

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. Cahill² 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. Turlington⁵. 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):

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

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

F_i^t is the cumulative individual and family characteristics for the i th 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 i th student

S_i^t school inputs relevant to the i th 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 APU⁷ 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.

