cost, therefore, is a composite of spending and pupils for all publicly-funded school children who reside there.[20]

This measure is available for school districts from 1988 to 2000 when the DOE discontinued the calculation.[21] Statewide statistics and trends in this measure of expenditure are discussed in detail in Section 7 of Chapter 2. Statistics for the 180 Sample School Districts are given in Table 3.07. The Standard Deviation measured as a percentage of the Mean has decreased in each period since 1992, which implies greater equity. The Minimum Expenditure as a percentage of the Mean has increased in each period since 1992, which also implies greater equity.

---

[20] http://finance1.doe.mass.edu/statistics/pp01_intcost.html

[21] http://finance1.doe.mass.edu/statistics/pp02_intro.html

Table 3.07 - Trends in Integrated Cost Per Pupil Expenditures (180 Sample School Districts).							
Year	1988	1990	1992	1994	1996	1998	2000
Mean	4,196	4,930	5,007	5,274	5,805	6,398	7,202
Standard Deviation	730	859	928	905	847	893	992
Minimum	2,829	3,304	3,222	3,959	4,285	4,673	5,342
Maximum	6,735	8,013	8,381	8,896	9,550	9,863	11,715
Std. Dev. as Percent Of Mean	17.4	17.4	18.5	17.2	14.6	14.0	13.8
Min as Percent of Mean	67.4	67.0	64.4	75.1	73.8	73.0	74.2
Max as Percent of Mean	160.5	162.5	167.4	168.7	164.5	154.2	162.7
Percentage Growth In Mean		17.5	1.6	5.3	10.1	10.2	12.6
Source: Massachusetts Department of Education and Author's Calculations.							

3.3.2 Per Pupil Expenditures By Grade Level

Data summarizing school district expenditure per pupil by grade level is available for the period from 1988 to 2000. The grade levels summarized are "Elementary", "Middle" and "High". This data serves to show how school districts differ in their emphasis on expenditure as the children progress through grades.

Table 3.08 - Trends in Per Pupil Elementary School Expenditure (180 Sample School Districts).							
Year	1988	1990	1992	1994	1996	1998	2000
Mean	3,243	3,727	3,778	4,039	4,375	4,840	5,528
Standard Deviation	622	718	788	741	682	703	861
Minimum	2,174	2,622	1,927	2,845	3,101	3,626	3,979
Maximum	5,662	5,729	5,974	6,528	7,064	7,834	9,830
Std. Dev. as Percent Of Mean	19.2	19.3	20.9	18.3	15.6	14.5	15.6
Min as Percent of Mean	67.0	70.3	51.0	70.4	70.9	74.9	72.0
Max as Percent of Mean	174.6	153.7	158.1	161.6	161.4	161.9	177.8
Percentage Growth In Mean		14.9	1.3	6.9	8.3	10.6	14.2
Source: Massachusetts Department of Education and Author's Calculations.							

Average per pupil expenditure at the Elementary School level, up by 46.3 percent between 1992 and 2000, has grown more strongly than per pupil expenditure at the Middle School level (up 32.3 percent) and at the High School level (up 32.0 percent). The minimum expenditure level in 1992 looks to be an anomaly – refer to Table 3.08: but taking the Standard Deviation as a Percent of Mean as a better indicator of the change in equity points to greater equity being seen over time.

Table 3.09 - Trends in Per Pupil Middle School Expenditure (180 Sample School Districts).

Year	1988	1990	1992	1994	1996	1998	2000
Mean	4,076	4,678	4,667	4,829	5,171	5,510	6,177
Standard Deviation	890	1,151	1,193	1,102	1,038	1,072	1,491
Minimum	2,493	2,486	2,441	3,235	3,171	3,296	3,822
Maximum	8,493	10,142	9,506	11,164	8,714	9,663	16,366
Std. Dev. as Percent Of Mean	21.8	24.6	25.6	22.8	20.1	19.5	24.1
Min as Percent of Mean	61.2	53.1	52.3	67.0	61.3	59.8	61.9
Max as Percent of Mean	208.4	216.8	203.7	231.2	168.5	175.4	265.0
Percentage Growth In Mean		14.8	-0.2	3.5	7.1	6.5	12.1

Source: Massachusetts Department of Education and Author's Calculations.

On average more money is spent per Middle School Pupil than per Elementary School Pupil, with most money being spent per High School Pupil. Measured by the Standard Deviation as a Percent of the Mean, it would appear that equity at the Middle School level was monotonically greater in each year after 1992, except for 2000 when the trend appears to have reversed itself – refer to Table 3.09.

Table 3.10 - Trends in Per Pupil High School Expenditure (180 Sample School Districts).

Year	1988	1990	1992	1994	1996	1998	2000
Mean	4,305	5,222	5,245	5,565	5,876	6,247	6,922
Standard Deviation	761	1,063	1,154	1,151	1,088	1,106	1,324
Minimum	2,684	2,570	2,700	3,326	3,286	3,826	4,675
Maximum	6,887	9,671	9,321	10,187	10,183	10,035	16,366
Std. Dev. as Percent Of Mean	17.7	20.3	22.0	20.7	18.5	17.7	19.1
Min as Percent of Mean	62.3	49.2	51.5	59.8	55.9	61.2	67.5
Max as Percent of Mean	160.0	185.2	177.7	183.1	173.3	160.6	236.4
Percentage Growth In Mean		21.3	0.4	6.1	5.6	6.3	10.8

Source: Massachusetts Department of Education and Author's Calculations.

The trends in expenditure per High School pupil – see Table 3.10 – are similar to those seen for Middle School pupils.

3.3.3 Per Pupil Expenditures By Program

Data summarizing school district expenditure by program consist of per pupil expenditures for Special Education, Bilingual Education, Occupational Education and Regular Day. Only Special Education and Regular Day per pupil expenditures are available, consistently, for the period from 1988-2000.

Table 3.11 - Trends in Regular Day Per Pupil Expenditure 180 Sample School Districts).							
Year	1988	1990	1992	1994	1996	1998	2000
Mean	3,592	4,149	4,134	4,408	4,722	5,176	5,820
Standard Deviation	629	766	802	794	746	751	887
Minimum	2,509	2,715	2,615	2,969	3,023	3,986	4,211
Maximum	5,616	6,524	6,745	7,452	7,944	8,533	10,875
Std. Dev. as Percent Of Mean	17.5	18.5	19.4	18.0	15.8	14.5	15.2
Min as Percent of Mean	69.9	65.4	63.3	67.4	64.0	77.0	72.4
Max as Percent of Mean	156.4	157.2	163.2	169.1	168.3	164.8	186.8
Percentage Growth In Mean		15.5	-0.4	6.6	7.1	9.6	12.4
Source: Massachusetts Department of Education and Author's Calculations.							

The results of analysis of Regular Day Per Pupil Expenditures – see Table 3.11 – are similar to those for Per Pupil Integrated Costs.

Table 3.12 - Trends in Per Pupil Special Education Expenditure (180 Sample School Districts).							
Year	1988	1990	1992	1994	1996	1998	2000
Mean	5,603	6,600	6,957	7,648	8,585	9,555	11,172
Standard Deviation	965	1,265	1,518	1,785	1,628	1,799	2,028
Minimum	3,326	4,013	3,481	5,033	3,832	5,023	7,563
Maximum	8,932	13,615	14,049	19,842	15,741	18,174	18,518
Std. Dev. as Percent Of Mean	17.2	19.2	21.8	23.3	19.0	18.8	18.2
Min as Percent of Mean	59.4	60.8	50.0	65.8	44.6	52.6	67.7
Max as Percent of Mean	159.4	206.3	201.9	259.4	183.4	190.2	165.8
Percentage Growth In Mean		17.8	5.4	9.9	12.3	11.3	16.9
Source: Massachusetts Department of Education and Author's Calculations.							

On average more money is spent per pupil in Special Education than per pupil in any other category – see Table 3.12. Measured by the Standard Deviation as a Percent of the Mean: it would appear that equity in Special Education Expenditure per pupil was greater in each year after 1992.

3.4 Socio-Economic Status

Hauser and Warren (1997) describe Socio-Economic Status as follows:

> Socioeconomic status is typically used as a shorthand expression for variables that characterize the placement of persons, families, households, census tracts, or other aggregates with respect to the capacity to create or consume goods that are valued in our society.

Using various search engines to search for "Socio Economic Status" or "SES" a list of 78 papers was compiled. The survey is both random in the sense that no judgment was applied in the choice of papers and non-random in the sense that it depended on what was in the search engines and readily available. Survey results are given in Appendix F.

Table 3.13 lists the most popular variables and the percentage of the 79 papers that referenced them. Education of the parents was the most used measure of Socio Economic Status in cases where the child's SES was being estimated. Occupation is the next most used measure with a number of Occupational Status / Prestige Indices being used. The index known as TSEI2, updated to cover 1990 Census Occupation categories was used as a scale for measurement of Occupations. Appendix G contains a short discussion of the various Occupational Status / Prestige Indices identified from the survey.

Table 3.13 - Popular Measures Of Socio-Economic Status.

Measure	Percent Using Measure
Education Father	68
Education Mother	68
Occupation Father	68
Occupation Mother	56
Family Income	51
Number of Parents	5
Number of Siblings	6
Number of Books in Home	4
Housing Tenure	10
Crime Rate	3
Poverty Rate	6
Population Density	1
Housing Density	3
Percent Urban	3

Source: Author's Survey.

The last 6 items in Table 3.13: Housing Tenure, Crime Rate, Poverty Rate, Population Density, Housing Density and Percentage Urban are infrequently used. This reflects the fact that most of the studies are concerned with measuring an individual's Socio Economic Status rather than that of a community. Taking this into consideration it becomes apparent that Housing is a variable frequently used to measure a community's Socio-Economic Status as are Poverty and Crime Rates.

Other indicators or measures of a community's SES that have been used include: Health[22], Drop Out Rates, Ethnicity, Divorce Rates, Car Ownership, Wealth, Unemployment Rates, Length of Service and Commuting Distance.

Education, Median Income, Occupation and Poverty were selected as the proxies for Socio-Economic Status. Each is discussed in turn in the following sub-sections.

3.4.1 Education

The 1980, 1990 and 2000 Censuses record the number of persons achieving the categories of education summarized in Table 3.14.

Taking the number of persons in each category, multiplying by the number of "years" and dividing the sum by the total number of people gives an index of a town's education "years".

For Massachusetts the resulting number of years of education are 12.65 years for 1980, 13.17 years for 1990 and 13.64 years for 2000. The top three towns in 2000 were Carlisle (16.87 years), Weston (16.56 years) and Dover (16.44 years). The bottom four towns were Chelsea (11.57 years), Lawrence (11.51 years), New Bedford (11.50 years) and Fall River (11.46 years).

---

[22] Oakes and Rossi (2003) trace a strong relationship between SES and health dating back to ancient Greece.

1980 Categories	Years*	1990 and 2000 Categories	Years*
Elementary (0 to 8 years) through High School 1 to 3 years	9	Less than 9th grade	8
		9th to 12th grade, no diploma	10
High School 4 years	12	High school graduate (includes equivalency)	12
		Some college, no degree	13
1 to 3 years College	14	Associate degree	14
4 years College	16	Bachelor's degree	16
5 or more years College	20	Graduate or professional degree	20

Table 3.14 - 1980, 1990 and 2000 Decennial Census Education Categories of Education.

Source: 1980, 1990, 2000 Decennial Census and Author's Calculations.

Note * Number of years used in calculations.

3.4.2 Median Incomes

The 1980, 1990 and 2000 Censuses record the median household income at the town level. The top ten towns in each year are given in Table 3.15.

There has been relatively little movement in the rankings with 7 towns, Weston, Sherborn, Dover, Carlisle, Sudbury, Boxford and Wellesley remaining in the top ten throughout.

Table 3.15 - Top Ten Towns In Massachusetts by Median Income, 1980, 1990 and 2000 Decennial Census.			
Rank	1980	1990	2000
1	Weston	Weston	Weston
2	Sherborn	Sherborn	Dover
3	Dover	Dover	Carlisle
4	Carlisle	Carlisle	Sherborn
5	Sudbury	Wellesley	Sudbury
6	Boxford	Sudbury	Wellesley
7	Wayland	Boxford	Boxford
8	Wellesley	Wayland	Harvard
9	Longmeadow	Concord	Southborough
10	Lexington	Lexington	Bolton
Source: 1980, 1990, 2000 Decennial Census			

The bottom ten towns in each year are given in Table 3.16. There has been relatively little movement at the bottom of the rankings with 5 towns, Holyoke, Lawrence, New Bedford, Fall River and Chelsea in the bottom ten in each Census year.

Table 3.16 - Bottom Ten Towns In Massachusetts by Median Income, 1980, 1990 and 2000 Decennial Census.			
Rank	1980	1990	2000
342	Holyoke	Chelsea	Adams
343	Oak Bluffs	Adams	Holyoke
344	Lawrence	Wellfleet	Springfield
345	New Bedford	Holyoke	Chelsea
346	Wendell	New Bedford	Fall River
347	Sunderland	Fall River	Lawrence
348	Fall River	Lawrence	North Adams
349	Chelsea	North Adams	New Bedford
350	Provincetown	Provincetown	Monroe
351	Aquinnah	Aquinnah	Gosnold
Source: 1980, 1990, 2000 Decennial Census			

There was little movement in the rankings by median income over two decades. The Pearson Correlations between the Median Incomes for 1980 with 1990 and 1990 with 2000 were 0.93 and 0.95 with p-values < 0.000.

The relative levels of education across the three censuses were also stable with Pearson Correlations between the "years" of Education for 1980 with 1990 and 1990 with 2000 being 0.95 and 0.95 with p-values < 0.000.

The relationship between Median Incomes and Education was less strong. Correlations between the factors for 1980, 1990 and 2000 were 0.59, 0.62 and 0.65 respectively with p-values < 0.000.

3.4.3 TSEI2 – An Occupational Index of SES

Occupational and Prestige indexes were identified by the survey of SES literature as proxies for SES. TSEI2 was selected for use in this research and updated to cover the 2000 Census Occupational Classifications – see Appendix H.

Massachusetts as a whole scored 34.29 in 1980, 35.49 in 1990, and 37.29 in 2000 on the TSEI. The standard deviation of the 351 scores was 2.91 for 1980, 2.79 for 1990, and 2.55 for 2000. The Pearson Correlations between the values in 1980 and 1990 and between 1990 and 2000 are 0.91 and 0.87 respectively with P-Value of 0.000 in each case. The Pearson Correlations between the rankings (1 to 351) are similar to those for the values, being 0.89 and 0.87 respectively with a p-value less than 0.000 in each case.

Although the index for 2000 was based on some assignments and averaging of categories and scores; the results are not inconsistent with the indices from 1980 and 1990.

Given that the Occupational Prestige / Status Indices combine income and education data to derive a score, a high level of correlation between the TSEI Indices and the income and education data was to be expected. Pearson Correlations of the TSEI Indices with Median Incomes were 0.90, 0.89 and 0.84 and with Education "Years" they were 0.71, 0.73 and 0.71 for the 1980, 1990 and 2000 data respectively with p-values < 0.000.

3.4.4 Poverty

For Massachusetts at the town level, the 1980, 1990 and 2000 Census present numbers of persons whose earnings are in 5 or can be aggregated to 5 bands of percentages of the Federal Poverty level. The bands are: below 75 percent of poverty level; between 75 and 124 percent of poverty level; between 125 and 149 percent of poverty level; between 150 and 199 percent of poverty level; and 200 percent of poverty level and above.

Taking the numbers of persons in each band and multiplying them by 0.375; 1.0; 1.375; 1.75 and 2.0 respectively and dividing the sum of the result by the total number of persons in the bands gives an index measure of poverty for 1980, 1990 and 2000 for each town in Massachusetts. The smaller the value of the index measure the more severe the poverty.

For Massachusetts as a whole the Poverty Index was 1.77 in 1980, 1.81 in 1990 and 1.80. Measured by the Poverty Index, Chelsea, Holyoke and Lawrence were the most poverty stricken towns in the Commonwealth in 2000 and Boxford, Norfolk and Topsfield were the three towns least affected.

The Poverty Indices are consistent across the three decades. The Pearson Correlation between the 1980 Poverty Index and the 1990 Poverty Index is 0.83, while that between the 1990 and 2000 Poverty Indices is 0.87. When considering the Town's Poverty Ranking the Correlations are 0.84 and 0.87, respectively. All the P-Values were less than 0.000.

The Poverty Indices are less consistent with the Town TSEI Indices. The Pearson Correlation between the 1980 Poverty Index and the 1980 Town TSEI is 0.55. For 1990 and 2000 the corresponding correlations are 0.55 and 0.54. Again all the p-values were less than 0.000.

There is a high degree of correlation between the Poverty Index and Median Incomes – 0.91, 0.90 and 0.89 for 1980, 1990 and 2000 respectively with p-values less than 0.000.

CHAPTER 4

EFFICIENCY AND DATA ENVELOPMENT ANALYSIS

This chapter has four objectives:

1. To justify – Section 1 – the use, in the analysis, of only one DEA model in its plainest and most simple version, when there are a very large number of different Data Envelopment Analysis models and model variants to choose from – refer to Appendix A,

2. To describe how efficiency is measured by the model chosen – the CCR model – and to describe the model itself – Sections 2, 3 and 4,

3. Through a series of simple examples to illustrate how structural properties in the data impact on the measurement of efficiency – Sections 5, 6 and 7, and

4. To identify how measurement of efficiency can be used to identify structural properties in the data such as correlation and dispersion – Section 8.

4.1 The Choice of Model

The CCR model was the first DEA model – Charnes et al. (1978). It was selected for use in this dissertation because the results from the model were not qualitatively distinct from the results obtained using three other DEA models. Price ratio constraints – see Section 3 of Appendix A – were tested and found to have no impact on the analysis. Models and model variants that allow for measurement of returns to scale were not used because the data – see Chapter 3 – are all scaled either to quantities per person or quantities per town.

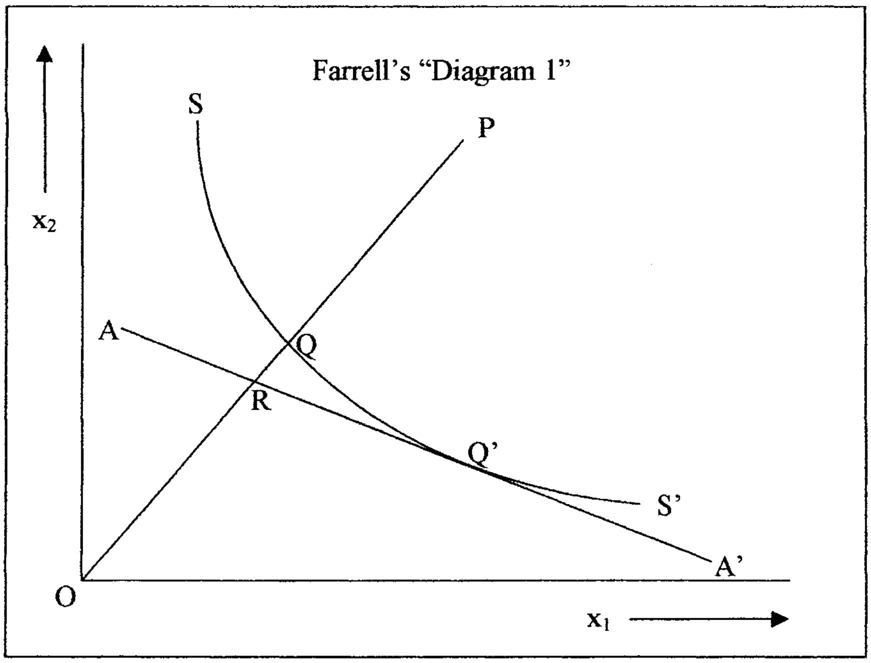
Since CCR model measurements of efficiency are the core of the analysis, an understanding of the nature of these measurements is key to an understanding of the analysis.

4.2 Farrell's Radial Measurement of Efficiency

Koopmans (1951) modeled an entire economy and provided a definition of a technically efficient production frontier. Technical efficiency is achieved when a producer is able to select the prices of inputs and outputs in such a way as to make a zero profit as all other producers make a zero profit or a loss. Taking the frontier and all points behind the frontier gives a production possibilities set – the union of the technically efficient DMUs and the technically inefficient DMUs.

Farrell (1957) was concerned with measurement of the difference between efficient and inefficient points. Farrell's "Diagram 1" is reproduced as Figure 4.01, with the modification that the axes representing the quantities of the primary factors have been labeled as x_1 and x_2 to conform with the notation used in this dissertation and set out in Section 3 of this Chapter. Primary factors are the raw materials: inputs to the production process.

Figure 4.01 – Farrell's "Diagram 1".



SS' is an isoquant representing combinations of the primary factors x_1 and x_2 required to produce a single output y_1 at a given level, q, with 100 percent technical efficiency.

AA' is a market price line whose slope is the ratio of the market prices of the primary factors. Production of quantity, q, of the output at Q' (where AA' is tangent to SS') is both technically efficient and price efficient. Q' represents the combination of the primary factors which results in the production of quantity, q, at the lowest cost.

A producer operating at P is both technically inefficient and price inefficient at the market prices implicit in AA'. P can improve its technical efficiency by reducing its usage of either or both of the primary factors. Assuming proportionate reductions in each primary factor, P would operate technically efficiently at Q. However to be as cost efficient as Q' at the market prices implied by AA' would require P to be operating at R if it were possible for P to further proportionately reduce its use of the primary factors.

Alternatively, P, operating at Q would be price efficient if AA' were tangent to SS' at Q. If a producer on SS' is allowed to select its own "market prices" such that AA' is tangent to SS' where the producer sits on SS' then the producer will be both 100 percent Technically and 100 percent Price Efficient. Put in another way, if a producer is allowed to choose its own market prices and it can achieve 100 percent Price Efficiency at those prices, then it is 100 percent Technically Efficient.

Farrell takes the ratio OR/OQ as the measure of Price Efficiency of P and OQ/OP as the measure of Technical Efficiency of P. The Overall Efficiency of P is the Price Efficiency * Technical Efficiency = OR/OQ * OQ/OP = OR/OP. Both measures are known as "radial" measures since they are derived from the ratio of the lengths of radii – lines from the geometric origin.

Reduction from P's usage to Q's usage involves reductions in x_2 and x_1 proportional to their actual usage, since PQO is a straight line. This quality is known as "equi-proportional reduction".

Actual examples of the production of output y_1 from primary factors x_1 and x_2 would be plotted as a series of points, rather than a continuous curve. Drawing a set of lines between a subset of the points, such that no point was closer to the origin than the lines, would give an estimation of SS'. Inefficient producers can then derive an estimated technical efficiency score by reference to the estimate of SS'.

Farrell addressed a single-output / dual-input case. As we shall see, the CCR model generalized this concept to multiple inputs and multiple outputs, but first the notation will be formalized in the context of Farrell and generalized.

4.3 Notation

A producer of outputs from primary factors will be known as a Decision Making Unit ("DMU").

The number of DMUs will be represented by n.

The subscript j will be used for DMUs numbered from 1 to n.

Output will be generalized and more than one output will be considered. The number of outputs will be represented by s.

The subscript r will be used for outputs numbered from 1 to s.

The quantities of each output used will be represented by the parameter y, so y_{rj} will represent the quantity of the rth output produced by the jth DMU.

Each DMU is allowed to choose the prices that result in its Price Efficiency being equal to its Technical Efficiency.

The output prices for a DMU are represented by u, which like the outputs themselves are indexed by r and by j, thus: μ_{rj}

The sum of the priced outputs will therefore be: $\sum_{r=1}^{s} \mu_{rj} y_{rj}$ for any given DMU, j.

In other contexts this might be thought of as a weighted average of the outputs.

The number of inputs will be represented by m. The subscript i will be used for inputs numbered 1 to m. Inputs will be represented by the parameter x, so x_{ij} will represent the quantity of the ith input used by the jth DMU.

The input prices for a DMU are represented by v, which, like the inputs themselves, are indexed by i and by j, thus: v_{ij}

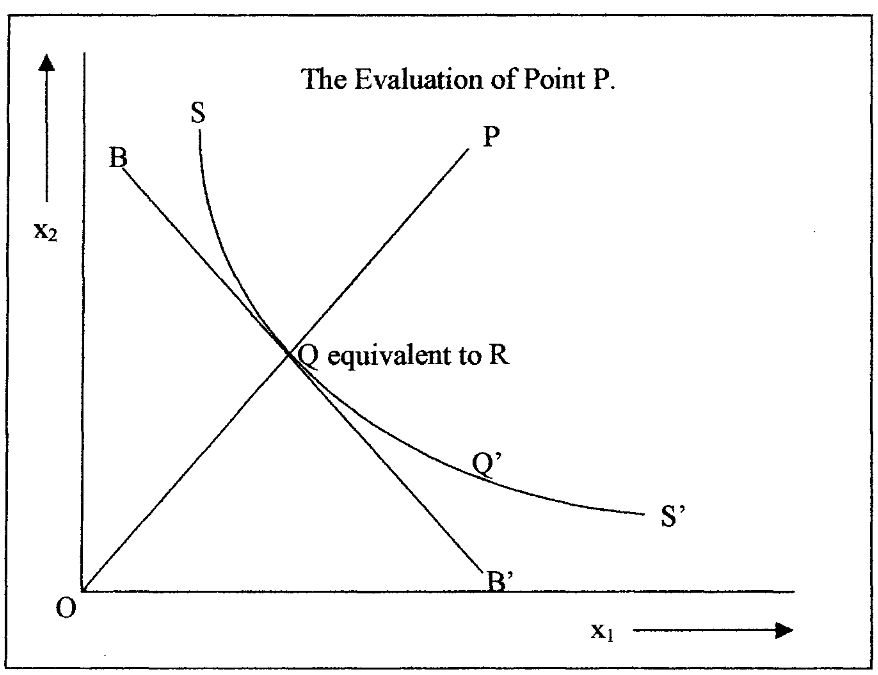
The sum of the priced inputs will therefore be: $\sum_{i=1}^{m} v_{ij}x_{ij}$ for any given DMU, j.

Each DMU is evaluated separately, so when the parameter or price is specifically the parameter or price for the DMU under consideration, the subscript "0" replaces "j".

4.4 The CCR Model

In Section 2, point P was evaluated at prices (AA') that allowed DMU Q' on SS' to be both Technically and Price Efficient. Prices that would allow P to operate as efficiently as Q – if P managed to reduce input usage of x_1 and x_2 proportionately to be operating at the same point as Q – are given by BB' the line tangent to SS' at Q – See Figure 4.02. Q, the Technically Efficient point is also R, from Figure 4.01, the Price Efficient point at which P would operate: if it could.

Figure 4.02 – The Evaluation of Point P.



Koopmans (1951) defined efficiency as arising when the sum of the Priced Outputs equals the sum of the Priced Inputs. The ratio is one for an efficient DMU. Using the notation described in Section 2 and substituting Q for the subscript j we can express the efficiency of Q, an efficient DMU, as: $h_Q = \frac{\sum_{i=1}^{m} v_{iQ} x_{iQ}}{\sum_{r=1}^{s} u_{rQ} y_{rQ}} = 1$. This expression of efficiency allows for up to s outputs: only one output is assumed in Farrell's model.

Evaluation of P at the market prices for Q gives P an efficiency[23] of:

$$h_P = \frac{\sum_{i=1}^{m} v_{iQ} x_{iP}}{\sum_{r=1}^{s} u_{rQ} y_{rP}} \geq 1$$

The Technical Efficiency and Price Efficiency of P, OP/OQ, is equivalent to h_P/h_Q. So, if both P and Q are evaluated at the "best" prices for them: P will be judged inefficient in relation to Q to the extent OP/OQ.

In order to solve for the efficiency of all of the DMUs requires a solution for *each* DMU of two problems:

---

[23] or we can take the reciprocal $\frac{1}{h_P} = \frac{\sum_{r=1}^{s} u_{rQ} y_{rP}}{\sum_{i=1}^{m} v_{iQ} x_{iP}} \leq 1$ which has the intuitively pleasing property of making 1 equivalent to full efficiency and numbers less than 1 and greater than zero represent less than full efficiency. The average of a set of reciprocals is equivalent to the reciprocal of the average of a set of values, so representing efficiency and average efficiencies from the CCR model on a scale from zero to one is simply a matter of calculating an average and taking the reciprocal.

1. What prices allow the DMU being evaluated to achieve the highest level of efficiency relative to all the other DMUs?

2. What is an Isoquant when there is more than one output?

Each DMU, in turn, is evaluated against its own Isoquant, which is to say at prices that allow the sum of the priced inputs to be equal to 1. $\sum_{r=1}^{s} u_{r0} y_{r0} = 1$. The DMU being evaluated then compares itself to the other DMUs at this Isoquant.

If we let the sum of the Priced Outputs for any given DMU, 0, be equal to 1:

$\sum_{r=1}^{s} u_{r0} y_{r0} = 1$, then DMU 0 is efficient if: $h_0 = \frac{\sum_{i=1}^{m} v_{i0} x_{i0}}{\sum_{r=1}^{s} u_{r0} y_{r0}} = \frac{\sum_{i=1}^{m} v_{i0} x_{i0}}{1} = \sum_{i=1}^{m} v_{i0} x_{i0} = 1$

but DMU, 0, is inefficient if: $\sum_{i=1}^{m} v_{i0} x_{i0} = h_0 > 1$.

For any given DMU, 0, minimizing $\sum_{i=1}^{m} v_{i0} x_{i0} = h_0$, while at the same time ensuring that:

1. no other DMU can do better than DMU, 0, at the prices for DMU, 0, and

2. $\sum_{r=1}^{s} u_{r0} y_{r0} = 1$

will yield the prices for DMU, 0, that allow it to achieve its highest level of efficiency relative to the other DMUS and will allow calculation of an efficiency score. This gives an algorithm that can be applied to the n DMUs as follows:

For 0 from 1 to n, solve:

$$\min h_0 = \sum_{i=1}^{m} v_{i0} x_{i0}$$

$s.t.$

(1) $-\sum_{r=1}^{s} \mu_{r0} y_{rj} + \sum_{i=1}^{m} v_{i0} x_{ij} \geq 0 \quad \forall j$

(2) $\sum_{r=1}^{s} \mu_{r0} y_{r0} = 1$

(3) $\mu_{r0}, v_{i0} \geq 0$

(4) $y_{rj}, x_{ij} \geq 0 \quad \forall i, r$ and j

The result is an efficiency score h_0 for each DMU and a set of input prices and a set of output prices for each DMU.

Consider P and Q in this framework. There are two DMUs: so n = 2. For Q one first evaluates Q against P at Q's prices:

$$\min h_Q = \sum_{i=1}^{m} v_{iQ} x_{iQ}$$

$s.t.$

(1Q) $-\sum_{r=1}^{s} \mu_{rQ} y_{rQ} + \sum_{i=1}^{m} v_{iQ} x_{iQ} \geq 0$

(1P) $-\sum_{r=1}^{s} \mu_{rQ} y_{rP} + \sum_{i=1}^{m} v_{iQ} x_{iP} \geq 0$

(2) $\sum_{r=1}^{s} \mu_{rQ} y_{rQ} = 1$

(3) $\mu_{rQ}, v_{iQ} \geq 0$

Expanding the constraints as (1Q) and (1P) make it explicit that Q is being evaluated against P at Q's prices.

Then one evaluates P against Q at P's prices:

$$\min h_P = \sum_{i=1}^{m} v_{iP} x_{iP}$$

$s.t.$

$$(1Q) \quad -\sum_{r=1}^{s} \mu_{rP} y_{rQ} + \sum_{i=1}^{m} v_{iP} x_{iQ} \geq 0$$

$$(1P) \quad -\sum_{r=1}^{s} \mu_{rP} y_{rP} + \sum_{i=1}^{m} v_{iP} x_{iP} \geq 0$$

$$(2) \quad \sum_{r=1}^{s} \mu_{rP} y_{rP} = 1$$

$$(3) \quad \mu_{rP}, v_{iP} \geq 0$$

The model, in effect, chooses the prices at which a DMU does the best it can: relative to the other DMUs at those prices.

4.5 The Radial Measurement of Efficiency

Example One will show how the CCR model adjusts for pure scale differences in the inputs and outputs of two DMUs whose inputs and outputs are positively correlated. Example Two will show how the CCR model adjusts for pure scale differences where inputs and outputs are negatively correlated. Average efficiency in Example One is high and average efficiency in Example Two is low. In other words high efficiency can be a measure of positive correlation and low efficiency can be a measure of negative correlation between sets of parameters: inputs and outputs.

These two examples also demonstrate the impact of a total lack of statistical independence between the values of the inputs for any given DMU and also between the values of the outputs for any given DMU.

4.5.1. Example One

Take two DMUs, A and B, with inputs highly positively correlated to outputs – Example One:

- Assume A has inputs of 1 and 1 and outputs of 1 and 1.

  $\left(x_{1A}=1,\ x_{2A}=1,\ y_{1A}=1,\ y_{2A}=1\right)$

- Assume B has inputs of 10 and 10 and outputs of 10 and 10.

  $\left(x_{1B}=10, x_{2B}=10, y_{1B}=10, y_{2B}=10\right)$

If A adopts prices of ½, $\left(u_{1A}=\frac{1}{2}, u_{2A}=\frac{1}{2}, v_{1A}=\frac{1}{2}, v_{2A}=\frac{1}{2}\right)$ then the sum of the priced inputs is $\sum_{i=1}^{m} v_{iA}x_{iA}=\frac{1}{2}1+\frac{1}{2}1=1$, and the sum of the priced outputs is $\sum_{r=1}^{s} u_{rA}y_{rA}=\frac{1}{2}1+\frac{1}{2}1=1$.

If B uses the same prices, then the sum of the priced inputs is $\sum_{i=1}^{m} v_{iA}x_{iB}=\frac{1}{2}10+\frac{1}{2}10=10$, and the sum of the priced outputs is $\sum_{r=1}^{s} u_{rA}y_{rB}=\frac{1}{2}10+\frac{1}{2}10=10$.

In each case, A and B, the ratio of priced inputs to priced outputs will be 1 – the two DMUs are equally efficient. When evaluating DMU A, constraint (2) will apply to DMU A, which could choose prices of ½. When evaluating DMU B, constraint (2) will now apply to DMU B, which will choose prices of 1/20. Note that the ratio of the prices is the same in either case: both A and B are evaluated against a price line with a slope of minus one.

B's greater scale has been incorporated into the prices it selects for itself in the CCR model. High positive correlation is evidenced in the high efficiency scores because the CCR model adjusts for pure scale differences between the DMUs.

Example Two illustrates the way in which the CCR model adjusts when the scale differences imply inefficiency or negative correlation between the inputs and the outputs.

4.5.2 Example Two

Take two DMUs, C and D, with inputs highly negatively correlated to outputs:

- Assume C has inputs of 1 and 1 and outputs of 10 and 10.
- Assume D has inputs of 10 and 10 and outputs of 1 and 1.

In this case the prices make no difference to the outcome. If C adopts prices of ½ for inputs and 1/20 for outputs, then for C, the efficiency score is given by:

$$h_C = \frac{\sum_{i=1}^{m} v_{iC} x_{iC}}{\sum_{r=1}^{s} u_{rC} y_{rC}} = \frac{\frac{1}{2}*1 + \frac{1}{2}*1}{\frac{1}{20}*10 + \frac{1}{20}*10} = 1$$

Scaling D to C's level of output assumes multiplication throughout by 10, so, for D the efficiency score, at C's prices is given by:

$$h_D = \frac{\sum_{i=1}^{m} v_{iD} 10 x_{iD}}{\sum_{r=1}^{s} u_{rD} 10 y_{rD}} = \frac{\frac{1}{2}*100 + \frac{1}{2}*100}{\frac{1}{20}*10 + \frac{1}{20}*10} = \frac{100}{1} = 100$$

The effect of this scaling difference is reflected in the efficiency scores.

4.5.3 Example Three

For Example Three – consider two DMUs E and F with inputs highly negatively correlated to outputs but more variation in the quantities. The input parameters and the output parameters for each DMU continue to be statistically dependent on each other, but not as absolutely as in Examples One and Two. This allows the two DMUs some freedom to choose two distinct sets of prices and reduces the measured inefficiency of the second DMU.

- Assume E has inputs of 2 and 1 and outputs of 10 and 8.
- Assume F has inputs of 8 and 10 and outputs of 1 and 2.
- Assume E and F are the only DMUs

If both E and F adopt prices of 0, ½, 0.1 and 0 respectively then for E and F the efficiency scores are given by:

$$h_E = \frac{\sum_{i=1}^{m} v_{iE} x_{iE}}{\sum_{r=1}^{s} u_{rE} y_{rE}} = \frac{0*2+\frac{1}{2}*1}{\frac{1}{10}*10+0*8} = 1 \text{ , and}$$

$$h_F = \frac{\sum_{i=1}^{m} v_{iE} x_{iF}}{\sum_{r=1}^{s} u_{rE} y_{rF}} = \frac{0*8+\frac{1}{2}*10}{\frac{1}{10}*1+0*2} = \frac{5}{0.1} = 50$$

Alternatively, if both E and F adopt prices of 2, 0, 0, and 1/2 respectively then for F and E the efficiency scores are given by:

$$h_F = \frac{\sum_{i=1}^{m} v_{iF} x_{iF}}{\sum_{r=1}^{s} u_{rF} y_{rF}} = \frac{2*8+0*10}{0*1+\frac{1}{2}*2} = \frac{16}{1} = 16 \text{ , and}$$

$$h_E = \frac{\sum_{i=1}^{m} v_{iF} x_{iE}}{\sum_{r=1}^{s} u_{rF} y_{rE}} = \frac{2*2+0*1}{0*10+\frac{1}{2}*8} = \frac{4}{4} = 1$$

16 is closer to 1 than is 50: so F will "choose" evaluation at prices 2, 0, 0, and 1/2 respectively. E is indifferent between the two sets of prices.

Comparison of the parameter values in Example Three with those in Example Two and then of the two sets of efficiency scores shows the extent to which the small variation between the parameters for each DMU impacts on efficiency scores.

C had inputs of 1 and 1 and outputs of 10 and 10 and had efficiency of 1. D had inputs of 10 and 10 and outputs of 1 and 1 and had efficiency of 100.

E had inputs of 2 and 1 and outputs of 10 and 8 and efficiency of 1. F had inputs of 8 and 10 and outputs of 1 and 2 and had efficiency of 16.

4.5.4 Example Four

When the standard deviations for each input parameter are relatively lower than in the previous examples, the DMUs may be able to choose three sets of prices at which they are both efficient. The more sets of prices they can choose, the more unlikely it is that the addition of a third DMU, of similar scale, will cause either one of them to be deemed inefficient. This makes it desirable to have a large number of DMUs in a model.

Example Four considers two DMUs G and H. Both are evaluated as 100 percent efficient at three different sets of prices. Applying these prices in turn allows G to be considered first to be twice the size of H; then the same size as H, and finally, half the size of H.

- Assume G has inputs of 2 and 1 and outputs of 1 and 2.
- Assume H has inputs of 1 and 2 and outputs of 2 and 1.
- Assume G and H are the only DMUs

If both G and H adopt prices of 1/3, 1/3, 1/3 and 1/3 respectively then G and H have efficiency scores given by:

$$h_G = \frac{\sum_{i=1}^{m} v_i x_{iG}}{\sum_{r=1}^{s} u_r y_{rG}} = \frac{\frac{1}{3}*2+\frac{1}{3}*1}{\frac{1}{3}*1+\frac{1}{3}*2} = 1 \text{ , and } h_H = \frac{\sum_{i=1}^{m} v_i x_{iH}}{\sum_{r=1}^{s} u_r y_{rH}} = \frac{\frac{1}{3}*1+\frac{1}{3}*2}{\frac{1}{3}*2+\frac{1}{3}*1} = 1$$

If both G and H adopt prices of 0, 1, 1 and 0 respectively then G and H have efficiency scores given by:

$$h_G = \frac{\sum_{i=1}^{m} v_i x_{iG}}{\sum_{r=1}^{s} u_r y_{rG}} = \frac{0*2+1*1}{1*1+0*2} = 1, \text{ and } h_H = \frac{\sum_{i=1}^{m} v_i x_{iH}}{\sum_{r=1}^{s} u_r y_{rH}} = \frac{0*1+1*2}{1*2+0*1} = 1$$

If both G and H adopt prices of 1, 0, 0 and 1 respectively then G and H have efficiency scores given by:

$$h_G = \frac{\sum_{i=1}^{m} v_i x_{iG}}{\sum_{r=1}^{s} u_r y_{rG}} = \frac{1*2+0*1}{0*1+1*2} = 1 \text{ , and } h_H = \frac{\sum_{i=1}^{m} v_i x_{iH}}{\sum_{r=1}^{s} u_r y_{rH}} = \frac{1*1+0*2}{0*2+1*1} = 1$$

G and H can both be efficient at three different sets of prices. With the first set of prices, they are both assumed to be operating at the same scale. At the second set of prices G produces one unit of output from one unit of input and H produces two units of output from one unit of input. At the third set of prices the G produces twice as much as H.

If G and H were being evaluated on three inputs and three outputs the number of sets of prices at which they both would be evaluated as efficient would increase. Consider H with inputs (1, 2, and 1) and outputs of (2, 1 and 2), the inputs pricing choices are: (1, 0, 0) or (0, ½, 0) or (0, 0, 1) or (¼,¼,¼) or (½, 0, ½) or (0, ⅓, ⅓) or… with a similar set of pricing choices for the outputs.

The greater the number of inputs parameters and the greater the number of outputs parameters the greater the probability that a DMU will be evaluated as fully efficient. The addition of more DMUs may limit the individual DMU's ability to use combinations of prices to "achieve" full efficiency.

So, the average efficiency scores depend on the distributions of the parameters, the number of the parameters and the number of DMUs evaluated.

4.6 Units of Measurement and Translation of Parameters

The CCR model is *not* invariant to a translation of the values of the parameters either in terms of the scale of the efficiency scores or in the rank order of efficiency of the DMUs. The CCR model is not "Translation Invariant". In other words, the distribution of the parameters, their means and standard deviations, impacts on the efficiency scores. Translation of the data changes the means, without changing the standard deviations.

In order to assess the impact on efficiency scores of differential scaling of means and standard deviations, the parameters are split into their means and standard deviations in the following analysis.

Recall that the CCR efficiency score h_o for any given DMU, 0, is given by the ratio of priced inputs to priced outputs:

$$h_o = \frac{\sum_{i=1}^{m} v_{io} x_{io}}{\sum_{r=1}^{s} u_{ro} y_{ro}} .$$

Multiplying through by the denominator on the right hand side gives:

$$h_o * \sum_{r=1}^{s} u_{ro} y_{ro} = \sum_{i=1}^{m} v_{io} x_{io}$$

The values of the inputs and outputs can be decomposed into their means, standard deviations and z-scores, for the values from all of the DMUs from 1 to n, such that for the "0"th DMU: $x_{io} = \mu_i + (z_{io}\sigma_i)$ and $y_{ro} = \mu_r + (z_{ro}\sigma_r)$

If all the inputs are multiplied by n, this is the same as multiplying the means and the standard deviations of the inputs by n, the prices of the inputs adjust by dividing by 1/n and the efficiency score is unaffected, thus:

$$h_o * \sum_{r=1}^{s} u_{ro} y_{ro} = \sum_{i=1}^{m} \frac{v_{io}}{n} n x_{io} = \frac{1}{n} \sum_{i=1}^{m} v_{io} n x_{io} = \frac{n}{n} \sum_{i=1}^{m} v_{io} x_{io} = \sum_{i=1}^{m} v_{io} x_{io}$$

If the standard deviations of the inputs are multiplied by 13.33 and the means of the inputs are multiplied by 5.65, then the prices adjust, but they do so by finding new relationships between the prices, rather than adopting a new scale, thus:

$$h_o * \sum_{r=1}^{s} u_{ro} y_{ro} = \sum_{i=1}^{m} \frac{v_{io}}{n_{io}} (5.65\mu_i + 13.33 * z_{io}\sigma_i) \neq \sum_{i=1}^{m} v_{io} x_{io},$$ where n_{io}

is the adjustment to the price for input i for unit o.

Even if one knew the distribution of the z-scores it would still be difficult to know what the impact on the prices and the efficiency scores of all the DMUs would be.

The "units of measurement effect" can be separated from the "translation effect". If the multiple of the standard deviation is split into two components, 5.65 and 7.68: the following results:

$$h_o * \sum_{r=1}^{s} u_{ro} y_{ro} = \sum_{i=1}^{m} \frac{v_{io}}{n_{io}} (5.65 * \mu_i + (5.65 + 7.68) * (z_{i0} * \sigma_i))$$

$$h_o * \sum_{r=1}^{s} u_{ro} y_{ro} = \sum_{i=1}^{m} \frac{v_{io}}{n_{io}} (5.65 * \mu_i + 5.65 * z_{i0} * \sigma_i + 7.68 * z_{i0} * \sigma_i)$$

$$h_o * \sum_{r=1}^{s} u_{ro} y_{ro} = \sum_{i=1}^{m} \frac{v_{io}}{n_{io}} (5.65 * (\mu_i + z_{i0} * \sigma_i) + 7.68 * z_{i0} * \sigma_i)$$

From this it is clear that if the σ_i are small relative to the μ_i then the impact of the translation component of the scaling $(7.68 * z_{i0} * \sigma_i)$ should be small relative to the units component of the scaling $(5.65 * (\mu_i + z_{i0} * \sigma_i)) = (5.65 * x_{i0})$.

The key components are the z-scores for each DMU. The relationships between the inputs z-scores and between the outputs z-scores can have a large impact. Consider DMU 0, with 4 inputs with z-scores of –2, -1, 1, and 2 and plug these values into the calculation of efficiency. Assume that the standard deviations are the same for each input parameter.

$$h_o * \sum_{r=1}^{s} u_{ro} y_{ro} = \sum_{i=1}^{m} \frac{v_{io}}{n_{io}} \left(5.65 * (\mu_i + z_{i0} * \sigma_i) + 7.68 * z_{i0} * \sigma_i\right)$$

and since $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$ is assumed σ replaces σ_i

and $(\mu_i + z_{i0} * \sigma_i) = x_{i0}$

$$h_o * \sum_{r=1}^{s} u_{ro} y_{ro} = 5.65 \sum_{i=1}^{m} \frac{v_{io} x_{i0}}{n_{io}} + 7.68 * \sigma * \sum_{i=1}^{m} \frac{v_{io}}{n_{io}} (* z_{i0})$$

$$h_o * \sum_{r=1}^{s} u_{ro} y_{ro} = 5.65 \sum_{i=1}^{m} \frac{v_{io} x_{i0}}{n_{io}} + 7.68 * \sigma * \left(\frac{v_{1o}}{n_{1o}} * (-2) + \frac{v_{2o}}{n_{2o}} * (-1) + \frac{v_{3o}}{n_{3o}} * 1 + \frac{v_{4o}}{n_{4o}} * 2 \right)$$

If the prices $\frac{v_{1o}}{n_{1o}} = \frac{v_{2o}}{n_{2o}} = \frac{v_{3o}}{n_{3o}} = \frac{v_{4o}}{n_{4o}}$ then the impact of the distribution of the z-scores will be to nullify the impact of the translation component of scaling

since: $7.68 * \sigma_i * \left(\frac{v_{1o}}{n_{1o}} * (-2) + \frac{v_{2o}}{n_{2o}} * (-1) + \frac{v_{3o}}{n_{3o}} * 1 + \frac{v_{4o}}{n_{4o}} * 2 \right) = 0$

But, the prices are unlikely to equal each other. With respect to test scores a better assumption might be that they are not statistically independent and therefore; the test score z-scores for a DMU are likely to be of a similar scale and so the translation component is likely to have some effect.

This analysis decomposed the inputs side of the equation $h_o * \sum_{r=1}^{s} u_{ro} y_{ro} = \sum_{i=1}^{m} v_{io} x_{io}$.

The same analysis can be applied to the output side of the equation.

4.7 Panel Data

This dissertation looks at trends in education over time. In other words it deals with "Panel Data". Assuming that, over time, there are no scale changes in the data being analyzed; then there are two main approaches to analysis of panel data with the CCR model.

In the first approach, each period is analyzed in a separate model and the results are compared across time. In this approach each DMU is compared with every other DMU from its own time period.

In the second approach, all of the periods are combined in a single model. Every DMU in every time period is compared with every DMU in all time periods.

A simple three-year and two-DMU example will illustrate the issues inherent in the two approaches.

Taking the first approach, first, and looking at the efficiency scores from three individual analyses, one for each year for two DMUs: A and B. Assume that: in year 1, A produced 4 from 3, and B produced 4 from 4. (What the input is and what the output is: is not important). A did better than B since 3/4 < 4/4. The model prices A such that A's priced ratio = 1 to conform to Koopmans definition of efficiency. So A's efficiency score is 1 and B's efficiency score is 1.333. Average efficiency is 1.166.

Assume B improves in year 2, such that it also produces 4 from 3: both A and B have efficiency scores of 1. Average efficiency improved (8 comes from 6) to 1.

Assume, that in year 3, A had become less efficient producing 3 from 3 but B reverts to producing 3 from 4. A is now as inefficient as B, but the model shows them to have efficiency scores of 1. Average efficiency in the third year is 1, an apparent improvement over the first year, although A and B are now producing 7 from 7 as opposed to 8 from 6 in the second year or 8 from 7 in the first year.

The average efficiency score is revealing something about the change in relative efficiency between DMU A and DMU B. A is more efficient than B in year 1, so the average efficiency score is higher (reflecting a lower level of efficiency) than it is for the two years in which A and B are equally efficient.

Now, taking the second approach and looking at the efficiency scores from a single analysis of both DMUs and three years. There are six "DMUs". A1, B1, A2, B2, A3 and B3. A1, A2, B2 now define full efficiency: each getting 4 from 3. B1 scores 1.333 as before. A3 and B3 now score 1.333 (relative to A1, A2 and B2). Average efficiency in year one is 1.1666. Average efficiency in year two is 1. Average efficiency in year three is 1.333.

From this simple example one can conclude that if the issue at stake is absolute change in efficiency over time, then the evaluation against the six "DMUs" provides good measurements. If, on the other hand, the issue is relative efficiency in each year and how this changes, then evaluating each year separately provides good measurements.

It is also instructive to note that average efficiency can increase in the case where some efficient DMUs loose efficiency or in the case where some inefficient DMUs become more efficient.

4.8 Conclusions

The eight properties or characteristics in the data that impact on the measurement of efficiency using the CCR model are:

1. The degree and direction of the correlation between the inputs (as priced) and the outputs (as priced),
2. The distributions of the data – the z-scores or the degree of statistical independence between the inputs parameters, and the statistical independence between the outputs parameters,
3. The number of DMUs evaluated,
4. The number of input parameters and the number of output parameters,
5. The freedom that some DMUs have to choose prices that are zero for some of the inputs and some of the outputs,
6. The units or absolute scale of the data – the means,
7. The standard deviations of the data, and
8. The treatment of Panel Data.

These factors can be used good effect by controlling some and thus measuring others.

CHAPTER 5

IS SOCIO-ECONOMIC STATUS A MAIN DETERMINANT OF EDUCATIONAL OUTCOMES?

The one strong relationship revealed or assumed by Education Production Function Research – see Section 5 in Chapter 1 – is that between Educational Outcomes and Socio-Economic Status, which are strongly positively correlated. If MERA was a success in providing a higher basic level of education, and if this led to better outcomes in poorer school districts, then the correlation between Socio-Economic Status and Educational Outcomes should have weakened over time as the reforms took effect.

Correlation between multiple input variables and multiple output variables is measured in this chapter using average efficiency scores from the CCR model with proxies for Socio-Economic Status as inputs and Test Scores as outputs. Strong correlation between Socio-Economic Status and Test Scores is shown to exist. If a trend can be identified from the results it is a strengthening of the relationship rather than a weakening of the relationship, which implies that MERA has not been successful.

5.1 Measuring Correlation Using The CCR Model

Statistical Correlation is measured on a scale ranging from 1 to minus 1. An average efficiency score of zero or less cannot be derived from the CCR model since at least one DMU will be 100 per cent efficient and negative efficiency scores do not arise as a ratio of positively priced positive values. The approach, referred to as the "sorting strategy", is to establish a scale for each set of data by sorting the data in two ways.

1. All the inputs and all the outputs are sorted into the ascending order.

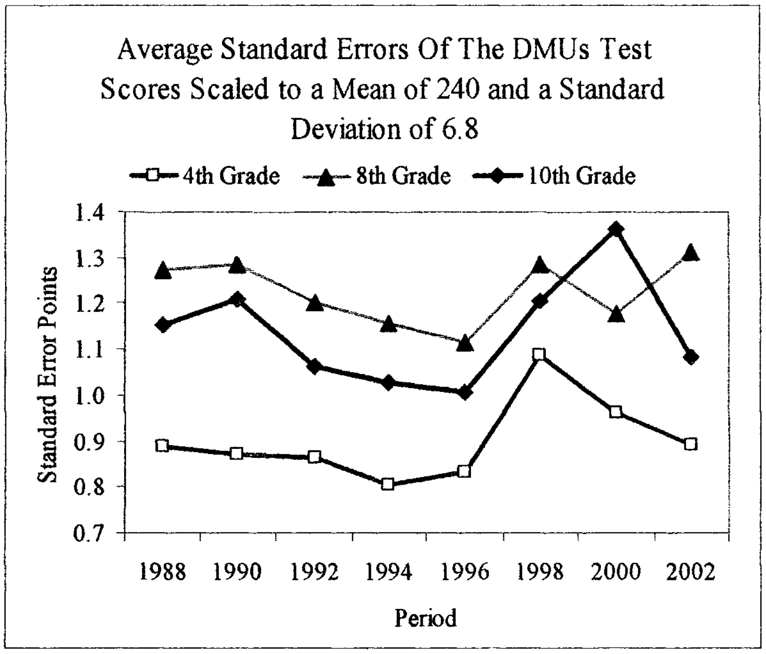
2. All the inputs are sorted into ascending order and all the outputs into descending order.

The first sorting gives a set of data that is highly positively correlated, while the second sorting gives a set of data that is highly negatively correlated. Average efficiency scores from the CCR model applied to the two sets of data render a range of average efficiency scores between highly positively correlated and highly negatively correlated. Using the actual data values means that the distributions of the parameters (SES Proxies and Test Scores) remain constant between the calculations of efficiency for the highly positively correlated, the highly negatively correlated and the actual unsorted data. The sorting strategy also controls for the number of parameters used on either side: inputs or outputs. In other words Factors 3 through 8 from Section 8 in Chapter 4 are kept constant.

This leaves Factors 1 and 2. Factor 1 is precisely the effect that is to be measured – correlation. Test scores for 4 disciplines for 180 DMUs have relationships that can be vertical – Math Scores have a normal distribution with a mean of x and a standard deviation of y, so an individual math score can be described by the distribution and the z-score – or the relationship can be horizontal, among the Math, Reading, Science and History scores for a particular DMU. If a school is a "good" school the expectation is that the scores in all disciplines will be high, if a "bad" school then the expectation is that the scores in all disciplines will be low. There should be little statistical independence among the scores for the different disciplines. Sorting a column (or sorting vertically) does not change the distribution or the z-score of any of the values in the column, but it does upset the horizontal relationships. If the effect of sorting is to increase the degree of statistical independence then average efficiencies from the CCR model may increase. Conversely if the degree of statistical independence decreases the CCR model may give higher average efficiency scores for the sorted than for the actual data because there is less scope for a DMU to use pricing to achieve higher efficiency scores in the sorted data – see Examples One, Two and Three in Section 5 of Chapter 4. So highly correlated data, which has some degree of statistical independence can result in efficiency scores that are higher than those obtained from sorting the same data into ascending order by inputs and outputs.

Unfortunately, although the effect of statistical independence between the parameters should be taken into account, it is not quantifiable. Measuring the statistical independence among the test scores using the average of the standard deviations of the test scores for each school district divided by the square root of the number of disciplines tested (the average of the standard errors) revealed very small standard errors and very little variability in these standard errors over time – see Figure 5.01. Thus although the inability to quantify the effect of, or control for, the degree of statistical independence makes analysis of trends over time imprecise there are reasons to think that the effect is limited.

Figure 5.01 – Average Standard Errors of The DMUs' Test Scores.

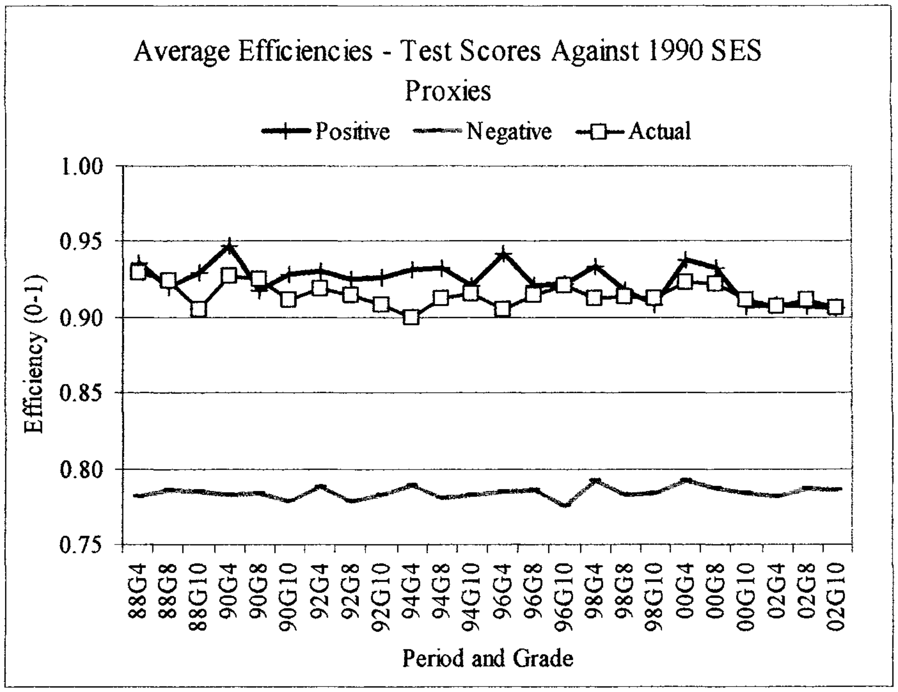


In summary, the approach to the estimation of correlation used in this chapter is to hold six of the eight factors constant across eight time periods, and to use the sorting strategy to identify ranges of efficiency scores that represent high positive correlation and high negative correlation. Comparing the efficiency scores derived from the actual data with these ranges allows the strength and direction of correlation to be estimated.

5.2 SES and Education Outcomes Are Highly Correlated

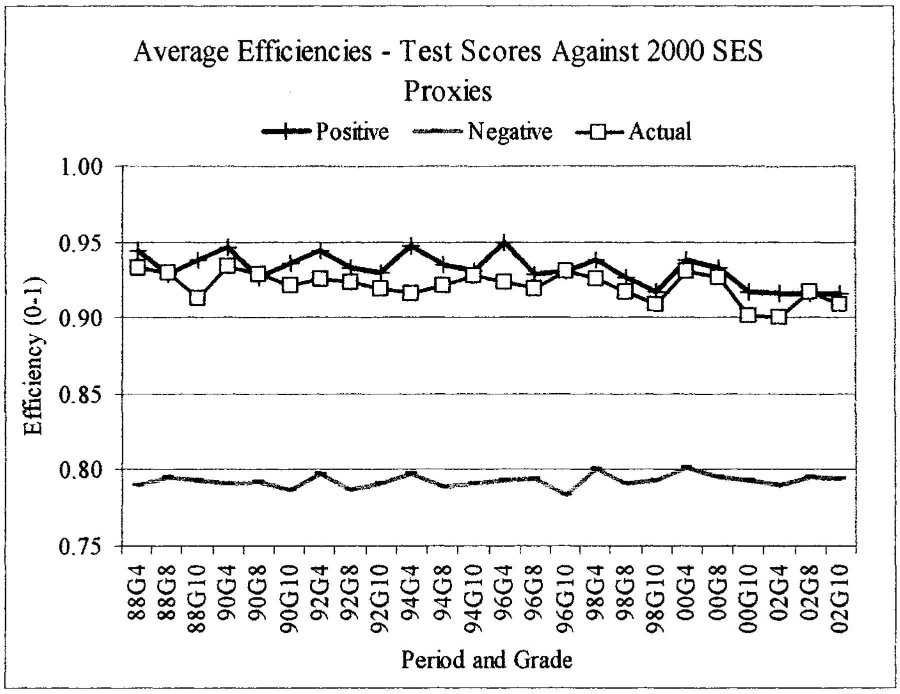
The proxies for Socio-Economic Status used were: Education, Median Income, TSEI2 and Poverty. Two sets of proxies were calculated: one from 1990 census data and the other from 2000 census data. Test scores came from 4th, 8th and 10th grade tests in 1988, 1990, 1992, 1994, 1996, 1998, 2000 and 2002. 24 models were run covering three grades and eight time periods. Models were also run using 1990 SES proxies for all periods and using 2000 SES Proxies for all periods.

Figure 5.02 – All Average Efficiency Scores – Test Scores Against 1990 SES Proxies.



The results from models using 1990 SES proxies are plotted as Figure 5.02. The Figure shows clearly that the average efficiencies from the models using the actual data, are significantly closer to the average efficiencies from the models using the actual data sorted in such a way as to create Positive Correlation, than they are to the average efficiencies from the models using actual data, sorted in such a way as to create Negative Correlation. Figure 5.03 shows the average efficiencies based on 2000 SES proxies. Again, as might be expected, actual data average efficiencies are very much nearer to the Positive data average efficiencies than to the Negative data average efficiencies.

Figure 5.03 – All Average Efficiency Scores – Test Scores Against 2000 SES Proxies.



These results strongly support the proposition that Socio-Economic Status was highly correlated with Educational Outcomes in Massachusetts between 1988 and 2002.

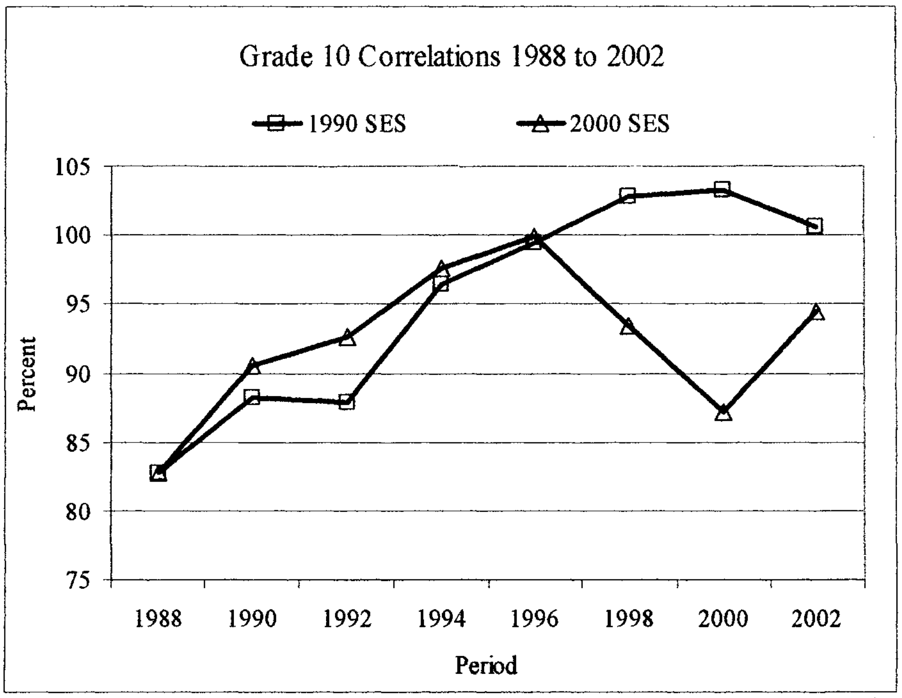
5.3 Change Over Time in The Relationship Between SES and Test Scores

If MERA was successful in reducing inequity in funding and if greater equity in turn was reflected by improvement in the educational outcomes in districts with lower Socio-Economic Status measures, then greater equity should results in a loosening of the correlation between Socio-Economic Status and Educational Test Scores.

When the range between the average efficiency derived from the actual data sorted to be Positively Correlated and the average efficiency derived from the actual data is divided by the range between the average efficiency derived from the actual data sorted to be Negatively Correlated and the average efficiency derived from the actual data, this yields a percentage between 0 and 100 percent – and sometimes higher as a consequence of the impacts of sorting on Factor 2, discussed in Section 1 of this Chapter. Values around zero imply very high negative correlation in the data and values around 100 imply very high positive correlation in the data.

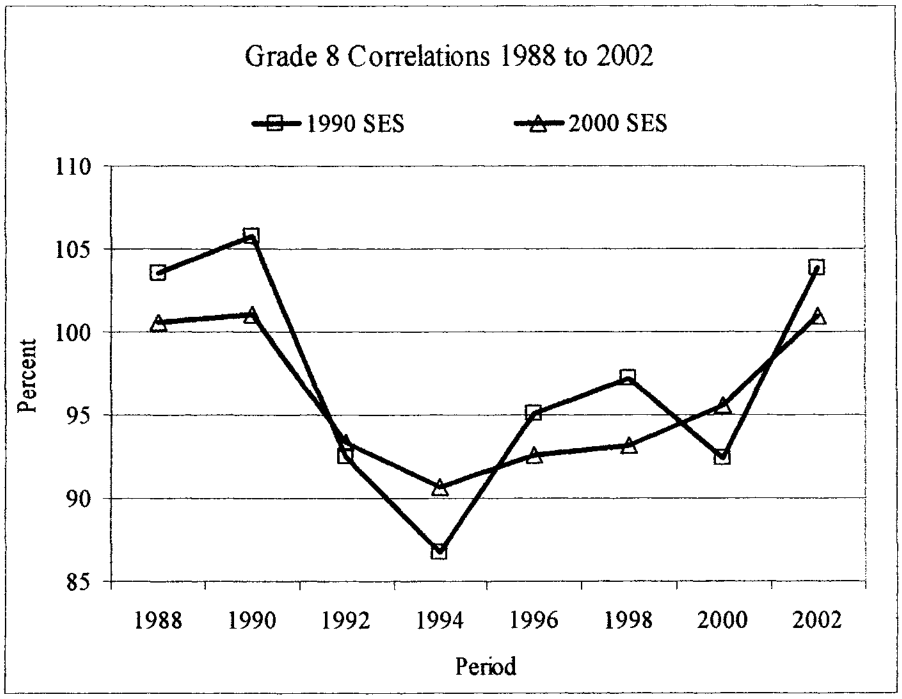
For efficiencies based on Grade 10 tests scores there is a clear trend towards greater correlation between SES and test scores over time – see Figure 5.04, which shows the results using both 1990 SES proxies as inputs and 2000 SES proxies as inputs.

Figure 5.04 – Trend in "Correlations" Between 1990 and 2000 SES Proxies and Grade Test Scores.



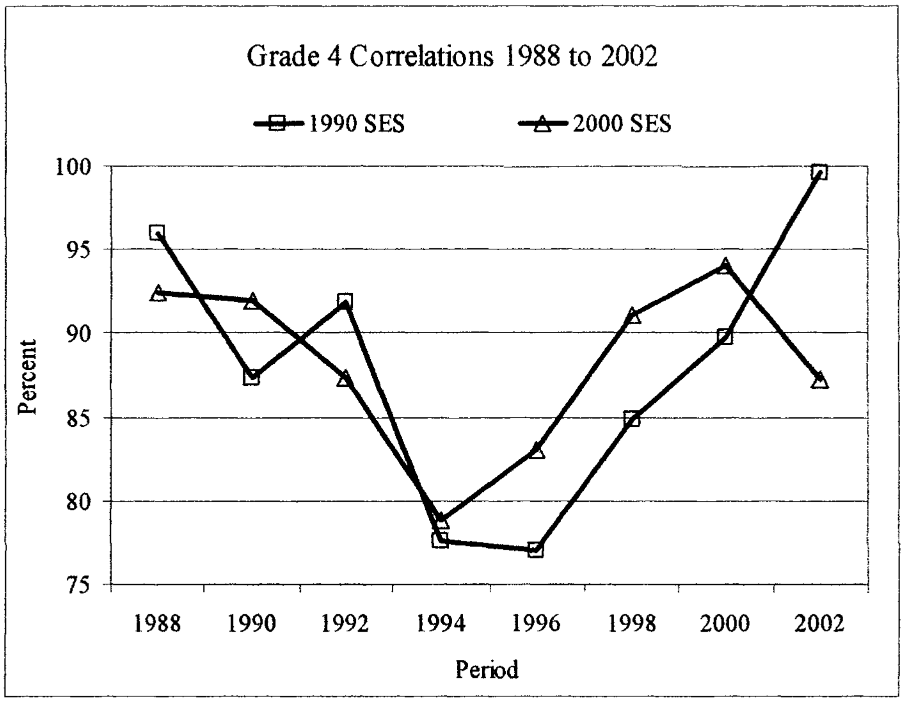
For efficiencies based on Grade 8 tests scores the picture is more ambiguous – see Figure 5.05. Both 1990 and 2000 SES proxy based "Correlations" suggest increasingly stronger positive correlation between test scores and SES after 1994, when MERA came into effect.

Figure 5.05 – Trend in "Correlations" Between 1990 and 2000 SES Proxies and Grade 8 Test Scores.



For efficiencies based on Grade 4 tests scores the picture is also ambiguous – see Figure 5.06. Again correlation between SES and test scores appears to have become more positive between 1994 and 1998. Then based on the 2000 SES proxies, the relationship falls off some way toward levels seen in 1996.

Figure 5.06 – Trend in "Correlations" Between 1990 and 2000 SES Proxies and Grade 4 Test Scores.

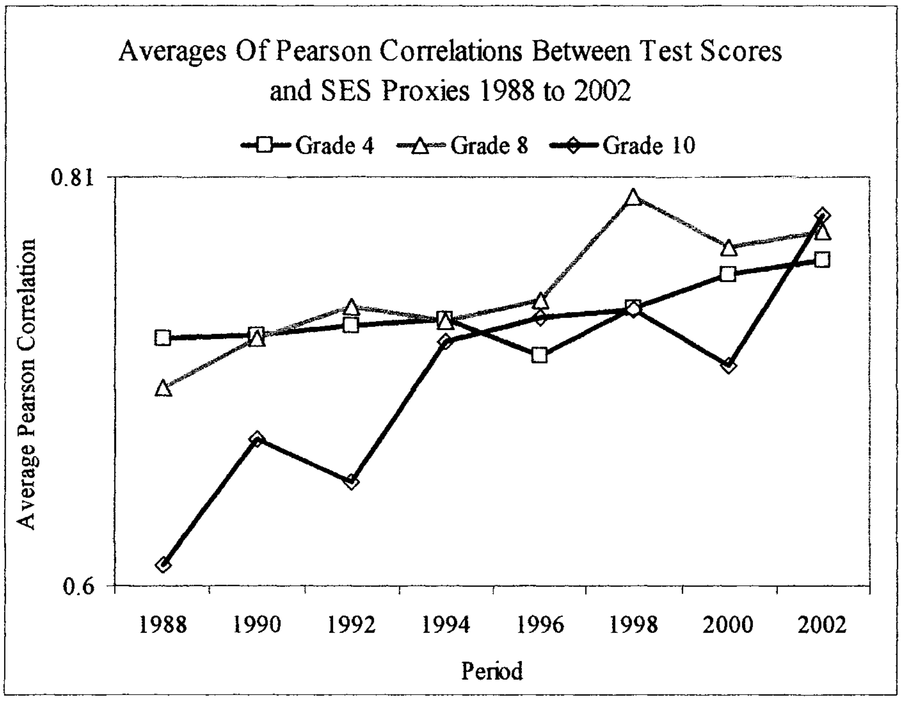


It is not possible to draw any categorical conclusions from this analysis. Looking at the period between 1994 and 2000 it would be possible to argue that, contrary to expectation, the relationship between SES and Educational Outcomes strengthened with the implementation of MERA. The Grade 10 results seem to show this as a fairly consistent trend since1988. The results for Grades 4 and 8 cast some doubt on this interpretation. In particular the sensitivity of the Grade 8 results to the scaling of the test scores calls the whole process into question.

5.4 Averages of Pearson Correlations

The averages of Pearson Rank Correlations between each of the SES Proxies and each of the test scores were calculated for all periods and grades and the trends in these averages were plotted as Figure 5.07. The results seem to confirm the results of the average efficiencies from the CCR models in that there is a clear trend in Grade 10 towards a higher positive correlation between test scores and SES.

Figure 5.07 – Trend in Averages of Pearson Correlations Between SES Proxies and Test Scores.



The average Pearson Correlations based on Grade 8 and Grade 4 test scores also trend upwards although less steeply. All P-values were less than 0.000. This analysis lends support to the results in Section 3 of this Chapterand suggests that the CCR model is estimating correlation reasonably well in spite of the complication introduced by the inability to adjust for, or control for, Factor 2.

5.5 Conclusions

There is no doubt from the Pearson Correlations or from the CCR model average efficiency scores that there is a strong positive correlation between measures of Socio-Economic Status and Educational Outcomes in general.

On balance the trends in the strength of the positive correlation (Pearson or CCR model based) seem to be towards a strengthening rather than a weakening of the relationship between Socio-Economic Status and Educational outcomes in Massachusetts both before MERA and since MERA. This implies that MERA has not had the effect of making educational outcomes more equal.

CHAPTER 6

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

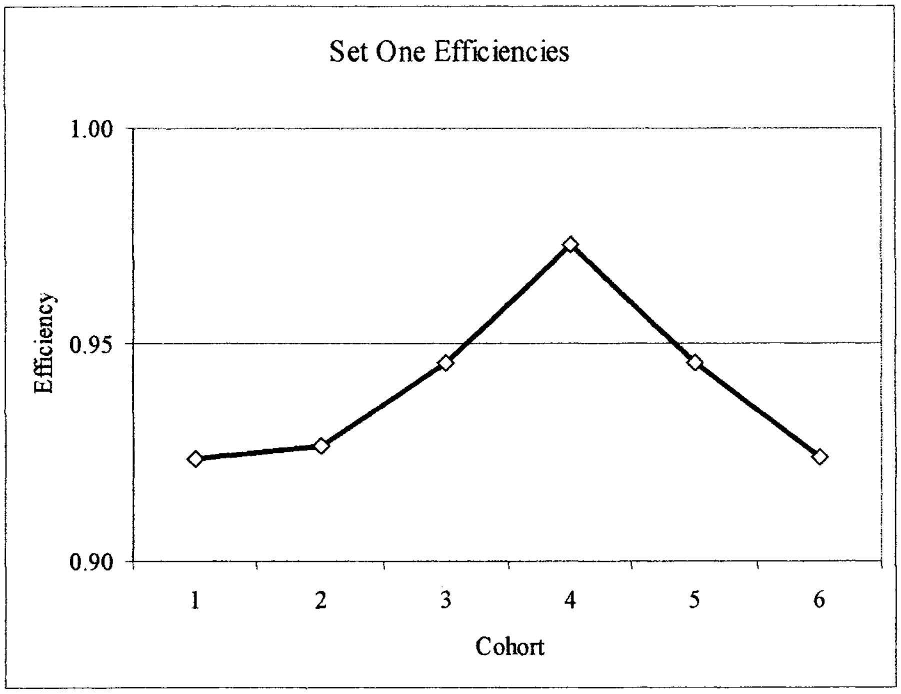
This chapter documents two sets of experiments. The first undertaken to test the proposition that "education standards have been raised and educational outcomes have improved as a consequence of MERA" – see Section 1 to 5 – and the second to test the proposition that opportunity was made more equal –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

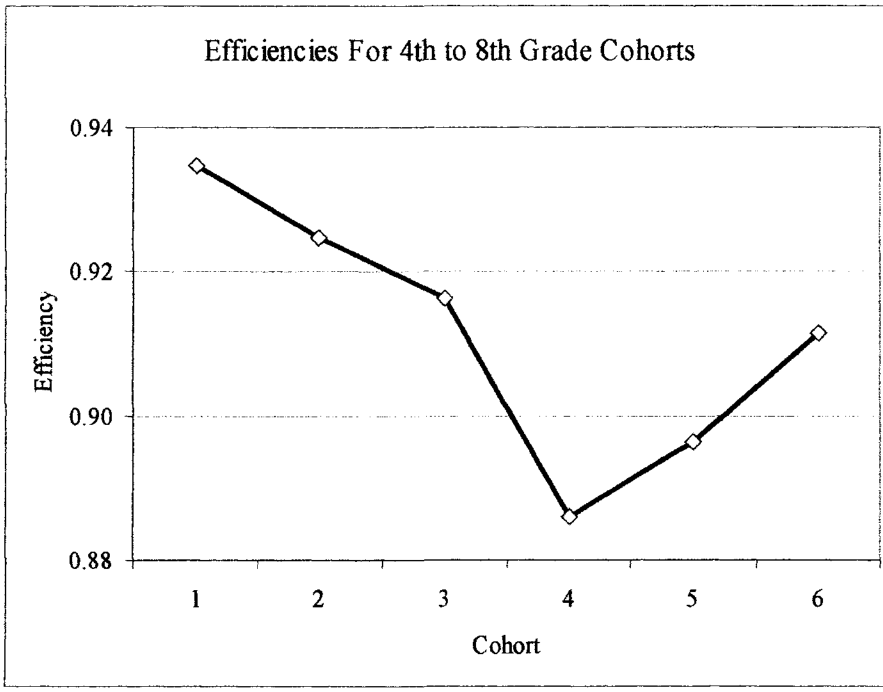
Figure 6.01 Average Efficiency Over Time From Test Data Set One.



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 4^{th} to 8^{th} 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 9^{th} grade next semester. Your other sibling is four years older. You all attended the same K-12 school district.

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

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

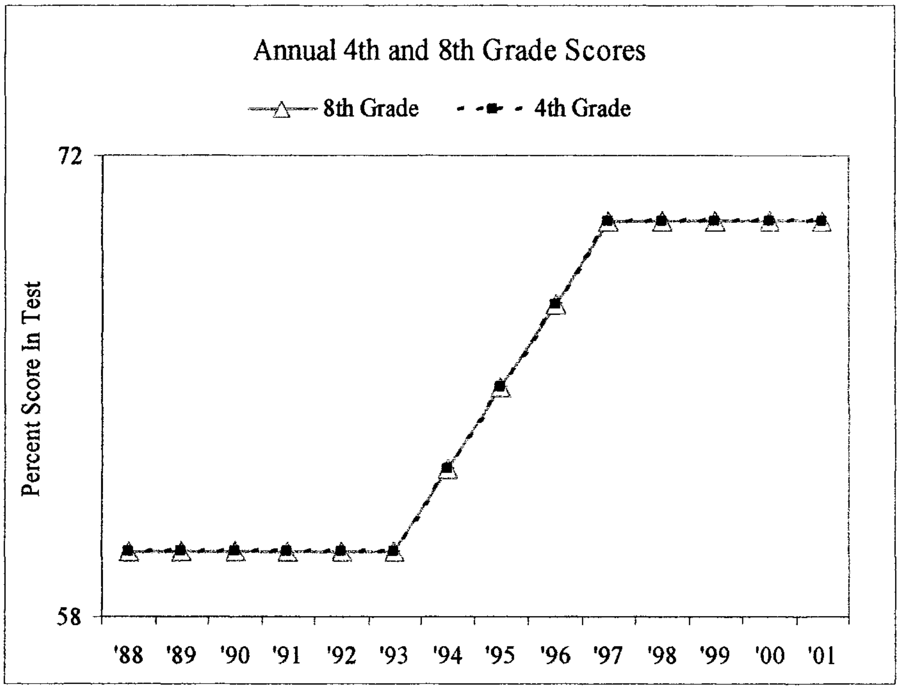
Treating your 4th grade score as an input and your 8th grade score as an output and applying a pricing to the input score that would make the ratio of your priced scores equal to one implies that the price, p, is equal to 7/6. ($70/60*p = 1 \Rightarrow p = 7/6$).

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 4th grade scores into 8th grade scores shows you to have been more efficient than either your younger or older sibling.

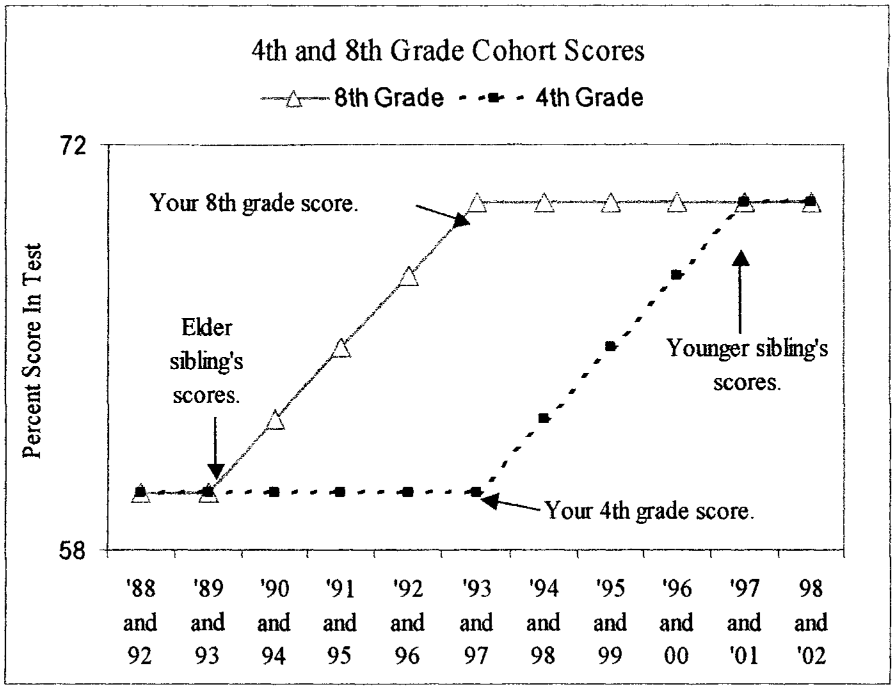
Figure 6.03 shows your test scores improving over three periods in a one off improvement starting in 1993 and ending in 1997.

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



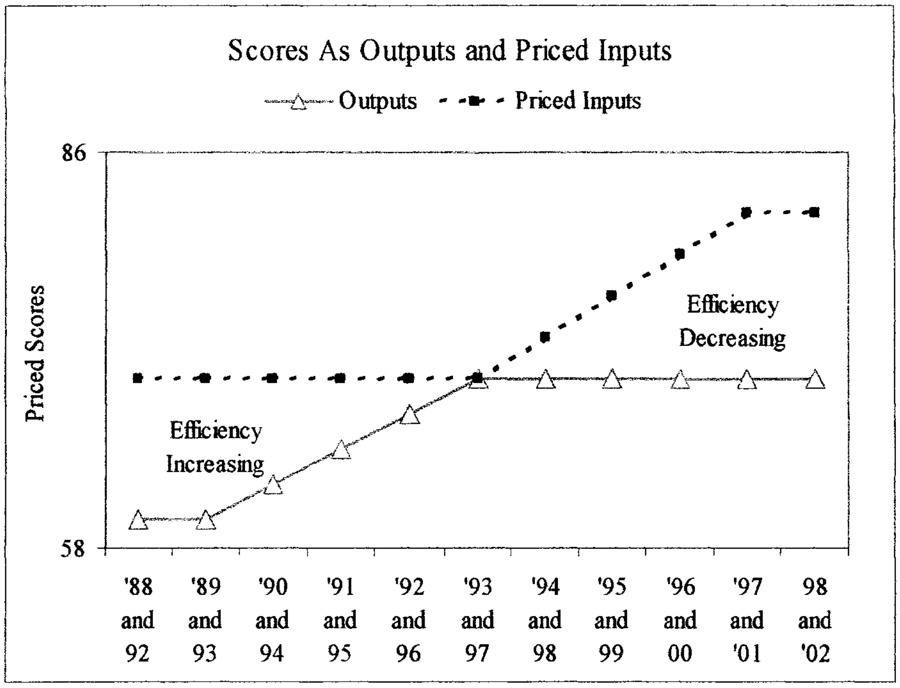
Imagine that your older sibling took tests in 1989 and 1993, you took tests in 1993 and 1997 and your younger sibling took tests in 1997 and 2001. Measuring your efficiency with your 4th grade score as an input and your 8th grade score as an output is easier if the two scores are lined up vertically.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

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



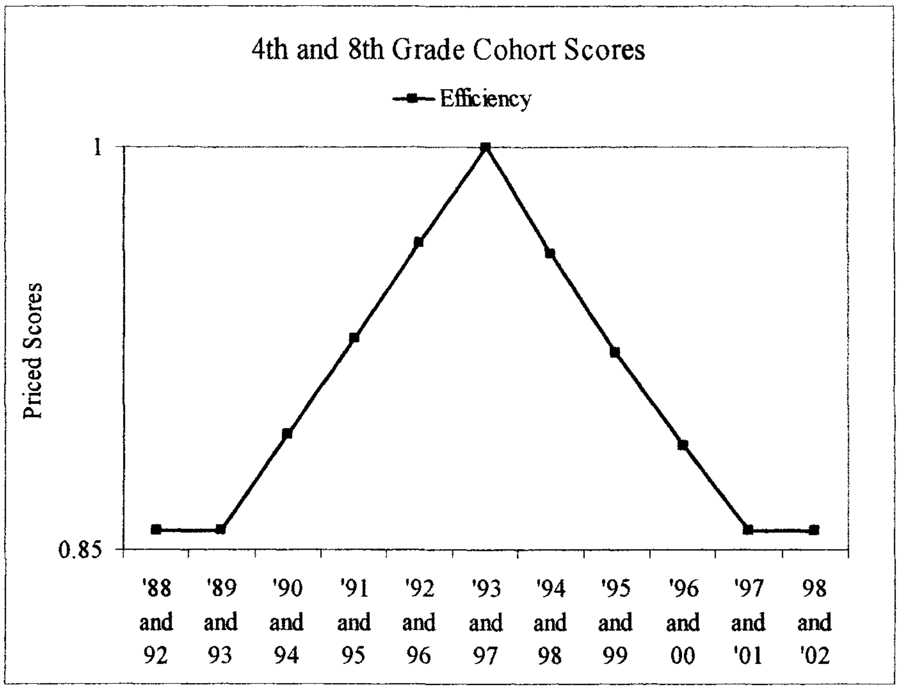
You have the minimum input to the maximum output, so when inputs are priced at 7/6 as calculated above, the inputs lie above the outputs for all cohorts except for the one that you are in. In other words all the others are inefficient – see Figure 6.05. Note that efficiency scores improve up to your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

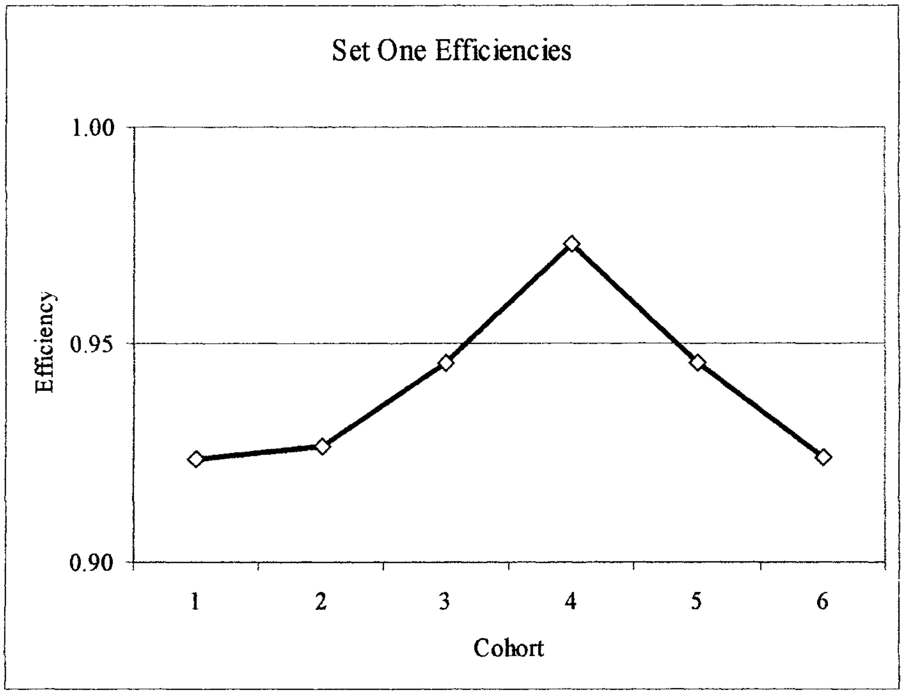
To test the proposition two sets of data were synthesized for even years from 1988 to 2002. Set One assumed a one off improvement evidenced over four years allowing for mean scores of 200 in each of 3 subjects for 1988, 1990, 1992 and 1994 in both the 4th and 8th grades. The mean increased to 205 for 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".

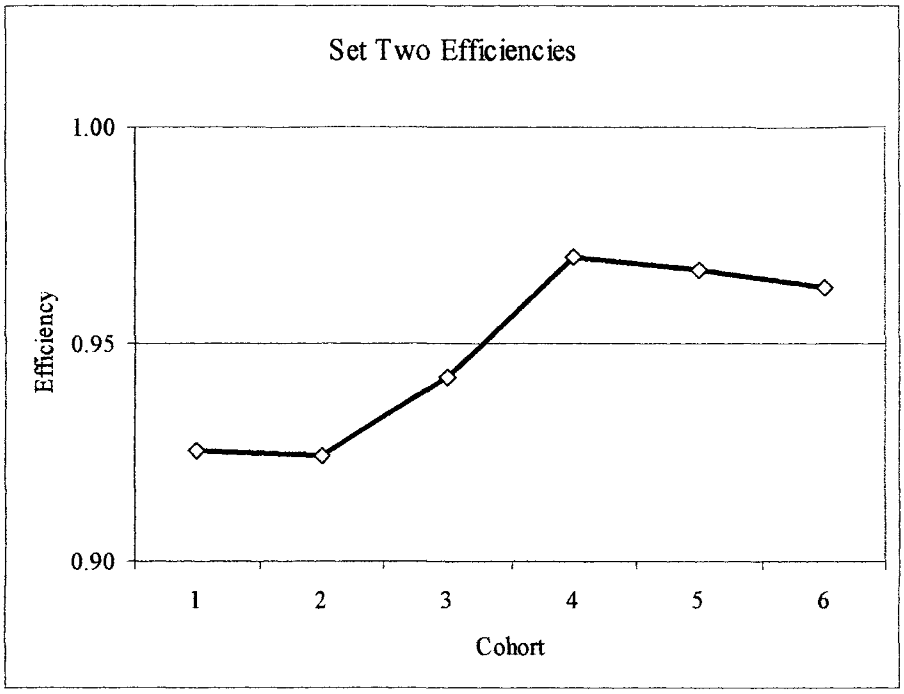
The results of these experiments confirmed the proposition that a one off improvement in Education should first result in increasing efficiency and then in decreasing efficiency – refer to Figure 6.07. This graph shows that the average efficiency first increased and then decreased as predicted.

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:

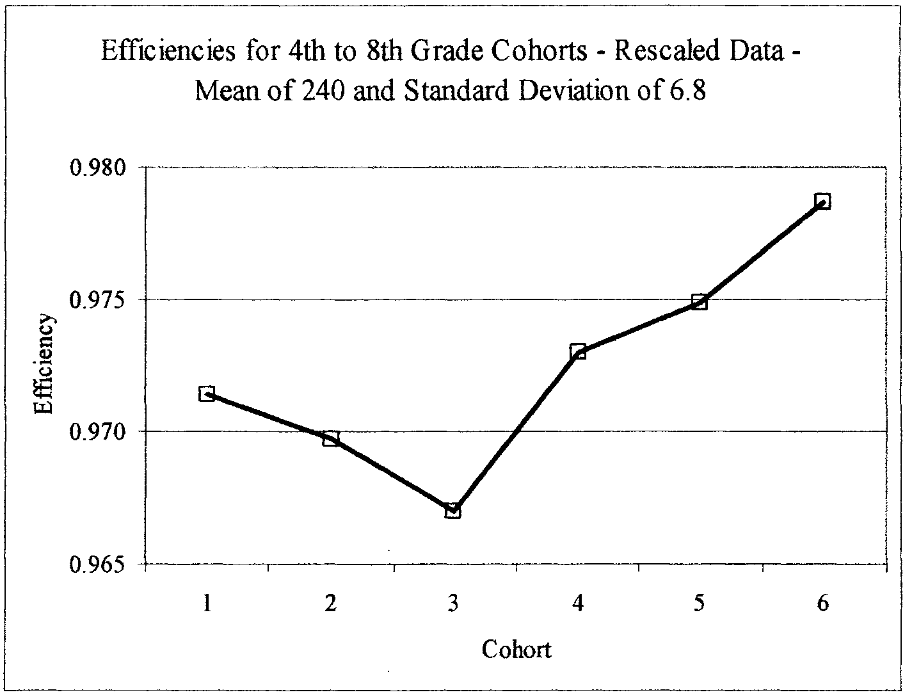
For MEAP Years (1988 to 1996), M=Mathematics; R=Reading; S=Science, and H=Social Studies.

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)

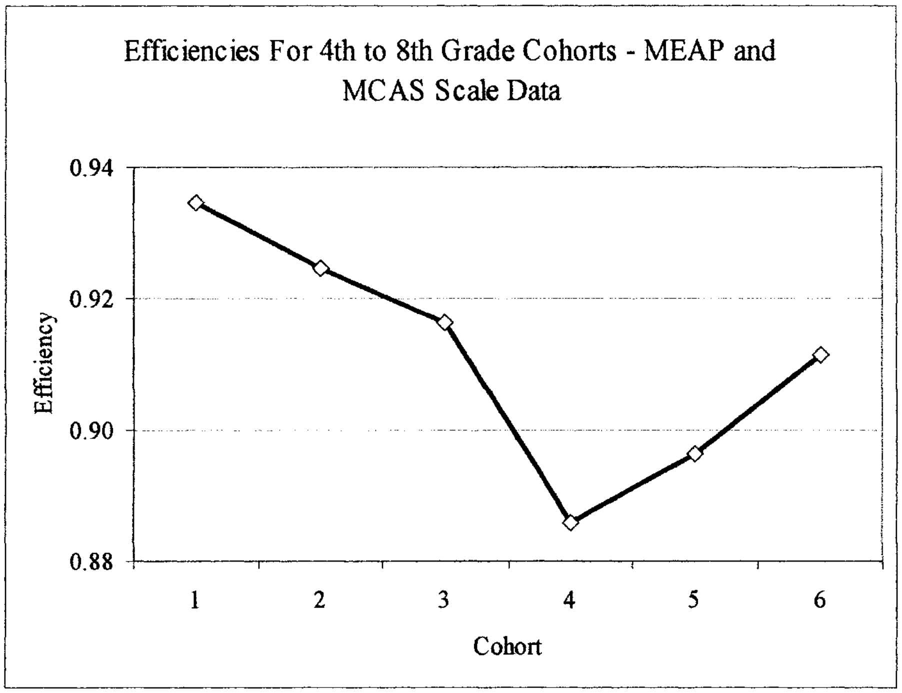
Figure 6.09 – Efficiencies for the 4th to 8th Grade Cohorts.



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.

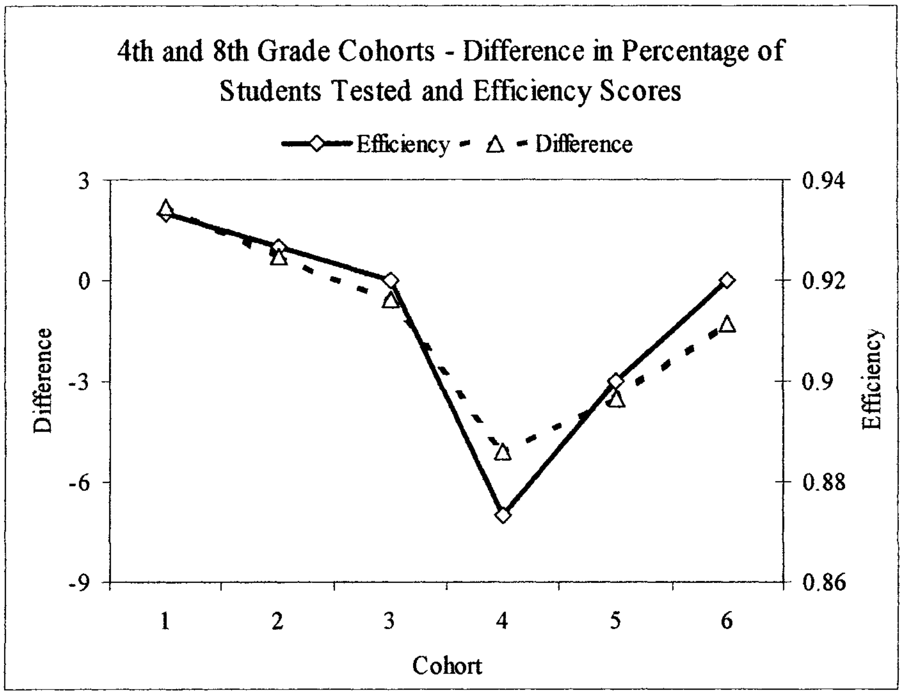
A table giving the percentages of enrolled students tested in each grade was given in Section 2 of Chapter 3 – Table 3.04. On the assumption that those who would avoid testing if possible would be the less able students, the impact of higher percentages taking a test should be to depress the average 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 4^{th} to 8^{th} 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 4^{th} Grade	90	90	89	90	90	97
Percent 8^{th} Grade	88	89	89	97	93	97
4^{th} less 8^{th}	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.



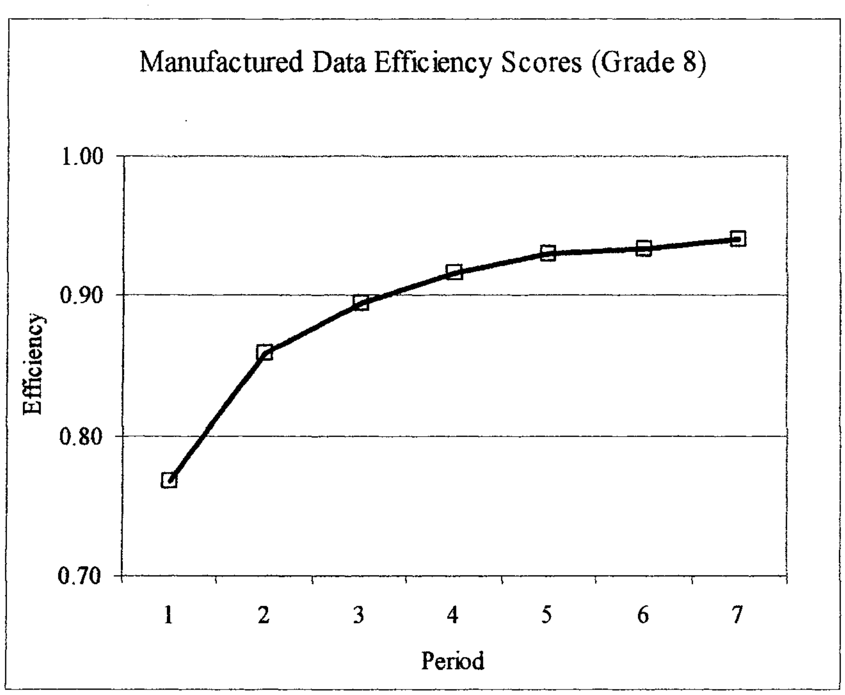
It appears that little can be said with any certainty about the impact of MERA on education in Massachusetts based on the analysis undertaken. The experiments described in Section 2 showed that if an even percentage of students had taken the tests over time and if an improvement had resulted from MERA 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.

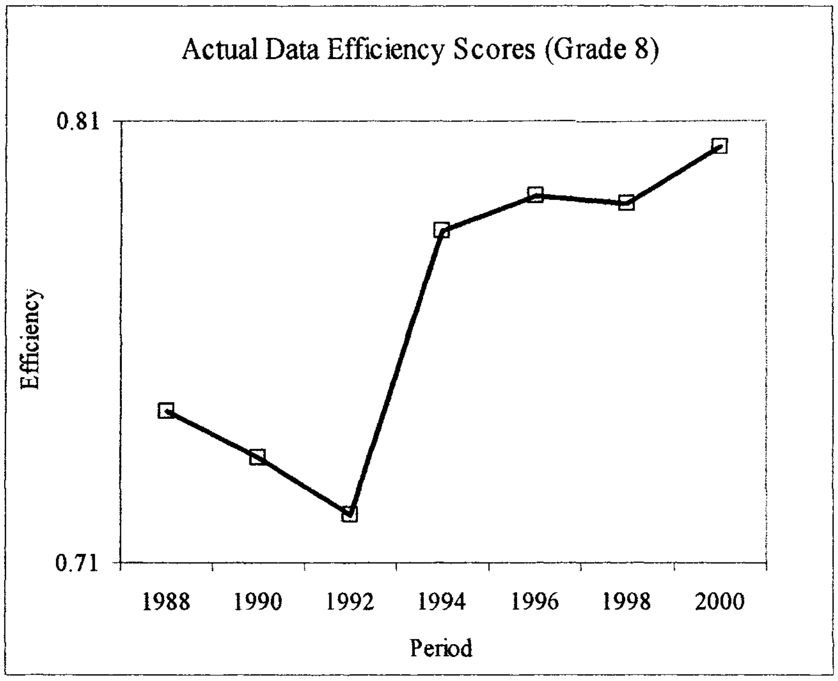
The proposition, set out in Section 7, is that if the base level of resources per pupil increases, then the efficiency with which poorer districts convert money into outcomes should decrease. To test this proposition data was synthesized to conform to the assumptions. The CCR model showed efficiency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

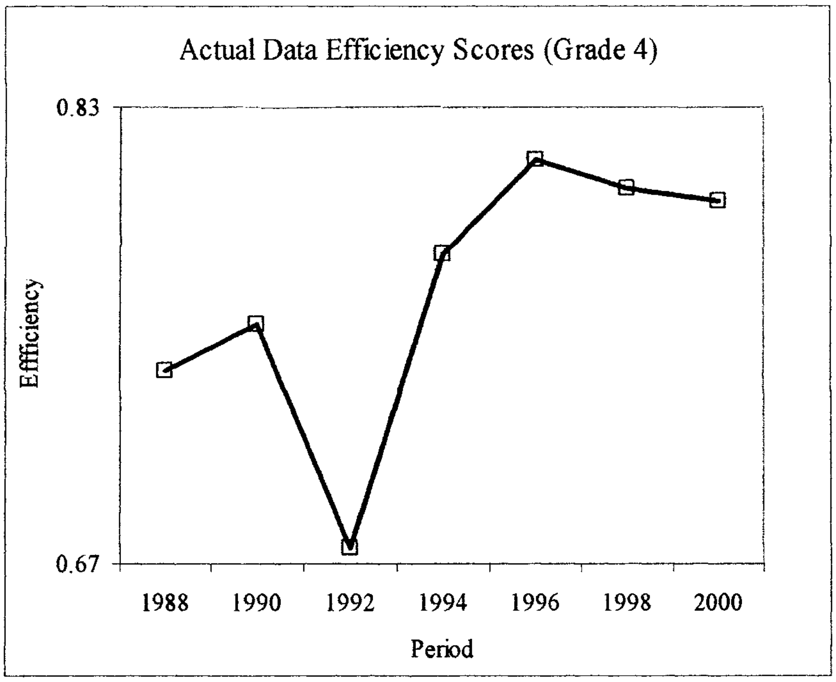
Consider two school districts, A and B, with the same tests scores. A spends twice as much as B. A is therefore 50% as efficient as B. Average efficiency is 0.75. Now increase B's budget to 80% of the amount spent by A. A is now 80% efficient against the benchmark set by B and the average efficiency 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

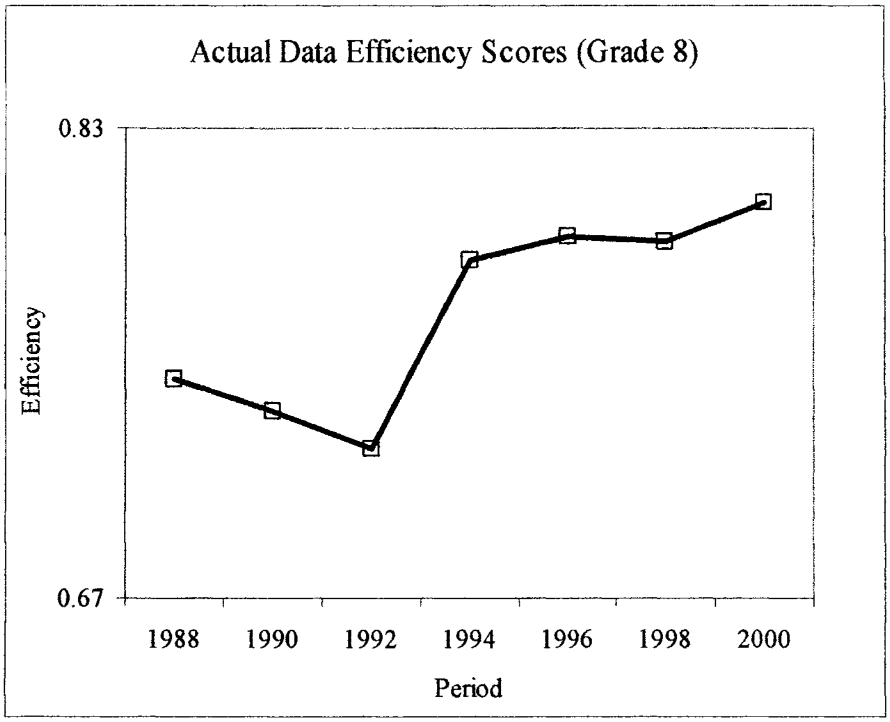
In the previous section it was established that the CCR model should show increasing efficiency accompanying greater equity when Per Pupil Expenditures are used as inputs to models that take test scores as outputs. The result derived from applying the process to actual data – see Figures 6.14, 6.15 and 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.



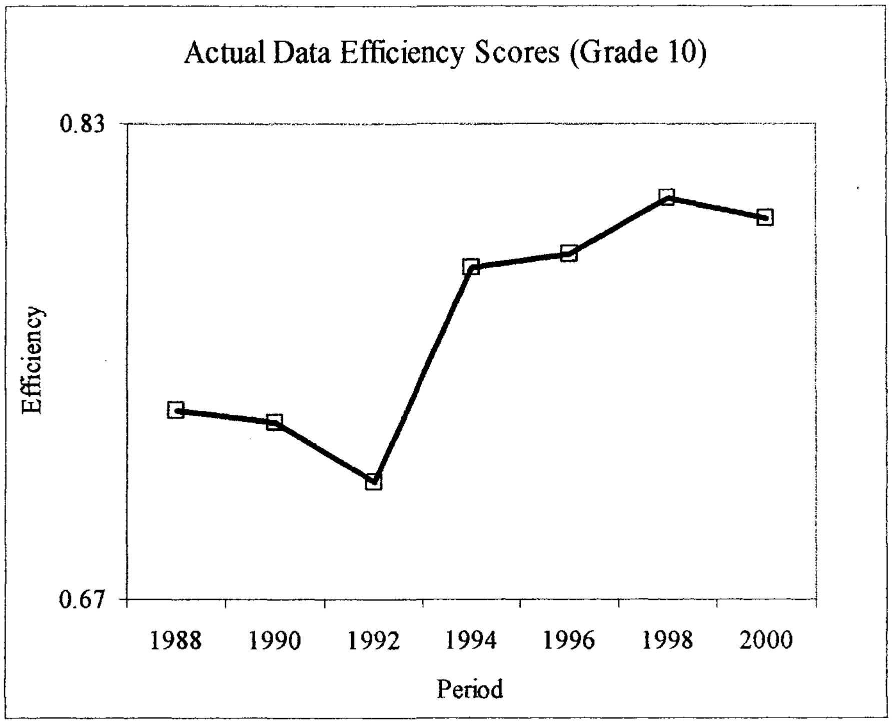
From the low point in 1992, before MERA, efficiency based on Elementary School Expenditures and Grade 4 Test Scores increased dramatically surpassing the levels seen in the years prior to MERA.

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.

CHAPTER 7

CONCLUSIONS

The conclusions drawn as a result of the three experiments conducted with data are reviewed first and then the conclusions drawn from the Massachusetts experience of implementing standards reform are presented.

7.1 The Three Experiments

The objective of the first experiment was to test the strength and direction of the correlation between the Socio-Economic Status of a school district and the educational outcomes for the school district in Massachusetts. A very strong positive "correlation" was found between Socio-Economic Status and outcomes, both from using DEA to provide a single measure between multiple independent and dependent variables and also as a result of taking the average of many Pearson Correlations between pairs of individual variables.

Taking the results over time since MERA from 1993 to 2002 the trend in "correlation" between Socio-Economic Status and outcomes was to higher levels of positive "correlation". This implies that, since MERA, better Socio-Economically endowed school districts have been achieving relatively better outcomes than the less well Socio-Economically endowed school districts. This is troubling because the School Finance reforms were intended to improve education in poorer school districts by allowing them to spend more money.

Seven Common Core subjects were identified by the Commission on the Common Core of Learning and seven Curriculum Frameworks were defined over a period of ten years. Testing has only taken place, consistently, in two subjects English

Language Arts and Mathematics: so one interpretation of these results could be that the better Socio-Economically endowed schools have concentrated on these subjects to the exclusion of other material. Another interpretation might be that the better off school districts are better able to "teach to the test". Yet another interpretation might be that better off parents encouraged their children to spend more time on subjects for which they would be given test score results – a qualification of sorts.

The objective of the second experiment was to establish whether or not there had been an overall improvement in education as a consequence of MERA. The analysis of the test score data to identify trends in educational achievement suggests that, rather than improving, the outcomes may have worsened. The picture is clouded by the fact that a higher proportion of the students in a grade have been forced to take the tests since MERA than had to take the tests before MERA. Under the assumption that the less academically able students were the ones avoiding the tests where possible it would be expected that the results should become worse as a higher proportion of students took the tests.

Another problem with an assessment of overall progress is that the tests have not covered all of the subjects defined in the Common Core. It might be expected that schools would concentrate more on the subjects being tested and that, with fewer subjects being tested after MERA than before MERA, performance in those subjects would improve. So the test results are not a measurement of educational progress other than for the subjects tested.

The final and most severe problem with measuring educational progress from tests score data results from the nature of the tests and the scoring: the test scores are scaled scores. Scaling, also known as "Norm-Referencing" takes raw scores and scales them to a distribution – in the case of both MEAP and MCAS a normal distribution with a statewide mean and standard deviation. This is convenient because it avoids the degree of subjectivity and expense inherent in the process of developing separate tests for each testing cycle that are of equal difficulty. Unfortunately norm-referencing renders cross period comparisons of overall progress meaningless since the mean is the mean in all periods and there are no established criteria by which to make comparisons between periods.

The relative performance of school districts can be inferred over time from the positions in the rankings, but great care needs to be taken in interpretation of the movements in the rankings since a statistically insignificant change in a school district's scores may result in a large movement in its ranking, particularly for school districts whose scores are around the mean where more similar scores will be concentrated by virtue of the normal distribution used in the scaling.

It is simply not possible to draw any hard and fast conclusions about the effect of MERA on overall educational performance in Massachusetts because the test score data does not support cross year comparisons.

The objective of the third experiment was to establish whether the changes made as a consequence of MERA to the state aid formulas and the use of the foundation budget and maximum contribution calculations allowed all school districts to spend more on education. The results of the DEA analysis showed that base expenditures increased and that poorer school districts spent proportionately more per pupil after 1994 than before.

Total School Related State Aid in real dollars did not return to the levels seen in 1989 until 1996. There was a steady increase in School Related State Aid from 1992 onward, but the increase in Total Town School Expenditures was faster than the increase in Total School Related State Aid. Much of this extra expenditure probably resulted from the booming economy in Massachusetts from 1994 to 2000. The fact that the increases in Total School Related State Aid actually took place may also have derived from the effect of the booming economy on State tax receipts.

The rate of growth in enrollment seems to determine the changes in the school district rankings by average expenditure. School districts seem to be less able to increase per pupil expenditure when the number of pupils is also increasing and more able to do so when enrollment is falling or stable.

Additional funding – however it was achieved – does not appear to have loosened the relationship between Socio-Economic Status and educational outcomes. If anything the results of the analysis suggest a continued strengthening of this relationship since before MERA.

7.2 Standards Reform and MERA

Implementation of the standards reform components of MERA has fallen foul of all the problems that appear to be intrinsic to standards reform generally. Simply stated, standards reform defies the laws of political gravity. All the participants must co-ordinate, common goals must be agreed and Curricula must be defined and "owned" by all the participants. There must be a fair system of testing that measures actual progress and the incentives and disincentives in the system must be fair and rational and no one should game or cheat the system.

MERA was an extraordinary good faith attempt to implement comprehensive standards reform. A lot of extra money was made available and State Aid was more redistributive. A Common Core was agreed, but the process started to fall apart when it came to the agreement on Curricula. Curricula are controversial and keenly fought over because they seek to define what is taught to children. "Creationism" vs "Darwinism" is just one example of the kinds of topic that can and do raise the temperature of the debate. Only two subjects have been tested in every period since 1993, because the seven curricula took 10 years to develop and individual curricula have been revised along the way.

Political realities have tempered the impact of the competency determination standard which has been fudged in Massachusetts with special re-tests scaled differently to the mainstream tests to allow an acceptable pass rate.

The failure of standards reform in Massachusetts follows earlier failures of similar reforms such as the scientific accountability reforms in Michigan in the early 1970's about which Murphy and Cohen (1974) concluded:

> It may be that with time, research and modest field trials, things can be improved. Certainly an effort ought to be made. But if Michigan is any guide, at this point scientific accountability hardly merits full-scale implementation.

Neither money nor standards reforms appear to be panaceas for education. The one relationship that appears to be certain is between Socio-Economic Status and educational outcomes and it would seem that the future of educational research should lie in understanding this relationship and in the development of good criterion referenced testing programs to provide meaningful measurements upon which to base assessment of the effectiveness of different programs and reforms.

APPENDICES

APPENDIX A

DATA ENVELOPMENT ANALYSIS

This Appendix examines at the models and variants to models that taken as a whole are considered to be Data Envelopment Analysis ("DEA").

A.1 Koopmans' efficient frontier and Farrell's Radial Measure of Efficiency

Koopmans (1951) modeled an entire economy as a convex polyhedral cone. He defined an efficient point and a frontier derived from repeated use of this definition. This frontier is referred to as a Pareto-Koopmans efficient frontier in DEA literature. Koopmans gives an economic interpretation of these efficiency conditions in the following theorem:

> THEOREM 4.7: A necessary and sufficient condition that the activity vector x shall lead to an efficient point y = Ax in the commodity space is that there exists a vector p of positive prices such that no activity in the technology permits a positive profit and such that the profit on all activities carried out at a positive level be zero.

At any particular point, j, on the frontier, s outputs y_{rj} (for r = 1 to s), are derived from the use of m inputs x_{ij} (for i = 1 to m). If the vector p of prices is split between outputs prices μ_r and inputs prices ν_i then the theorem translates into:

$\text{Profit}_j = 0 = \sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i=1}^{m} \nu_i x_{ij}$ for a unit, j, on the efficient frontier.

If Koopmans was concerned with defining the efficient surface of a production possibilities set, then Debreu (1951) and Farrell (1957) were concerned with measurement of the difference between efficient and inefficient points in a production possibilities set estimated by reference to actual observations.

A.2 The Charnes Cooper Rhodes ratio model. (CCR)

Charnes et al. (1978) formulated an approach to the solution to Farrell's measurement of efficiency for each of a number of observed units using Farrell's approximation to the theoretical efficient frontier based upon Koopmans definition of efficiency. The ability to use the simplex algorithm to derive efficiency rankings based upon observed inputs and outputs proved to be wildly popular. By 1996 Seiford (1997) had published a bibliography for DEA, which contained over 800 published articles and dissertations related to DEA, excluding working papers and technical reports.

A.3 Other DEA Models and Extensions

In "Components of efficiency evaluation in data envelopment analysis" Ali et al. (1995) give a very clear, readable and useful framework for the classification of the various DEA models using three components of the models:

1. The form of envelopment surface (constraints in the dual)
2. The orientation (objective function in the primal)
3. The pricing mechanism implicit in the multiplier (price) lower bounds.

We will look at each of these in turn and discuss the models and extensions to models that have been presented:

A.3.1 The Envelopment Surface

The envelopment surface in the CCR model is a convex polyhedral cone similar to Koopmans surface. It rests on actual observations and comprises the convex combination of these observations in the smallest set possible. It assumes or implies constant returns to scale.

Models allowing for increasing, decreasing, constant or varying returns to scale result in different definitions of the envelopment surface. The second DEA model, Banker et al. (1984), is known as the BCC model after the authors of the paper that ushered it in; Banker, Cooper and Charnes. It relaxed the assumption of constant returns to scale. Hyperplanes that form this envelopment are allowed to avoid passing through the origin, so that, rather than being a convex polyhedral cone, the envelopment surface is a piecewise convex hull. The BCC model assumes increasing, constant and decreasing returns to scale in that order, as scale increases. See Tone (1996), and Seiford and Zhu (1999a) for reviews of the different methods proposed for measurement of returns to scale.

According to Golany and Yu (1997), the BCC model assumes a unique optimal solution to the LP formulation for each of the units under consideration. However, as Seiford and Zhu (1998) point out, degeneracy arises and when the efficient units are at the corners of many hyper planes on the efficient surface, each has a set of possible virtual multipliers sets associated with it. Since the estimates of returns to scale in DEA depend on $\sum_{j=1}^{n} \lambda_j^*$, the sum of the virtual multipliers, this impacts on the estimates of RTS.

The Multiplicative Model of Charnes et al. (1983a) transforms the observed data using logarithms, which, as with Logistic Regression, results in a linear estimation which when transformed back is a non-linear estimation. The CCR process is applied to the transformed observed data and therefore this model suffers many of the shortcomings of the CCR model.

Charnes et al. (1983b) present a units invariant version of the Multiplicative model. Units invariance is the desirable property that the ratio of weighted inputs and weighted outputs resulting from the linear program should be the same, regardless of the units of measurement used for the observations. See Pastor and Knox-Lovell (1995), and Coelli (1998) for more details. Banker et al. (1981), present a Bi-Extremal version of the Multiplicative model. The Multiplicative is useful in cases where the Cobb-Douglas form of technology is assumed.

Petersen (1990) suggested relaxing the convexity postulate and replacing it with two separate convexity postulates one for the input space and the other for the output space. Convexity of the technology requires the two spaces to be convex. Convexity of both spaces however does not imply convexity of the technology. With both input and output spaces, separately convex, the technology can be quasi-convex. Bogetoft (1996) found problems with Petersen's implementation, which allows a frontier with a series of switches between types of returns to scale. He provides an alternative method, as does Chang (1999), who forces quasi-convexity by first ordering units by input/output level. Bogetoft et al. (2000) present a recursive method. Banker and Thrall (1992) partition the optimal frontier into three parts corresponding respectively to increasing, constant and decreasing returns to scale thereby achieving a "quasi-convex" analysis of the data.

Andersen and Petersen (1993) use a concept of super-efficiency, which leaves each unit, as it is being evaluated, out of the production possibilities set, such that: if the unit would be on the efficient frontier in the regular CCR model, it is compared to the convex combination of its nearest neighbors and returns a score of more than 100% efficient. Super-Efficiency models are not always feasible. Thrall (1996), Zhu (1996), and Seiford and Zhu (1999b) provide a set of theorems which demonstrate when a Super-Efficiency Model will or will not be feasible. Dula and Hickman (1997) also provide insights into the issues inherent in the super-efficiency approach.

Bessent et al. (1988) present Constrained Facet Analysis ("CFA") in which the non fully dimensional facets are identified through analysis and adjustments are made for inefficient units. Lang et al. (1995) present Controlled Envelopment Analysis to address the issue that not all units will be either naturally or quasi-enveloped, as required by CFA.

The Free Disposal Hull (FDH) model of Deprins et al. (1984) consists of a yet smaller production possibilities set since it relaxes Postulates 3, 4 and 6. The surface is full of right angles, like a surface built with Lego™ bricks in a three dimensional space. Thrall (1999) makes the point that FDH is not compatible with a real world economy where values (prices and costs) are considered important. Note that all FDH prices are either zero or infinite.

A.3.2 Orientation

Implicit in Farrell's use of an isoquant is the idea that efficiency in input use is measured relative to a fixed level of output, or that input use is allowed to vary whilst output is held constant. Using an isocost would have implied that the level of output achievement is measured relative to a fixed level of input or that output use is allowed to vary whilst input is held constant.

Transforming the fractional programming problem to a linear programming problem by setting the weighted value of observed inputs (or outputs) equal to one; results in a linear program which has an orientation towards measurement of efficiency relative to outputs (or to inputs). The CCR model therefore comes in two variants, the input-orientated model and the output-orientated model. The measurement of efficiency from one is the mathematical inverse of the measurement of efficiency from the other. The rank ordering is the same from either model, but the prices and price ratios may be different.

Orientation is given effect in the linear programs by the objective function and by forcing either the weighted input or outputs to equal one. Again, recall that the objective function of the CCR dual given in Section 2 was the weighted sum of the inputs.

One non-radial measure is the non-proportional Russell Measure introduced by Färe and Knox-Lovell (1978). It measures the maximum sum of individual input reduction consistent with observed production, which places the unit on the efficient frontier. Implicit in the Russell measure is the assumption that all inputs have equal input weights. Thanassoulis and Dyson (1992), and Ruggiero and Bretschneider (1998), use a first stage regression to estimate factor weights to be included in the objective function of the "Weighted Russell Measure".

A.3.3 Non-Orientated Models

Non-oriented models are also non-radial measures in that efficiency is not measured by reference to radii from the geometric origin. Charnes (1985) introduced a non-oriented model, which came to be known as the Additive Model. Slacks were maximized to minimize resource usage and maximize output at one and the same time. The weighted outputs (or inputs) were no longer scaled to one. Green et al. (1997) gave a non-linear modification to this model, which produces efficiency scores between zero and one. In 1996, Bardhan et al. (1996) presented a mixed integer model in which the comparison of inefficient units is with the actually observed efficient units rather than with convex combinations of units. Brockett et al. (1998) suggested a model in which the adjustment to the slack in the original Additive Model's objective function is not the absolute amount out the factor, but the range between the maximum and minimum observations of the factor:

Pastor et al. (1999) introduced the Russell Graph Measure of Technical Efficiency that extends the previous Russell Measures to simultaneously account for the inefficiency in both inputs and outputs. It is a non-linear program whose computation is hard and whose solution is not easily obtained. The authors suggest the use of a different objective function, which allows LP computation in what they will call the Enhanced Russell Graph Efficiency Measure.

Tone (2001) presents an oriented model, which is units invariant, monotone and reference set dependent, but not translation invariant. Cooper et al. (2000) allow the slack variables to be "free" in the Additive model in order to obtain a model in which improvement in some inputs or outputs can be offset by worsening in other inputs or outputs beyond the frontier established by the convex combinations of observed units.

A.3.4 The Pricing Mechanism

Whether they are called price constraints, weights restrictions or bounds; ever since the introduction of the non-archimedian infinitesimal, epsilon, which had the effect of placing lower bounds on the weights in the dual of the CCR; researchers have desired, for one reason or another, to restrict the values that the weights can take on in oriented models. For a good review of this topic see Allen et al. (1997).

In 1988 Dyson and Thanassoulis (1988) placed lower bounds on the prices or weights to try to ensure that all the factors in their model contributed to the result of the analysis. Charnes et al. (1989) and (1990) generalized the CCR model with the Cone Ratio Data Envelopment model. Beasley and Wong (1990) expressed the weights restrictions as the ratios of the weights. Thompson et al. (1990) define Assurance Regions ("AR"), which are weights restrictions with upper and lower bounds that exclude vectors with unreasonable input and output multipliers. ARI consists of separate output cones and input cones. ARII adds in linkage constraints that link together prices in the output space and the input space. These linkages make it possible to define a meaningful single numeraire and thus to have a single unit of account for the prices of the inputs and of the outputs. Roll et al. (1991) developed their own model and more importantly they discuss the issues to be taken into account when selecting bounds.

Cook et al. (1992) use lower bounds on prices to "break the tie" between technically efficient units, in what amounts to Multi-Criteria Decision Making. They also present a model with weak ordinal weights constraints, seeking to force variables into the model in a particular order of importance. They refine this model, further, to maximize the difference between the weights whilst maintaining the ordinal ranking.

Rather than placing constraints on the prices or weights in order to impact the efficiency ratings emerging from the model, Doyle and Green (1994) and (1995) evaluate the efficiency of each DMU relative to the average efficiency of the each unit evaluated at the prices of all the other units. Green et al. (1996) develop an idea from Chang and Guh (1991) to implement price constraints by replacing the epsilon in the CCR model with “a data dependent finite magnitude” and present a three-stage process that achieves this.

Huang et al. (1997) provide a method by which efficient hyperplanes can be identified and a means of estimating the slopes on the hyperplanes. Friedman and Sunuany-Stern (1997) use canonical correlation analysis to find weights. Thanassoulis and Allen (1998) show the equivalence, for CCR and BCC models, of adding a reduced set of “dummy” DMUs to the model to various assurance region constraint sets. Wei and Yu (1997) use K-cones to place restrictions on DEA solutions.

A.4 Decomposition of Efficiency Scores

Farrell’s measure of Overall Efficiency comprised Allocative or Price Efficiency, and, Technical Efficiency. Byrnes et al. (1984) decomposed Farrell’s measure of Technical Efficiency into Scale Efficiency; Congestion[27]; and, Purely Technical Efficiency. Hjalmarsson and Veiderpass (1992) describe the Malmquist Index, which can be decomposed into two components – see Färe et al. (1992) for more details of the decomposition:

- MC – the change in efficiency relative to the frontier(s).
- MF – the change in the frontier itself over time.

Sueyoshi (1995) looked at panel data and gave a new measure referred to as 'Overall Time Efficiency', which is further broken down into four efficiency concepts: overall efficiency, price efficiency, scale efficiency and time efficiency. He proposes a DEA approach to empirically measure these new efficiency concepts.

---

[27] Congestion occurs when reducing usage of a proper subset of inputs holding constant the usage of all remaining inputs, generates an increase in output.

A.5 Computational Issues

A lot of effort has gone into defining better DEA models. A lot of effort has also gone into implementing computational strategies that improve on the implementation of these models or which themselves improve on DEA. Ali and Seiford (1993) addressed the selection of ε for the CCR model. Ali (1993) provides a very clear description of degeneracy in DEA models. Pitaktong et al. (1998), addressed the shortcomings in the Simplex Algorithm with respect to DEA. Where the solution is degenerate, i.e. at corners of the production possibilities set; the objective is to select the solution on a facet that is "fully dimensional". Coelli (1996) developed a multi stage method, which is described by Brown et al. (1999).

Pastor and Knox-Lovell (1995) describe units invariance as:

> A fundamental property of an efficiency measure embedded in a DEA model, or any other performance analysis model, is that it is independent of the units in which the input and output variables are measured.

And translation invariance as:

> A useful property of an efficiency measure is that it be independent of an affine translation of the input an output variables.

Ali and Seiford (1992) show that translation invariance is important to solution models that pre-process the data and in particular because it allows zero values in the model data. Scaling and zeros problems can be addressed by using an affine displacement of the data. Pastor and Knox-Lovell (1995) knew of no model, which satisfied both properties, so they proposed two new models the normalized weighted additive DEA Model, which is both units and translation invariant and the normalized weighted BCC DEA model, which is completely units invariant, and preserves its partial translation invariance; being invariant to a translation of inputs or outputs, but not both.

A.6 Evaluation of DEA

How good an estimate, of a known production function, does DEA provide? There are a number of studies that have used Monte Carlo techniques to set up experimental data with "known" levels of efficiency. This data has then been analyzed using DEA and other methods such as Corrected Ordinary Least Squares Regression ("COLS") and Stochastic Frontier Models.

Banker et al. (1993) compared COLS with DEA and found that COLS performs better for classical distributions and DEA performs better for non-classical distributions in the case of two inputs and one output. Banker et al. (1996) obtained similar results with simulation of a 4 Piece Cobb-Douglas function. Resti (2000) worked with two inputs and three outputs and showed that CCR and BCC models produced good estimates of "actual" efficiency.

Riddington and Cowie (1994) comment that if there are few units, DEA leads to problematic results. Zhang and Bartels (1998) showed that the more units under consideration the lower the average level of efficiency. Tauer and Hanchar (1995) show a similar result with 3 outputs and 15 inputs. When they used 200 firms, 50 percent were found to be 100% efficient and when they used 25 firms 70 percent were found to be 100% efficient in DEA. Smith (1997) used a Cobb-Douglas function as did Pedraja-Chaparro et al. (1999) who concluded that a large sample size (160 units in their case) and a small number of factors (4 factors in their case) were needed before DEA yielded acceptably accurate results.

Gong and Sickles (1992) compare the performance of Stochastic Frontier ("SF") models with DEA models. They find that the choice of functional form appears to be crucial to the performance of SF and that DEA does well with simple technologies, but deteriorates rapidly as the technology becomes more complex.

As a consequence of these results, the largest possible sample of units was used in this dissertation. In addition no more than four inputs and four outputs were used in any of the models.

APPENDIX B

DEA AND K-12 EDUCATION

As we saw in Chapter One, very little can be said about K-12 education based upon the results of Education Production Function studies, from the Coleman Report onward, that used simple regression analysis to relate an outcome to a series of different variables.

In response to this failure some researches adopted more sophisticated parametric econometric techniques such as that proposed by Aigner and Chu (1968). In particular Chizmar and Zak (1983) use a Cobb-Douglas function with two outputs and p inputs:

$$\sum_{i=1}^{2}\alpha_i \log y_i = \beta_0 + \sum_{j=1}^{p}\beta_j \log x_j + \varepsilon$$

together with canonical regression to estimate the parameters to the Cobb-Douglas function.

Montmarquette and Mahseredjian (1989) use a Nested-Error Components analysis of Montreal Schools and come up with the same conclusion, as Coleman, that Socio-Economic Status is the key determinant of outcomes.

Other researchers turned to Data Envelopment Analysis, which was first used by Charnes et al. (1981) to evaluate Program Follow Through ("PFT"), a large-scale social experiment in public school education. The research was designed to test the advantages of PFT relative to designated Non-Follow Through counterparts in various parts of the U.S. They used the CCR model to distinguish between "management efficiency" and "program efficiency".

Bessent and Bessent (1980) applied the original CCR model to data from 55 elementary schools in an urban school district. 31 schools were found to be efficient and

24 were found to be inefficient. School management was encouraged to look at the information available from slack and to consider the opportunity costs of these slacks.

In common with the statistical production function studies this study had no educational theory as a foundation for the choice of outputs and inputs. The Output Measures used came from the California Achievement Test:

- Y1 the median percentile reading achievement for only those pupils in attendance at the school for a full year; and,
- Y2 the median percentile mathematics achievement test score for only those pupils in attendance for a full year.

The Input Measures used were

- (X1 and X2) the results of the same tests from the previous year;
- proxy measures for neighborhood and home conditions – X3 percent Anglo-American students, X4 percent students not from low income families, – X5 percent in average daily attendance, and X6 mobility index – (total enrollment – number entered late or withdrawn)/ total enrollment;
- proxy measures for within school conditions – X7 the number of professional staff per 100 pupils and X8 the total per pupil expenditure for instruction;
- School organizational climate indicators – X9 esprit – job satisfaction. X10 intimacy – how much social interaction exists among teachers, X11 trust-teachers motivated by principal's personal example of work orientation, and X12 consideration-the principal's friendliness and cooperativeness with teachers;
- Measure of classroom instructional processes – X13 – the total individualized instruction index.

According to Zomorrodian (1990):

> Splitek (1981) used DEA for 497 elementary schools in Texas; 28% were efficient and 72% were inefficient. For inefficient schools he identified three areas of inefficiency: first, financial resources as measured by instructional expenditures; second, the student resources as measured by the percent of non-title I students in the school; and finally, teaching resources as measured by the level of experience that teachers bring to schools.

Bessent et al. (1982) contributed another study. The objectives of the study were to assess the relative efficiency of each of the 167 schools in the Houston Independent School District and to:

> provide management with information like the following:
>
> (1) An efficiency evaluation of individual schools that would include the productivity of professional staff of the school while making allowance for the conditions under which they were operating.
>
> (2) Targeted output goals and identification of needed input modifications.
>
> (3) Identification of areas in which efficiency could be increased.

Two separate sets of analyses were conducted; one looking at children in 3rd grade and the other at children in the 6th grade. This time the outputs were 3rd and 6th Grade Iowa Test of Basic Skills (ITBS) composite scores, with the first two inputs being the 2nd and 5th grade scores of the children. The other inputs were similar to those in Bessent and Bessent (1980) listed above. As before, the CCR Ratio Model was used. The study revealed no significant differences between the measured efficiency and effectiveness of each school at the 3rd and 6th grades. Besent et al. (1982) remarked that:

> The Validity of DEA was assessed in an informal test in which the General Superintendent and his staff first identified 'trouble' schools and 'outstanding' schools before looking at the DEA results. In the two-hour session approximately 40 of the 167 schools were reviewed in detail and a check of administrative assessment against DEA solutions resulted in 100% correct classification; i.e., the 'trouble' schools were all found by DEA to be inefficient to some degree and the 'outstanding' schools were all found to be efficient. Equally important, the reason's for a school's status based on known local conditions generally coincided with the DEA slack values.
>
> Again, according to Zomorrodian (1990):
>
> Bessent, Bessent, Charnes, Cooper and Thorogood (1983) used DEA to measure the efficiency of 22 programs in San Antonio College in Texas; of the 22 programs, 8 were efficient and 14 inefficient. The researchers provided samples of how DEA may be used in planning and decision making, augmenting existing programs, starting new programs, and combining existing programs.

Zomorrodian (1990) used MEAP scores to assess how to improve efficiency in elementary schools in western Massachusetts. He used MEAP Reading, Math, Science and Social Studies average scores for fourth grade students in 1988 as the outputs and 16 inputs chosen by reference to educational input/output studies (i.e. the usual suspects).

In an analysis of 122 school districts in Connecticut, Ray (1991) used the 9th grade district average scores for students in the State Wide Proficiency tests administered in October 1980 and DEA to produce an index of efficiency of Connecticut's public school districts, which he used as the dependent variable and the socio economic factors as the independent variables in a regression analysis. The result is that DEA Scores are explained in large measure by socio economic factors. So, Ray was able to confirm the Coleman Report's main assumption and the often unstated assumption of all production function research in education: that Socio Economic Status is the key determinant of educational outcomes.

Duncombe and Yinger (1993) used DEA to assess scale economies and technical efficiency in the 692 schools forming the New York State public school system. The outputs were 6th grade math and reading scores from Pupil Evaluation Program (PEP) tests administered to children by the New York State Department of Education. They differentiated between discretionary inputs, those under the school's control, and fixed factors and used a BCC "Input Oriented" model with fixed inputs, which allows factors which are deemed to be outside of the control of the school be included in the model.

The discretionary inputs were – Numbers of Aides and Assistants; Books (per 10,000 pupils); Computers in 3rd Grade; Number of Administrators; Number of Para-professional Staff; Percentage of Teachers in 6th Grade with advanced education; Percentage of Teachers in 3rd Grade with advanced education; Number of Classrooms; Number of years of teaching experience, and Number of Videos. The fixed factors were – Income in district; English proficiency level; Percent non-minority; Percent Not Chapter 1, and Percent not on welfare.

In the standard "Input Oriented" BCC model "slack variables for all inputs appear in the objective function. In the modified model, input slack variables appear only if the corresponding input is discretionary. Further, the constraints for the fixed inputs are modified to eliminate the dependency of the fixed inputs on the efficiency measure". The result is that a district is compared to districts that collectively face an environment as harsh as their own and more districts are identified as efficient.

In the United Kingdom, Thanassoulis and Dunstan (1994), and Thanassoulis (1996a) showed how data envelopment analysis (DEA) can be used to guide secondary schools to improved performance through role-model identification and target setting in a way that recognizes the multi-outcome nature of the education process and reflects the relative desirability of improving individual outcomes. The inputs were: Mean Verbal Reasoning score per pupil and Percentage not receiving school meals. The outputs were: Average GCSE score per pupil and Percentage students not employed after GCSE's. The 1978 CCR model was used to identify "Peer" schools and to set performance targets for different schools.

Arnold et al. (1994) took another look at Texas public schools. They looked at 647 secondary schools of which 100 were found to be fully efficient.

Bonnesronning and Rattso (1994) looked at the relationship between resource use and student achievement in 34 Norwegian high schools. The study is an education production function study benefiting from student level as well as school level data. It measures the change between achievement at junior school and high school. It splits students into Low Achievers/High Achievers and Schools into Low Value Added/High Value Added. The results are consistent with all the other educational production function studies.

Oleson and Petersen (1995) looked at including quality into DEA analysis using education as a vehicle to motivate the approach. Units are allowed to use more inputs to produce higher quality, which is represented in the model as either:

1. the cumulative quantities of outputs produced at or above any given level of quality;

2. the cumulative probabilities that units of a given output can be produced at or above any given level of quality.

They then incorporate bounds on the virtual multipliers that in effect require older students to do better than younger students. The modeling is impressive, but the educational conclusions are not.

Chalos and Cherian (1995) used DEA to develop a process by which the determination of financial policy in education is analyzed. Analysis was of 207 school districts in Illinois. Outputs were Reading Score Grade 6; Reading Score Grade 8; Math Score Grade 6, and Math Score Grade 8. Inputs were percent non-minority; percent non-low income; attendance rate; operating expenditure per pupil, and percent of teachers with a master's degree.

The authors were concerned that a system of control should consider the causality, substitutability and controllability of both input and output indicators. They point to several problems that bedevil attempts to meaningfully interpret input-output indicators of performance. Test score comparisons of affluent and low-income school districts for example make cross-sectional production function comparisons invidious. Equally important, they say, is the politicization of public sector expenditures. They point to public choice literature and say (citing Horton 1987, p. 380) that it suggests that public sector managers are more inclined to meet budgets or creatively manage budgetary slack rather than pursue cost minimization and output maximization.

> The results, have both operational and public policy implications. First, efficiency measures provide worthwhile information above and beyond output measures of performance. Such measures are extremely useful, as illustrated, in setting budgetary operational goals and providing feedback with respect to resource utilization relative to other comparable socio-economic units.
>
> Second, the results provide additional evidence with respect to the ongoing controversy surrounding the issue of fiscal equity in education. Local property revenues and tax bases were positively associated with district efficiency and effectiveness. This suggests possible inequities (corroborated by demographic trends) in educational opportunities across school districts.

Thanassoulis (1996a) and (1996b) proposed models that would compare the distribution of educational outcomes not just the school or school district's average scores. The model seeks to compare cohorts with 11+ exam (taken at 11 years of age) results as inputs and GCSE (taken at 16 years of age) exams as outputs. He did not have any data to run the proposed models at the time of publication, so the results are moot, although the idea is clearly interesting.

Arnold et al. (1996) used five outputs and eight inputs from 638 secondary schools in Texas to illustrate new uses of DEA and statistical regression analysis. A copy of this document could not be located, so reliance is placed on the abstract in which the authors claim that the "results are consistent with what might be expected from economic theory and are informative for educational policy uses."

Ruggiero (1996a) is interested in the effect of including environmental variables in his modified model and comparing this with an existing model which he says "purportedly controls for exogenous factors". Perhaps not surprisingly, he finds that his own model is "more efficient in controlling for exogenous factors".

Ruggiero (1996b) identifies Koopmans Efficient units using DEA. Then he applies Canonical Regression to the Koopmans efficient units in a process that he says contributes a multi-frontier DEA model that fits the special nature of public production in a way that could not be identified from the paper.

Bates (1997) shows us how little progress has been made in the forty years since the Coleman Report. He provides a method to adjust for socio-economic status that reduces the number of efficient schools reported by DEA. It seems that what is being explained – the probability of high/low scores – is being incorporated in the explanation.

Ruggiero and Bretschneider (1998) use education to evaluate the weighted Russell measure of technical efficiency.

Anderson et al. (1998) emphasize the overall performance of the entire set of elementary schools in the Chicago Public School System, following the fundamental school reforms of 1988-1989, rather than on the resource use in each school. Regression Analysis provided measures of effectiveness from grade equivalents from The Iowa Test of Basic Skills in reading, math, and vocabulary on the same cohort in the previous years for the third and eighth grades. DEA provided measures of efficiency. The results suggest encouraging improvement in schooling outcomes over the period of study.

Noulas and Ketkar (1998) used the 1978 CCR model, in their study of public schools in the state of New Jersey, and found that wealthier school districts had an average efficiency score of 0.88 while the most needy school districts had an average efficiency score of 0.63. When they adjusted for socio-economic factors the difference between the two groups' average efficiency scores became smaller. The results are entirely predictable.

Kirjavainen and Loikkanen (1998) studied efficiency differences in 291 Finnish Senior Secondary Schools using four DEA model variants and Tobit Analysis. Schools with small classes and heterogeneous student bodies were inefficient whereas school size did not affect efficiency. Surprisingly, private schools were inefficient relative to public schools. When parents' educational level was only included in the Tobit model, it affected efficiency positively:

> Usually the input variables used in the DEA models are such that they are controllable by the school or school district depending on the level of analysis. However, one of the most significant and robust results of input-output studies has been that students' socio-economic status affects achievement.

Ruggiero (1999) again addresses the issue of non-discretionary inputs in the public sector. He develops a flexible non-parametric methodology for estimating cost and cost efficiency and extends DEA to measurement of cost for 584 New York school districts. It is found that nearly 64% of districts are cost inefficient, spending on average $1,200 per pupil above the cost minimizing level. In addition, it is estimated that the average school district faces environmental costs of over $1,700 per pupil.

Grosskopf et al. (1999) approached education with a slightly different slant using LP techniques commonly used in DEA. "Revenue Maximization" subject to a budget constraint is used to estimate the desirable redistribution of inputs. They modeled the situation prior to reform using an output distance function to determine the best practice frontier given observed allocations of inputs. Then they modeled the situation after reform using the cost indirect output distance function allowing schools to choose their allocations of inputs using only a budget constraint:

Our analysis of urban, public school districts in Texas indicates that most school districts would realize substantial gains from reform. The simulation also indicates that school administrators, teachers, and professional staff (such as counselors) are likely to lose employment, while teacher aides are likely to gain employment. Finally, the simulation reveals that school districts that would not gain from reform have a greater proportion of minority and low-income students, less property wealth per pupil, and lower per pupil expenditures.

Ruggeiro and Vitaliano (1999) compare DEA and stochastic frontier regression using data from 520 New York school districts. Mean inefficiency is 14% using either method, and the rank order correlation coefficient of inefficiency between the two methods is 0.86. Student test scores are output variables, and socio-economic control variables are included to adjust for the harshness of the educational environment in which districts operate. In other words, a vanilla education production function study.

Mancebon and Molinero (2000) looked at schools, in Southampton and Portsmouth, in the United Kingdom, to assess the factors that influence their productive efficiency. The data set included 19 variables on 176 schools and was analyzed by means of Data Envelopment Analysis. Contextual variables, not included in the efficiency analysis, were used to explain the sources of inefficiency. It was found that religious orientation, parental influence and level of exclusions all impacted on the ability of a school to deliver the best possible results in standard assessment tests. This study is a pretty standard education production function study with DEA.

Ruggiero (2000) uses a DEA model that allows for environmental factors (a form of non-discretionary) by allowing more than one production frontier to be estimated in a model. The programming models used for public sector applications, however, are based on standard private sector production theory. In the public sector environmental variables have a significant impact on the provision of public services. Without controlling for these environmental factors, point estimates, of efficiency and returns to scale, will be biased. The method is applied to the provision of educational services in New York State school districts for illustrative purposes.

I will leave it to Kirjavainen and Loikkanen (1998) to articulate a sense of the value of most of these investigations to the progress of education.

> As we pointed out in the introduction, it is not quite obvious what the inputs and outputs of educational process are and at what stage (timing) they should be measured. Because of this one should pay attention to the robustness of the results with respect to the choice of input and output variables. In most of the studies, the selection of variables seems to be based more on data availability than any other reason and only the final results are reported.

So, as with educational production function studies using regression analysis, DEA studies of education seem to lack for a theoretical basis. This is a major handicap. As first the bath, then the bathwater and then the baby are added to the mix; the results from DEA get worse rather than better. Applying the lessons of Appendix A, Section 6, we can see that studies with ten or more variables and tens of schools or units of analysis are not going to produce meaningful rankings of efficiency and little else has emerged from the research.

APPENDIX C

THE SAMPLE TOWNS AND REGIONS

The final sample contained the following 141 Towns and 39 Regions.

The towns were Abington, Agawam, Andover, Arlington, Ashland, Attleboro, Auburn, Avon, Bedford, Belchertown, Bellingham, Belmont, Beverly, Billerica, Bourne, Braintree, Brockton, Brookline, Burlington, Canton, Chelsea, Chicopee, Cohasset, Danvers, Dartmouth, Dedham, Douglas, Dracut, Duxbury, East Bridgewater, East Longmeadow, Easthampton, Easton, Fall River, Framingham, Georgetown, Gloucester, Grafton, Granby, Greenfield, Hadley, Hanover, Harvard, Hatfield, Haverhill, Hingham, Holbrook, Holliston, Holyoke, Hopedale, Hopkinton, Hudson, Ipswich, Lawrence, Leicester, Lenox, Lexington, Littleton, Longmeadow, Lowell, Ludlow, Lynn, Lynnfield, Marlborough, Marshfield, Maynard, Medfield, Medway, Methuen, Middleborough, Milford, Millbury, Millis, Milton, Monson, Nantucket, Natick, Needham, Newton, North Andover, North Attleborough, North Brookfield, North Reading, Northbridge, Norton, Norwell, Norwood, Oxford, Palmer, Peabody, Pittsfield, Quincy, Randolph, Reading, Revere, Rockland, Rockport, Salem, Sandwich, Saugus, Scituate, Seekonk, Sharon, Shrewsbury, Somerville, South Hadley, Southbridge, Springfield, Stoneham, Stoughton, Sutton, Swansea, Taunton, Tewksbury, Tyngsborough, Uxbridge, Wakefield, Walpole, Waltham, Ware, Wareham, Watertown, Wayland, Webster, Wellesley, West Boylston, West Bridgewater, West Springfield, Westborough, Westfield, Westford, Weston, Westport, Westwood, Weymouth, Wilmington, Winchendon, Winchester, Winthrop, Woburn, and Worcester.

The Regions were Acton-Boxborough; Acushnet, Fairhaven and New Bedford; Adams, Cheshire and Savoy; Athol-Royalston; Berkley and Somerset; Berlin-Boylston; Blackstone-Millville; Brewster, Eastham, Orleans, Provincetown, Truro and Wellfleet; Bridgewater-Raynham; Carlisle-Concord; Central Berkshire Regional comprising Becket, Cummington, Dalton, Hinsdale, Peru, Washington and Windsor; Clarksburg, Florida, Monroe, and North Adams; Hancock, Lanesborough, New Ashford, Richmond, and Williamstown; Dighton-Rehoboth; Dover-Sherborn; Essex, Hamilton, Manchester and Wenham; Falmouth and Mashpee; Freetown-Lakeville; Frontier Regional comprising Conway, Deerfield, Sunderland and Whately; Gateway Regional comprising Blandford, Chester, Huntington, Middlefield, Montgomery, Russell and Worthington; Gill-Montague with Erving; Berkshire Hills Regional comprising Great Barrington, Stockbridge and West Stockbridge; Groton-Dunstable; Hampden-Wilbraham; Lincoln-Sudbury; Masconomet comprising Boxford Middleton and Topsfield; Mendon-Upton; Nahant and Swampscott; Narrangasett Regional comprising Phillipston and Templeton; Northboro-Southboro; Old Rochester Regional comprising Marion, Mattapoisett and Rochester; Pioneer Valley Regional comprising Bernardston, Leyden, Warwick and Northfield; Plymouth-Carver; Quabbin Regional comprising Barre, Hardwick, Hubbardston, New Braintree and Oakham; Quaboag Regional comprising Warren and West Brookfield; Southern Berkshire Regional comprising Alford, Egremont, Monterey, Mount Washington, New Marlborough and Sheffield; Spencer-East Brookfield; Tantasqua Regional comprising Brimfield, Brookfield, Holland, Sturbridge and Wales; and Whitman-Hanson.

APPENDIX D

DETAILED DESCRIPTION OF SCHOOL DISTRICTS

This appendix gives details of the school districts with detailed descriptions of the reasons for inclusion or exclusion from the sample.

D.1 Current Academic Regional School Districts

Current Academic Regional School District geographies have been retained and used, where possible, to define the geographic units used in the analysis. The next four pages contain descriptions or these regions. If test score or other data was missing this is reported and the region was dropped from the analysis.

- Ashburnham-Westminster has combined the K - 12 students from the two towns since before 1988 and sends vocational students to Montachusett Voc Tech Regional.
- Athol-Royalston has combined the K - 12 students from the two towns since before 1988 and sends vocational students to Montachusett Voc Tech Regional.
- Great Barrington, Stockbridge and West Stockbridge have formed Berkshire Hills Regional since before 1988 and participate in no Vocational or Agricultural Region.
- Blackstone-Millville has combined the K - 12 students from the two towns since before 1988 and sends vocational students to Blackstone Valley Regional.
- Lincoln-Sudbury has combined the grade 9 - 12 students from the two towns since before 1988. Both towns send vocational students to the Minuteman Vocational Tech Regional district.

- Bridgewater-Raynham combines the K - 12 students from the two towns and sends vocational students to Bristol-Plymouth Voc Tech. Raynham sends students to the Bristol Country Agricultural Regional. In 1988 and 1990 the towns combined at the High School level only.
- Dennis-Yarmouth combines the K - 12 students from the two towns and sends vocational students to Cape Code Region Voc.
- Dighton-Rehoboth combines the K - 12 students from the two towns. They do not form part of a Vocational Regional, but send students to Bristol County Agricultural.
- Dudley-Charlton combines the K - 12 students from the two towns, which form part of the Southern Worcester County Voc Tech. Grade 4 scores from 1988 and 1990 MEAP are missing so Dudley Charlton was dropped from the analysis.
- Masconomet Regional School District has provided grades 9 - 12 to Boxford and Topsfield since 1988, and to Middleton since 1990. All three send vocational students to the North Shore Regional Voc Tech.
- King-Philip Regional has provided for middle and high school students from Norfolk, Plainville and Wrentham since before 1988. The three towns send vocational students to the Tri County Vocational Tech Region.
- Blandford, Chester, Huntington, Middlefield, Montgomery, Russell and Montgomery have formed Gateway Regional for grades K - 12 since before 1988. They do not form part of either a Vocational or an Agricultural Regional.

- Groton-Dunstable has combined the K - 12 students from the two towns since before 1988. Dunstable sends vocational students to Greater Lowell Voc Tech and Groton sends students to Nashoba Valley Tech.
- Hampden-Wilbraham has combined the K - 12 students from the two towns since Financial Year 1995. Prior to that the region provided high school education for the two towns. Neither town sends students to a Vocational Regional or and Agricultural Regional.
- Mendon-Upton has combined the K - 12 students from the two towns since before 1988. They send vocational students to the Blackstone Valley Regional Voc Tech.
- Phillipston and Templeton have formed Narrangasett Regional for students in grades K - 12 since before 1988. They send Vocational students to Montachusett Voc Tech Regional.
- Bolton, Lancaster and Stow have formed the Nashoba Regional School district since before 1988 and send Vocational students to Minuteman Voc Tech Regional.
- Ashby, Pepperell and Townsend have formed North Middlesex Regional for grades K - 12 since before 1988. Ashby sends vocational students to Montachusett Voc Tech Regional and Pepperell and Townsend send students to Nashoba Valley Tech.
- Groveland, Merrimac and West Newbury have formed Pentucket Regional for grades K - 12 since before 1988 and send students to Whittier Voc Tech.

- Bernardston, Leyden, Warwick and Northfield have formed Pioneer Valley Regional since before 1988 for grades K - 12 and send students to Franklin County Regional.
- Barre, Hardwick, Hubbardston, New Braintree and Oakham have formed Quabbin Regional at grades K - 12 since before 1988. Barre and Hubbardston send students to Montachusett Voc Tech Regional and Hardwick and New Braintree send students to Pathfinder Voc Tech. Oakham is not associated with a Vocational Regional.
- Warren and West Brookfield have formed Quaboag Regional for grades 7 - 12 since before 1988 and for grades K - 12 since 1999. They are not associated with a Vocational or an Agricultural Regional District.
- Spencer-East Brookfield has combined the K - 12 students from the two towns since before 1988. Neither town sends students to a Vocational Regional or and Agricultural Regional.
- Acton-Boxborough has combined grades 7 - 12 from the two towns since before 1988. Both towns send students to the Minuteman Vocational Regional.
- Newbury, Rowley and Salisbury have formed Triton Regional School District for grades K - 12 since before 1988; and send students to Whittier Voc Tech.
- Holden, Paxton, Princeton, Rutland and Sterling have formed Wachusett Regional School District for grades K - 12 since before 1988. Holden, Princeton and Sterling send vocational students to Montachusett Vocational Technical Regional and Rutland sends vocational students to Southern Worcester County Vocational Technical. Paxton is not affiliated with a vocational regional.

- Whitman-Hanson combines the K - 12 grades from the two towns, which send vocational students to the South Shore Regional Vocational. In 1988 and 1990 the two towns shared the high school only and MEAP scores were recorded at the town level for 4^{th} and 8^{th} grade students.
- Halifax, Kingston, Pembroke and Plympton have formed Silver Lake Regional School District for Grades 7 - 12, since before 1988. Pembroke – 46% of enrollment in Grade 4 is missing 1988 MEAP Grade 4 Scores so Silver Lake was dropped from the analysis.
- Marion, Mattapoisett and Rochester have formed Old Rochester Regional School District for Grades 7 - 12, since before 1988.
- Conway, Deerfield, Sunderland and Whately have formed Frontier Regional School District for Grades 7 - 12, since before 1988.
- Brimfield, Brookfield, Holland, Sturbridge and Wales have formed Tantasqua Regional for Grades 7 - 12 since before 1988.
- Alford, Egremont, Monterey, Mount Washington, New Marlborough and Sheffield have formed Southern Berkshire Regional for grades K - 12 since before 1988 and do not send students to a Vocational or Agricultural Regional. The six towns have been treated as a single region.
- Becket, Cummington, Dalton, Hinsdale, Peru, Washington and Windsor form the Central Berkshire Regional. They do not form part of either a Vocational or an Agricultural Regional. For the purposes of this analysis it makes sense to treat these towns as a single region.

D.2 Historic Regions or Unions of School Districts

Other "regions" were constructed so that data from historical regions or unions of school districts or other informal arrangements could be used consistently over time. These constructed regions are described in the next four pages. Again, where data was missing, this is reported and the "region" was dropped from the analysis.

- In the late 1980s and early 1990s Acushnet, Fairhaven and New Bedford combined at the high school level.
- In the late 1980s and early 1990s Ayer, Lunenberg and Shirley combined at the high school level so they were treated as a single region.
- In the late 1980s and early 1990s Mashpee sent high school students to Falmouth, so Falmouth and Mashpee were treated as a single region.
- Gill-Montague has combined the K - 12 grades from the two towns since before 1988 and sends vocational students to Franklin County Region Voc. Erving has sent grades 7 - 12 to Gill-Montague since before 1988 so Gill-Montague and Erving was treated as a single region.
- Plymouth-Carver de-regionalized in Financial Year 1993. Prior to that it combined the students from the two towns in middle and high school. Plymouth and Carver have been combined for the purposes of analysis.
- In the late 1990s Berkley sent high school students to Somerset. Berkley and Somerset have been combined into one region for the analysis. Berkley sends vocational students to Bristol-Plymouth Voc Tech and Somerset to the Greater Fall River Voc Tech Regional.

- Brewster, Eastham, Orleans, Truro and Wellfleet now form Nauset School District for grades 6 - 12. In 1988 and 1990 Truro sent its students to Provincetown. Brewster, Eastham, Orleans, Provincetown, Truro and Wellfleet have been treated as a single region.
- Amherst-Pelham has combined the students in grades K - 12 from Amherst and Pelham since before 1988. New Salem and Wendell form New Salem-Wendell for grades K – 6 and send Vocational Students to Franklin County Region Voc. Ralph C. Mahar Regional comprises New Salem, Orange, Petersham and Wendell for grades 7 - 12. Prior to Financial Year 1994, Leverett and Shutesbury sent K – 6 students to New Salem. Leverett and Shutesbury now send high school students to Amherst-Pelham. It would make sense for the purposes of this analysis to treat Amherst, Leverett, Orange, New Salem, Pelham, Petersham, Shutesbury and Wendell as a single region but data is missing for 1988 so these towns were dropped from the analysis.
- Chesterfield, Goshen, Southampton, Westhampton and Williamsburg have formed Hampshire Regional for grades 7 - 12 since before 1988. Chesterfield and Goshen send K - 6 students to Chesterfield-Goshen Regional and they do not form part of either a Vocational or an Agricultural Regional. Southampton, Westhampton and Williamsburg have run their own schools for grades K – 6. For the purposes of the analysis it makes sense to treat Chesterfield, Goshen, Southampton, Westhampton and Williamsburg as a single region. For 1988 the scores for 17 students from Chesterfield-Goshen are missing from 125 students in total for Hampshire at Grade 4. In 1996 the scores for 24 students out of 142

students are also missing. The region is also impacted by the Hilltown Cooperative Charter School and as a consequence it was dropped from the analysis.

- Southwick-Tolland combines the K - 12 students from the two towns. Prior to Financial Year 1991 the district took students from Sandisfield and Granville. Since Financial Year 1999 the district has taken students from Granville. Otis and Sandisfield form the Farmington River Regional for grades K – 6 and send grades 7 - 12 to Lee. Otis and Lee formed a union in the late 1980s. Tyringham sends its students to Lee. None of the towns sends students to a Vocational Regional or and Agricultural Regional. Many MEAP Scores for 1988 and 1990 are missing, so these towns were dropped from the analysis.
- Hamilton-Wenham has combined the K - 12 students from the two towns since before 1988. They send vocational students to the North Shore Regional Vocational. In the late 1980s and early 1990s Essex sent high school students to Hamilton-Wenham, but now Manchester-Essex combines the K - 12 from Essex and Manchester. Essex, Hamilton, Manchester and Wenham all send vocational students to the North Shore Regional Vocational. For the purposes of the analysis it makes sense to treat these four towns as a single region.
- Ashfield, Buckland, Colrain, Heath, Plainfield and Shelburne form the Mohawk Trail Regional. Mohawk Trail became a K - 12 regional school district in 1995. Neither Ashfield nor Plainfield form part of either a Vocational or an Agricultural Regional. Buckland, Colrain, Heath and Shelburne send vocational students to Franklin County Region Voc. Hawley and Charlemont formed Hawlemont

Regional for grades K - 6 and send grades 7 - 12 to Mohawk Trail Regional. Rowe has sent students to Mohawk Trail Regional in grades 7 - 12 since before 1988. Few scores were recorded for the 1988 MEAP or for the 1990 MEAP at grade 4 so these towns were dropped from the analysis.

- Adams, Clarksburg, Florida, Monroe, North Adams, Savoy and Williamstown send vocational students (grades 9 - 12) to the Northern Berkshire Vocational Regional. Savoy sent students to Clarksburg in the late 1980 and early 1990s, but now sends them to Adams-Cheshire. Clarksburg has sent high school students to North Adams since before 1988. Florida has sent students to Clarksburg since before 1988. Mount Greylock Regional provides Grades 7 - 12 to students from Lanesborough, New Ashford and Williamstown. New Ashford sends grades K - 6 to Lanesborough. In the late 1980s and early 1990s Richmond and Hancock formed a union with Lanesborough. Adams-Cheshire has combined the K-12 students from the two towns since before 1988. Adams sends vocational students to Northern Berkshire Vocational Regional, but Cheshire does not. For the purposes of this analysis it makes sense to combine Clarksburg, Florida, Monroe, and North Adams as one single region; Hancock, Lanesborough, New Ashford, Richmond, and Williamstown as another single region and Adams, Cheshire and Savoy (whose total enrollment in 2000-01 was 48) as a third single region.

- Northampton sends vocational students to Northampton-Smith. Northampton-Smith has no MEAP scores (at 12th Grade) for 1988. Northampton has a high percentage of students at the Hilltown Cooperative Charter School and as a consequence Northampton and Northampton-Smith were dropped from the analysis.
- Worcester sends vocational students to Worcester Trade and these two were treated as a single region for the purposes of this analysis. Worcester Trade consolidated with Worcester Public Schools in 1999.
- In the late 1990s Nahant sent high school students to Swampscott and both towns send vocational students to the North Shore Regional Vocational so for the purposes of this analysis it makes sense to treat the two towns as a single region.

D.3 Vocational and Agricultural School Districts

Vocational and Agricultural School Districts operate for grades 9 - 12 only and impact the 12^{th} grade and 10^{th} grade test scores. MEAP Scores are not recorded for Vocational and Agricultural School Districts in 1988, 1990 and 1992. An adjustment was made using the percentage that each adjusted score in 1994 was of the unadjusted score in that year for each town for each discipline. Both adjusted and unadjusted scores were carried forward into the analysis. Little difference in the trends in efficiency was found between scores with or without adjustment for Vocational School districts and the results from adjusted data are not presented. In 1994 and 1996 MEAP scores are recorded for all Vocational Districts, as were MCAS scores for 1998, 1999, 2000, 2001 and for 2002.

No adjustments were made for the three Agricultural School Districts as their catchment areas were wide and so their likely impact on an individual school district is very small. The Vocational Regions are described in the next three pages.

- Bourne, Falmouth, Marion, Sandwich and Wareham send students to Upper Cape Cod Voc Tech Regional.
- Barnstable, Brewster, Chatham, Dennis, Eastham, Harwich, Mashpee, Orleans, Provincetown, Truro, Wellfleet, and Yarmouth send students to Cape Cod Regional Voc Tech.
- Avon, Braintree, Canton, Dedham, Holbrook, Milton, Norwood, Randolph, and Westwood send students to Blue Hills Vocational Region.
- Berkley, Bridgewater, Middleborough, Raynham, and Taunton send students to Bristol-Plymouth Voc Tech.
- Andover, Lawrence, Methuen, and North Andover send students to Greater Lawrence Regional Voc Tech.
- Dracut, Dunstable, Lowell, and Tyngsborough send students to Greater Lowell Voc Tec.
- Acton, Arlington, Belmont, Bolton, Boxborough, Carlisle, Concord, Dover, Lancaster, Lexington, Lincoln, Needham, Stow, Sudbury, Wayland, and Weston send students to Minuteman Vocational Technical.
- Franklin, Medfield, Medway, Millis, Norfolk, North Attleborough, Plainville, Seekonk, Sherborn, Walpole, and Wrentham send students to Tri-County Voc Tech Regional.

- Ashburnham, Ashby, Athol, Barre, Fitchburg, Gardner, Harvard, Holden, Hubbardston, Lunenburg, Petersham, Phillipston, Princeton, Royalston, Sterling, Templeton, Westminster, and Winchendon send students to Montachusett Voc Tech Regional
- Chelmsford, Groton, Littleton, Pepperell, Shirley, Townsend, and Westford send students to Nashoba Valley Voc Tech
- Beverly, Boxford, Danvers, Essex, Gloucester, Hamilton, Lynnfield, Manchester, Marblehead, Middleton, Nahant, Rockport, Salem, Swampscott, Topsfield, and Wenham send students to North Shore Regional Voc Tech.
- Chelsea, Malden, Melrose, North Reading, Reading, Revere, Saugus, Stoneham, Wakefield, Winchester, Winthrop, and Woburn send students to Northeast Metro Vocational.
- Adams, Clarksburg, Florida, Monroe, North Adams, Savoy, and Williamstown send students to Northern Berkshire Vocational.
- Acushnet, Carver, Lakeville, Mattapoisett, and Rochester send students to Old Colony Voc Tech.
- Belchertown, Granby, Hardwick, Monson, New Braintree, Palmer, and Ware send students to Pathfinder Vocational Tech.
- Bedford, Billerica, Burlington, Tewksbury, and Wilmington send students to Shawsheen Valley Voc Tech.
- Ashland, Framingham, Holliston, Hopkinton, and Natick send students to South Middlesex Regional Voc Tech.

- Berlin, Hudson, Marlborough, Maynard, Northborough, Southborough, and Westborough send students to Assabet Valley Vocational
- Amesbury, Georgetown, Groveland, Haverhill, Ipswich, Merrimac, Newbury, Newburyport, Rowley, Salisbury, and West Newbury send students to Whittier Vocational.
- Brockton, East Bridgewater, Easton, Foxborough, Mansfield, Norton, Sharon, Stoughton, and West Bridgewater send students to Southeastern Regional Voc Tech.
- Auburn, Charlton, Dudley, Oxford, Rutland, Southbridge, and Webster send students to Southern Worcester County Voc Tech.
- Dartmouth, Fairhaven, and New Bedford send students to Greater New Bedford Regional Voc Tech.
- Fall River, Somerset, Swansea, and Westport send students to Greater Fall River Vocational Regional.
- Abington, Cohasset, Hanover, Hanson, Norwell, Rockland, Scituate, and Whitman send students to South Shore Vocational Regional.
- Bellingham, Blackstone, Douglas, Grafton, Mendon, Milford, Millbury, Millville, Northbridge, Sutton, Upton, and Uxbridge send students to Blackstone Valley Regional.
- Bernardston, Buckland, Colrain, Conway, Deerfield, Erving, Gill, Greenfield, Heath, Leyden, Montague, New Salem, Northfield, Orange, Shelburne, Sunderland, Warwick, Wendell, and Whately send students to Franklin County Vocational Regional.

- Essex Agricultural Technical School serves the whole of Massachusetts.
- Avon, Bellingham, Braintree, Brookline, Canton, Cohasset, Dedham, Dover, Foxborough, Franklin, Holbrook, Medfield, Medway, Millis, Milton, Needham, Norfolk, Norwood, Plainville, Quincy, Randolph, Sharon, Stoughton, Walpole, Wellesley, Westwood, Weymouth, and Wrentham send agricultural students to Norfolk Country Agricultural.
- Acushnet, Attleboro, Berkley, Dartmouth, Dighton (non-op), Easton, Fairhaven, Fall River, Freetown, Mansfield, New Bedford, North Attleborough, Norton, Raynham (non-op), Rehoboth (non-op), Seekonk, Somerset, Swansea, Taunton, and Westport send agricultural students to Bristol Country Agricultural.

D.4 Charter School Districts

Charter Schools took 1.3% of Massachusetts' public school students in 2004, but they are unevenly distributed geographically so they potentially have a much greater impact on individual towns or regions. If a geography had more than 2% of its students at a charter school[28], the geography was dropped from the sample unless it was possible to aggregate the Charter School test scores into the scores for the geography.

Section 1 of Chapter 2 described the lawsuits that led to the 1993 Massachusetts Education Reform Act. These lawsuits involved 16 students from Brockton, Belchertown, Berkeley, Carver, Hanson, Holyoke, Lawrence, Leicester, Lowell, Lynn, Rockland, Rowley, Salisbury, Springfield, Whitman and Winchendon. Efforts were made to ensure that these towns were retained in the sample even if, as in the case of

[28] Public School District Enrollment Data from the DOE gives the number of students in Charter Schools. In the case of Lawrence, Lowell, Springfield and Worcester this allowed aggregation of test score data from these schools with the towns' test scores.

Brockton, this meant retaining the town although it had more than 2% of its students at a charter school or schools. In the cases of Lawrence, Lowell and Springfield it was possible to calculate weighted averages of the Charter School test scores and the other scores from the town.

The remainder of this section describes the impact of the significant and operative Charter Schools on the selection of the sample starting with the Academy of Strategic Learning Horace Mann Charter School ("HMCS")

Amesbury has the Academy of Strategic Learning HMCS, with 45 students in grades 7 - 12 in 2004 against 1,009 students in total in these grades. Amesbury is also served by the River Valley Charter School, a regional public Montessori school, based in Newburyport, serving students from Amesbury, Newbury, Newburyport, Salisbury, and West Newbury[29]. The River Valley Charter School had an enrollment of 287 students in grades K-8 in 2004, against 1,916 for Amesbury and 1,612 for Newbury, Newburyport, Salisbury and West Newbury in the same grades. Charter school students represent 4.5 percent of high school students in Amesbury and 8.13 percent of K-8 grade students in Amesbury, Newbury, Newburyport, Salisbury, and West Newbury. Neither Amesbury nor Newburyport is part of an academic region and they were dropped from the analysis. West Newbury is part of Pentucket Academic Region. Newbury and Salisbury are part of Triton. Charter school enrollment is 4.26 percent of the total K – 8 enrollment of 6,739 for Amesbury, Pentucket and Triton and so these Amesbury, Pentucket and Triton were dropped from the analysis.

---

[29] www.rivervalleycharter.org

Boston has the Boston Evening Academy HMCS with 201 students in grades 11 and 12 in 2004; the Boston Renaissance Charter School with 1,136 students in grades K – 8; the City on a Hill Charter School with 248 students in grades 9 - 12; the South Boston Harbor Academy Charter School with 346 students in grades 5 - 12; the Conservatory Lab Charter School with 126 students in grades K-5; the Media and Technology Charter High School with 170 students in grades 9 - 12 and the Health Careers Academy HMCS with 182 students in Grades 9 - 12.

Dorchester is a Boston neighborhood. The Neighborhood House Charter is based in Dorchester and had 221 students from grades K – 8 in 2004. According to the school's web pages[30], most NHCS students live in Dorchester (77%) with the remainder living in Roxbury (6%), Mattapan (5%) and other Boston neighborhoods (9%). Codman Academy Charter District is also based in Dorchester and had 81 students in grades 9-11 in 2004.

Academy of the Pacific Rim Charter is in Hyde Park, a Boston Neighborhood. It had 327 students in grades 6 - 12 in 2004. The Frederick Douglass Charter School is also in Hyde Park. It had 268 students in grades 6 – 9 in 2004. Roxbury Preparatory Charter School has 179 students in grades 6 – 8.

The total number of Charter School students in Boston is 3,485 or 5.8 percent of the total enrollment of 60,150 students. Boston was therefore dropped from the analysis.

---

[30] www.NHCSonline.org

The Champion HMCS, based in Brockton had 87 students in grades 9 - 12 in 2004 against 4,363 public school students in these grades. Champion's represents 2 percent of Brockton's high Brockton also send a negligible number of students to Foxboro Regional Charter – see below. Champion was ignored from the point of view of the analysis and Brockton was retained in the sample.

Cambridge hosts the Benjamin Banneker Charter School, with 329 students in grades K – 8 making up 7.2 percent of the total public school enrollment of 4,554 students in these grades. Cambridge was dropped from the analysis.

In Chelmsford, the Murdoch Middle School had 263 students in grades 5 – 8 and made up 14.86 percent of the total public school enrollment of 1,770 in these grades. Chelmsford was therefore dropped from the analysis.

Devens is home to the Francis W Parker Charter School with an enrollment of 356 students in grades 7 - 12. According to the school's web pages[31] Parker's region includes 46 school districts in eastern and central Massachusetts, so it is difficult to quantify the likely effect of this school on the results of any given school district and so the effect was ignored and no adjustments have been made to the data.

[31] www.parker.org

The Atlantis Charter School is based in Fall River and in 2004 it served 693 students in grades K – 8 with preference being given in admissions to siblings and to children who reside in the city of Fall River[32]. Assuming that the 693 students are from Fall River they represented 7.9 percent of the total public school enrollment of 8,812 for Fall River. Enrollment has grown from 252 in 1996 when the school was first chartered and MCAS test scores are available from 1998 to 2002. Fall River was not dropped from the analysis and the Atlantis Charter School scores were aggregated into the Fall River test scores.

Fitchburg is home to the North Central Charter Essential School District, which had 305 students in grades 7 – 10 in 2004. The web page[33] describes admissions eligibility as "living within the designated towns of the school's region (Fitchburg, Leominster, Gardner, Ashburnham, Ashby, Townsend, Lunenburg, Lancaster, Clinton, Sterling, Princeton, Westminster) – students outside the region are eligible if open spaces remain." Enrollment represents about 4 percent of the total of approximately 7,600 (Ashby and Townsend are part of North Middlesex Region; Lancaster is part of the Nashoba Region; Princeton and Sterling are part of the Wachusett Region, so precise numbers are not known). No MCAS scores are recorded for the North Central Charter Essential School District so Ashburnham-Westminster, North Middlesex, Nashoba, and Wachusett Regions were dropped from analysis together with the towns of Clinton, Fitchburg, Gardener, Leominster and Lunenburg.

---

[32] www.atlantiscs.org

[33] www.ncces.org

Foxboro Regional Charter is based in Foxboro and had 875 students in grades K-12 in 2004. It is open to all students who reside in the communities of Attleboro, Avon, Brockton, Canton, Easton, Foxborough, Mansfield, Medfield, Medway, Millis, Norfolk, North Attleboro, Norton, Norwood, Plainville, Sharon, Stoughton, Walpole, West Bridgewater, and Wrentham.[34] These communities had 72,540 K-12 students in 2004 so the percentage in charter schools was 1.2 percent. The School's Annual Report for 2001-2002 gives a breakdown of the origin of students. The percentage of each towns' student population at Foxboro Regional Charter was calculated, as follows: Attleboro (1.2%), Avon (0.9%), Brockton (0.2%), Canton (0.1%), Easton (0.4%), Foxborough (3.8%), Mansfield (3.0%), Medfield (0.1%), Medway (0.2%), Millis (0.8%), Norfolk (0.6%), North Attleborough (1.9%), Norton (1.9%), Norwood (0.5%), Plainville (4.2%), Sharon (0.4%), Stoughton (0.4%), Walpole (1.1%), West Bridgewater (0.2%), Wrentham (0.9%). Foxborough and Mansfield were dropped from the analysis. In order to drop Plainville from the analysis, the King-Phillip region of which it is a part was dropped from the analysis.

The Benjamin Franklin Charter School had 381 students in grades K – 8 in 2004, of whom according to the school web pages[35] 339 students were from Franklin; which had 4,473 students in these grades. Charter School enrollment is 7.6 percent of the total public school enrollment and so Franklin was dropped from the analysis.

---

[34] myschoolonline.com/folder/0,1872,2133-185669-2-49479,00.html

[35] www.bfccs.org

The Pioneer Valley Performing Arts Charter School is based in Hadley and had 329 students in grades 8-12 in 2004. According to the school's annual report for 2001-2002 students come from 55 towns in western Massachusetts. Given the nature of the school, performing arts, it was assumed that this charter school would have no significant impact on the analysis.

The Hilltown Cooperative Charter School, based in Haydenville served 146 students in grades K – 8 from 25 towns, in 2004. Northampton has 72 students at the school representing 3.5 percent of Northampton's public school enrollment in grades K – 8. Williamsburg has 10.2 percent of its enrollment in these grades at this school. Accordingly the town of Northampton and the region comprising Chesterfield, Goshen, Southampton, Westhampton and Williamsburg were dropped from the analysis.

The South Shore Charter School is based in Hull and had 366 students in grades K - 12 in 2004 from 22 towns. According to the school's 2001-2002 annual report, 192 of the students in that year were from Hull representing 12.1 percent of Hull's public school population. Hull was therefore dropped from the analysis. No other town sending students to the school sent more than 1 percent of its students to the school.

The Sturgis Charter School in Hyannis had 325 students in grades 9 - 12 in 2004. From the 2001-2002 School Report it was calculated that students at the school represented 6 percent of Dennis-Yarmouth public school students; 5.3 percent of Barnstable public school students and 2.6 percent of Harwich public school students and so these regions and towns were dropped from the analysis.

514 students in grades K – 8 attended Lawrence Family Development Charter in 2004. According to the school web pages,[36] admission to the school is open to all eligible students who are residents of the City of Lawrence. The school opened in 1995 with 178 students in K-3. The Community Day Charter School, which opened in 1995, is also based in Lawrence and it had 306 students in grades K – 8 in 2004. According to the School's 2001-2002 annual report, one hundred percent (100%) of the students came from Lawrence. Lawrence charter school students represent 8.2 percent of the 10,007 public school students in grades K-8. MCAS scores are available for the two schools from 1998 onwards, which means that the missing information is 4th grade MCAS scores for 1996 when the enrollment in the two schools was less than 350 and there were no students in the 8th grade.

Aggregation of the Regular and Charter school scores for Lawrence will not make very much difference to the scores for that town as evidenced by the weighted averages for the 2000 MCAS scores given in Table D.01. In both the fourth grade and eighth grade scores the presence of the two charter schools results in less than 0.2 percent change in the scores.

---

[36] www.lfdcs.org

Table D.01 - Lawrence MCAS Scores for 2001 adjusted for Charter Schools.				
School District	Average MCAS Scores 4th Grade			
	ELA	MATH	SCIENCE	
COMMUNITY DAY	230	232	239	
LAWRENCE	222	219	224	
LAWRENCE FAMILY	222	217	223	
Weighted Averages	222.21	219.25	224.35	
Percent Difference	0.095%	0.114%	0.156%	
	Average MCAS Scores 8th Grade			
	ELA	MATH	SCIENCE	HISTORY
COMMUNITY DAY	247	236	227	227
LAWRENCE	226	210	209	209
LAWRENCE FAMILY				
Weighted Averages	226.34	210.42	209.29	209.29
Percent Difference	0.150%	0.200%	0.139%	0.139%
Source: The Massachusetts Department of Education.				

Lowell has the Lowell Community Charter School (opened in 2000, it first reported MCAS scores in 2002) with 542 students in grades K – 6 in 2004 and the Lowell Middlesex Academy Charter High with 108 students in grades 9 –12 in 2004. According to the 2001-2002 school report Lowell Community Charter had 402 students 392 of whom were from Lowell and the Lowell Middlesex Academy Charter High, which opened in 1995 and has reported MCAS scores since 1988, gives preference to students from Lowell. In 2004 Lowell had 15,117 public school students, so the two charter schools accounted for 6.24 percent of the public school population in Lowell. The Charter School test scores were aggregated with the rest of Lowell's test scores in a manner similar to that employed for Lawrence.

Mystic Valley Advantage Regional in Malden was opened in 1998 and has reported MCAS scores since 1999. In 2002 it had 652 students in grades K – 8. According to the school's web pages,[37] it services the towns of Malden, Medford, Melrose, Wakefield, Everett and Stoneham, which together have 18,098 students in grades K – 8 and so the school accounted for 3.6 per cent of the public school enrollment in that year. Using the 2003-2004 Mystic Valley school report Stoneham and Wakefield have less than 1 percent of their students at Mystic; Melrose, Medford and Everett have around 5 percent and Malden has 11 percent of its K – 8 public school population at Mystic. Everett, Malden, Melrose, and Medford were dropped from the analysis.

Marblehead Community Charter Public School was chartered in September 1995 and had 176 students in grades 5 – 8 in 2002 of which according to the school's 2001-2002 report 138 were Marblehead residents representing 16.27 percent of the 848 public school students in the town in those grades. Accordingly, Marblehead was dropped from the analysis.

Founded in 1994, the Cape Cod Lighthouse Charter School (CCLCS) is a public middle school serving Cape Cod students in grades 6, 7, and 8. One-hundred seventy-five students from nearly every town on the Cape attend.[38] No breakdown of the students by town has been located and so no action was taken to account for this school in the analysis.

---

37 www.mvrcs.com/

38 www.lighthouse.chtr.k12.ma.us/mission.htm

Rising Tide Charter School opened, for the 1998-99 school year, in Plymouth. All students from the surrounding region in grades 5 – 8 are eligible for enrollment at the Rising Tide Charter School.[39] In 2002 it had 147 students in attendance 91 from Plymouth-Carver (2.6 percent of Plymouth-Carver's enrollment in these grades); 15 from Kingston (2.3 percent); 6 from Duxbury (0.6 percent) and 9 from Middleboro (0.7 percent) with the other 26 from eight other towns. No adjustments were made to account for the Rising Tide Charter School in the analysis.

The Prospect Hill Academy Charter School in was founded as The Somerville Charter School in 1996 and has reported MCAS scores since 1998. It had 733 students in grades K - 12 in 2004. According to the schools web pages,[40] students come from 33 cities and towns, mainly Somerville (47%), Medford (14%), Cambridge (11%), Everett (7%), Malden (5%), and Boston (1%). 5.9 percent of Somerville's public school enrollment and 2.3 percent of Medford's public school enrollment are at the Prospect Hill Academy. Adjustments were made to Somerville's and Medford's scores to take this school into account in the analysis. Boston, Everett and Malden had already been dropped from the analysis.

Springfield is home to three charter schools. The New Leadership HMCS was formed in 1997 or 1998, first in West Springfield and now in Springfield. In 2002 it had 156 students in grades 6 –10. It is assumed that students come from the Springfield and West Springfield school districts. The Robert M. Hughes Charter had 121 students in

---

39 www.risingtide.org

40 www.prospecthillacademy.org

grades K – 6 in 2002. The mission statement[41] implies that all the students come from the Springfield district. The SABIS International Charter School had 1,138 students in grades K - 12 in 2002.

The 1,415 Springfield students in Charter Schools in 2002 represented 5.64 percent of the public school students. It is desirable to keep Springfield in the analysis as it was one of the towns from which students in McDuffy v. Robertson came. Fortunately SABIS, which opened in 1995, has been reporting test scores since the 1996 MEAP tests and Robert Hughes, which opened in 1999-2000, has been reporting MCAS scores since 2000; so aggregate scores have been calculated for Springfield in a manner similar to that applied to Lawrence and Lowell.

The Abby Kelley Foster Regional Charter School opened in 1998 with Grades K-5 in a renovated mill building.[42] MCAS Scores were first reported in 1999. It had 639 students in grades K – 8 in 2002 about 70 percent coming from Worcester. The 2001-2002 annual report gave the numbers of students applying from the following towns and from these the percentages of each town's public school enrollment attending Abby Kelly Foster, given in parentheses, were estimated. Auburn (1.7 percent), Holden (0.0 percent), Leicester (0.6 percent), Millbury (2.2 percent), Oxford (1.0 percent), Shrewsbury (0.4 percent), Sutton (0.0 percent), West Boylston (1.6 percent) and Worcester (2.6 percent). Abby Kelly Foster scores were aggregated with those for the regular day scores for Millbury and Worcester.

---

[41] www.rmhughes.org

[42] www.akfcs.org/abt/history.php

In September 1996 Seven Hills Charter School opened its doors to serve the students of Worcester in Grades Kindergarten through grade 7. The school expanded to include grade 8 in the fall of 1997. In 2002 it served 662 students in grades K – 8. MCAS Scores have been recorded since 1998. The Seven Hills Charter scores were aggregated back into the Worcester regular day scores together with the Abby Kelly Foster scores.

D.5 Small School Districts

Gosnold has consistently been too small for testing data to be reported and was dropped from the analysis. Chatham is missing year 2000 English Language Arts MCAS scores and so was dropped from the analysis.

APPENDIX E

SAMPLE NORMALITY PLOTS BASED ON 1988 AND 2002 TEST SCORE DATA

Figure E.01 – Normality Plot 1988 4th Grade Reading

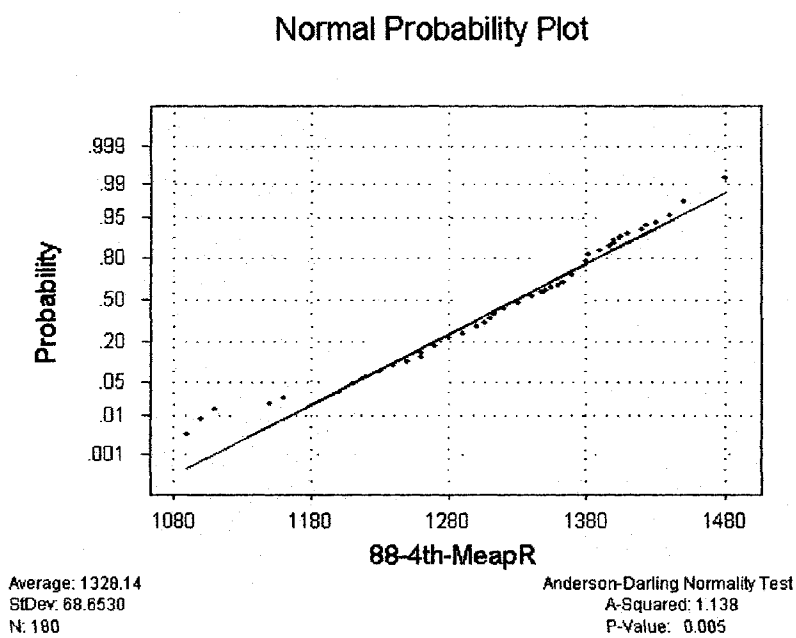


Figure E.02 – Normality Plot 1988 8th Grade Science

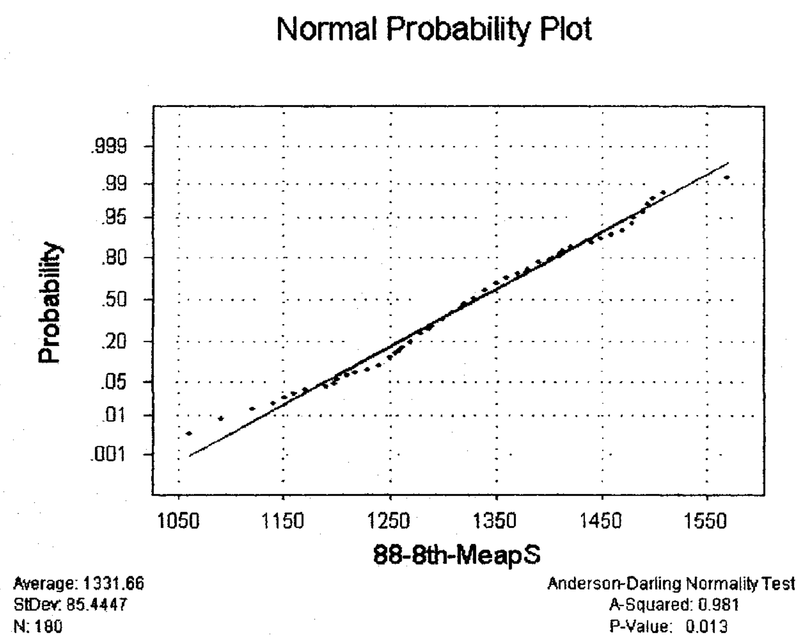


Figure E.03 – Normality Plot 1988 12th Grade Mathematics

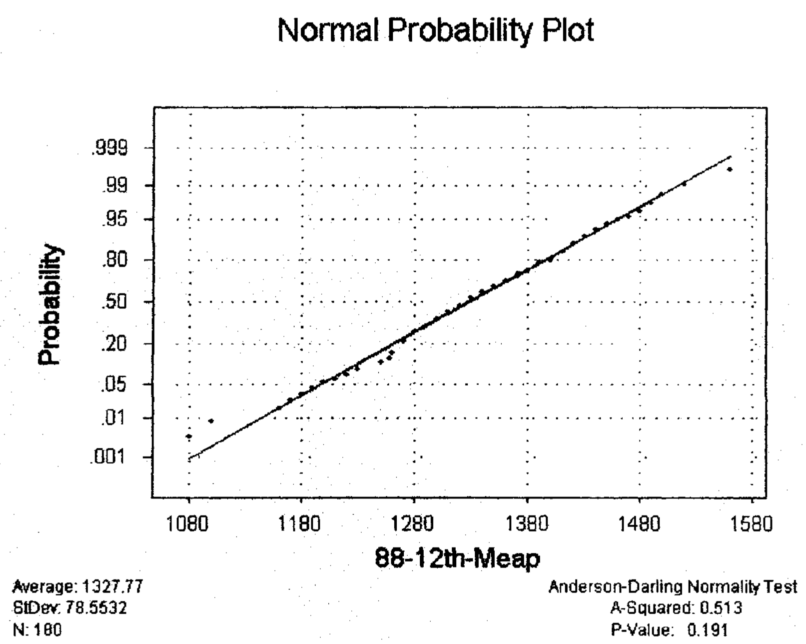


Figure E.04 – Normality Plot 2002 4th Grade English Language Arts

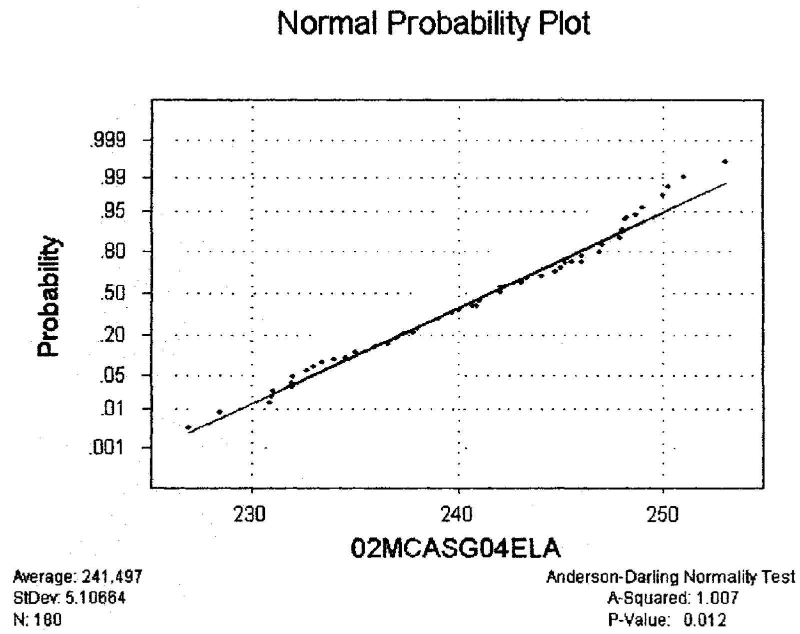


Figure E.05 – Normality Plot 2002 7th Grade English Language Arts

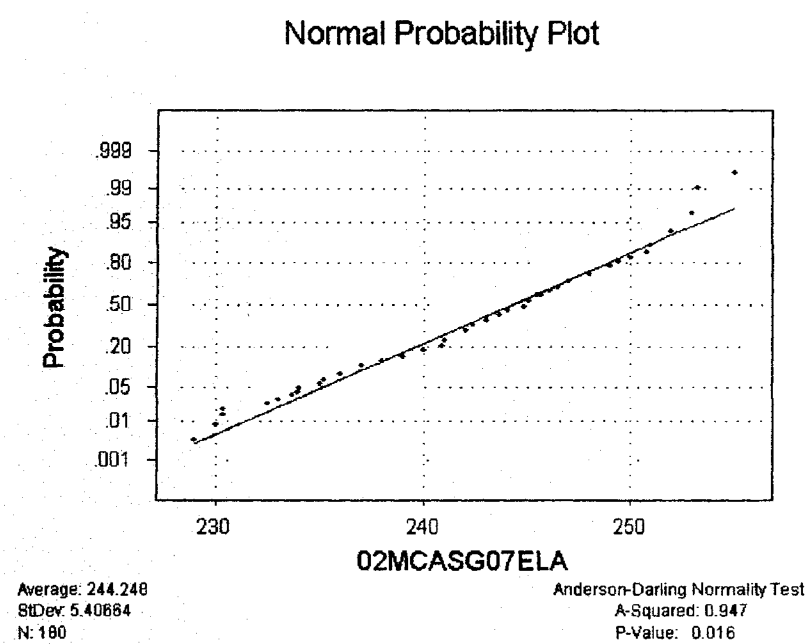
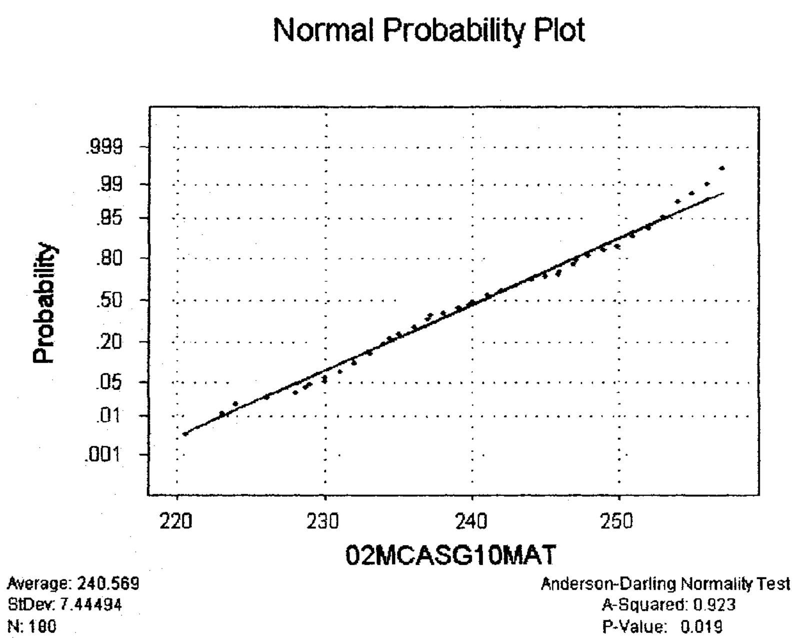


Figure E.06 – Normality Plot 2002 10th Grade Mathematics



APPENDIX F

SOCIO-ECONOMIC STATUS FACTORS SURVEY

Paper Ref.	Educ. Father	Educ. Mother	Occupation Father	Occupation Mother	Family Income	Number of Parents	Number of Siblings	Number of Books in Home	Housing Tenure	Crime Rate	Poverty Rate	Pop. Density	Housing Density	Percent Urban	Notes
01	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	None
02	Y	Y	Edw	N	N	N	N	Y	N	N	N	N	N	N	N
03	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N	None
04	Y	Y	N	N	Y	N	N	N	N	N	Y	N	N	N	Meal Subsidy
05	N	N	Occ1	N	Y	N	N	N	N	N	Y	N	N	N	N
06	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	Y	Ethnicity
07	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	N
08	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	N	N
09	N	N	Occ4	N	N	N	N	N	N	N	N	N	N	N	N
10	Y	Y	N	N	Y	N	N	N	N	N	N	N	N	N	N
11	Y	Y	Occ2	Occ2	Y	N	N	N	N	N	N	Y	N	N	Note4
12	N	N	Y	Y	N	Y	N	N	N	N	N	N	N	N	None
13	Y	Y	N	N	Y	N	N	N	N	N	N	N	N	N	N
14	N	N	Y	Y	Y	N	N	N	Y	N	N	N	N	N	N
15	N	N	BoC & N&P	BoC & N&P	N	N	N	N	N	N	N	N	N	N	N
16	Y	Y	SEI	SEI	Y	N	N	N	N	N	N	N	N	N	Note6
17	Y	Y	N	N	Y	N	N	N	N	N	N	N	N	N	N
18	Y	Y	N	N	Y	N	N	N	N	N	N	N	N	N	N
19	N	N	SEI	SEI	N	N	N	N	N	N	N	N	N	N	N
20	Median Years	Median Years	Occ3	N	Median Income	N	N	N	N	N	N	N	N	N	N
21	N	N	N	N	N	N	N	N	N	N	N	N	N	N	Note2
22	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	N
23	Y	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N

Paper Ref #	Educ. Father	Educ. Mother	Occupation Father	Occupation Mother	Family Income	Number of Parents	Number of Siblings	Number of Books in Home	Housing Tenure	Crime Rate	Poverty Rate	Pop. Density	Housing Density	Percent Urban	Notes
24	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N	Note3
25	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	N
26	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	N	N
27	Y	Y	N	N	Y	N	N	N	N	Y	Y	N	N	N	None
28	N	N	N	N	N	N	N	N	N	N	N	N	N	N	Note5
29	N	N	H4	H4	N	N	N	N	N	N	N	N	N	N	N
30	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N
31	Y	Y	Y	Y	N	N	Y	N	N	N	N	N	N	N	N
32	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	None
33	Y	Y	N	N	Y	N	N	N	N	N	N	N	N	N	N
34	N	N	Occ1	N	N	%Single Mothers	N	N	Y	Y	N	N	N	Y	N
35	Y	Y	Y	Y	Y	N	N	N	Y	N	N	N	Y	N	N
36	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N	N
37	Y	Y	N	N	Y	N	N	N	N	N	N	N	N	N	N
38	N	N	Y	N	N	N	N	N	N	N	N	N	N	N	N
39	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	N
40	Y	Y	Y	N	N	N	N	Y	N	N	N	N	N	N	Note9
41	N	N	SEI & SP	SEI & SP	N	N	N	N	N	N	N	N	N	N	N
42	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	N
43	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	N
44	Y	Y	Y	N	N	N	N	N	N	N	N	N	N	N	None
45	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	N
46	N	N	N	N	N	N	Y	N	N	N	N	N	N	N	N
47	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N	N
48	Y	Y	Y	Y	Y	N	N	N	Y	N	N	N	N	N	N
49	Y	Y	Y	Y	Y	N	N	N	Y	N	N	N	N	N	N

Paper Ref #	Educ. Father	Educ. Mother	Occupation Father	Occupation Mother	Family Income	Number of Parents	Number of Siblings	Number of Books in Home	Housing Tenure	Crime Rate	Poverty Rate	Pop. Density	Housing Density	Percent Urban	Notes
50	Y	Y	COPM	COPM	Y	Y	N	N	N	N	N	N	N	N	N
51	Y	Y	H7	H7	Y	N	N	N	N	N	N	N	N	N	N
52	Y	Y	Y	Y	N	N	N	N	N	N	Y	N	N	N	N
53	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	N
54	N	N	Y	Y	N	N	N	N	N	N	N	N	N	N	N
55	N	N	SEI & N&P	SEI & N&P	N	N	N	N	N	N	N	N	N	N	None
56	N	N	Y	Y	N	N	Y	N	N	N	N	N	N	N	None
57	N	N	Y	Y	N	N	N	N	N	N	N	N	N	N	N
58	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N	None
59	Y	Y	SEI	SEI	N	N	N	N	Y	N	N	N	N	N	N
60	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N	N
61	Y	Y	Y	Y	Y	N	N	Y	N	N	N	N	N	N	Note8
62	N	N	SEI & SP	SEI & SP	N	N	N	N	N	N	N	N	N	N	N
63	Y	Y	N	N	Y	N	N	N	N	N	N	N	N	N	N
64	Y	Y	Y	N	Y	N	Y	N	N	N	N	N	N	N	Note7
65	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	N	N
66	Y	Y	N	N	Y	N	N	N	N	N	N	N	N	N	N
67	Y	Y	N	N	Y	N	N	N	N	N	N	N	N	N	N
68	N	N	Y	Y	N	N	N	N	N	N	N	N	N	N	N
69	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	N
70	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	N
71	Y	Y	Occ4	Occ4	N	N	N	N	Y	N	N	N	Y	N	Note10
72	N	N	N	N	Y	N	N	N	N	N	N	N	N	N	N
73	N	N	Y	Y	Y	N	N	N	Y	N	N	N	N	N	Note1

Paper Ref #	Educ. Father	Educ. Mother	Occupation Father	Occupation Mother	Family Income	Number of Parents	Number of Siblings	Number of Books in Home	Housing Tenure	Crime Rate	Poverty Rate	Pop. Density	Housing Density	Percent Urban	Notes
74	N	N	N&P	N&P	N	N	N	N	N	N	N	N	N	N	N
75	N	N	Scaled	Scaled	N	N	Y	N	N	N	N	N	N	N	None
76	N	N	Y	Y	N	N	N	N	N	N	N	N	N	N	N
77	Y	Y	N	N	Y	N	N	N	N	N	N	N	N	N	N
78	Y	Y	N	N	Y	N	N	N	N	N	N	N	N	N	Note11
79	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N	N

In the survey table, given on the previous pages, the following codes were used to reference the method employed to classify the Father or the Mother's Occupation.

Occupation	
SEI	Duncan's SEI
N&P	Nam & Powers Occupational Status Scores
SP	Seigel's PRESTIGE
COPM	Canadian Occupational Performance Measurement
Edw	Edward's (1943) Census Classification of Occupational Status
BoC	Bureau of the Census (1963)
H4	Hollingshead 4 Factor
H7	A Hollingshead-type seven category code of occupations.
Occ1	Percent Professionals and Managers
Occ2	Percent Employed Civilian Labor Force; Percent Labor Force in Military; %Labor Force in Professional Occupations. Percent Labor Force in Clerical, Crafts' Occupations; %Labor Force in Operative, Fabrication and Labor.
Occ3	Percent of the male civilian employed labor force in the upper status, white-collar occupations of professional, technical and kindred workers and managers, proprietors and officials, excluding farm.
Occ4	1 if white-collar, 0 otherwise

The following table contains the explanation of the notes referenced in the notes column of the survey table.

Notes	
Note1	Wealth, Car Ownership, Environmental Hazards and Levels of Violence.
Note2	"Standard Department of Education SES Variable"
Note3	"Standard of Living"
Note4	Percent Urban; Median Age; Percent Females in Labor Force; Percent Nonwhites; Reside Different State 5 Years Ago; Percent Government Employees; Per Capita Education Expenditure; Per Capita Welfare Expenditure; Per Capita Health Expenditure; Percent Over 64 Years; Median Value Occupied Housing; Percent Gross Farm Income of Total Personal Income; Income from Property Per Capita; Standard Deviation of Years of School Completed.
Note5	Uses the "Home Index" as measure of SES. Citation "Gough, H. 1971 'A cluster analysis of home index status items.' Psychological Reports 9(28), 23-29."
Note6	Length of service at Job. Commuting distance to work: (1) number of air miles and (2) number of road miles between place of residence and the plant site.
Note7	Child's IQ; Child's Birth Rank; Father's IQ; Mother's IQ; Natural Mother's Age
Note8	Value of the home owned by the student's family, the possession of certain articles in the home
Note9	Educational attainment of oldest older sibling.
Note10	Percent Dropouts; Percent Unemployed; Percent Poor families (less than $5,000 income); Percent Rich Families ($15,000 or more income); Percent Owner-Occupied Housing valued at less than $10,000; Percent Owner-Occupied Housing valued at $25,000 and over; Percent Renter-occupied units with gross monthly rent of $150 or more; Ethnicity; Percent persons 5 years old and over who live in the same house as 5 years ago.
Note11	Own Room and Own Desk

SES References

[01] - Adda, Jerome, Chandala, Tarani, and Marmot, Michael. 2003. Socio-economic Status and Health: Causality and Pathways, *Journal of Econometrics* 112: 57-63.

[02] - Alexander, Karl L., Fennessey, James, McDill, Edward L. and D'Amico, Ronald J. 1979. School SES Influences - Composition or Context?, *Sociology of Education* 4(52) (October): 222-237.

[03] - Alexander, Karl L., and Pallas, Aaron M. 1983. Bringing the Arrows Back In: On the Recursivity Assumptions in School Process Models, *Social Forces* 62(1) (September): 32-53.

[04] - Alexander, Karl L., Entwisle, Doris R., and Bedinger, Samuel D. 1994. When Expectations Work: Race and Socioeconomic Differences in School Performance, *Social Psychology Quarterly* 4(57) (December): 283-299.

[05] - Aneshensel, Carol S., and Sucoff, Clea A. 1996. The Neighborhood Context of Adolescent Mental Health, *Journal of Health and Social Behavior* 4(37) (December): 293-310.

[06] - Athanasou, James A. 2001. Young people in transition: Factors influencing the educational-vocational pathways of Australian school-leavers, *Education & Training* 3(43): 132-138.

[07] - Battle, Juan. 1997. The Relative Effects of Married Versus Divorced Family Configuration and Socioeconomic Status on the Educational Achievement of African American Middle-Grade Students, *Journal of Negro Education* 1(66) (Winter): 29-42.

[08] - Baer, Judith. 1999. The Effects of Family Structure and SES on Family Processes in Early Adolescence, *Journal of Adolescence* 22: 341-354.

[09] - Baigi, Amir, Marklund, Bertil, and Fridlund, Bengt. 2001. The Association between Socio-Economic Status and Chest Pain, Focusing on Self-Rated Health in a Primary Health Care Area of Sweden, *European Journal of Public Health* 4(11) (December): 420-424.

[10] - Bayer, Alan E. 1997. Sexual Permissiveness and Correlates as Determined through Interaction Analyses, *Journal of Marriage and the Family* 1(39) (February): 29-40.

[11] - Braun, Denny. 1998. Multiple Measurements of U.S. Income Inequality, *Review Of Economic Statistics* 3(70) (August): 398-405.

[12] - Brutsaert, Herman. 1998. Home and school influences on academic performance: State and Catholic elementary schools in Belgium compared, *Educational Review* 1(50) (February): 37-43.

[13] - Campbell, Karen E., and Lee, Barrett A. 1992. Sources of Personal Neighbor Networks: Social Integration, Need, or Time?, *Social Forces* 4(70) (June):1077-1100.

[14] - Chandola, Tarani. 2000. Social Class Differences in Mortality Using the New UK National Statistics Socio-Economic Classification, *Social Science & Medicine* 5(50) (March): 641-649.

[15] - Chiricos, Theodore G., and Waldo, Gordon P. 1975. Socioeconomic Status and Criminal Sentencing: An Empirical Assessment of a Conflict Proposition, *American Sociological Review* 6(40) (December): 753-772.

[16] - Clemente, Frank, and Summers, Gene F. 1975. The Journey to Work of Rural Industrial Employees, *Social Forces* 1(54) (September): 212-219.

[17] - Clemente, Frank, and Sauer, William J. 1976. Life Satisfaction in the United States, *Social Forces* 3(54) (March): 621-631.

[18] - Coburn, David, and Pope, Clyde R. 1974. Socioeconomic Status and Preventive Health Behaviour, *Journal of Health and Social Behavior* 2(15) (June): 67-78.

[19] - Cohn, Elchanan, and Kiker, B.F. 1986. Socioeconomic Background, Schooling, Experience and Monetary Rewards in the United States, *Economica* 2(53) (November): 497-503.

[20] - Collver, Andrew, and Semyonov, Moshe. 1979. Suburban Change and Persistence, *American Sociological Review* 3(44) (June): 480-486.

[21] - Daly, John A and Kreiser, Pamela O. 1994. Question-Asking Comfort: Explorations of the Demography of Communication in the Eighth Grade Classroom, *Communication Education* 1(43) (January): 27-41.

[22] - Erbe, William. 1964. Social Involvement and Political Activity: A Replication and Elaboration, *American Sociological Review* 2(29) (April): 198-215.

[23] - Gordon, Robert A., and Rudert, Eileen E. 1979. Bad News Concerning IQ Tests, *Sociology of Education* 3(52) (July): 174-190.

[24] - Gore, Susan, Aseltine, Robert H., Jr., and Colten, Mary Ellen. 1992. Social Structure, Life Stress, and Depressive Symptoms in a High School-Age Population, *Journal of Health and Social Behavior* 2(33) (June): 97-113.

[25] - Griffin, Larry. 1976. Specification Biases in Estimates of Socioeconomic Returns to Schooling, *Sociology of Education* 2(49) (April): 121-139.

[26] - Guppy, Neil, and Goyder, John C. 1984. Consensus on Occupational Prestige: A Reassessment of the Evidence, *Social Forces* 3(62) (March): 709-725.

[27] - Gyimah-Brempong, K., and Gyapong, A.O. 1991. Characteristics of Educational Production Functions: An Application of Canonical Regression Analysis, *Economics Of Education Review* 10(1): 7-17.

[28] - Haller, Emil J., and Davis, Sharon A. 1981. Teacher Perceptions, Parental Social Status and Grouping for Reading Instruction, *Sociology of Education* 3(54) (July): 162-174.

[29] - Hanson, Shirley M.H. 1986. Healthy Single Parent Families, *Family Relations* 1(35) (January): 125-132.

[30] - Hauser, Robert M., and Warren, John Robert. 1997. Socioeconomic Indexes for Occupations: A Review, Update, and Critique, *Sociological Methodology* 27: 177-298.

[31] - Hayes Bernadette C. 1996. Gender differences in religious mobility in Great Britain, *British Journal of Sociology*, 4(47) (December): 643-656.

[32] - Hoffer, Thomas, Greeley, Andrew M., and Coleman, James S. 1985. Achievement Growth in Public and Catholic Schools, *Sociology of Education* 58(2) (April): 74-97.

[33] - House, James S., Lepkowski, James M., Kinney, Ann M., Mero, Richard P., Kessler, Ronald C. and Herzog, A. Regula. 1994. The Social Stratification of Aging and Health, *Journal of Health and Social Behavior* 3(35) (September): 213-234.

[34] - Jaggia, Sanjiv, and Kelly-Hawke, Alison. 1999. An Analysis of the Factors That Influence Student Performance: A Fresh Approach to an Old Debate, *Contemporary Economic Policy* 17(2) (April): 189-198.

[35] - Jalovaara, Marika. 2001. Socio-Economic Status and Divorce in First Marriages in Finland 1991-93, *Population Studies* 2(55) (July): 119-133.

[36] - John, Nancy H. St., and Lewis, Ralph G. 1975. Race and the Social Structure of the Elementary Classroom, *Sociology of Education* 3(48) (Summer): 346-368.

[37] - John, Nancy St. 1971. The Elementary Classroom as a Frog Pond: Self-Concept, Sense of Control and Social Context, *Social Forces* 4(49) (June): 581-595.

[38] - Kandel, Denise and Lesser, Gerald S. 1970. School, Family, and Peer Influences on Educational Plans of Adolescents in the United States and Denmark, *Sociology of Education* 3(43) (Summer): 270-287.

[39] - Kennedy, Eugene. 1995. Correlates of Perceived Popularity Among Peers: A Study of Race and Gender Differences Among Middle School Students, *Journal of Negro Education* 2(64) (Spring): 186-195.

[40] - Kerckhoff, Alan C., and Parrow, Alan A. 1979. The Effect of Early Marriage on the Educational Attainment of Young Men, *Journal of Marriage and the Family* 1(41) (February): 97-107.

[41] - Kerckhoff, Alan C., Campbell, Richard T., Trott, Jerry M., and Kraus, Vered. 1989. The Transmission of Socioeconomic Status and Prestige in Great Britain and the United States, *Sociological Forum* 2(4) (June): 155-177.

[42] - Kessler, Ronald C. 1982. A Disaggregation of the Relationship between Socioeconomic Status and Psychological Distress, *American Sociological Review* 6(47) (December): 752-764.

[43] - Konstantopoulos, Spyros, Modi, Manisha, and Hedges, Larry V. 2001. Who Are America's Gifted?, *American Journal of Education* 109(3): 344-382.

[44] - Kostakis, Anastasia. 1987. Differences Among School Outputs and Educational Production Functions, *Sociology of Education* 60(4) (October): 232-241.

[45] - Kposowa, Augustine J., Breault, K.D., and Singh, Gopal K. 1995. White Male Suicide in the United States: A Multivariate Individual-Level Analysis, *Social Forces* 1(74) (September): 315-325.

[46] - Kupinsky, Stanley. 1971. Non-Familial Activity and Socio-Economic Differentials in Fertility, *Demography* 3(8) (August): 353-367.

[47] - Link, Bruce G., Dohrenwend, Bruce P., and Skodol, Andrew E. 1986. Socio-Economic Status and Schizophrenia: Noisome Occupational Characteristics as a Risk Factor, *American Sociological Review* 2(51) (April): 242-258.

[48] - Makela, Pia. 1999. Alcohol-Related Mortality as a Function of Socio-Economic Status, *Addiction* 6(94) (June): 867-886.

[49] - Martikainen, Pekka. 1995. Mortality and Socio-economic Status Among Finnish Women, *Population Studies* 1(49) (March): 71-90.

[50] - McColl, Mary Ann, Paterson, Margo, Davies, Diane Doubt, Lorna, and Law, Mary. 2000. Validity and community utility of the Canadian Occupational Performance Measure, *Canadian Journal of Occupational Therapy* 1(67) (February): 22-30.

[51] - McLeod, Jane D., and Kessler, Ronald C. 1990. Socioeconomic Status Differences in Vulnerability to Undesirable Life Events, *Journal of Health and Social Behavior* 2(31) (June): 162-172.

[52] - Mheen, H., Dike van de Stronks, Karien, and Mackenbach, Johann P. 1998. A Lifecourse Perspective on Socio-Economic Inequalities in Health: The Influence of Childhood Socio-Economic Conditions and Selection Processes, *Sociology of Health and Illness* 5(20) (September): 754-777.

[53] - Miech, Richard A., Caspi, Avshalom, Moffitt, Terrie E., Wright, Bradley R. Entner, and Silva, Phil A. 1999. Low Socioeconomic Status and Mental Disorders: A Longitudinal Study of Selection and Causation During Young Adulthood, *American Journal Of Sociology* 4(104) (January): 1096-1131.

[54] - Murphy, M. J. 1985. Marital Breakdown and Socio-Economic Status: A Reappraisal of the Evidence from Recent British Sources, *British Journal of Sociology* 1(36) (March): 81-93.

[55] - Mutchler, Jan E. and Poston, Dudley L. Jr. 1983. Do Females Necessarily Have the Same Occupational Status Scores as Males? A Conceptual and Empirical Examination of the Duncan Socioeconomic Status Index and Nam-Powers Occupational Status Scores, *Social Science Research* 4(12) (December): 353-362.

[56] - Natriello, Gary, and McDill, Edward L. 1986. Performance Standards, Student Effort on Homework, and Academic Achievement, *Sociology of Education* 59(1) (January): 18-31.

[57] - Novik, T.S. 2000. Child Behavior Checklist item scores in Norwegian children, *European Child & Adolescent Psychiatry* 1(9) (March): 54-60.

[58] - O'Leary, Richard and Finnas, Fjalar. 2002. Education, Social Integration and Minority-Majority Group Intermarriage, *Sociology* 2(36) (May): 235-254.

[59] - Portes, Alejandro, and MacLeod, Dag. 1996. Educational Progress of Children of Immigrants: The Roles of Class, Ethnicity, and School Context, *Sociology of Education* 4(69) (October): 255-275.

[60] - Ross, Catherine E., and Wu, Chia-Ling. 1996. Education, Age, and the Cumulative Advantage in Health, *Journal of Health and Social Behavior* 1(37) (March): 104-120.

[61] - Russ-Eft, Darlene, Sprenger, Marlene, and Beever, Anne. 1979. Antecedents of Adolescent Parenthood and Consequences at Age 30, *Family Coordinator* 2(28) (April): 173-179.

[62] - Rytina, Steven. 2000. Is Occupational Mobility Declining in the U.S.?, *Social Forces* 4(78) (June): 1227-1276.

[63] - Sampson, Robert J. 1986. Effects of Socioeconomic Context on Official Reaction to Juvenile Delinquency, *American Sociological Review* 6(51) (December): 876-885.

[64] - Scarr, Sandra, and Weinberg, Richard A. 1978. The Influence of 'Family Background' on Intellectual Attainment, *American Sociological Review* 5(43) (October): 674-692.

[65] - Schnabel, Kai U., Alfeld, Corinne, Eccles, Jacquelynne S., Koller, Olaf and Baumert, Jurgen. 2002. Parental Influence on Students Educational Choices in the United States and Germany: Different Ramifications-Same Effect?, *Journal of Vocational Behavior* 2(60) (April): 178-198.

[66] - Stanfiel, James D. 1973. Socioeconomic Status as Related to Aptitude, Attrition, and Achievement of College Students, *Sociology of Education* 4(46) (Autumn): 480-488.

[67] - Stein, Judith A., Fox, Sarah A., and Murata, Paul J. 1991. The Influence of Ethnicity, Socioeconomic Status, and Psychological Barriers on Use of Mammography, *Journal of Health and Social Behavior* 2(32) (June): 101-113.

[68] - Symons, Douglas K., and McLeod, Peter J. 1993. Maternal Employment Plans and Outcomes after the Birth of an Infant in a Canadian Sample, *Family Relations* 4(42) (October): 442-446.

[69] - Thiede, Michael and Traub, Stefan. 1997. Mutual Influences of Health and Poverty. Evidence from German Panel Data, *Social Science & Medicine* 6(45) (September): 867-877.

[70] - Ulbrich, Patricia M., Warheit, George J., and Zimmerman, Rick S. 1989. Race, Socioeconomic Status, and Psychological Distress: An Examination of Differential Vulnerability, *Journal of Health and Social Behavior* 1(30) (March): 131-146.

[71] - Villemez, Wayne J. 1980. Race, Class, and Neighborhood: Differences in the Residential Return on Individual Resources, *Social Forces* 2(59) (December): 414-430.

[72] - Wicks, Jerry W., and Stockwell, Edward G. 1984. A Comment on the Neonatal Mortality-Socioeconomic Status Relationship, *Social Forces* 4(62) (June): 1035-1039.

[73] - Williams, David R., and Collins, Chiquita. 1995. US Socioeconomic and Racial Differences in Health: Patterns and Explanations, *Annual Review of Sociology* 21: 349-386.

[74] - Williams, David R., Takeuchi, David T., and Adair, Russell K. 1992. Socioeconomic Status and Psychiatric Disorder among Blacks and Whites, *Social Forces* 1(71) (September): 179-194.

[75] - Willms, Douglas J. 1986. Social Class Segregation and Its Relationship to Pupils' Examination Results in Scotland, *American Sociological Review* 5(2) (April): 224-241.

[77] - Wood, Michael. 1998. Socio-economic status, delay of gratification, and impulse buying, *Journal of Economic Psychology* 3(19) (June): 295-320.

[78] - Zhang, Li-fang, and Postiglione, Gerard A. 2001. Thinking styles, self-esteem, and socio-economic status, *Personality & Individual Differences* 8(31) (December): 1333-1346.

[79] - Zipp, John F., Landerman, Richard, and Luebke, Paul. 1982. Political Parties and Political Participation: A Reexamination of the Standard Socioeconomic Model, *Social Forces* 4(60) (June): 1140-1153.

APPENDIX G

OCCUPATIONAL STATUS / PRESTIGE INDICES

Oakes and Rossi (2003), give a comprehensive review of SES measurement. They point to a long history of SES indices based upon occupation, starting with Stevenson, a British Census worker, who, in 1913, relied on "expert" knowledge to develop the first Registrar General's Social Class scale:

> An alternative American approach was to define SES through the 'objective' characteristics of educational levels and income associated with occupations. This approach held that the face validity of education requirements for and income derived from an occupation were sufficient to define SES (Nam & Powers, 1965). The idea was that educational attainment determined who could be considered for entry into occupations and the incomes from jobs were the rewards given for investments in education. Accordingly, occupational status – a proxy for SES – was a simple function of educational attainment and income derived from a given occupation. The latent construct came to be known as status.

Nam-Powers occupational status scores were calculated using an equally weighted combination of median education and median income levels for each occupation and the results were mapped onto a score range from zero to one hundred, starting at the bottom with the lowest ranked occupation.

Yet another index is Siegel's PRESTIGE (1971), which, according to Hauser and Warren (1997), was updated by Nakao and Treas in 1994, by regressing their prestige ratings on the characteristics of male and female occupational incumbents in the 1980 Census.

Duncan's SEI also classified occupation according to income and education, but with a more subjective basis than that employed by Nam & Powers. Duncan's SEI is criticized by Mutchler and Poston (1983) as being biased towards "Male" occupations since it is based on 1950 Census occupational classifications. Hauser and Featherman (1977) developed TSEI2 to meet this objection in the 1970's. TSEI2 was updated by Stevens and Cho (1985) to cover the 1980 US Census Occupational Classifications and further updated by Hauser and Warren (1997) to cover the 1980 and the 1990 U.S. Census Occupational Classifications.

For completeness, I mention that Haller and Davis (1981) use Gough's (1971) "Home Index". There is the Hollingshead (1965) Index of Social Position, which was published as four-factor index. Hauser and Warren (1997) refer to the latter as a classification of selected occupational titles into seven occupational grades and about which they remark:

> There must be something appealing about it, for its use has persisted over more than 30 years, and, as far as we know, it has never been published formally.

Two indices have been developed to allow international comparisons, in the first according to Oakes and Rossi (2003):

> Treiman (1975) combined data from 55 national studies of occupational prestige to construct his Standard International Occupational Prestige Score (SIOPS). The innovation here was the inclusion of multinational data

The second from Ganzeboom et al. (1992) was called the International Socio-Economic Index (ISEI) of occupational status. Rather than using prestige as a criterion, they explicitly constructed a set of scores that best account for the correlation between occupational education and occupational income.

The scales are not without their critics for example, Hauser and Warren (1997) say:

> The main problem with occupational prestige ratings is that they lack criterion validity. Prestige is not as highly correlated with other variables as are other measures of occupational social standing-specifically, measures of the socioeconomic status of occupations, as indicated by the average educational attainment and income of occupational incumbents.

Other researchers use simpler, more aggregated scales such as the following simple scale employed by the United Kingdom Office of Population Censuses and Surveys employed a simple scale, reproduced in Table G.01, during the 1980s; or Erickson and Goldthorp's Class Schema and Wright's Class Typology which use 11 and 12 categories of occupation, respectively.

Table G.01 - United Kingdom Office of Population Censuses and Survey Scale.

No	Job Class	Score
I.	Professional	72.85
II.	Intermediate	59.82
III.	Skilled	
	(N) non-manual,	45.94
	(M) manual	40.17
IV.	Partly Skilled	32.13
V.	Unskilled	20.17
VI.	No occupation or unclassified	34.53

Source: United Kingdom Office of Population Censuses and Surveys.

This simpler approach is also implicit in many of my sample research papers.

APPENDIX H

TSEI MAPPINGS ONTO THE 2000 CENSUS OCCUPATIONAL CLASSIFICATIONS

For Massachusetts, at the Town level, the 1980 and 1990 Census present the numbers of persons in 13 major occupational groups and TSEI2 provides a score for each of the groups. The groups and their TESI2 scores are listed in Table H.01.

Table H.01 - 1990 Census Occupational Classifications at Town Level With Associated TSEI2 Scores for Persons in Group.

Census Major Occupational Group with Sub Groups Ranges in Parentheses.	TSEI2 Score
Executive, administrative, and managerial occupations (000-042)	40.22
Professional specialty occupations (043-202)	60.92
Technicians and related support occupations (203-242)	47.05
Sales occupations (243-302)	36.24
Administrative support occupations, including clerical (303-402)	32.24
Private household occupations (403-412)	16.87
Protective service occupations (413-432)	39.29
Service occupations, except protective and household (433-472)	21.96
Farming, forestry, and fishing occupations (473-502)	23.34
Precision production, craft, and repair occupations (503-702)	31.51
Machine operators, assemblers, and inspectors (703-802)	22.58
Transportation and material moving occupations (803-863)	26.50
Handlers, equipment cleaners, helpers, and laborers (864-902)	22.22
Source: Hauser and Warren (1997)	

The 2000 Census used a different set of classifications to the 1980 and 1990 Censuses. In the 2000 Census, the numbers of persons reported at a town level in Massachusetts were given for 46 categories and sub-categories.

It was possible to isolate 16 categories that covered all employed persons and to either map these directly onto 1990 groups or to calculate "TSEI-2000" scores based on simple averages of the TSEI2 scores of the component groups in the 1990 classification, as follows:

1. "Management occupations, except farmers and farm managers" in 2000 were mapped to: 1990 Executive, administrative, and managerial occupations (000-042), giving a score of 40.22 to this category.
2. "Professional and related occupations" in 2000 were mapped to: 1990 Professional specialty occupations (043-202), giving a score of 60.92 to this category.
3. "Healthcare support occupations" in 2000 were mapped to: 1990 Sub-Groups; 445,446 and 447. "TSEI-2000" calculated as simple average of the Sub-Group TSEI2 of 28.65, 25.80 and 24.29 respectively, giving a score of 27.15 to this category.
4. "Protective service occupations" in 2000 were mapped to: 1990 (413-432) Protective Service Occupations, giving a score of 39.29 to this category.
5. "Food preparation and serving related occupations" in 2000 were mapped to: 1990 Food Preparation and Service Occupations (433-444). "TSEI-2000" calculated as simple average of the Sub-Group TSEI2 of 27.23, 26.39, 23.05, 16.42, 16.29, 18.02, 18.02 and 15.54 respectively, giving a score of 20.12 to this category.

6. "Building and grounds cleaning and maintenance occupations" in 2000 were mapped to: 1990 Cleaning and Building Service Occupations, Except Household (448-455). "TSEI-2000" calculated as simple average of the Sub-Group TSEI2 of 28.23, 13.85, 19.56, 22.61 and 27.22 respectively, giving a score of 22.29 to this category.
7. "Personal care and service occupations" in 2000 were mapped to: 1990 Personal Service Occupations (456-472) and Private Household Occupations (403-412). "TSEI-2000" Calculated as simple average of the Sub-Group TSEI2 of 33.01, 25.15, 26.02, 29.30, 33.39, 27.12, (33.39 + 47.23), (27.12 + 30.27), (47.23 + 28.43), (30.27 + 21.79), (28.43 + 24.56), 23.09, (23.02 + 26.10) and 16.87 (Covering the 5 groups in Sub-Groups 403-412), giving a score of 25.85 to this category.
8. "Sales and related occupations" in 2000 were mapped to: 1990 Sales Occupations (243-302), giving a score of 36.24 to this category.
9. "Office and administrative support occupations" in 2000 were mapped to: 1990 Technicians and Related Support Occupations (203-242) (1/3rd of 47.05), and Administrative Support Occupations, Including Clerical (303-402) (2/3^{rds} of 32.24), giving a score of 37.18 to this category.
10. "Farming, fishing, and forestry occupations" in 2000 were mapped to: 1990 Farming Various (473-502). "TSEI-2000" Calculated as simple average of the Sub-Group TSEI2 of 23.34, 27.6, 32.77, 28.88, 33.99, 25.95, 14.92, 19.17, 31.30, 21.30, 25.19, 10.51, 27.16, 30.15, 29.07, 20.55, 33.36, 26.72 and 29.79 respectively, giving a score of 25.88 to this category.

11. "Supervisors, construction and extraction workers" together with "Construction trades workers" in 2000 were mapped to: 1990 Construction Trades (553-612). "TSEI-2000" Calculated as simple average of the Sub-Group TSEI2 of 30.96, 34.47, 42.90, 29.93, 35.13, 35.84, 25.40, 23.04, 27.52, 24.86, 28.31, 24.54, 24.26, 36.95, 27.31, 38.27, 25.83, 31.74, 24.48, 31.42, 26.13, 23.52, 27.09, 27.67, 21.76, 21.46, 28.87, 31.36, 24.86 and 24.19 respectively, giving a score of 28.67 to this category.
12. "Extraction workers" in 2000 were mapped to: 1990 Extractive Occupations (613-627). "TSEI-2000" Calculated as simple average of the Sub-Group TSEI2 of 38.39, 24.19, 26.30, 29.75 and 27.37 respectively, giving a score of 29.20 to this category.
13. "Installation, maintenance, and repair occupations" in 2000 were mapped to: 1990 Mechanics and Repairers (503-552). "TSEI-2000" Calculated as simple average of the Sub-Group TSEI2 of 39.80, 26.16, 28.09, 27.09, 42.01, 25.60, 23.63, 39.72, 30.10, 24.12, 29.81, 28.43, 38.38, 46.65, 31.49, 41.91, 41.80, 38.14, 31.85, 36.47, 31.18, 37.75, 33.48, 40.40, 34.91, 31.44 and 30.26, giving a score of 33.73 to this category.
14. "Production occupations" in 2000 were mapped to: 1990 Precision Production, Craft, and Repair (628-702) ($3/7^{ths}$ of 31.51) and Machine Operators, Assemblers and Inspectors (703-802) ($4/7^{ths}$ of 22.58), giving a score of 26.41 to this category.

15. “Transportation and material moving occupations” in 2000 were mapped to: 1990 Transportation and Material Moving Occupations (803-863), giving a score of 26.50 to this category.
16. The military and unemployed persons were not given a TSEI2 Score in Hauser and Warren (1997), so a score of 35 was allocated to the military and a score of 10 was allocated to unemployed persons.

Calculating a Town “TSEI-2000” and TSEI2 from the data in each Census was then simply a matter of multiplying the score in the category by the TSEI-2000 or TSEI2 score for the category and dividing by the total number of persons in all the categories including the military and the unemployed.

The number of unemployed persons for 2000 was taken from LAUS data, rather than from the Census, because the Census data seemed to err in the case of Wenham in recording 304 women and 418 men as unemployed making a 40%+ unemployment rate when the LAUS showed 26 people unemployed: The Census Bureau – Summary File 3 Data Note 4 – Updated June 2004 – says:

> In July 2002, the Census Bureau issued the following Data Note 4 regarding the Census 2000 Summary File 3 (SF3) data: The Census Bureau is aware there may be a problem or problems in the employment-status data of Census 2000 Summary File 3 (including tables P38, P43-46, P149A-1, P150A-I, PCT35, PCT69A-1, and PCT 70A-1). The labor force data for some places where colleges are located appear to overstate the number in the labor force, the number unemployed, and the percent unemployed, probably because of reporting or processing errors. The exact cause is unknown, but the Census Bureau will continue to research the problem.

APPENDIX I

NORMALITY PLOTS FOR SAMPLE SOCIO-ECONOMIC PROXY VARIABLES

Figure I.01 – Normality Plot 1990 Education Years

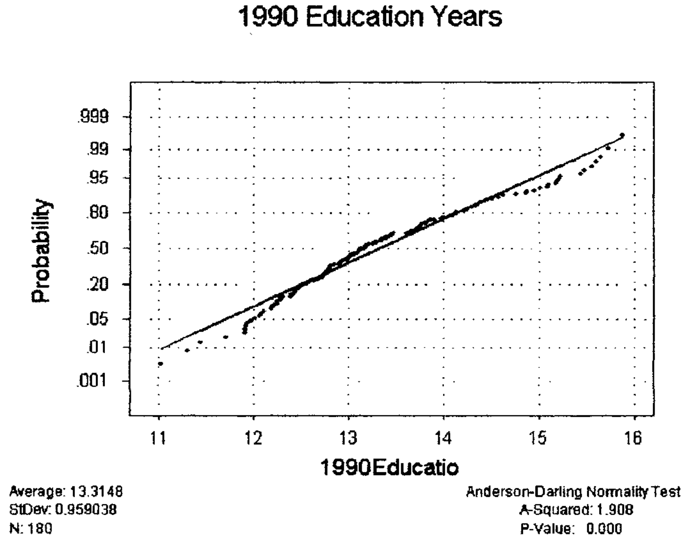


Figure I.02 – Normality Plot 2000 Education Years

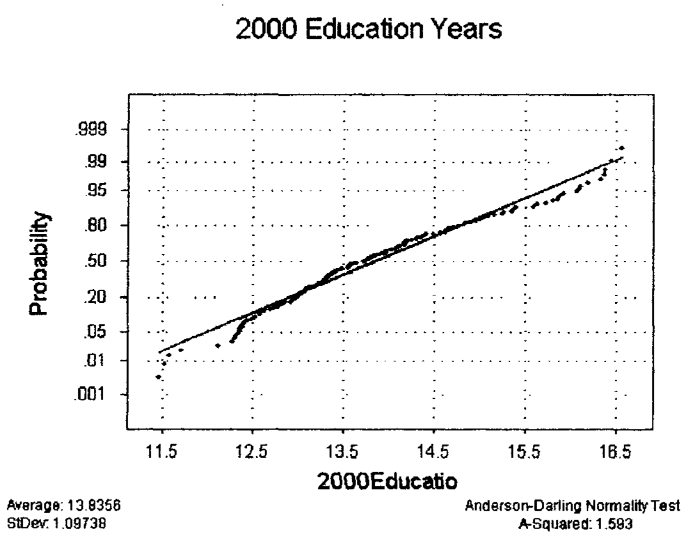


Figure I.03 – Normality Plot 1990 Median Income

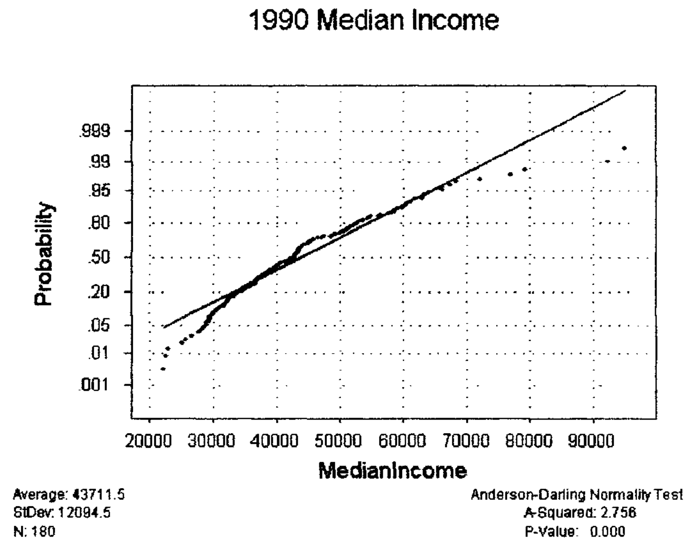


Figure I.04 – Normality Plot 2000 Median Income

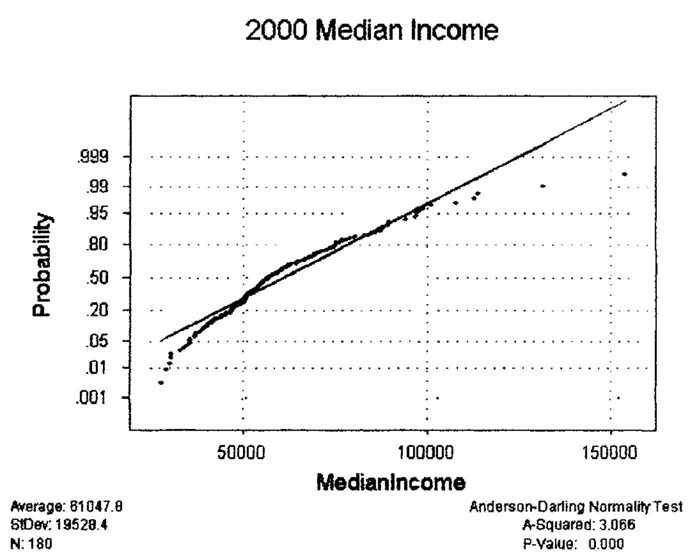


Figure I.05 – Normality Plot 1990 Poverty Index

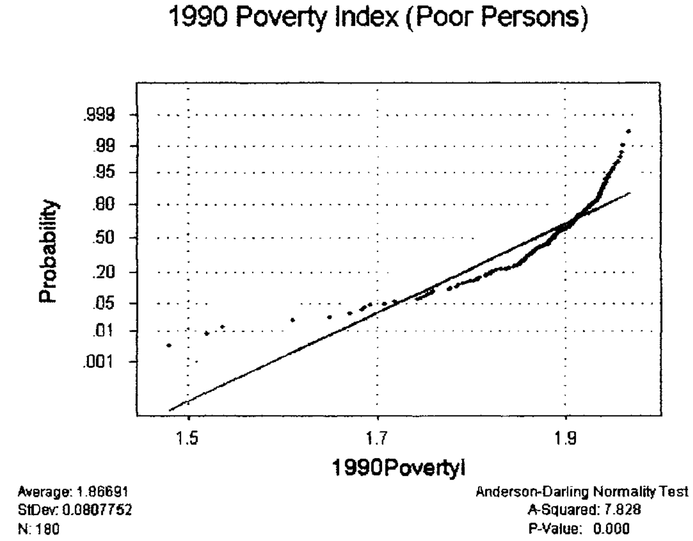


Figure I.06 – Normality Plot 2000 Poverty Index

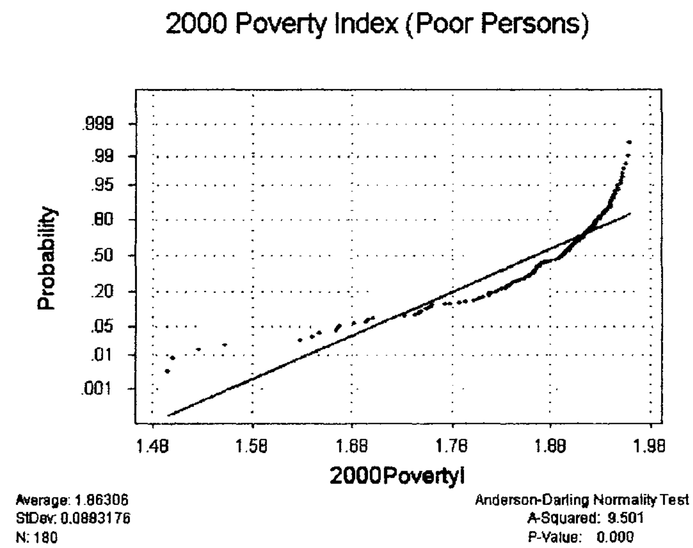


Figure I.07 – Normality Plot 1990 TSEI

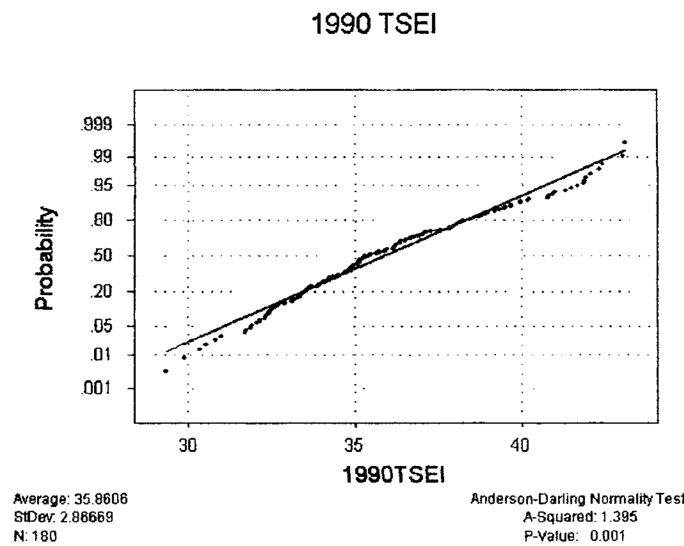
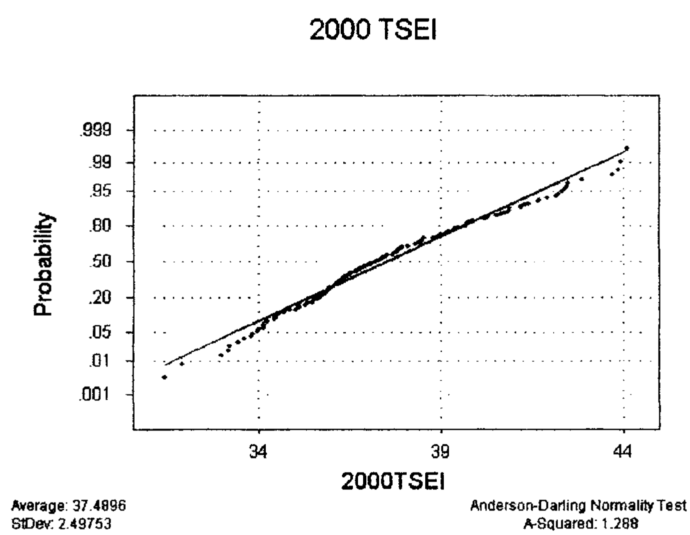


Figure I.08 – Normality Plot 2000 TSEI



APPENDIX J

DATA GENERATED FOR SECTION 6.2

One set of stochastic data, reflecting the assumption that there should be a one off improvement in test scores, was generated using Minitab – Set One. It was assumed that average test scores at school district level are normally distributed.

Another set of data was generated on the assumption of a continual improvement in outcomes from 1994 onward – Set Two.

In each grade, year and Set, three columns of 180 values were generated to represent three different tests and the 180 different geographic units in the sample of Massachusetts' geographies described and selected in Chapter 4.

The normality of test scores was shown in Chapter 5. The mean and standard deviations used to seed the generation of data by Minitab, for each year, grade and Set are given in Table J.01. The range of the standard deviations of the actual 2002 scores was from 5 to 7.5. These two values were used as the seed standard deviations in the generation of Set Two and Set One respectively.

Table J.01 - Means and Standard Deviations Used in Data Generation.								
Set One								
Grade	1988	1990	1992	1994	1996	1998	2000	2002
4th	200 (5)	200 (5)	200 (5)	200 (5)	205 (5)	210 (5)		
8th			200 (5)	200 (5)	205 (5)	210 (5)	210 (5)	210 (5)
Set Two								
	1988	1990	1992	1994	1996	1998	2000	2002
4th	200 (5)	200 (5)	200 (5)	200 (5)	205 (5)	210 (5)		
8th			200 (5)	200 (5)	205 (5)	210 (5)	215 (5)	220 (5)
Note: Mean and (Standard Deviation)								

540 values were generated for each grade in each year and each set in order to give a "score" for 180 units in three "disciplines". Table J.02 gives the means and standard deviations of each group of 540 values generated. Tables J.01 and J.02 are very similar which is to be expected, but there is some variation between the seed values and the values calculated from the results of data generation.

Table J.02 - Means and Standard Deviations Of The Data Generated.								
Set One								
Grd.	1988	1990	1992	1994	1996	1998	2000	2002
4^{th}	200.2 (4.94)	199.8 (4.98)	200.1 (5.08)	199.7 (4.88)	205.2 (5.18)	209.8 (4.91)		
8^{th}			200.1 (5.02)	200.3 (4.98)	204.8 (4.99)	210.4 (5.23)	210.2 (5.11)	210.2 (5.04)
Set Two								
	1988	1990	1992	1994	1996	1998	2000	2002
4^{th}	199.7 (5.06)	200.1 (4.87)	200.2 (5.06)	199.9 (4.96)	205.2 (4.92)	210.1 (4.58)		
8^{th}			200.2 (4.90)	200.1 (4.96)	204.3 (5.00)	210.3 (4.78)	215.3 (4.92)	219.9 (4.78)
Note: Mean and (Standard Deviation)								

APPENDIX K

ADJUSTMENTS FOR VOCATIONAL REGIONS

Vocational Regions each cover many more towns than do Academic Regions and there are 115 towns whose only involvement in a Region is to send high school students to a Vocational Region. Combining Towns and Academic Regions up to the level of Vocational Regions would have resulted in too small a sample and would have diluted the Socio Economic variance between the sample geographies. Instead the Vocational Region scores were disaggregated back into the towns and Academic Regions.

For example the Town of Uxbridge sends students to Blackstone Valley Vocational Region[43]. In 1996 Uxbridge had 126 students in grade 10 who scores 1,320 in the Reading portion of the MEAP tests. In addition, the town sent 27 students to Blackstone Valley where the average score on the Reading portion of the MEAP tests was 1,180; this gave Uxbridge an average grade 10 Reading MEAP score of 1,295.29, thus:

$$\frac{(126*1320)+(27*1180)}{126+27}=1295.29$$

The scores adjusted for Vocational Region scores were on average 99.03 percent of the unadjusted scores. The minimum percentage was 94.40 percent and the maximum was 100.16 percent, so on the whole the adjustment made little difference to the scores and the results based on test scores adjusted for Vocational Education are not presented.

---

[43] The DOE supplied Town Enrollment, provides, at a town level and by Grade, the total number of students in the following categories: In Local Public Schools; At Academic Regional Schools; At Vocational Regional Schools; In Collaboratives; At "Out of District" Public Schools; at "In State" Private Schools, and At "Out of State" Private Schools.

APPENDIX L

ASSESSMENT IN MASSACHUSETTS 1988 TO 2002

L.1 Assessment Systems

Three, state sponsored, systems of assessment were used in Massachusetts between 1985 and 2003. The Massachusetts Basic Skills Testing Program was abandoned in 1991. The other two systems of assessment ran, in series, continuously from 1986 to date, starting with the Massachusetts Educational Assessment Program (MEAP) and continuing with the Massachusetts Comprehensive Assessment System (MCAS).

MEAP tested students at grades 4, 8 and 12 every other year from 1988 to 1996 in Reading, Mathematics, Science and Social Studies – see Table L.01. MCAS was supposed to test each of the seven curriculum frameworks in at least three grades each year – refer to Chapter 2 section 3.

In 1998, 1999 and 2000 three subjects were tested and scores reported at grades 4, 8 and 10 – English Language Arts, Mathematics and Science and Technology/Engineering. History was tested in the 8th grade from 1999 to 2002.

In 2001 and 2002 two subjects were tested: English Language Arts, in grades 4,7 and 10 (and at grade 8 in 2001); and Mathematics, in grades 4,8,10 and 6.

Table L.01 - Massachusetts Student Assessment 1988 to 2002. Subjects Assessed and Reported at Grade.

Assessment and Year	Subjects and Grades Assessed and Reported				
MEAP	Reading	Mathematics	Science	Social Studies	
1988	4,8,12	4,8,12	4,8,12	4,8,12	
1990	4,8,12	4,8,12	4,8,12	4,8,12	
1992	4,8,12	4,8,12	4,8,12	4,8,12	
1994	4,8,10	4,8,10	4,8,10	4,8,10	
1996	4,8,10	4,8,10	4,8,10	4,8,10	
MCAS	English Language Arts.	Mathematics	Science and Technology / Engineering	History and Social Science	
1998	4,8,10	4,8,10	4,8,10		
1999	4,8,10	4,8,10	4,8,10	8	
2000	4,8,10	4,8,10	4,8,10	8	
2001	4,7,8,10	4,8,10,*6*		8	
2002	4,7,10	4,8,10,*6*		8	
Source: Massachusetts Department of Education					

The published scores for 2003 and 2004 comprise the percentages of students in each test achieving one of four levels of proficiency rather than the average scaled scores reported for previous years. This data could not be compared with the data from previous years and 2003 and 2004 were dropped from the analysis. Finally, for convenience, odd years MCAS data were ignored so that the test scores used would come from even years from 1988 to 2002.

L.2 MEAP

MEAP was mandated by Chapter 188, in 1985, to provide comparisons of performance at school, district, state and national levels and to improve curriculum and instruction. MEAP exempted certain students with special needs or limited English proficiency.

The tests were in a "matrix sampling" format in which the questions in a subject area were distributed across 13 different test forms. It required 13 students to take the assessment for all the topics in a subject to be assessed and allowed highly reliable results at the building level. "Matrix Sampling" also allowed a minimal assessment time as each student was randomly assessed on less than the full curriculum. Assessment time of between twelve and twenty hours was reduced to 90 minutes. The tests were multiple-choice in 1986, 1988, and 1990.

Average scaled scores were reported at state, school and district levels. Proficiency levels were identified and the percentage of students reaching each level was also published.

In 1990 the Mathematics Framework was adjusted to reflect the emphasis placed by the National Council of Teachers of Mathematics on problem solving and reasoning. The Reading Framework was expanded to capture the role that reading strategies play in constructing meaning.

For the 1992 MEAP, one test form at each grade level and 30 percent of the points were allocated to a set of open-ended, essay type, questions, requiring students to provide a written response up to one page long. The remainder of the points went on multiple-choice questions.

The 1994 MEAP remained unchanged from 1992, except that students were tested in Grade 10 rather than in grade 12. Assessment in each content area was modified to reflect this change. The 1996 MEAP was unchanged from 1994.

L.3 MCAS

MCAS, an annual assessment, began in 1998. MERA required that the MCAS tests should be based on the academic learning standards contained in the Massachusetts Curriculum Frameworks, which were only completed in October 2003. The assessments are supposed measure how well students, schools and districts are doing on meeting the academic learning standards.

Since part of the objective was to assess individual students, as well as schools, districts and the state as a whole, assessments are no longer completely of a "matrix sample" design.

At each grade and subject there were 12 Student Test Booklets, featuring "Common Questions" (the same in each booklet) and "Matrix Questions" (different in each booklet). The assessments comprised: Multiple-Choice Questions; Short Answer Questions; Open-Response Questions, and Writing Prompts.

<u>All</u> public school students were and are required to participate. A Spanish-language version of the MCAS is available, so only those, non-English and non-Spanish speaking students who have been enrolled in U.S. Schools for three or fewer years and those home-schooled are exempt.

APPENDIX M

THE SCHOOL DISTRICT SAMPLE

The sample selected, listed in Appendix C is indistinguishable from the state when compared on a number of factors.[44]

Taking the age ranges first it is clear, from Table M.01, that the 5,624,242 people in the sample towns were virtually identical to the 6,349,097 people in the state in age distribution in the 2000 Decennial Census and (not shown) also in 1980 and 1990.

Table M.01 - Sample and State by Age Ranges.

Age	Sample	State
Under 7	8.853%	8.832%
7 to 9 years	4.198%	4.180%
10 to 13 years	5.455%	5.476%
14 to 18 years	6.393%	6.434%
19 to 29 years	14.782%	14.532%
30 to 44 years	24.677%	24.683%
45 to 59 years	18.421%	18.600%
60 to 74 years	10.461%	10.503%
75 or more years	6.760%	6.760%

Source 2000 Decennial Census

Taking average years of education the sample is marginally less well educated with an average 13.609 years in 2000 compared with 13.639 years for the state – refer to Table M.02.

[44] Where numbers and percentages are presented for a particular year they are representative of the differences between the sample and the state in the data available for the years from 1988 to 2002.

Table M.02 - Average Educational Attainment In Years.			
Year	1980	1990	2000
Sample	12.622	13.139	13.609
State	12.649	13.166	13.639
Source - Decennial Census, 1980, 1990 and 2000			

Unemployment is slightly higher in the sample towns at 4.62 percent as against 4.55 percent for the state – refer to Table M.03.

Table M.03 - Employment and Unemployment.		
	Percent Employed	Percent Unemployed
Sample	95.21%	4.62%
State	95.29%	4.55%
Source - 2000 Decennial Census.		

The sample and the state have very similar levels of poverty. The Poverty Index calculated from the 2000 Decennial Census was 1.737 for the sample towns and 1.741 for the state. (Refer to the next section for the derivation of this index). Taking DOE data on the numbers of children from Low Income families, the state had 25.34% defined as low income in 2001-02 and the sample had 23.21% defined as low income.

The sample was slightly less likely to be owner-occupier with 60.58% owner occupation in 2000 against 61.72% owner occupation for the state. The sample had a slightly lower average "TSEI2000" (refer to the next section) in 2000 at 36.49 as compared to 36.56 for the state as a whole.

The sample was slightly less White and more Hispanic than the state as a whole according to the 2000 Decennial Census – see Table M.04.

Table M.04 - Racial and Ethnic Composition - 2000.		
	State	Sample
White	84.50%	83.58%
African American	5.31%	5.72%
American Indian	0.24%	0.24%
Asian	3.75%	3.88%
Other	0.03%	0.03%
Mixed	6.16%	6.55%
Hispanic	6.73%	7.23%
Source - 2000 Decennial Census		

The K - 12 population is more diverse than the general population with only 75.74% of the state's K - 12 students being white – see Table M.05. Unfortunately, the sample under-represents African American students who make up 5.82% of the sample towns' students and 8.60% of the state's students. This is due to the fact that Boston's student population was 47.57% African American in 2002 and that Boston was dropped from the sample because of its high Charter School Enrollment.

Table M.05 - Racial and Ethnic Composition of Students 2002.		
	State	Sample
White	75.74%	79.11%
African American	8.60%	5.82%
Native American	0.33%	0.30%
Asian	4.54%	4.36%
Hispanic	10.79%	10.41%
Source Massachusetts Department of Education		

The sample towns had a lower drop out rate in 2000, 3.16 percent compared with 3.43 percent for the state. The number of students to each teacher was slightly higher for the state in 2001 at 13.62 students per teacher as against 13.50 students per teacher in the sample towns.

In 2000 Integrated Per-Pupil Expenditure was lower in the sample than for the state at $7,874 against $7,703 for the sample. This reflects the absence of Boston where the Integrated Per-Pupil Expenditure was $10,026 in 2000.

The State also has a higher proportion of Special Education students who made up 15.41% of the state student population and 15.09% of the sample student population. Boston, which had 20.26 percent Special Education students, was again the main factor behind this disparity.

Although there were some issues with the sample, these were not significant and the 180 Towns and Regions were accepted as the geographical units forming the basis of the analyses.

BIBLIOGRAPHY

Education and Socio-Economic Status

Baron, Joan Boykoff, and Sergi, Theodore S. 1979. Minimum Competency Testing: The Role of the State Education Agency, *Education and Urban Society* 1(12) (November): 81-88.

Behrman, Jere R. and Stacey, Nevzer. eds. 1997. *The Social Benefits of Education*, University of Michigan Press, Ann Arbor, Michigan 266 pages.

Betts, Julian R., and Costrell, Robert M. 2001. *Incentives and Equity under Standards-Based Reform*, Brookings Papers on Education Policy, pages 9-74.

Bowles, Samuel. 1974. Towards an Educational Production Function. In *Education, Income and Human Capital - Towards an Education Production Function*, Hansen, W. Lee. Ed. NBER, Chicago.

Bracey, Gerald W. 1997. What happened to America's public schools? (Cover Story), *American Heritage* 7(48) (November): 38-46.

Breckheimer, Steve, Armistead, Rhonda and Armistead, Leigh. 2001. *Grading the North Carolina Student Accountability Standards. Student Accountability Standards and High-Stakes Testing in North Carolina: A Position Statement and Supporting Paper*, North Carolina School Psychology Association, Charlotte (January).

Brown, George D. 1994. Binding advisory opinions: a federal courts perspective on the state school finance decisions. (Symposium: Issues in Education Law and Policy), *Boston College Law Review* 3(35) (May): 543-568.

Cardman, Michael. 2002. Federal judge boots MCAS lawsuit to state courts. (Massachusetts Comprehensive Assessment System) (Brief Article), *Education Daily* 228(35) (December): page 3.

Coleman, James Samuel. 1966. *Equality of Educational Opportunity (The "Coleman Report.")*, Washington: U.S. Dept. of Health, Education, and Welfare, Office of Education.

Cohn, Elchanan. 1968. Economies of Scale in Iowa High School Operations, *Journal Of Human Resources* 4(3) (Fall): 422-434.

Commonwealth of Massachusetts. 1993. *The Education Reform Act of 1993*, Commonwealth of Massachusetts.

Congressional Quarterly Inc. 1981. *Education in America, quality vs. cost*, Congressional Quarterly, Inc., Washington, D.C., 202 pages.

Editorial Research Reports. 1978. *Competency Tests*, Washington, D.C., Editorial Research Reports, Congressional Quarterly Inc. (August).

EyeOnEducation. 2002. *Reform*, EyeOnEducation, EyeOnEducation.

Futrell, Mary Hatwood and Brown, Walter A. 2000. Should African Americans Support the Current Education Reform Standards Movement?, *Journal of Negro Education* 69(4) (Fall): 288-304.

Finn, Chester E. Jr., and Kanstoroom, Marci. 2001. *State Academic Standards*, Brookings Papers on Education Policy, pages 131-179.

Galles, G.M., and Sexton, R.L. 1998. A Tale Of Two Tax Jurisdictions: The Surprising Effects Of California's Proposition 13 And Massachusetts' Proposition 2½, *American Journal Of Economics And Sociology* 2(57) (April): 123-133.

Gaudet, Robert D. 1994. The Impact of School Spending on Student Achievement: Results of MEAP Statewide Tests, *New England Journal of Public Policy* 1(10) (Summer-Fall): 9-28.

———. 1999. *Education Reform Over the Years*, Education Benchmarks, University of Massachusetts. (http://www.edbenchmarks.org/)

Gauthier, William J. Jr., and Pecheone, Raymond L. 1985. Schools Can Become More Effective, *Journal of Negro Education* 3(54) (Summer): 388-408.

Gipps, Caroline. 1988. The Debate over Standards and the Uses of Testing, *British Journal of Educational Studies* 1(36) (February): 21-36.

Gough, H. 1971. A cluster analysis of home index status items, *Psychological Reports* 9(28): 23-29.

Guthrie, James W. ed. 1980. *School Finance Policies and Practices*, Ballinger: Cambridge MA, pages 3-46.

Hanushek, Eric A. 1994. Making America's schools work. (Improving the public school system), *Brookings Review* 4(12) (Fall).

———. 1996. Measuring Investment in Education (in Symposia: Primary and Secondary Education), *Journal of Economic Perspectives* 4(10) (Autumn): 9-30.

Hanushek, Eric A., and Raymond, Margaret E. 2001. The confusing world of educational accountability, *National Tax Journal* 2(54) (June): page 365.

Hatward, Steven. 1998. The Tax Revolt Turns 20, *Policy Review* 90 (July-August): 9-12.

Hauser, Robert M., and Featherman, David L. 1977. *The Process of Stratification Trends and Analyses*, New York: Academic Press.

Hollingshead, A.B. 1965. *Four factor index of social status,* New Haven, CT, Yale Station.

Jordan, K. Forbis, and Lyons, Teresa S. 1992. Courts and School Finance. In *Financing Public Education In An Era of Change*, USA, Bloomington, Ind., Phi Delta Kappa Educational Foundation.

Keisling, Herbert. 1967. Measuring a Local Government Service: A Study of School Districts in New York State, *Review Of Economic Statistics* 3(49) (August): 356-367.

Kreft, Ita G.G. 1993. Using Multilevel Analysis to Assess School Effectiveness: A Study of Dutch Secondary Schools, *Sociology of Education* 2(66) (April): 104-129.

Levin, Henry M., Froomkin, Joseph T., Jamison, Dean T., and Radner, Roy. eds. 1976. *Education as an Industry - Economic Efficiency and Educational Production*, NBER: Cambridge MA, pages 149-198.

Levinson, Meira. 1999. *The demands of liberal education*, Oxford, New York, Oxford University Press, 237 pages.

Link, Charles R., and Ratledge, Edward C. 1979. Student Perceptions, I.Q., and Achievement, *Journal Of Human Resources* 1(14) (Winter): 98-111.

Massachusetts Department of Education. 1987. *The Massachusetts Educational Assessment Program: description of test content and reporting categories 1987*, Massachusetts Department of Education.

———. 1988. *Chapter 188 State Report on High School Graduation Requirements*, Massachusetts Department of Education (August).

———. 1991. *Report on the Condition of the Public Schools in Holyoke, Lawrence, Brockton and Chelsea*, Massachusetts Department of Education (October 15).

———. 1992. *A guide to the 1992 Massachusetts Educational Assessment Program*, Massachusetts Department of Education.

———. 1993. *Reorganization plan*, Massachusetts Department of Education: Quincy, Massachusetts, 50 pages.

———. 1994a. *Massachusetts Common Core of Learning*, Massachusetts Department of Education: Malden, Massachusetts, 1994, 1-13.

———. 1994b. *Background Paper: Eliminating the General Track*, Massachusetts Department of Education: Malden, Massachusetts (June).

———. 1994c. *First Annual Education Reform Implementation Report*, Massachusetts Department of Education (Fall).

———. 2000. *2000 MCAS Composition Writing Prompts, September 2000*, Massachusetts Department of Education: Malden, Massachusetts, (September): 1-19.

———. 2002. Lawsuit claims MCAS are unfair, *District Administration* 11(38) (November): page 14.

Massachusetts Department of Education and Massachusetts Business for Education. 1991. *Every Child a Winner! A Proposal for a Legislative Action Plan for Systemic Reform of Massachusetts` Public Primary and Secondary Education System*, Massachusetts Department of Education and Massachusetts Business for Education, Worcester (July).

Massachusetts Department of Revenue. 2002. *First Findings The Summative Report Of The Educational Management Accountability Board On The Audits Of Massachusetts School Districts And The Impact Of The Education Reform Act*, Education Management Accountability Board, Massachusetts Department of Revenue (March): 1-53.

McDuffy v. Robinson. 1993. *Jami McDuffy and others v. Secretary of the Executive Office of Education & others*, Massachusetts Supreme Court (June) (415 Mass. 545; 615 N.E.2d 516; 1993 Mass.)

McNeely, Donald R. 1985. School Effectiveness Training: An Education Association's Initiative for Instruction Improvement, *Journal of Negro Education* 3(54) (Summer): 462-472.

Michelson, Stephen, and Levin, Henry M. 1970. *Do Teachers Make a Difference?* United States Office of Education, U.S. GPO: Washington DC, 1970, 120-168.

Mosteller, Frederick, Moynihan, Daniel P. eds. 1972. *On Equality of Educational Opportunity*, Random House: NY, 547 pages.

Murnane, Richard J. and Levy, Frank. 2001. Will Standards-based Reforms Improve the Education of Students of Color?, *National Tax Journal* 2(54) (June): 401-415.

Murphy, Jerome T and Cohen, David K. 1974. Accountability in Education--the Michigan Experience, *The Public Interest* 36 (Summer): 53-81.

Nuttall, Ronald, Hell, Regina Jean-Van. 2001. *Poverty, Courses Taken and MCAS Test Scores in Mathematics and Science*, Paper presented at the Annual Meeting of the American Educational Research Association Seattle, WA, April 10-14, 2001.

Oakes, Michael J. and Rossi, Peter H. 2003. The Measurement of SES in Health Research: Current Practice and Steps Toward a New Approach, *Social Science & Medicine*, 56:769-784.

Parker, Francis W. 1997. *Overview of the Massachusetts Education Reform Act of 1993*, Fairtest Organization (http://www.fairtest.org/care/Overview_of_MA_Ed_Reform.html)

Phillips S.E. 1993. Legal defensibility of standards: Issues and policy perspectives, *Boston Bar Journal* 3(98) (January-February): page 16.

Raymond, Richard. 1968. Determinants of the Quality of Primary and Secondary Public Education in West Virginia, *Journal Of Human Resources* 4(3) (Fall): 450-470.

Rossman, Gretchen B., and Anthony, Patricia G. 1994. *The Massachusetts Education Reform Act: What Is It and Will It Work?* U.S. Department of Education, Office of Educational Research and Improvement (OERI), National Library of Education (NLE), Educational Resources Information Center (ERIC).

Sebold, Frederick D., and Dato, William. 1981. School Funding and Student Achievement: An Empirical Analysis, *Public Finance Quarterly* 1(9) (January): 91-105.

Sexton, T.A., Sheffrin, S.M., and O'Sullivan, A. 1999. Proposition 13. Unintended Effects And Feasible Reforms, *National Tax Journal* 1(52) (March): 99-111.

Sims, Edith R. 1988. Successful Programs, Policies, and Practices Employed at Corliss High School (in Case Studies), *Journal of Negro Education* 3(57) (Summer): 394-407.

Smith, Marshall S., Stevenson, David L., and Li, Christine P. 1998. Voluntary national tests would improve education, *Educational Leadership* 6(55) (March): 42-44.

Town of Shrewsbury, MA. 2000. *Property Taxes and Proposition 2 and a half.* A description of Proposition 2½ from the town of Shrewsbury Web Site (http://www.shrewsbury-ma.gov/assessor/taxinfoprop2.asp).

Thro, William E. 1994. Judicial analysis during the third wave of school finance litigation: the Massachusetts decision as a model. (Symposium: Issues in Education Law and Policy), *Boston College Law Review* 3(35) (May): 597-617.

United States Department of Education. 1986. *What Works: Research About Teaching and Learning*, United States Department of Education.

Vogler, Kenneth E. 2002. The Impact of High-Stakes, State-Mandated Student Performance Assessment on Teachers' Instructional Practices, *Education* 1(123) (Fall): 39-55.

West, Edwin G. 1983. *Are American Schools Working? Disturbing Cost and Quality Trends*, Policy Analysis, Cato Institute, 1000 Massachusetts Avenue NW, Washington, D.C. 20001 (August).

Weisman, Michael D., Bloom, Sara Miron, and Simonoff, Mark A. 1993. McDuffy v. Robertson. (School finance), *Boston Bar Journal* 2(37) (March-April): 14-19.

Weisman, Michael D., and Simonoff, Mark A. 1993. McDuffy v. Secretary of the Executive Office of Education: a new era for public school children in Massachusetts, *Boston Bar Journal* 1(38) (January-February): 4-7.

Wheelock, Anne. 2002. *School Awards Programs and Accountability in Massachusetts. Misusing MCAS scores to assess school quality*, U.S. Department of Education, Office of Educational Research and Improvement (OERI), National Library of Education (NLE), Educational Resources Information Center (ERIC), 2002 March.

Wilkins, Douglas H., Blumenthal, Robert H., and Connaughton, Mary C. 1993. The defendants' brief in the school finance case: McDuffy v. Robertson: an excerpt and a summary, *Boston Bar Journal* 2(37) (March-April): 18-20.

Woodhall, Maureen and Blaug, Mark. 1968. Productivity Trends in British Secondary Education, 1950-63, *Sociology of Education* 1(41) (Winter): 1-35.

Wirtz, Willard. 1977. *On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline*, College Entrance Examination Board: NY, pages 1-48.

Young, Rufus Jr. 1988. A Process for Developing More Effective Urban Schools: A Case Study of Stowe Middle School, *Journal of Negro Education* 3(57): 307-334.

MEAP and MCAS

Abar, Sylvia H. 1996. The perceived influences that prompt teachers to initiate changes in curriculum and instruction, Thesis (Ed.D.), University of Massachusetts at Amherst, 295 pages.

Gaudet, Robert D. 1994. The Impact of School Spending on Student Achievement: Results of MEAP Statewide Tests, *New England Journal of Public Policy* 1(10) (Summer-Fall): 9-28.

———. 2002. The Fourth Annual Report, *Education Benchmarks*, University of Massachusetts (January) (http://www.edbenchmarks.org/).

Luna, Catherine and Turner, Cara Livingstone. 2001. The Impact of the MCAS: Teachers Talk about High-Stakes Testing, *English Journal* 1(91) (September): 79-87.

Massachusetts Department of Education. 1986. *Using the school report / Massachusetts Educational Assessment Program 1986*, Massachusetts Department of Education (September): 22 pages.

———. 1987. *The Massachusetts Educational Assessment Program: description of test content and reporting categories 1987*, Massachusetts Department of Education.

———. 1988. *MEAP Summary of District Performance, 1986 and 1988*, Massachusetts Department of Education.

———. 1988. *Using the school report / Massachusetts Educational Assessment Program 1988*, Massachusetts Department of Education (November): 30 pages.

———. 1990a. *Massachusetts Basic Skills Tests: 1989 Results*, Massachusetts Department of Education (March): 21 pages.

———. 1990b. *Description of the MEAP Proficiency Scales. Massachusetts Educational Assessment Program*, Massachusetts Department of Education (November).

———. 1990c. *Using the school report / Massachusetts Educational Assessment Program 1990*, Massachusetts Department of Education (November).

———. 1992. *A guide to the 1992 Massachusetts Educational Assessment Program*, Massachusetts Department of Education.

———. 1994. *The Massachusetts Educational Assessment Program: description of test content and reporting categories 1994*, Massachusetts Department of Education.

———. 1996. *1996 MEAP Statewide Summary*, Massachusetts Department of Education (October).

———. 1998a. *Overview of the MCAS Program*, Massachusetts Department of Education.

———. 1998b. *Guide to Interpreting the 1998 MCAS Reports for Schools and Districts*, Massachusetts Department of Education, 64 pages.

———. 1999. *Guide to Interpreting the 1999 MCAS Reports for Schools and Districts*, Massachusetts Department of Education, 45 pages.

———. 2000a. *Alternate Assessment for Students with Disabilities: Educator's Manual for the MCAS Alternative Assessment*, Massachusetts Department of Education.

———. 2000b. *MCAS Update March 6th 2000*, Massachusetts Department of Education (March): 1-3.

———. 2000c. *Report of 2000 Massachusetts and Local School District MCAS Results by Gender*, Massachusetts Department of Education (September): 1-19.

———. 2001. *2001 Spring MCAS tests: summary of district performance*, Massachusetts Department of Education (November).

———. 2002a. *Progress report on the class of 2003: percentage of students who have earned a competency determination statewide and by district*, Massachusetts Department of Education (May): 1-15.

———. 2002b. *Progress report on the class of 2003: percentage of students who have earned a competency determination statewide and by district*, Massachusetts Department of Education (September): 1-12.

———. 2002c. *Evaluating School Performance - Stage 2 - The Panel Review Process*, Massachusetts Department of Education (October): 1-13.

———. 2003. *Requirements for the participation of students with disabilities: a guide for educators and parents. Spring 2003 Update*, Massachusetts Department of Education.

Russell, Michael, and Plati, Tom. 2000. *Mode of Administration Effects On MCAS Composition. Performance for Grades Four, Eight, and Ten. A Report Of findings Submitted to the Massachusetts Department of Education. NBEETP Statements World Wide Web Bulletin*, National Board on Educational Testing and Public Policy, Chestnut Hill, MA (May).

DEA

Aigner, Dennis J., and Chu, Sing Fat. 1968. On Estimating the Industry Production Function, *American Economic Review* 4(58) (September): 826-839.

Ali, Agha Iqbal, and Seiford, Lawrence M. 1990. Translation invariance in data envelopment analysis, *Operations Research Letters* 6(9): 403-405.

———. 1993. Computational accuracy and infinitesimals in data envelopment analysis, *INFOR Journal* 4(31): page 290.

Ali, Agha Iqbal. 1994. Computation Aspects of DEA. In *Data Envelopment Analysis: Theory, Methodology and Applications*. Charnes, Abraham, Cooper William W., Lewin, A. and Seiford, Lawrence M. eds. Boston, Kluwer Academic Publishers, pages 63-88.

Ali, Agha Iqbal, Lerme, Catherine S., and Seiford, Lawrence M. 1995. Components of efficiency evaluation in data envelopment analysis, *European Journal Of Operational Research* 3(80): 462-473.

Allen, R., Athanassopoulos, Antreas D., and Dyson, R.G. 1997. Weights restrictions and value judgements in data envelopment analysis: Evolution, development and future directions, *Annals Of Operations Research* 73:13-34.

Andersen, Per, and Petersen, Niels Christian. 1993. A Procedure for Ranking Efficient Units in Data Envelopment Analysis, *Management Science* 10(39) (October): 1261-1264.

Anderson, L., Walberg, H.J., and Weinstein, T. 1998. Efficiency and effectiveness analysis of Chicago public elementary schools: 1989, 1991, 1993, *Educational Administration Quarterly* 4(34) (October): 484-504.

Arnold, Victor L., Bardhan, Indranil R., and Cooper, William W. 1994. Excellence and efficiency in Texas public schools, *Texas Business Review* (February): 5-7.

Arnold, Victor L., Bardhan, Indranil R., Cooper, William W., and Kumbhakar, Subal C. 1996. New uses of DEA and statistical regressions for efficiency evaluation and estimation - With an illustrative application to public secondary schools in Texas, *Annals Of Operations Research* 66: 255-277.

Banker, Rajiv D., Charnes, Abraham, and Cooper, William W. 1984. Some Models for Estimating Technical and Scale Efficiencies in Data Envelopment Analysis, *Management Science* 9(30) (September): 1078-1092.

Banker, Rajiv D., Charnes, Abraham, Cooper, William W., and Schinnar, A.P. 1981. Bi-Extremal Principle For Frontier Estimation and Efficiency Evaluations, *Management Science* 12(27) (December): 1370-1382.

Banker, Rajiv D., and Thrall, Robert M. 1992. Estimation of Returns to Scale Using Data Envelopment Analysis, *European Journal Of Operational Research* 1(62): 74-84.

Banker, Rajiv D., Gadh, Vandana M., and Gorr, Wilpen L. 1993. A Monte Carlo comparison of two production frontier estimation methods: corrected ordinary least squares and data envelopment analysis, *European Journal Of Operational Research* 3(67) (June 25): 332-341.

Banker, Rajiv D., Chang, Hsi-Hui., Cooper, William W. 1996. Simulation studies of efficiency, returns to scale and misspecification with nonlinear functions in DEA, *Annals Of Operations Research* 66:233-253.

Bardhan, Indranil R., Bowlin, William F., Cooper, William W., and Sueyoshi, Toshiyuki. 1996. Models and measures for efficiency dominance in DEA 1 Additive models and MED measures, *Journal Of The Operations Research Society Of Japan* 3(39) (September): 322-332.

Bates, J. M. 1997. Measuring predetermined socioeconomic 'inputs' when assessing the efficiency of educational outputs, *Applied Economics* 1(29) (January): 85-94.

Beasley, J.E., Wong, Y-H.B. 1990. Restricting weight flexibility in data envelopment analysis, *Journal Of the Operational Research Society* 9(41): 829-835.

Bessent, Authella M., and Bessent, E. Wailand. 1980. Determining the Comparative Efficiency of Schools Through Data Envelopment Analysis, *Educational Administration Quarterly* 2(16): 57-75.

Bessent, Authella M., Bessent, E. Wailand, Kennington J., and Reagan B. 1982. An Application of Mathematical Programming to Assess Productivity in the Houston Independent School District, *Management Science* 12(28) (December): 1355-1367.

Bessent, Authella M., Bessent, E. Wailand, Elam, J., and Clark, T. 1988. Efficiency frontier determination by constrained facet analysis, *Operations Research* 5(36): 785-796.

Bogetoft, Peter. 1996. DEA on relaxed convexity assumptions, *Management Science* 3(42) (March): 457-466.

Bogetoft, Peter, Tama, Joseph M., and Tind, Jorgen. 2000. Convex Input and Output Projections of Nonconvex Production Possibility Sets, *Management Science* 6(46) (June): page 858.

Bonnesronning, Hans, and Rattso, Jorn. 1994. Efficiency Variation among the Norwegian High Schools: Consequences of Equalization Policy, *Economics Of Education Review* 4(13) (December): 289-304.

Brockett, Patrick L., Cooper, William W., Golden, L.L., Rousseau, John J., and Wang, Y. 1998. DEA evaluations of the efficiency of organizational forms and distribution systems in the US property and liability insurance industry, *International Journal Of Systems Science* 11(29): 1235-1247.

Brown, Rayna, Brown, Rob, and O'Connor, Ian. 1999. Efficiency, Bond of Association and Exit Patterns in Credit Unions: Australian Evidence, *Annals Of Public & Cooperative Economics* 1(70) (March): 5-23.

Byrnes, Patricia E., Färe, Rolf, and Grosskopf, Susan. 1984. Measuring Productive Efficiency: An Application to Illinois Strip Mines, *Management Science* 6(30) (June): 671-681.

Chalos, Peter, and Cherian, Joseph. 1995. An application of data envelopment analysis to public sector performance measurement and accountability, *Journal Of Accounting and Public Policy* 2(14) (Summer): 143-160.

Chang, Kuo-Ping, and Guh, Yeah-Yuh. 1991. Linear production functions and the data envelopment analysis, *European Journal Of Operational Research* 2(52): 215-223.

Chang, Kuo-Ping. 1999. Measuring efficiency with quasiconcave production frontiers, *European Journal Of Operational Research* 3(115): 497-506.

Charnes, Abraham, Cooper, William W., and Rhodes, E.L. 1978. Measuring the Efficiency of Decision Making Units, *European Journal Of Operational Research* 6(2): 429-444.

———. 1979. Measuring the Efficiency of Decision Making Units: Short Communication, *European Journal Of Operational Research* 4(3): page 339.

———. 1981. Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through, *Management Science* 6(27) (June): 668-697.

Charnes, Abraham, Cooper, William W., Seiford Lawrence M., and Stutz, J. 1983a. A multiplicative model for efficiency analysis, *Socio-Economic Planning Sciences* (forthcoming 1983). Referenced from Charnes et al. (1983b)

———. 1983b. Invariant Multiplicative Efficiency and Piecewise Cobb-Douglas Envelopments, *Operations Research Letters* 3(2): 101-103.

Charnes, Abraham. 1985. Foundations of Data Envelopment Analysis for Pareto-Koopmans Efficient Empirical Production Functions, *Journal Of Econometrics*, 1-2(30) (October-November): 91-107.

Charnes, Abraham, Cooper, William W., Wei, Quanling, and Huang, Zhimin. 1989. Cone ratio data envelopment analysis and multi-objective programming, *International Journal Of Systems Science* 7(20): 1099-1118.

Charnes, Abraham, Cooper, William W., Huang, Zhimin, and Sun, D.B. 1990. Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks, *Journal Of Econometrics* 1-2(46) (October-November): 73-91.

Chizmar, John F. and Zak, Thomas A. 1983. Modeling Multiple Outputs in Educational Production Functions, *American Economic Review* 2(73) (May): 18-22.

Coelli, Tim. 1996. *A guide to DEAP Version 2.1: A data envelopment analysis (computer) program*, CEPA Working Paper 96/08, Centre for Efficiency and Productivity Analysis, University of New England. Referenced from Brown et al. (1999).

———. 1998. A multi-stage methodology for the solution of orientated DEA models, *Operations Research Letters* 3-5(23): 143-149.

Cook, Wade D., Kress, Moshe, and Seiford, Lawrence M. 1992. Prioritization models for frontier decision making units in DEA, *European Journal Of Operational Research* 2(59): 319-323.

Cooper, William.W., and Pastor J.T. 1995. Global efficiency measurement in DEA, Working Paper, Departamento de Estadística e Investigacíon Operativa. Universidad de Alicante, Alicante. Referenced from Pastor et al. (1999).

Cooper, William W., Park, D.K.S, and Ciurana, P.J.T.P. 2000. Marginal rates and elasticities of substitution with additive models in DEA, *Journal Of Productivity Analysis* 2(13) (March): 105-123.

Debreu, Gerard. 1951. The Coefficient of Resource Utilization, *Econometrica* 3(19) (July): 273-292.

Deprins, D., Simar, L., Tulkens, H. 1984. The Performance of Public Enterprises. In *The Performance of Public Enterprises*, Marchand, M., Pestieu, P., Tulkens, H. eds. North-Holland, Amsterdam.

Doyle, John R., and Green, Rodney H. 1994. Efficiency and cross-efficiency in DEA: Derivations, meanings and uses, *Journal Of the Operational Research Society* 5(45): 567-578.

———. 1995. Cross-Evaluation In DEA - Improving Discrimination Among DMUs, *INFOR Journal* 3(33) (August): 205-222.

Dula, J.H., and Hickman, B.L. 1997. Effects of excluding the column being scored from the DEA envelopment LP technology matrix, *Journal Of the Operational Research Society* 10(48): 1001-1012.

Duncombe, William, and Yinger, John. 1993. *Scale Economies and Technical Efficiency in New York Public Schools*, Metropolitan Studies Program, Syracuse University (August).

Dyson, R.G., and Thanassoulis, Emmanuel. 1988. Reducing Weight Flexibility In Data Envelopment Analysis, *Journal Of the Operational Research Society* 6(39): 563-576.

Färe, Rolf, and Li, Sung Ko. 1998. Inner and outer approximations of technology: A data envelopment analysis approach, *European Journal Of Operational Research* 3(105): 622-625.

Färe, Rolf, Grosskopf, S., Lindgren, B., Roos, P. 1992. Productivity changes in Swedish pharmacies 1980-1989: A non-parametric approach, *Journal of Productivity Analysis* 3:85-101.

Färe, Rolf, and Knox-Lovell C.A. 1978. Measuring the technical efficiency of production, *Journal of Economic Theory* 19:150-162.

Färe, Rolf, and Grosskopf, Shawna. 2002. Two Perspectives On DEA: Unveiling The Link Between CCR And Shephard, *Journal Of Productivity Analysis* 1-2(17) (January): 41-47.

Farrell, M.J. 1957. The Measurement of Productive Efficiency, *Journal Of the Royal Statistical Society* 120(A): 253-290.

Friedman, Lea, and Sinuany-Stern, Zilla. 1997. Scaling units via the canonical correlation analysis in the DEA context, *European Journal Of Operational Research* 3(100) (August 1): 629-637.

Golany, Boaz, and Yu, Gang. 1997. Estimating returns to scale in DEA, *European Journal Of Operational Research* 1(103): 28-37.

Gong, Byeong-Ho, and Sickles, Robin C. 1992. Finite sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data, *Journal Of Econometrics* 1-2(51) (January-February): 259-284.

Green, Rodney H., Doyle, John R., and Cook, Wade D. 1996. Efficiency bounds in data envelopment analysis, *European Journal Of Operational Research* 3(89): 482-490.

Green, Rodney H., Cook, Wade D., and Doyle, John R. 1997. A note on the additive data envelopment analysis model, *Journal Of the Operational Research Society* 4(48) (April): 446-448.

Grosskopf, Shawna, Hayes, Kathy J., Taylor, Lori L., and Weber, William L. 1999. Anticipating the consequences of school reform: a new use of DEA, *Management Science* 4(45) (April): page 608.

Hjalmarsson, Lennart, and Veiderpass, Ann. 1992. Productivity in Swedish Electricity Retail Distribution, *Scandinavian Journal Of Economics. Supplement 1992(94)*: 193-205.

Huang, Zhimin, Li, Susan X., and Rousseau, John J. 1997. Determining rates of change in data envelopment analysis, *Journal Of the Operational Research Society* 6(48): 591-599.

Kirjavainen, Tanja, and Loikkanen, Heikki A. 1998. Efficiency Differences of Finnish Senior Secondary Schools: An Application of DEA and Tobit Analysis, *Economics Of Education Review* 4(17) (October): 377-394.

Koopmans, Tjalling C. 1951. Analysis of Production as an Efficient Combination of Activities. In *Activity Analysis of Production and Allocation*, Koopmans, Tjalling C. ed. John Wiley & Sons, Inc., New York.

Lang, P., Yolalan, O.R., and Kettani, O. 1995. Controlled envelopment by face extension in DEA (data envelopment analysis), *Journal Of the Operational Research Society* 4(46) (April): 473-491.

Lerme, Catherine S. 1992. Efficiency and frontier analysis with extension to strategic planning, Ph.D. Dissertation, University of Massachusetts at Amherst, 255 pages.

Mancebon, Maria-Jesus, and Molinero, Cecilio Mar. 2000. Performance in primary schools, *Journal Of the Operational Research Society* 7(51): 843-854.

Mehrabian, Saeid, Jahanshahloo, Gholam R., Alirezaee, Mohammad R., and Amin, Gholam R. 2000. An assurance interval for the non-Archmidean epsilon in DEA models (Data Envelopment Analysis) (Statistical Data Included), *Operations Research* 2(48) (March-April): 344-347.

Montmarquette, Claude, and Mahseredjian, Sophie 1989. Does School Matter for Educational Achievement? A Two-Way Nested-Error Components Analysis, *Journal of Applied Econometrics* 2(4) (April): 181-193.

Noulas, Athanasios G., and Ketkar, K.W. 1998. Efficient utilization of resources in public schools: a case study of New Jersey, *Applied Economics* 10(30) (October): 1299-1306.

Olesen, Ole B., and Petersen, Niels Christian. 1995. Incorporating Quality into Data Envelopment Analysis: A Stochastic Dominance Approach, *International Journal Of Production Economics* 1-2(39) (April): 117-135.

Pastor, Jesus T., and Knox-Lovell, C.A. 1995. Units invariant and translation invariant data envelopment analysis models, *Operations Research Letters* 3(18): 147-151.

Pastor, Jesus T., Ruiz, Jose L., and Sirvent, Inmaculada. 1999. An enhanced DEA Russell graph efficiency measure (data envelopment analysis), *European Journal Of Operational Research* 3(115) (June 16): 596-607.

Pedraja-Chaparro, Francisco, Salinas-Jimenez, Javier, and Smith, P. 1999. On the quality of the data envelopment analysis model, *Journal Of the Operational Research Society* 6(50): 636-644.

Petersen, Niels Christian. 1990. Data envelopment analysis on a relaxed set of assumptions, *Management Science* 3(36) (March): 305-314.

Pitaktong, U., Brockett, Patrick L., Mote, J. R., and Rousseau, John J. 1998. Identification of Pareto-efficient facets in Data Envelopment Analysis, *European Journal Of Operational Research* 3(109): 559-570.

Post, Thierry. 2001. Estimating non-convex production sets - imposing convex input sets and output sets in data envelopment analysis, *European Journal Of Operational Research* 1(131): 132-142.

Ray, Subhash C. 1991. Resource-use efficiency in public schools: a study of Connecticut data, *Management Science* 12(37) (December): 1620-1628.

Resti, Andrea. 2000. Efficiency measurement for multi-product industries: a comparison of classic and recent techniques based on simulated data, *European Journal Of Operational Research* 3(121): 559-578.

Riddington, Geoff, and Cowie, Jonathan. 1994. Performance assessment using DEA - a cautionary note. (response to E. Thanassoulis, J. of the Operational Research Society, 44, page 1129, 1993), *Journal Of the Operational Research Society* 5(45) (May): 603-605.

Roll, Yaakov, Cook, Wade D., and Golany, Boaz. 1991. Controlling factor weights in data envelopment analysis, *IIE Transactions* 1(23): 2-9.

Ruggiero, John. 1996a. On the measurement of technical efficiency in the public sector, *European Journal Of Operational Research* 3(90): 553-565.

———. 1996b. Efficiency of educational production: An analysis of New York school districts, *Review Of Economics and Statistics* 3(78) (August): 499-509.

———. 1999. Nonparametric analysis of educational costs, *European Journal Of Operational Research* 3(119): 605-612.

———. 2000. Nonparametric estimation of returns to scale in the public sector with an application to the provision of educational services. (Statistical Data Included), *Journal Of the Operational Research Society* 8(51) (August): 906-912.

Ruggiero, John, and Bretschneider, Stuart. 1998. The weighted Russell measure of technical efficiency, *European Journal Of Operational Research* 2(108) (July 16): 438-451.

Ruggiero, John, and Vitaliano, Donald F. 1999. Assessing the Efficiency of Public Schools Using Data Envelopment Analysis and Frontier Regression, *Contemporary Economic Policy* 3(17) (July): 321-331.

Seiford, Lawrence M. 1997. A bibliography for data envelopment analysis (1978-1996), *Annals Of Operations Research* 73:393-438.

Seiford, Lawrence M., and Zhu, Joe. 1998. On alternative optimal solutions in the estimation of returns to scale in DEA, *European Journal Of Operational Research* 1(108): 149-152.

———. 1999a. An investigation of returns to scale in data envelopment analysis, *Omega* 1(27) (February): 1-11.

———. 1999b. Infeasibility of super-efficiency data envelopment analysis models, *INFOR Journal* 2(37): 174-187.

Shephard R.W. 1974. *Indirect production functions*, Verlag Anton Hain, Meisenheim Am Glad.

Smith, P. 1997. Model misspecification in data envelopment analysis, *Annals Of Operations Research* 73:233-252.

Splitek, D.F. 1981. A Study of the Productive Efficiency of Texas Public Elementary Schools, Ph.D. Dissertation, The University of Texas at Austin. Referenced in Zomorrodian (1990).

Sueyoshi, Toshiyuki. 1995. Production analysis in different time periods: an application of data envelopment analysis, *European Journal Of Operational Research* 2(86) (October 19): 216-230.

Tauer, L.W., and Hanchar, J.J. 1995. Nonparametric technical efficiency with K firms, N inputs, and M outputs: a simulation, *Agricultural and Resource Economics Review* 2(24): 185-189.

Thanassoulis, Emmanuel. 1996a. Assessing the effectiveness of schools with pupils of different ability using data envelopment analysis, *Journal Of the Operational Research Society* 1(47): 84-97.

———. 1996b. Altering the bias in differential school effectiveness using data envelopment analysis, *Journal Of the Operational Research Society* 7(47): 882-894

Thanassoulis, Emmanuel, and Dunstan, P. 1994. Guiding schools to improved performance using data envelopment analysis: An illustration with data from a local education authority, *Journal Of the Operational Research Society* 11(45): 1247-1262.

Thanassoulis, Emmanuel, and Allen, R. 1998. Simulating weights restrictions in data envelopment analysis by means of unobserved DMUs, *Management Science* 4(44) (April): 586-594.

Thanassoulis, Emmanuel, and Dyson, R.G. 1992. Estimating preferred target input-output levels using data envelopment analysis, *European Journal Of Operational Research* 1(56): 80-97.

Thompson, Russell G., Langemeier, L., Lee, C.T., Lee, E., and Thrall, Robert M. 1990. The Role of Multiplier Bounds in Efficiency Analysis with an Application to Kansas Farming, *Journal Of Econometrics* 1-2(46) (October-November): 93-108.

Tone, Kaoru. 1996. A simple characterization of returns to scale in DEA, *Journal Of The Operations Research Society Of Japan* 4(39) (December): 604-613.

———. 2001. A slacks-based measure of efficiency in data envelopment analysis, *European Journal Of Operational Research* 3(130): 498-509.

Thrall, Robert M. 1996. Duality, classification and slacks in DEA Extensions and new developments in data envelopment analysis, *Annals Of Operations Research* 66:109-138

———. 1999. What is the Economic Meaning of FDH?, *Journal of Productivity Analysis* 11:243-250.

Tulkens, Henry, and Vanden Eeckaut, Philippe. 1995. Non-parametric efficiency, progress and regress measures for panel data: methodological aspects, *European Journal Of Operational Research* 3(80): 474-499.

Varian, H.R. 1984, The nonparametric approach to production analysis, *Econometrica* 52:579-597.

Wei, Quanling, and Yu, Gang. 1997. Analyzing Properties of K-Cones in the Generalized Data Envelopment Analysis Model, *Journal Of Econometrics* 1(80) (September): 63-84.

Yu, Gang, Wei, Quanling, Brockett, Patrick L., and Zhou, Li. 1996. Construction of all DEA efficient surfaces of the production possibility set under the generalized data envelopment analysis model, *European Journal Of Operational Research* 3(95) (December): 491-510.

Zhang, Yun, and Bartels, Robert. 1998. The Effect of Sample Size on the Mean Efficiency in DEA with an Application to Electricity Distribution in Australia, Sweden and New Zealand, *Journal Of Productivity Analysis* 3(9) (May): 187-204.

Zhu, Joe. 1996. Robustness of the efficient DMUs in data envelopment analysis, *European Journal Of Operational Research* 3(90): 451-460.

Zomorrodian, Mohammad Reza. 1990. Guidelines for improving efficiency in elementary schools in western Massachusetts, Thesis (Ed.D.), University of Massachusetts at Amherst, 173 pages.