THE COST-QUALITY DEBATE IN SCHOOL FINANCE LITIGATION: 
DO DOLLARS MAKE A DIFFERENCE?

JOHN E. MCDERMOTT* AND STEPHEN P. KLEIN†

INTRODUCTION

Since the appearance in 1966 of the Coleman Report, Equality of Educational Opportunity,¹ social science researchers have been attacking one of America's most cherished beliefs: that education may be used as an instrument to reduce social inequality.² Social scientists of national reputation, including the Coleman team members of the faculty seminar at Harvard which reanalyzed the Coleman Report findings,³ and Christopher Jencks⁴ and others, claim that beyond some minimum per pupil expenditure level, statistical relationships between dollars spent per pupil and student performance on standardized tests do not exist. Put another way, higher expenditures do not produce higher student achievement. Home background and other socioeconomic factors account for variations in pupil achievement, not inequalities in school expenditures.

These research findings have not been relegated to some dusty corner of academia.⁵ As the cost of public education has soared and the demand for accountability by state legislatures gathered momentum, the Coleman Report and similar research has been cited by policymakers who question whether increasing school expenditures will improve the quality of education. The issue has also emerged in courtrooms across the country in lawsuits attempting to remedy intrastate and intradistrict disparities in per pupil expenditures. If money does not matter, so the argument goes, then fiscal disparities from district to district or school to school do not result in unequal educational opportunities afforded school children.

The traditional concept that the efficacy of schools is linked substantially to expenditures is, however, not without its defenders. Those defenders concede that American society has been naive in its faith that education can cure

*Attorney, Western Center on Law and Poverty, Los Angeles, California; Lecturer, University of Southern California Law Center. Mr. McDermott was chief trial attorney for the plaintiffs in the Serrano v. Priest school finance litigation in which a trial court decision in favor of plaintiffs was issued on April 10, 1974. Mr. McDermott also is co-counsel in Olsen v. Oregon, a case of the same nature.

†Department of Education and Center for the Study of Evaluation, University of California at Los Angeles. Dr. Klein was an expert witness for the plaintiffs in Serrano.

⁴ C. Jencks, Inequality (1973).
⁵ That an attempt was made to do so is suggested in Moynihan, Sources of Resistance to the Coleman Report, 38 Harv. Educ. Rev. 23 (1968).
every social ill, but urge that a reprieve be granted before discarding the traditional importance attached to schooling. They also point to countervailing social science research which supports the existence of a relationship between cost and quality.

The focus of this article is on the appropriateness of using the judicial process as a forum for resolving the cost-quality debate in educational finance. The relationship between social science and public policy has always been less than amicable, which is perhaps as it should be. But the courtroom is a narrower and more rigidly circumscribed context in which to resolve the tension between science and policy than are the legislative and administrative arenas. Principled adjudication adorns itself with techniques which avoid infringing upon legislative prerogatives. Where, as here, satisfactory definitions of equal educational opportunity are elusive and where the impenetrable jargon of science threatens the proper functioning of the adversary process, courts justifiably should be loath to intervene in matters relating to the financing of public education. The challenge to the school finance reform lawyer, and the purpose of this essay, is to discover some intermediate ground, some judicially manageable standard, which permits the judicial machinery to operate, but avoids settling the scientific dispute. That search here proceeds in three steps: first, existing judicial standards of educational opportunity are examined; second, the viability of pupil performance on standardized tests as a measure of school output is discussed; and finally, litigation strategies for resolving the cost-quality debate are suggested.

I

JUDICIAL STANDARDS OF EQUALIZATION OF EDUCATIONAL OPPORTUNITY

Defining equality of educational opportunity is a hazardous undertaking where there exists no social consensus as to the proper role of schools.\(^6\)

Judicial response under such circumstances is predictably poor—courts in school finance litigation, in the absence of Coleman Report-type evidence on school outputs, often assume a cost-quality relationship;\(^7\) other courts accept such evidence at face value in rejecting plaintiffs' claims.\(^8\) Seldom does the adversary process work effectively. Nevertheless, past judicial attempts to construct standards of equality of educational opportunity offer an introduction to the available options. Eight different standards have been considered: (1) equal dollars per pupil; (2) dollars adjusted according to pupil needs; (3) lack of judicially manageable standards; (4) maximum variable ratio; (5) negative standard; (6) inputs; (7) outputs; and (8) minimum adequacy.\(^9\)

---

\(^6\) Educators have not satisfactorily defined the concept of equal educational opportunity, perhaps because the concept is not static, but evolutionary. Professor Coleman has noted the change in expectations of schools: increasingly, our society has come to rely on schools to solve complex social problems such as poverty, crime, and racial discrimination. Coleman, *The Concept of Equality of Educational Opportunity*, 38 HARV. EDUC. REV. 7 (1968).

\(^7\) "Presumably students receiving a $1000 education are better educated than those acquiring a $600 schooling." McInnis v. Shapiro, 293 F. Supp. 327, 331 (N.D. Ill. 1968), aff'd mem. sub nom. McInnis v. Ogilvie, 394 U.S. 322 (1969).


\(^9\) In 1968, Arthur Wise suggested nine definitions of equality of educational opportunity:
A. McInnis v. Shapiro

1. Equal Dollars Versus Educational Needs

The first three standards are amply illustrated in McInnis v. Shapiro, where plaintiffs attacked the interdistrict expenditure inequalities of the Illinois school financing system. Plaintiffs claimed that the fourteenth amendment commanded that expenditures be distributed only on the basis of the educational needs of pupils. The court first rejected any rigid dollar equality standard of educational opportunity, noting that expenditures are not "the exclusive yardstick of a child's educational needs." Perhaps the only point of agreement on the part of educators, courts, and laymen is that a "one-dollar one-scholar" measure of equality is inappropriate. Absolute dollar equality would ignore variations in costs and educational needs from district to district. Fear was also expressed in McInnis that leveling high-expenditure districts downward would produce uniform mediocrity and stifle local experimentation. No court has ever adopted such a standard; even courts which find a relationship between cost and quality, and regard dollar expenditures as one relevant criterion for measuring equal educational opportunity, disavow that an equal-dollar-per-pupil solution is in any way legally required. Moreover, because the reasons for appropriate expenditure disparities are several, a "needs only" standard of expenditure distribution was also unacceptable in McInnis. A standard was needed which was flexible enough to accommodate competing fiscal claims, a standard in other words which was consistent with some acceptable concept of proportional equality (the principle of justified inequality).

2. Lack of Judicially Manageable Standards

Yet the articulation of such a standard was thought in McInnis to be "a basic policy decision more appropriately handled by a legislature than a court." Complaining that educational need was a "nebulous concept" beyond judicial competence to define, the court refused to decide the case on its merits. A similar result was reached in Burrus v. Wilkerson, where the court, in upholding Virginia's system of school financing declared, "courts have neither the knowledge nor the means nor the power to allocate the public monies to

(1) negative, (2) full-opportunity, (3) foundation, (4) minimum-attainment, (5) leveling, (6) competition, (7) equal-dollars-per-pupil, (8) maximum variance ratio, and (9) classification. A. Wise, Rich Schools, Poor Schools 143-59 (1968). Wise's standards have been criticized as being too vague for judicial use, Kirp, Book Review, 78 Yale L.J. 908, 915 (1969), yet courts again and again have utilized these standards, or variations thereof.

11 293 F. Supp. at 335.
12 Id. at 331 n.11.
13 In holding the state's system of financing schools unconstitutional, a lower court in Idaho said that the state is not "obligated to insure that all districts have the same dollar input per pupil. The state may recognize differences in educational costs so long as the differences are based on relevant economic and educational factors." Thompson v. Engelking, Civil No. 47,055 (Idaho Dist. Ct., Nov. 16, 1973).
14 293 F. Supp. at 332.
15 Id. at 335.
fit the varying needs of these students throughout the state." Thus, to be successful, school finance litigants must construct a standard for measuring equality of educational opportunity which permits appropriate expenditure variations but which is also judicially manageable.

B. Maximum Variable Ratio

One solution is a maximum variable ratio standard for measuring equality. Such a standard, while essentially arbitrary, would permit expenditure variations within a specified range, thereby providing school districts with some financial flexibility in responding to local problems. One court suggested, but did not order, that variations in operating expenditures due to local initiative might be permissible "to the extent of 10 percent or 15 percent of the level of income guaranteed for the district by the state in any year." The variable ratio approach can be combined with existing categorical aid programs so as to apply only to ordinary operating expenditures. Hence, in Hobson v. Hansen, a case involving between-school disparities in per pupil expenditures within a single school district, Judge J. Skelly Wright, ordered that disparities in per pupil expenditures for teachers' salaries not exceed a five per cent variation except for "adequate justification," which was defined to include compensatory and special education programs. Yet a maximum variable ratio approach is fraught with problems. The standard assumes the adequacy of the state-guaranteed minimum above which expenditures may vary; if the assumption is false and the ratio is set too low, then the policy demand would become intense. The standard is probably unworkable unless special categorical programs and capital outlays are excluded. There is also the problem of deciding upon an appropriate ratio. If the ratio is set too high, serious inequalities will result; if set too low, the system will be unresponsive. Because of the lack of standards with which to make these judgments, a judicial solution of this nature, involving as it must the imposition of a particular method of school finance upon the legislature, raises serious separation of powers issues.

C. Negative Standards

A possible way of dealing with the issue of equality of educational opportunity in school finance litigation is to define what equality of educational opportunity is not, limiting one's examination to what a state may not do in the distribution of educational revenues, rather than ordering what a state must do. The "fiscal neutrality" principle of Professors Coons, Clune and Sugarman is the outstanding example of the negative standard approach. Fiscal neutrality means that expenditure disparities resulting from differences in local district taxable wealth are unlawful, or in the terminology of its proponents, "[t]he quality of public education may not be a function of wealth other than

---

17 Id. at 574.
the wealth of the state as a whole." A negative fiscal neutrality standard satisfies the requirement of flexibility because it does not require exact dollar equality. Although the standard implicitly assumes a cost-quality relationship with respect to the dollar variance among districts caused by wealth differences, expenditures may fluctuate among districts for any reason other than district wealth. The question of what differences other than wealth are legitimate is left unanswered, a real advantage to a court which desires to avoid deciding that larger question. Fiscal neutrality, in other words, satisfies the requirement of a judicially manageable standard, for a court need not itself define what equal educational opportunity is in all of its dimensions. Judicial concern is concentrated solely upon wealth-produced expenditure disparities.

Yet the fiscal neutrality standard, although achieving some success in state and lower federal courts, was rejected by the United States Supreme Court in *San Antonio Independent School District v. Rodriguez* and has been the subject of sharp criticism from commentators, partly for not going far enough, partly for going too far, and partly because the solution tendered by the Coons team—"district power equalizing"—was unacceptable. A common criticism is that under a power equalized school finance system, expenditures will still vary, not according to wealth, but according to tax rate. Whereas Professor Coons would justify what he admits would be qualitative differences in a power equalized system on the basis of local control, others have found that claim inconsistent with the asserted fundamentality of education, for it would allow a child's education to be poorer where the community was feckless. Perhaps a more satisfactory approach would be to expand the negative standard to include other inequities. In any event, the negative definition of equality of educational opportunity, so long as there is consensus on what does not justify expenditure disparities, should be regarded as a viable judicial approach to measuring equality.

D. Inputs

A sixth measure of equal educational opportunity is "inputs"—equality in the level of educational resources (which may vary in price between districts) brought to bear on a child's education. Justice Marshall, dissenting in *Rodriguez*, suggested such a standard: "the question of discrimination in educational quality must be deemed to be an objective one that looks to what the State provides its children, not to what the children are able to do with what they

---

21 *Id.* at 2.
receive." Marshall, noting the disparities in teacher training, teacher-student ratio, and other resource inputs, argued that "[d]iscrimination in the opportunity to learn that is afforded a child must be our standard."

Marshall's focus upon objective educational inputs is firmly rooted in separate-but-equal cases such as Sweatt v. Painter where inputs, both tangible and intangible, were the standard of equality in comparing the University of Texas School of Law with the separate black law school:

[W]e cannot find substantial equality in the educational opportunities offered white and Negro law students by the State. In terms of number of the faculty, variety of courses and opportunity for specialization, size of the student body, scope of the library, availability of law review and similar activities, the University of Texas Law School is superior. What is more important, the University of Texas Law School possesses to a far greater degree those qualities which are incapable of objective measurement but which make for greatness in a law school. Such qualities, to name a few, include reputation of the faculty, experience of the administration, position and influence of the alumni, standing in the community, traditions and prestige. It is difficult to believe that one who had a free choice between these schools would consider the question close.

More recently, Judge Wright in Hobson v. Hansen declared that "if whites and blacks, or rich and poor, are to be consigned to separate schools pursuant to whatever policy, the minimum that the Constitution will require and guarantee is that for their objectively measurable aspects these schools be run on the basis of real equality." Although an objective approach is certainly desirable, an inputs standard, without more, not only fails to respond to countervailing scientific research as to the educational consequences of input disparities, but is subject to the same objections as an equal dollar standard. Equality of inputs is either a rigid standard which does not permit variations to meet special needs or costs, or, if variations are to be permitted, leaves the judiciary without adequate standards to determine when equality has been achieved.

E. Outputs

A seventh definition of equality of educational opportunity is "outputs"; the effects of different educational investments are gauged in terms of pupil performance on standardized achievement tests. On the basis of social science research which finds an insignificant relationship between variation in expenditures and variation in test scores, the argument is made that disparities in expenditures do not result in unequal educational opportunities. Courts which have been confronted with social science evidence of this sort have demonstrated an inability to deal with it adequately, largely because such evidence often goes unrebutted. Seldom has the adversary process worked effectively so as to permit serious and thoughtful judicial examination of the cost-quality issue. The evidence presented is often only the tip of the social science iceberg,

---

27 411 U.S. at 84.
28 Id.
and often not analyzed carefully or expertly by lawyers on either side or by judges. Judicial opinions treat the social science evidence curtly and shallowly, and demonstrate a crucial lack of appreciation for the myriad issues inherent therein. The alien and imposing terminology of science has often found lawyers woefully unsuited to perform their adversary role and led judges to take the seeming certainty of numbers in blind faith. The intersection of law and social science is one of tension, nowhere more so than in educational inequality litigation. Before an outcome definition of equality is deemed acceptable, a more careful examination is due that intersection.

F. Minimum Adequacy

A final standard of equal educational opportunity is the one employed by the majority in *Rodriguez*—minimum adequacy. In most states, a foundation program financing scheme is utilized, based on the premise that the state should guarantee every pupil with a “foundation program for adequate financing of all educational services.” Above that level, districts are free to spend as they wish, but the revenue must be raised locally. A minimum adequacy definition predicates the foundation program as the limit of the state's responsibility for providing equality in its educational offering. So long as the state provides every child with a minimally adequate education, differences in spending beyond that specified minimum are not legally significant. The adequacy standard is manageable, it does not forbid expenditure differences, and it allows courts neatly to avoid the necessity for resolving the cost-quality debate. No research challenges the importance of a basic minimum level of expenditure; the controversy relates to the educational consequences of spending inequalities above that basic level. Justice Powell, influenced by the existence of this controversy, and the lack of judicial tools for resolving it, found an adequacy standard to be the solution to a vexing problem. Yet, the minimal adequacy perception of equality is not without its difficulties. First, the adequacy standard does not enable courts to avoid the qualitative issue. As Justice Marshall's dissent correctly notes, the majority opinion does not specify applicable standards for determining what level of expendi-

---

32 411 U.S. at 25, 37, 45.
34 [I]t must be stressed again that we are not assessing the absolute effect of schools on achievement, but rather the effect of schools on variation in achievement levels. We have no doubt that schools have an overall baseline effect. This baseline effect is probably fairly uniform for all schools, because school factors may be relatively equalized with respect to minimum standards. But, given variation in verbal achievement, especially that between black and white students, we are searching for factors which might reduce the difference. School factors may be important for a certain basic level of achievement, but this does not necessarily mean that improving those factors will help bring black achievement closer to white achievement. If one wants to reduce this differential, Coleman's and our analyses point to family-background factors as the more promising area for improvement.


35 Justice Powell declared that the judiciary would be “well advised to refrain from interposing on the states inflexible constitutional restraints that could circumscribe or handicap the continued research and experimentation so vital to finding even partial solutions to educational problems and to keeping abreast of ever-changing conditions.” 411 U.S. at 43.
ture is constitutionally adequate. Further, what may be adequate for one type of pupil may not be for another. Disadvantaged or handicapped students may require expenditures in addition to the foundation program guaranteed minimum expenditure level, in order to receive an adequate education. Determining what those varying levels of adequacy are will require articulation of manageable standards. To rely upon an equal dollar definition of adequacy (and the foundation program level of support is generally an equal dollar result) is no more defensible than to rely upon an equal dollar definition of quality. Both standards are rigid and fail to permit justifiable variations. Marshall's criticism of Powell's double standard treatment is telling: "One would think that the majority would heed its own fervent affirmation of judicial self-restraint before undertaking the complex task of determining at large what level of education is constitutionally sufficient."

There is an even more fundamental difficulty with the adequacy definition. The Texas system of financing public education was upheld because that system served a rational state purpose, that is, local fiscal control. Yet there is an inconsistency in using an adequacy standard to avoid resolving the social science debate while at the same time employing a justification for permitting relative spending disparities that itself presumes the existence of a cost-quality relationship. Justice Powell described the importance of local control in the following terms:

In part, local control means... the freedom to devote more money to the education of one's children. Equally important, however, is the opportunity it offers for participation in the decision-making process that determines how those local tax dollars will be spent. Each locality is free to tailor local programs to local needs. Pluralism also affords some opportunity for experimentation, innovation, and a healthy competition for educational excellence.

Yet local control over financing and, thus, over program planning, occurs only above the state-guaranteed foundation program minimum expenditure level. If local control means all that Justice Powell says it does, then a relationship between cost and quality must exist. Local control implies control over something that has meaningful content. If there is no relationship between cost and quality, then local control has no meaning and expenditure inequalities which use local control for their justification can have no rational basis. For the system to be "rational" (at least if local control is the rational basis offered), courts must go beyond an adequacy definition of equality. Justice Powell, in short, has employed a double standard, refusing to decide the cost-

---

36 Id. at 89.
37 Id.
38 Id. at 49-50.
39 While the correlation between expenditure per pupil and the quality of education may be open to argument, the Court must assume here that it is high. To do otherwise would be to hold that in those wealthy districts where the per pupil expenditure is higher than some real or imaginary norm, the school boards are merely wasting the taxpayers' money. The Court is not willing to so hold, absent some strong evidence. Even those who staunchly advocate that the disparities here complained of are the result of local control and that such control and taxation with the resulting inequality should be maintained would not be willing to concede that such local autonomy results in waste or inefficiency. Van Dusart v. Hatfield, 394 F. Supp. 870, 874 (D. Minn. 1971).
quality issue and then presuming the existence of the challenged relationship in order to justify the system's inequalities.

Because Rodriguez is law for the federal courts, there remains the question of what measure of inequality resulting from a state's school finance system will successfully support a federal claim for relief. One can speculate that in a state where the average per pupil expenditure substantially exceeds the foundation program minimum, where evidence supports the qualitative inadequacy of the minimum expenditure level, where there are demonstrated inadequacies in the educational programs of low wealth districts, and where there exists little local option—factors that can currently be demonstrated in many states—a federal suit might succeed and still be consistent with Rodriguez. State forums, however, are more likely to offer enhanced opportunities for successful school finance litigation. Whatever the forum, a more satisfactory approach to the problem of defining and measuring inequality of educational opportunity needs development.

II

SOCIAL SCIENCE RESEARCH ON THE COST-QUALITY RELATIONSHIP

The judicial standards for equality of educational opportunity provide the conceptual context in which courts utilize social science research findings such as those in the Coleman Report. In this section, we provide a critique of that

---

40 Focusing upon a specific group of districts in which the programs and opportunities afforded are inadequate and in which access to the fiscal resources to remedy those inadequacies is effectively foreclosed will bring an important particularity to interdistrict school finance litigation lacking in earlier litigation based on the fiscal neutrality principle. The educational process will become paramount and fiscal issues secondary. The group of districts and children will be smaller in number than that conceived in previous litigation, but more compelling.

41 The challenge to the effectiveness of different expenditure levels in public education is a relatively recent development, having its genesis in the 1966 Coleman Report. The Equality of Educational Opportunity Survey (EEOS), undertaken in 1965 in accordance with a mandate from Congress to the U.S. Office of Education to conduct a survey “concerning the lack of availability of equal educational opportunities for individuals by reason of race, color, religion, or national origin in public educational institutions,” Civil Rights Act of 1964, § 402, 42 U.S.C. § 2000c-1 (1970), collected an extensive array of data from a sample of 4,000 schools across the country. Several aspects of school inequality were investigated: (1) inequality in tangible school inputs (expenditures, facilities, and so on); (2) degree of racial segregation; (3) inequality in intangible school inputs (for example, teacher expectations); (4) inequality in school outputs as measured by pupil performance on standardized achievement tests relative to non-school factors; and (5) inequality of school output as measured by pupil performance on standardized achievement tests without regard to non-school factors. See Coleman, supra note 6, at 16-17. The Report was published in 1966 and its most controversial finding related to the relatively insignificant effect that differential school resources had upon student outcomes independent of other factors such as home background.

One implication stands out above all: That schools bring little influence to bear on a child's achievement that is independent of his background and general social context; and that this very lack of an independent effect means that the inequalities imposed on children by their home, neighborhood, and peer environment are carried along to become the inequalities with which they confront adult life at the end of school. For equality of educational opportunity through the schools must imply a strong effect of schools that is independent of the child’s immediate social environment, and that strong independent effect is not present in American schools.

Coleman Report 325.
research—which relies largely on pupil performance on standardized achievement tests—and the utility of such research in deciding cases. The critique concerns four major reasons why input-output statistical analyses using pupil performance on standardized achievement tests provide an inappropriate measure of educational output: (1) defects in the tests themselves; (2) the failure of correlational analyses to match inputs with outputs; (3) inadequacies in data analysis techniques; and (4) other methodological defects inherent in correlational analysis.

A. Pupil Performance on Standardized Achievement Tests as an Index of Educational Output\textsuperscript{42}

An assumption which is implicit in any method of analyzing the relationship between expenditures and achievement is that pupil performance on standardized achievement tests provides an index of educational "output" which is useful for comparisons of educational opportunity among school districts. While there are other educational output measures such as school drop-out rates, later life income, job status, and delinquency rates, most social science "output" research relies upon student test scores. Yet there are a range of problems in the use of such scores to evaluate program and expenditure effectiveness.

One is that the educational goals measured by these achievement tests, typically in reading and mathematics, are very limited. The cognitive academic domain is the only area tested. No tests are administered in the "psychomotor" domain, which relates to physical skills used in physical education and vocational education. No tests are administered in the "affective" domain, which includes such educational objectives as citizenship, self-concept, political socialization, attitudes toward school, maturity, and interpersonal skills. Even within the cognitive academic domain, many subject areas and goals are not tested.

A second and more important problem with using standardized achieve-
ment tests to evaluate program effectiveness among districts is that achievement tests commonly do not overlap well with the program objectives of schools even within the cognitive domain purportedly being tested. An achievement test in a particular subject, for example, may cover aspects not taught in a particular school, while on the other hand, a school may cover aspects pertinent to the subject matter area that are not covered by the test. Hence, the degree of overlap between course content and test content will vary from school to school and district to district. As a result, the test instrument is not especially sensitive to the effects of instruction.

A third problem in using test scores to measure educational quality among school districts is attributable to poor test design. Confusing directions, instructions which strain a child's attention span, lack of uniformity in the administration of tests, and ambiguous format all undermine the validity of test results. The "standardization" of tests is similarly impaired by teaching to the test, promotion of students to grade levels beyond their ability, and differences in student test-taking abilities.

The fourth and most serious problem with test results is that of validity, that is, whether tests measure what is purported to be measured. A reading test, for example, may be measuring skills other than reading, such as reasoning or general intellectual ability. Test questions measuring reasoning ability may be heavily influenced by non-school factors such as home background and innate ability, and should not be included in a reading test (as distinguished from an intelligence test) if the "reading test" purports to compare the quality of instruction in reading among school districts. Test publishers often, consciously or unconsciously, select test questions which involve reasoning ability in order to spread students out so as to have variations in test scores. Items which are sensitive to instruction, but which do not produce test score variations, are excluded. In sum, most standardized achievement tests were not constructed for the purpose of evaluating program effectiveness; rather, they were designed to measure general intellectual ability for purposes of making individual counseling, selection, and classification decisions. Achievement

43 The problem of poor overlap is magnified by the fact that the degree of importance attached to certain content objectives in standardized achievement tests may vary considerably from the degree of emphasis placed on the same items within any school. For example, the statewide grade one mathematics test used in California covers eight content objectives, yet forty-four out of fifty-five items on the test are devoted to four of the areas, while the other four areas are measured by only eleven items. As a result, the test does not measure all eight objectives equally well. Even more disconcerting is the variation in the average difficulty of the items among different content objectives. On the grade one mathematics test referred to above, two content areas containing twenty-three items are comparatively easy, while two other areas totalling twenty-one items are comparatively difficult. Only a total score for all eight content objectives is reported as the pupil's performance on that test. The test, then, does not really measure how well a student does at mathematics; it measures how well he or she can answer easy items in two of the eight content areas and difficult items in two of the other areas. The State Department of Education in California is now moving away from statewide standardized achievement test scores to a sampling-based testing procedure to measure attainment of broad educational objectives on a school level rather than a district level, leaving each district with the responsibility and duty of carrying out its own evaluation of its own unique objectives. See CALIFORNIA ASSEMBLY EDUCATION COMMITTEE, REPORT OF ADVISORY COMMITTEE ON STATEWIDE TESTING PROGRAM 20 (March, 1972).

44 The problem can be illustrated by the following sample items from a commonly used reading test:
tests, to the extent that they are really intelligence tests measuring non-school factors over which the school has no control, are not valid instruments for the purpose of assessing the outcomes of instruction, because the characteristics which such tests measure are not necessarily those that are the outcomes of instruction.45

The usefulness of test scores for comparative purposes is also impaired by cultural bias.46 On a very simple level, tests are culturally biased in the sense

READ THIS:

A GAME

You say “box” You say “bean”
I say “animal” I say “girl”
You say “fox” You say “Jean”

<table>
<thead>
<tr>
<th>You say “men”</th>
<th>“four”</th>
<th>“hen”</th>
<th>“ten”</th>
</tr>
</thead>
<tbody>
<tr>
<td>I say “number”</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>You say . . .</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>You say “head”</td>
<td>“bread”</td>
<td>“cake”</td>
<td>“red”</td>
</tr>
<tr>
<td>I say “food”</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>You say . . .</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cooperative Primary Tests, Reading, Form 12A (Educational Testing Service, 1965). These items do not measure how well a child can read and understand the meaning of words, but how well he reasons. Tests, in other words, measure many non-school factors such as innate ability and home background, and these non-school factors may have more influence than school factors on test performance.

45 Because of the validity problem, it is difficult, if not impossible to interpret what a test score means. Two children could both receive a score of 27 on a reading test, but one child could answer 20 reading items correctly and 7 reasoning-type items correctly, while the other child answers only 12 reading items correctly but correctly answers 15 reasoning items. Because the children are behaving quite differently, the score of 27 means different things for those two children and it is, therefore, very difficult to evaluate the effects of instruction on these children by referring to the total test score. This problem is magnified even further when the test scores of all children in a school district are averaged together to compute an average score for the entire district. When the district average test score is correlated with some other variable, such as average per pupil expenditures, any resulting correlation is very suspect because of the difficulty of determining whether the result reflects the effects of instruction or of reasoning ability or of some combination of both. Since standardized achievement tests are not very sensitive to instruction, and measure general intellectual ability, which is affected by non-school factors such as family background, it should not be surprising that family background should correlate higher with achievement scores than school expenditures.

that items on the tests may refer to things such as sleds or mittens which are not present in the cultural experience and background of many children taking the test. More significantly, however, standardized achievement tests are norm-referenced in the sense that the test scores of children or school districts are compared with the scores of a national or state sample of the normative or typical student population. As a result, standardized tests generally reflect the cultural characteristics of white middle class suburban students, partly because large urban school districts with heavy concentrations of low income and minority students often decline to be included in the reference sample. Thus, tests may not be as appropriate for certain minority or ethnic groups as they are for the main body of white students. In addition, the content of the tests reflects the cultural biases of the test designers, who are also largely white middle class people. The most egregious type of cultural bias in tests occurs when the scores of children with English language disabilities are compared with students whose primary language is English. When that occurs, the test is no longer standardized and is a qualitatively and quantitatively different test for different groups of children, based upon their proficiency in English.

In sum, standardized achievement tests which are the basis for social science input-output studies are misused when devoted to the purpose of making comparative evaluations of school district programs. The tests are often not even designed for program evaluation or measuring the effectiveness of expenditures. It is simply inappropriate to use just a few of the currently available standardized tests to measure all of the important outputs of education.

B. Inadequacy of Expenditure Index: Failure to Match Specific Inputs and Specific Outputs

A second major difficulty with input-output research is the inadequacy of the expenditure index used in correlating expenditures with achievement. Obviously, one does not use a reading test to evaluate the effectiveness of expenditures relating to the mathematics curriculum. Yet, typically, the attempt is made to correlate the average expenditure per child for instruction in each school district with the average test score in that district on a particular achievement test. A sixth grade reading test score, however, is not an appropriate measure for all of the programs included in the average instructional expenditure per child since this figure reflects the per child expenditure on salaries for principals and vice-principals, supervisors, consultants, classroom teachers including substitutes, counselors, librarians, psychologists, instructional aides, secretaries, and clerks, as well as direct expenditures for reading instruction. A reading test can at best only measure the results obtained from a very small portion of the resources devoted to a district's educational program. To correlate performance on reading tests with expenditures for a whole host of programs, goals, services, and materials which have nothing to do with reading is to use the tests for purposes for which they were never designed. The expenditure indices used by researchers, then, are often inadequate because they

---

47 Hanushek & Kain, supra note 41, at 120.
fail to adequately match specific inputs with the specific outputs being measured.

A related problem in correlating average district instructional expenditures with average district test scores is that such an analysis assumes that the same expenditure per pupil in two districts will have the same effect upon test scores, no matter how differently funds are spent in the two districts, an assumption which is hardly justifiable. Many states provide additional funds to small districts to offset diseconomies of scale. Thus, a per pupil expenditure of $600 in each of two districts does not necessarily mean that the same amount of education was purchased.\(^4\)

Another difficulty is that the instructional expenditure per child is the average amount spent on pupils within a district. Yet, the amount spent per child varies widely from school to school within a district and from classroom to classroom within a school.\(^4\) Comparisons among school districts of per pupil expenditures fail to take into account these intradistrict and intraschool variations. When a district's average expenditure per pupil is correlated with its average achievement test scores, actual input-output relationships may thus be masked. Such data make it impossible to match specific school inputs to the individual student, and as a consequence, there can be little precision in any conclusions drawn from research about the impact of school inputs on student achievement.\(^5\)

C. Inadequacies of the Data Analysis Techniques

Social science research examines the statistical relationship between student test scores on standardized achievement tests and school expenditures. These two variables, expenditures and achievement, are compared with each other in order to determine the frequency with which higher expenditures occur together with higher achievement scores. The strength of that statistical association is expressed in a correlation coefficient, which may range from -1.0, a perfect negative or inverse correlation, to +1.0, a perfect positive correlation; a zero correlation implies no relationship at all between the two variables. Any correlation coefficient expressing the strength of association between expenditures and achievement represents a statistical relationship between the two variables, and not necessarily a causal relationship. A positive correlation between expenditures and achievement, for example, means only

\(^{48}\) The above discussion is a summary of the testimony of Dr. Klein, supra note 42.

\(^{49}\) In Los Angeles Unified School District, expenditures per pupil range from a low of $448.93 in Liberty Boulevard Elementary School to a high of $1,417.08 in 32nd Street Elementary School. Los Angeles Unified School District, Report, Expenditures Classified by School for the Fiscal Year Ended June 30, 1972 (March 29, 1973). Even within a school, more money is spent on some children than others because of different tracking schemes, special education programs, compensatory education, and so on.

\(^{50}\) This is one of the major defects of the Coleman Report, a measurement error which, according to Dr. James Guthrie, fails to "pick up the real world with sufficient precision to enable us to know much about the real world ... because such averaging of data disguises more reality than it discloses;" and is a research problem "so serious ... as simply to erode the validity of ... [any] findings." Testimony of Dr. James W. Guthrie, Serrano v. Priest, Civil No. 998,254 (Cal. Super. Ct., Apr. 10, 1974), Reporter's Transcript, vol. 38, at 1481-82. See also Bowles & Levin, The Determinants of Scholastic Achievement—An Appraisal of Some Recent Evidence, supra note 41, at 8-9.
that, as expenditures vary upward or downward, so also do achievement scores; that higher expenditures in fact cause higher achievement scores is not implied and does not necessarily follow from a positive correlation.

The correlation coefficient, which has been used as the criterion in the cost-quality issue, supposedly indicates the degree of relationship between expenditures and student performance. If expenditures and student performance were perfectly correlated with one another in the "real world," then three conditions would have to prevail in order for researchers to obtain a positive 1.00 correlation coefficient (any deviation from these assumptions would lower the correlation coefficient): (1) a perfect linear relationship between the indices of expenditures and achievement (for example, for every $0.03 spent on education, there is a corresponding increase in achievement of exactly 2 points); (2) reliable and valid indices of expenditures and student achievement, with no chance or systematic errors in the recording of these data prior to or during the correlation analysis; and (3) the absence of any influence by other intervening variables such as the sagacity of the decisions influencing how money is spent. If any one or a combination of these assumptions is not fully met, then the correlation coefficient between the indices of expenditures and achievement will necessarily be less than 1.00. Yet, because of the nonlinearity of the test scores in measuring student performance, it is unlikely that there will be a corresponding increase in test scores for each additional dollar spent. In addition, as demonstrated above, the indices of both expenditures and achievement are inadequate. It is also likely that a number of intervening variables will affect the relationship between expenditures and student performance. For example, it is reasonable to assume that educational decision-making varies in quality from school to school simply because all school administrators are not equally intelligent or experienced. One can imagine a situation in which there were high expenditures per pupil but low student performance as a result of relatively poor (or unlucky) decisions affecting how the money was spent. On the other hand, having a wise decision-maker with inadequate funds could also mean low student performance. Student performance will be high, therefore, only when both the decisions affecting how the money is spent are good (or lucky) and adequate financial resources are available to implement those decisions. Thus a high expenditure level may be a necessary, but not sufficient, condition for high levels of student performance.

Because the quality of decision-making influences student performance and because the degree of influence is related to expenditures in a systematic non-linear fashion, a "triangular" rather than a symmetrical distribution of school means is produced when the relationship of expenditures is plotted against student test scores. This type of distribution is illustrated in Figure I, in which it is assumed that expenditures and performance are measured with complete accuracy. If the figure were an accurate reflection of the relationship between test scores and expenditures, then the correlation coefficient associated with this relationship would be no higher than 0.50.
The addition of each intervening variable (a variable that has a linear or nonlinear effect on one or both of the two primary variables being correlated) will have the effect of decreasing the correlation coefficient between expenditures and student performance. Since there are a host of factors that might have an impact on the relationship between expenditures and performance, there is small likelihood of finding a statistically significant correlation between these expenditures and student test scores. Thus, even though there may be a very strong underlying relationship between expenditures and student performance, the presence of numerous chance and systematic (intervening) variables affecting this relationship essentially precludes obtaining a significant positive expenditure-performance correlation coefficient.

D. Methodological Defects

The accuracy of input-output research studies is further impaired by other substantial methodological defects. One such defect is the "cross-sectional" nature of the research; that is, the relationship between inputs and outputs is examined at one point in time, comparing, for example, average expenditures per
child with average eighth grade reading achievement for a given year. Such cross-sectional research by definition ignores the cumulative effect of school resources upon achievement, thereby assuming that the amount of resources received in each of the seven years prior to the year of testing is the same as that received in the year of testing. To state the point differently, cross-sectional research assumes that any differences in inputs received in prior years did not have any impact on the output measured in grade eight. A longitudinal analysis, by contrast, would compare the change in achievement between grades one and eight with the actual school resources that the student has received between grades one and eight. This type of analysis enables one to measure growth in achievement over time, which is not measurable with a cross-sectional analysis. A model of the educational process which ignores the school resources received by a student prior to the year of testing not only is defective in its description of the way the educational process operates, but also masks the cumulative effect of school resources on achievement.\footnote{Testimony of Dr. James W. Guthrie, \textit{id.} at 4167-68; Testimony of Dr. Nathan Gage, \textit{id.}, vols. 46-47, at 5000-184.} The few longitudinal analyses so far conducted have found significant relationships between school inputs and student achievement.\footnote{A study of third grade students in a California school district, see Hanushek, \textit{supra} note 41, at 88, which used achievement and school input data for the first three grades, classified students ethnically and by socioeconomic status, based on the occupation of the father. For white students, teachers' verbal ability and recentness of teachers' educational experience were significantly related to achievement. Another study which looked at eighth grade students in a small school district found the quality of teacher undergraduate institution, teacher experience, and teacher salary to be important. \textit{See} testimony of Dr. Donald Winkler, Serrano v. Priest, Civil No. 938,254 (Cal. Super. Ct., Apr. 10, 1974), Reporter's Transcript, vol. 40, at 4360-493.}

Another defect of input-output research is that it is typically conducted through the use of survey questionnaires which measure "status" variables rather than "process" variables. Rather than going into the classroom and examining how teachers actually behave, the researcher reports on certain characteristics of teachers such as the number of years of teaching experience, teachers' education, or teachers' verbal abilities, which are believed to correspond with how teachers behave. Teacher verbal ability is often used in determining the effect of teachers on pupil achievement. Yet verbal ability based on a multiple choice vocabulary test is primarily a measure of IQ or intellectual ability. In controlled experimental studies, teacher intelligence has proved to be poorly related to teacher effectiveness; thus such a status variable may be inappropriate to evaluate the effect of teachers on student achievement. Teacher behavior in the classroom is the more appropriate variable to be considered.

A third significant methodological problem is that of "multicollinearity." A simple correlation between two variables such as expenditures and achievement does not eliminate the effects which other unconsidered variables may be having on the results; in order to determine the independent effect which schools have on achievement, it is necessary to disentangle the effect which other variables such as home background, innate ability, peer group composition, and motivation have upon achievement. The statistical procedure often
used to attempt this disentanglement is a “regression analysis,” a technique which seeks to measure the independent effect of each variable upon achievement. One type of regression analysis employed by the Coleman research team, the “multiple stepwise regression analysis,” is not an appropriate technique for determining the unique effect of schools upon achievement because home and school environment are highly intercorrelated with each other, or “collinear,” and thus tend to measure the same things.\(^5\) In a stepwise regression analysis, variables are entered into the analysis one at a time; because of multicollinearity, however, the effect of selected variables upon pupil achievement is very sensitive to, and depends entirely upon, the order in which the variables are entered. For example, home environment variables, if entered first, may explain thirty-six per cent of the variance in student achievement, while school variables, if entered second, may explain only one and a half per cent, leading to the conclusion that school inputs are not very important in explaining differences in student achievement. If the order of entry is reversed, however, school variables may explain thirty-six per cent of the variation in student achievement. Because home and school environment are highly collinear, much of the influence which the variable entered second has on achievement will be “captured” by the variable entered first. The presence of collinearity thus acts as a bar to a determination of school factors upon achievement.

E. Conclusions

In sum, considering the substantial problems arising from the use of standardized achievement test scores as measures of educational opportunity among school districts, and the substantial methodological difficulties in attempting to correlate school inputs with school outcomes, it seems unwise to use an output measure of equality of educational opportunity among school districts.\(^4\)

For purposes of litigation on the issue of educational inequality, social science output research arguably can be deemed legally irrelevant. Under the equal protection clause, the concern is whether government treats people equally, not with making people equal and not with equality of results emanating from a distributed benefit, for the latter may be beyond the capacity and power of governments and schools to control. Where equality and equal protection analysis are concerned, the focus is upon the rationality and fairness of how government distributes benefits, not with what people do with those benefits. Input-output research is irrelevant to that inquiry. Before deciding

\(^{53}\) See Bowles & Levin, More on Multicollinearity and the Effectiveness of Schools, supra note 41, at 393.

\(^{54}\) Not all social science research on the relationship between expenditures and achievement is negative. Cross-sectional studies both before and subsequent to the Coleman Report have found significant positive relationships between certain school inputs and student achievement. See Guthrie, A Survey of School Effectiveness Studies, in DO TEACHERS MAKE A DIFFERENCE? 25 (1970). Several of these studies were reanalyses of Coleman Report data which compensated for some of the Report's statistical and methodological weakness. See generally Do TEACHERS MAKE A DIFFERENCE? (1970). Other reanalyses of the Coleman Report, however, have confirmed its conclusion. See, e.g., Armor, School and Family Effects on Black and White Achievement: A Reexamination of the USOE Data, in ON EQUALITY OF EDUCATIONAL OPPORTUNITY, supra note 3, at 168; Jencks, The Coleman Report and the Conventional Wisdom, in id. at 69; Smith, Equality of Educational Opportunity: The Basic Findings Reconsidered, in id. at 230.
that education itself is irrelevant, which is the logical result of saying dollars do not make a difference, important avenues of improvement must genuinely be exhausted. Until this is done, this society should not mortgage away tomorrow's human capital—and certainly not on the basis of exploratory research findings more defective than informative.

III

LITIGATION STRATEGIES FOR RESOLVING THE COST-QUALITY DEBATE

A. A Negative Inputs Standard

The preceding analysis of the social science research evidence may simply confirm the wisdom of regarding the qualitative issue, and the cases in which it is raised, as non-justiciable because of a lack of judicially manageable standards. Yet it would be unfortunate if this scientific debate removed disputes from the judicial process. Although courts are expected to reach correct and just decisions, a far more essential function of courts is to settle disputes as best they can. It is particularly important that arbitrary inequalities in the way government treats individuals not be removed from judicial inquiry simply because there exists an unresolved and perhaps unresolvable scientific dispute over the effects of that differential treatment. No one would suggest that de jure segregation should now be viewed as non-justiciable because recent social science research challenges the importance of any harmful effects that attending segregated schools may have upon children.55

Given the standards of educational opportunity delineated in Part I of this article, what judicial approach is most likely to achieve equity in light of the scientific dispute discussed in Part II? Straightforward definitions of equal educational opportunity must give way to hybrid solutions and a multiplicity of approaches developed to meet varying contexts.

Basically, we believe that a negative inputs or expenditures standard of equal educational opportunity should be adopted. An inputs standard defines and evaluates educational opportunities in terms of the programs, services, and facilities made available to children, and thus is a "school concept" of educational opportunities. Inputs are chosen because the focus is upon what the state provides to the child, not on what the child does with what the state provides him, as that is beyond the power of the school to control. The word "quality" should perhaps be avoided, for that connotes an attribute that goes beyond opportunity, beyond what the school itself provides.

An inputs standard alone would require an equal dollar revenue distribution; hence the negative standard is more appropriate. The focus is upon what the state must not do, not upon what it must do or how it must distribute dollars and resources. The concern is only with expenditure or input inequalities that lack rational justification, such as wealth-created disparities. The state is left free to vary expenditures and inputs in order to reflect variations in costs and needs.

We are not so disingenuous as to suggest that eliminating unjustified fiscal disparities will achieve equal educational opportunity in all of its dimensions. A negative inputs standard will, however, preserve the dispute-settling function of the judiciary while at the same time being mindful of judicial limitations. While schools cannot be held responsible for the test performance of children, they can be held accountable for the quality of the school services provided, the "inputs" to the educational process. Schools cannot make children equal, but what they provide to children can be made equal, with appropriate adjustments for special problems such as the need for compensatory, vocational, and special education. Money, of course, is not the whole answer, but with it a district can purchase the needed inputs to attain educational excellence. Opinions vary on how districts can most wisely and efficiently invest their fiscal resources to achieve the highest possible quality of educational opportunity. But there is not disagreement that the level of resources available to a district is an ever-present factor which constrains a district's effort to perform its mission. Equality of fiscal resources is a threshold requirement for achieving equality of educational opportunity.

B. Application of the Standard

Proof of the input-opportunity relationship in the affirmative case should be limited to demonstrating existing inequalities in inputs and to presenting testimony of school personnel on the educational consequences of input disparities. Social science research evidence should be avoided during the affirmative case and held for rebuttal, a strategy consistent with the notion that output research is not relevant to equal protection analysis. Input evidence should alone establish a prima facie case of inequality of educational opportunity. Plaintiffs should prevail where no social science evidence is offered to rebut the input-opportunity relationship or where such evidence is deemed legally irrelevant.

The cost-quality relationship may also be regarded as admitted, because of the conduct of the state in encouraging spending above the foundation plan, and the testimony of state and local defendants that expenditures beyond the foundation program can be important educationally if spent wisely. Assertions by the defendants regarding the value of local control also represent an admission on the cost-quality issue, for the very concept of local control presupposes a relationship between cost and educational quality.

Where social science research evidence is offered in school finance litigation to rebut the input-opportunity relationship, and that evidence is accepted by the trial court in resolving the cost-quality issue, the burden of proof can be shifted to defendants who are challenging the legislative declaration of fact underlying the entire financing structure, that is, that money matters. In effect, in school finance litigation plaintiffs are contending that there are no rational bases for those explicit and implicit statutory declarations. If the state is to be permitted to question what should be viewed as legislative factual

---

determinations, then the burden of proof in such an inquiry should properly rest with the party challenging the validity of such statutes.57

Defendants' research evidence should be rejected for still another reason. Wealth-created expenditure disparities have been defended on the ground that the school finance system advances the state interest in encouraging local fiscal control, an argument that was accepted by the United States Supreme Court in *Rodriguez*. Yet there is an inconsistency between defending the system on the basis of local control and contending that cost and quality are not related. As already argued, if expenditure disparities above some base level expenditure have no educational consequence, then local control has no meaning and there would be no state interest which would validate the financing system. If the state purports to defend the system's irregularities on the basis that expenditures above some minimum level are, in effect, wasted, then the local control justification ought not to be available to it or, alternatively, the state should bear the burden of proof in demonstrating the rationality of the local control justification.

**Conclusion**

The strategies offered here for measuring equality and for resolving the cost-quality issue are specific, practical, and operational—designed solely for the narrow context of the judicial process. No specific school finance system is proposed as a remedy; that is the function of the legislature. Our suggestions are intended simply to preserve the dispute-settling function of the legal process against a threatened erosion by the intimidating aura of scientific certainty.

---