

Education Working Paper Archive

The Alchemy of “Costing Out” an Adequate Education

September 14, 2006

Eric A. Hanushek

*Paul and Jean Hanna Senior Fellow, Hoover Institution
Stanford University*

(Revised version of paper prepared for the conference on *Adequacy Lawsuits: Their Growing Impact on American Education*, October 13-14, 2005, Kennedy School of Government, Harvard University)

Abstract

In response to the rapid rise in court cases related to the adequacy of school funding, a variety of alternative methods have been developed to provide an analytical base about the necessary expenditure on schools. These approaches have been titled to give an aura of a thoughtful and solid scientific basis: the professional judgment model, the state-of-the-art approach, the successful schools method, and the cost function approach. Unfortunately, none can provide a reliable and unbiased answer to the question 'how much do adequate schools cost?' Each is highly manipulable, generally satisfying the interested party commissioning the work to be done but not meeting the fundamental tenets of scientific inquiry. This paper reviews and critiques the methodology as applied in a substantial number of states.

The Alchemy of “Costing Out” an Adequate Education

Eric A. Hanushek¹

Holding schools accountable for student performance has highlighted a simple fact: Many students are not achieving at desired levels. Significant achievement gaps by race and income persist, and concerns abound about whether most schools are on an obvious path toward improving the achievement of all students. While people from diverse perspectives have offered reform plans and solutions, a prevailing argument is that the schools lack sufficient resources to support academic success, and a variety of parties have sued states to compel them to provide greater funding for education. A key question in these lawsuits – “What will it cost to improve student achievement?” – has led courts and legislatures to seek out a scientific determination of the amount of spending required by schools. And there has been no shortage of consultants prepared to provide one.

Consultants have developed four distinct methods for “costing out,” that is, estimating, the additional spending necessary to secure an adequate education. They are generally referred to as the “professional judgment,” “state-of-the-art,” “successful schools,” and “cost function” methods. These costing-out studies are frequently contracted for by plaintiffs or other interested parties who desire increased levels of spending for education, although defendants may commission one in an attempt to neutralize a rival study. This paper describes the main features of each method and explains why they all fall short of scientific standards of inquiry and validity.

The Origin of School Finance Lawsuits

The judiciary's involvement in the evaluation of education funding schemes has prompted a significant shift in policy discussions about school finance. All state constitutions mandate a statewide educational system and prescribe a legislative process for determining the state and local funding for elementary and secondary education (and the many other public services these governments provide). Nationwide, less than 10 percent of spending on education comes from the federal government, with the balance being roughly equally split between state and local governments. The exact distribution of fiscal responsibility differs significantly from one state to the next, but in almost all states, local governments, usually independent school districts, raise their share mainly through the local property tax. States generally distribute their funds so as to compensate, at least partially, for differences in local property values that affect the ability of local school districts raise funds.

Following the California court case of *Serrano v. Priest*, decided in 1971, a majority of states saw legal actions designed to equalize funding across districts. The plaintiffs in these cases argued that some school districts—by virtue of a low property tax base or unwillingness to support school funding—spent significantly less than other, more advantaged, districts. This situation presented an equity concern because children growing up in a low-spending jurisdiction could receive an inferior education

The outcomes of these suits, argued under separate state constitutions, were mixed, with some state courts finding disparities to be unconstitutional and others not.² Whether successful or not, the lawsuits tended to increase the state share of funding and

brought about more equalized funding within states, with many state legislatures acting without being pressured to do so by a court judgment.³ Interestingly, although school funding suits were motivated by the assumption that an inferior education disadvantaged students, until recently virtually no scholars have examined whether student test-score performances or other educational outcomes tended to be more equal after spending was equalized. In fact, the few investigations of this issue that have been conducted show that the spending increases produced by equity lawsuits have had little or no effect on student achievement.⁴

Adequacy Litigation

Beginning in the 1980s, some plaintiffs argued that children may not be getting a constitutionally acceptable level of education even when spending across a state was more or less equalized. Alabama, the target of the 1993 case, *ACE v. Hunt*, epitomized this situation; spending across districts was quite equal but student achievement levels were near the bottom of the nation. The juxtaposition of an equitable system and poor performance led to a new legal and policy goal, described as “adequacy.” The plaintiffs in adequacy lawsuits argue that students’ low achievement stems from insufficient public funding and ask the courts to correct this fiscal inadequacy.

This new focus on adequacy dovetailed with the accountability and standards movement, which has asked states to track student educational proficiency relative to state standards or goals. The federal No Child Left Behind Act of 2001 (NCLB) has reinforced and extended this movement, requiring testing in grades three through eight and once in high school that gives the public detailed information on how well students are performing in school. Plaintiffs engaged in adequacy litigation have been able to use

this information to assert that the state has failed to meet its constitutional obligations as described in the educational clauses of each constitution. They then find it easy to argue that states are not investing the necessary resources to ensure that students are reaching the proficiency standards the states themselves have set. Costing-out studies purport to show what it will cost for students to reach proficiency.⁴

Costing Out Approaches

In court, adequacy litigants present such costing-out studies as “scientific” evidence of the amount of money needed to obtain an adequate education. Such studies have been conducted in over thirty-three states, and the demand for such analyses has only continued to rise as adequacy lawsuits proliferate.⁵ Plaintiffs have discovered that there is a great value in presenting to the court and the public a specific “number” for total “required” state spending, which they want to be treated as the amount that is both necessary and sufficient.⁶ Courts have clearly been influenced by this strategy, as judges have been willing to write that specific number, derived from costing out studies, into the remedies they order.⁷ Legislatures also consistently use these studies to guide their appropriations.

Before describing and assessing the various costing-out methods, it is worth discussing some of the terminology they use and a fundamental problem common to them all. School finance discussions are punctuated by certain terms, whose meaning in this context often differs greatly from the meaning that is generally accepted. Most notably, the concepts of *cost* and *efficiency* have been redefined to suit the argument at hand. Ordinarily, the words imply finding the most inexpensive way of achieving one’s

objective, but, adequacy consultants have refashioned them in such a way as to help make the case that more money—indeed, as much money as is politically feasible—should be spent on education.

The overarching problem stems from the empirical evidence available to estimate the costs of adequate student proficiency. The consultants' work would be simple, if scholars had shown, repeatedly, something like the following: An additional expenditure of one thousand dollars per pupil will translate, on average, into a 15 percent gain in student proficiency. Unfortunately, such studies do not exist. Research has not shown a clear causal relationship between the amount schools spend and student achievement.⁸ After hundreds of studies, it is now generally recognized that how money is spent is much more important than how much is spent. This finding is particularly important for consideration of judicially ordered changes in school finances, because these alterations have little control over how any new moneys are spent.

A simplistic view of this argument – convenient as a straw man to be beaten down – is that 'money never matters.'⁹ The research of course does not say that. Nor does it say that 'money cannot matter.' It simply underscores the fact that there has historically been a set of decisions and incentives in schools that have blunted any impacts of added funds, leading to inconsistent outcomes. That is, more spending on schools has not led reliably to substantially better results on the tests that states use to determine whether students are proficient—the same tests plaintiffs use to document inadequacy in a state's educational system.

This fact also underscores the challenge facing the consultants who purport to describe the spending necessary to achieve adequate levels of student achievement.

Because looking at the state's schools – where spending a lot shows little relationship to the desired performance – is fraught with embarrassment, they must find some way around current reality. Each of the costing-out methods takes a different approach for dealing with this dilemma. As might be guessed, these methods fall far short of standards for scientific validity.

Professional Judgment

Perhaps the most commonly applied approach is the “professional-judgment” method.¹⁰ With a few nuances, the approach involves asking a chosen panel of educators—teachers, principals, superintendents, and other education personnel—to develop an educational program that would meet certain specified outcomes. Their efforts typically produce “model schools” defined in terms of class sizes, guidance and support personnel, and other programs that might be necessary. The analysts running the process then provide missing elements (e.g., central administration costs or computers and materials) and employ externally derived cost factors (e.g., average teacher or principal salaries) to determine the total cost of the model schools. The panel may or may not provide guidance on extra resources needed for disadvantaged children, special education, or the like.

Professional-judgment panels are generally instructed not to consider where revenues would come from or any other restrictions on spending. In other words, “dream big”—unfettered by any sense of realism or thoughts of trade-offs. Indeed, one motivation for filing adequacy law suits is to resolve financial questions in an arena other than provided by state legislatures or local school boards, which are not single issue

oriented and of necessity take such practicalities into account. If courts can be induced to ignore practical constraints, more money for education might well be obtained.

Augenblick, Palaich and Associates described the operation of the professional judgment panels in North Dakota:

We worked hard to push people to identify resources they thought were needed to help students meet state and federal standards in spite of their natural tendency to exclude items because local voters might not approve of them or schools could “get by” without them.¹¹

Admonitions to professional judgment panels to dream big amount to a fundamental redefinition of the term *cost*. Whether discussing the purchase of a car, home, or service, the term *cost* is usually understood to mean the *minimum* necessary expenditure to achieve a given outcome. The idea is to establish the desired quality level and determine the lowest amount of money required. By contrast, professional judgment panels are effectively encouraged to identify the maximum expenditure imaginable, in the hope that the amount will be enough to produce adequately proficient students. A 2004 New York study conducted by a consortium of researchers from the American Institutes for Research and Management Analysis and Planning, Inc. even used a two-stage process in which a super-panel was given the results from separate subpanels that had each estimated the desirability of some educational component. The super-panel then aggregated the results, input-by-input, from each of the subpanels. This design effectively maximized expenditure estimates by ensuring that any trade-offs between programs and resources made by the individual subpanels were ignored and that the resulting recommendation would be the maximum possible. The very design of the study, though couched in scientific terms, reflected the underlying policy goal of increased spending on education.

Courts relying on professional-judgment studies to mandate spending levels assume that the panelists' model school will produce the desired results just because that was the panel's charge. None of the reports ever test this assumption. In fact, the reports often admit that there is little reason to expect that students will achieve at the desired levels. The AIR/MAP team's November 2002 proposal to conduct its costing-out study promised the consultants would answer the question, "What does it actually cost to provide the resources that each school needs to allow its students to meet the achievement levels specified in the Regents Learning Standards?" Yet the 2004 study based on that proposal includes a disclaimer the courts apparently overlooked:

It must be recognized that the success of schools also depends on other individuals and institutions to provide the health, intellectual stimulus, and family support upon which public school systems can build. Schools cannot and do not perform their role in a vacuum, and this is an important qualification of conclusions reached in any study of adequacy in education.

Also, success of schools depends on effective allocation of resources and implementation of programs in school districts.¹²

A study conducted by Augenblick, Palaich and Associates and data from North Dakota illustrates the extent to which costing-out studies using the professional judgment method ignore empirical evidence.¹³ The authors of this study prescribe the necessary spending level for each of the K-12 districts in North Dakota in 2002. Two points are important: First, there is wide variation in the calculated needs of districts. Second, a number of districts were spending *more* in 2002 than the consultants (through their professional judgment panels) thought needed to achieve the full 2014 performance levels.

Because we have information on students' actual performance in North Dakota for 2002, we can see how performance is related to the fiscal deficits and surpluses that were

calculated by the professional judgment (PJ) model. (Here, spending lower than the study found necessary is termed a “PJ deficit”; spending higher than the study determined necessary is termed a “PJ surplus.”) It seems natural to expect that districts with PJ surpluses (spending more money than they “need”) would be performing above their panel’s achievement goals. It is also plausible to expect that districts with larger PJ fiscal deficits would be further from achievement goals than those with smaller PJ fiscal deficits. Such expectations are appropriate, since the methodology is designed to adjust for needs that arise from the concentration of a disadvantaged population, variation in school size, and the like.

Yet we observe exactly the opposite of what might reasonably be expected. A regression of reading or math proficiency percentages of North Dakota districts on the PJ deficits indicates a *positive* relationship between a PJ deficit and student achievement. In other words, the larger the PJ deficit, the higher is the student performance. (The positive relationship between deficits and achievement remains results remain the same even after trimming off all surpluses and deficits greater than \$2,000 to ensure that the analysis is not distorted by outliers. See figure 1.) Moreover, in terms of simple averages, those districts with PJ surpluses have student achievement significantly below that found in districts with a professional judgment (PJ) deficit. In other words, the PJ deficits give worse than no information about school performance.

Incredibly, Augenblick, Palaich and Associates actually discuss the lack of empirical validation of the professional-judgment method in their North Dakota study. “The advantages of the approach [professional judgment] are that it reflects the views of actual service providers and its results are easy to understand; the disadvantages are that

resource allocation tends to reflect current practice and there is *only an assumption, with little evidence, that the provision of money at the designated level will produce the anticipated outcomes.*"¹⁴

In sum, the professional-judgment model lacks all empirical grounding. The professional educators called upon for their judgment generally lack expertise in designing programs to meet objectives outside of their experiences. While they may have experience making trade-offs within current budgets, they do not have the research knowledge or personal experience to know how resources needs will change if they design a program for higher student outcomes or for different student body compositions. Most importantly, the direct conflicts of interest are palpable: The outcomes may directly affect participants' own pay and working conditions, creating an incentive for them to distort whatever judgments they might otherwise make. The professional judgment approach could be more accurately described as the *educators' wish list* model.

State-of-the-Art or Evidence-Based

If the professional-judgment model relies on self-interested experts, the second costing-out approach relies upon the judgments of the analysts themselves. This approach has been immodestly called "state-of-the-art" by the major firms using this approach.¹⁵ Seeking to give their study scientific cache, they also refer to it as the "evidence-based" method. The consultants involved sort through available research, select specific studies that relate to elements of a model school, and translate these studies into precise estimates for resource needs. A set of model schools that are subsequently costed-out in the same manner as the professional-judgment model schools.

The state-of-the-art approach purportedly relies on evidence about the effectiveness of different programs. The researchers search for studies that show a statistically significant impact of some resource on some outcome, but ignore the many that do not.¹⁶ As long as the relationship is statistically significant, the consultants seem to give no apparent consideration to the magnitude of the estimated relationship. Typically, little evidence is directly cited in the report, and considerable weight is placed on research that has not been published in peer-reviewed journals. Where there are citations, they demonstrate that consultants are picking and choosing among a wide range of studies and evidence rather than drawing a conclusion from the research literature as a whole. In other words, the authors often pick a particular study and take one estimated effect from a large range of estimates. Such a procedure does not meet scientific standards.

No attempt is made in these analyses to specify the expected quantitative effect on student achievement or other educational outcomes of the changes in inputs and programs the authors are costing out. For example, state-of-the-art reports commonly recommend across-the-board reductions in class size at different grades. The analyses on which these recommendations are based typically estimate the impact such reductions might be expected to have—and these estimated impacts never come close to taking the systems they analyze to meeting the state proficiency standards established under NCLB.¹⁷ For other common “research-based” recommendations, credible estimates of program impacts are entirely lacking. In short, the “state-of-the-art” consultants’ failure to relate their recommendations to specific levels of achievement reflects the lack of evidence that would allow them to do so.

Again, the methodology specifically eschews taking costs into account or attempting to calculate the minimum costs of any level of achievement. This facet may, however, simply reflect the difficulty they have in finding *any* programs that reliably relate to student outcomes, so considering trade-off at their level of generality is far beyond being feasible.

The only empirical bases for these state-of-the-art analyses come from a small number of selected research studies that do not necessarily reflect the experiences in the individual state being sued. And, most importantly, because these are studies are particular ones that have been selected from the research base to suit the consultant's own purposes, there is no reason to believe that they provide an unbiased estimate of the empirical reality more generally. Indeed, given the selected nature of the studies the consultants favor, the state-of-the-art model would more appropriately be termed the *consultants' choice*.

Successful Schools (or Districts)

The “successful-schools” approach begins by identifying schools or districts in a state that are effective at meeting educational goals. Various methodologies may be used to identify successful schools. Typically, the process concentrates on student achievement, occasionally with some allowance for the background of students.¹⁸ Spending on special programs—say, remedial education or special education—is stripped out of budgets in order to obtain a “base cost” figure for each district. Typically, then, particularly high- or low-spending schools are excluded, and the base costs for the remaining schools are averaged to develop a level of spending that can feasibly yield high

performance. To get the full costs of the school, expenditures on special programs are then added back in, based on the distribution of students with such special needs for each school.

The method used for selecting successful schools is obviously important. The typical method is to take the highest-performing schools in the state, defined by the level of student test scores and other educational outcomes. While this may seem appropriate, it ignores the many non-school factors that affect student performance, such as family background, peer relationships, and prior schooling experiences. When the consultants ignore such considerations, they can hardly conclude that the high performance in the successful schools is driven by the amount of spending taking place. There is no reliable evidence that equivalent spending in other social contexts would yield similar levels of student performance. Indeed, there is powerful evidence to the contrary.

Quite apart from these considerations, the successful schools approach attempts to estimate the future from what is known about the present. The consultants are asked to project future levels of student proficiency that would occur if spending were increased. Yet the methodology is rooted in the current operations of a state's schools. Therefore, it can say something about meeting the performance goals that states have established under NCLB *only if* some subset of schools is currently achieving at the level that NCLB requires. But since no district has yet reached the standards NCLB has set forth, that is most unlikely. Because the approach relies on the observed performance of one set of schools with a given level of success, it has no way to project the results to any higher performance level. Assume for illustration that the set of schools identified as successful

has 70 to 80 percent of students performing at the proficiency level; there is no way to extrapolate these results to a 95 percent proficient performance level.¹⁹

Policy decisions should be built on the joint consideration of program effectiveness and costs. This is the standard meaning of *efficiency*—achieving a given outcome, such as a given amount of learning, at the minimum cost. In education discussions, efficiency often has a bad name, in part because it is taken, wrongly, to mean least cost without regard to the outcome.²⁰ Utilizing an efficiency standard in education requires acknowledging that different schools operate at different levels of efficiency. Presumably, the court would want to compel only such additional expenditures as can and will be used efficiently. Yet the vary range in expenditure levels found among “successful” schools (those meeting a prescribed student output standard) implies that not all school systems are using their funds as effectively as others. Should the starting point of discussion be current spending, accepting whatever is being done, or should there be some attempt to deal with the efficiency issue?

The panel referees appointed by the trial court judge in the landmark *Campaign for Fiscal Equity (CFE) v. New York* case addressed the idea of efficiency, but their approach was little less than bizarre. The plaintiffs presented to the referees the professional judgment cost estimates of the AIR/MAP discussed above. The state, using much lower estimates provided by Standard & Poor’s School Evaluation Service, had suggested that it was reasonable to concentrate on the spending patterns of the most efficient of the successful schools—those with high levels of student performance terms at lower levels of expenditure. In their calculations the S&P analysts therefore excluded the top half of the spending distribution across the successful districts. But to reconcile

the state's recommendation of \$1.9 billion with the AIR/MAP estimates of over \$5 billion, the referees insisted on adding back in the higher spending successful districts, even when those districts did not produce better academic outcomes. After all, the referees reasoned, "there was no evidence whatsoever indicating that the higher spending districts...were in fact inefficient."²¹ In other words, spending more to achieve the same outcomes should not be construed as being inefficient. One might then ask, What would indicate inefficiency? The importance of this is clear: If spending must be sufficient to bring up achievement regardless of how efficiently resources are used, the answer is likely to be a very large number.

The successful schools approach calculates costs for a unique subset of successful schools. The chosen subset of schools conflates the various reasons why achievement may be high, including the family background of those attending the schools. This approach is better labeled the *successful students* model, because it does not separate the effects of school expenditures from other, external factors that are probably much more important.

Cost Function

The "cost-function" approach, sometimes also referred to as the "econometric" approach, relies on current spending and achievement patterns across the full set of schools in a state. In economics and other quantitative sciences, one variable is said to be a function of another if its level is shown to vary, whether positively or negatively, in response to changes in another variable. (When the price of gas increases, demand for gas goes down; demand for gas is therefore a function of price.) The cost-function label

reflects the assumption made in these studies that the level of required spending in a district varies predictably along with various observable characteristics of its students and the desired achievement level.

The methodology is similar to the successful schools analysis in its attempt to characterize districts that are meeting desired achievement standards. Consultants use statistical methods to estimate the relationships statewide between spending levels and different combinations of student achievement levels and student characteristics. They then use the results of this analysis to derive appropriate spending levels for each district. Cost-function studies may or may not attempt to distinguish between efficient and inefficient producers of outcomes—that is, between districts that spend more for some given level of achievement than others.²²

For all their scientific pretensions, however, all cost-function studies fail to adequately identify the causal relationship between student performance and spending. As noted above, there is a large body of statistical research examining the relationship between spending and achievement. This work examines how various measures of the resources available influence student achievement, taking into account differences in a range of background characteristics. This research has generally found little in the way of a consistent relationship between spending and student outcomes. Almost all estimates that do suggest a spending-achievement relationship typically show a very small effect of spending on student outcomes.²³ The obvious implication of this literature is that, absent other reforms that would make the education system more efficient, large spending increases are required to obtain a noticeable achievement gain.

Consultants conducting cost-function studies turn this analysis on its head. They begin by estimating a statistical relationship between spending (as the dependent variable) and achievement and characteristics of the student population (as the explanatory variables).²⁴ That is, they reverse the usual position of spending and achievement in standard evaluations of education policy, which typically predict achievement based on spending and a various other student characteristics.²⁵ Although consultants refer to these results as the “cost-function,” they actually just describe the existing spending patterns across districts with different achievement levels.²⁶ Unless one can assume that all districts are spending money wisely—an assumption broadly contradicted by existing research—these estimates cannot be interpreted as minimum costs.²⁷ They could simply indicate that the current pattern of spending is not very productive.

Yet this is just the most obvious of the problems plaguing these studies. Cost-function analyses have to deal with the fact that there are frequently no districts that achieve at the performance levels defined to be adequate. In such cases, the consultants typically assume that the relationship between spending and achievement remains the same regardless of the achievement level. That is, if they observe proficiency levels to be increasing by 10 percentage points for every additional \$1,000 per pupil spent among a set of districts with a maximum proficiency rate of 60 percent, they assume that relationship remains unchanged as districts near the target of 100 percent proficiency. There is of course no way to know whether this is true.

Finally, cost-function analyses also have to make analogous assumptions about the way in which various student characteristics, such as the percentage of low-income

students in a district, affect required costs. The cost-function studies' apparent strength—the fact that they draw on all the data available on performance and spending in a state—here becomes a weakness. It is unclear whether the evidence from Westchester County is at all informative about how to improve student achievement in the Bronx or precisely what adjustments would have to be made to account for the many differences in the two locations. Yet this is exactly the kind of analytic leap-of-faith that cost-function studies conducted in New York State are forced to make.

Cost-function approaches cannot identify the costs of an adequate education, as they do not even attempt to trace out the necessary cost of given performance levels. Instead, their name should reflect the fact that they simply capture the *expenditure function* for education – how much schools now spend to achieve at current levels.

Additional Cause for Concern

The four approaches to determining the costs of an adequate education each have some surface appeal, but the methodological flaws outlined above render their conclusions unreliable. Several additional issues—the process for choosing a method, the selection of an outcome standard, the assumptions utilized in developing cost estimates, and the lack of evidence that greater funding brings its intended results—raise further questions about the validity of these calculations.

Choosing the Method

The choice of approach for costing out is generally determined by the party requesting the work to be done. It appears that it might be a quite purposeful choice,

given that many costing-out studies are funded by parties with an interest in the outcome of the study.²⁸ For example, a review of analyses by Augenblick and Associates in four states where they applied both professional-judgment and successful-schools methods found that the professional judgment method yielded systematically higher estimates of “adequate” expenditure.²⁹ This apparently has influenced the choice of methodology by clients, who almost uniformly prefer to begin with the professional-judgment approach.³⁰

A recent compilation of estimates of necessary per-pupil expenditure for an adequate education across states and studies underscores the arbitrariness of these estimates.³¹ Even after adjusting for geographic cost differences across states and putting the estimates in real terms for 2004, they differ by more than a factor of three. If the methods systematically produce very different results when addressing the same question, they obviously cannot be taken as a reliable and unbiased estimate of the resources required. It is difficult to imagine what true underlying differences could drive such differences, given the many similarities in the school systems of different states. A more plausible explanation for the differences is that methods are chosen so as to provide politically palatable estimates for different state deliberations.

Defining an Outcome Standard

Organizations that commission costing-out studies appear to recognize the importance of the outcome standard chosen. The courts, in contrast, seldom focus on the standard employed by the consultant and instead tend to grasp onto the cost figures. Yet the outcome standards that are embedded in adequacy calculations clearly have a significant impact on the analysis of costs. For example, bringing all New York State

students up to the level of having an elite “Regents Diploma” is one of the loftiest goals of any state in the nation.³² This standard is substantively different from the constitutional requirement of a “sound basic education.” Each of the methods for costing out adequacy explicitly or implicitly bases its calculations on a definition of outcomes, yet the political judgments required are seldom admitted.

NCLB has only complicated matters. It is now popular to link costing out studies to achieving the goals of NCLB, even though they have no obvious relationship to the language in state constitutions that provides the legal basis for these lawsuits. By declaring proficiency the goal nationwide, the law would seem to have set in place a universal outcome standard. Yet although NCLB requires all states to ensure that every student is “proficient” by 2014, it leaves the task of defining proficiency to the states. As a result, proficiency in one state differs markedly from proficiency in another.

Before NCLB, some states chose to establish very high achievement standards—what might be termed aspirational goals. Others chose modest standards that were not a large stretch from what many students were already achieving. What achievement level constitutes “proficiency” is, then, a political choice that almost certainly changes over time. Thus, when adequacy suits are pinned to state proficiency levels, it is important to consider where the standards came from and how they should be interpreted.

The plaintiff in the New York City adequacy suit, the Campaign for Fiscal Equity, hired two consulting firms, AIR/MAP, to cost out an adequate education in New York City under the New York State constitutional requirement for providing a “sound basic education.”³³ The consultants chose instead to evaluate the costs of meeting the Regents Learning Standards. The Governor’s commission adopted a lower standard in its

estimation of costs, conducted with Standard & Poor's School Evaluation Service. The judicial referees, who were appointed by the court to advise it on the appropriate decision, were pleased by the consistency of the two estimates (after they made adjustments), even though they used different outcome standards and should not have been the same according to by the logic of costing out.³⁴ The referees even went on to recognize that the highest court said that the Regents Learning Standards were inappropriate, even as they ignored this in reviewing the cost estimates.³⁵

Three separate studies were conducted in Kentucky in 2003 by two firms: Verstagen and Associates and Picus and Associates (who conducted parallel studies using a professional judgment and a "state of the art" approach). Picus and Associates let the professional judgment panels interpret the seven constitutional requirements of education laid down by the Kentucky Supreme Court.³⁶ Verstegen and Associates added to these seven an extensive set of input and process requirements included in the current Kentucky school regulations.³⁷

An analysis in Augenblick, Myers, Silverstein, and Barkis was written into the judgment of the Kansas State Supreme Court and provides insight into the consultant's role in establishing an outcome standard:

A&M worked with the LEPC [Legislative Education Planning Committee] to develop a more specific definition of a suitable education. We suggested using a combination of both input and output measures. For the input measures, it was decided that the current QPA [Quality Performance Accreditation] requirements would be used, along with some added language provided by the LEPC.... Next...we determined which content areas and grade levels would be used.... A&M felt that the reading and math tests, which are given every year, gave us the most flexibility in setting the output measures.³⁸

Perhaps more interestingly, the definition of adequacy is not always related to outcomes. In North Dakota, Augenblick, Palaich and Associates, the successor firm to

Augenblick and Myers, noted that the state did not have explicit outcome standards but instead had input requirements. For their analysis, they layered on a set of outcomes that were related to state goals under No Child Left Behind.³⁹

Duncombe, Lukemeyer, and Yinger analyze the impacts of different goals on the estimated costs under alternative estimation approaches. They demonstrate that reasonable differences in the loftiness of the educational goal can lead to 25 percent differences in estimated costs within their cost-function analysis and 50 percent differences across alternative approaches to costing-out, including the professional-judgment approach.⁴⁰

No matter how one judges the analytical capabilities of the consultants, their expertise does not extend to deciding educational requirements of the state constitution. The plaintiffs and other interested parties can of course argue these matters in court, but they invariably attempt to submerge the centrality of these choices in the costing out studies.

Estimating the Cost of Quality Teachers

All approaches use information about current spending of schools—generally with important modifications—to estimate what resources are needed to bring students up to the desired level of proficiency. But using existing spending, as utilized within existing structures, with existing incentive systems, is a dubious way to begin. Nowhere is this more evident than what estimating the cost of obtaining higher quality teachers.

If one wished to hire higher quality teachers than currently are employed, what would it cost? The answer to this question depends markedly on whether one reproduces

the current single salary schedule (that pays teachers the same salary, except for differences in education and experience) that does not recognize differences in teacher's effectiveness in the classroom or whether one were to introduce a different pay and incentive scheme.

The same holds for often noted shortages, say in mathematics and science or language teachers. The "cost" of addressing these issues depends crucially on whether a district pays all teachers higher salaries in the hope of attracting those in shortage areas or whether the district just pays bonuses or higher salaries to fill the demand in the shortage areas.

The calculation of salaries is a particularly interesting point of comparison across different studies. Sometimes the consultants simply use the average salaries for existing teachers,⁴¹ or they may increase them by some amount (e.g., 10 percent in North Dakota in one study and 18 percent in Arkansas in another),⁴² vaguely arguing in terms of what other states spend. They then imagine that such increments will improve teacher quality. In other cases, the consultants imagine a bonus for teachers.

While the widely varying teacher salary factor has obvious and powerful effects on any cost estimates, none of these studies provides any evidence about the current quality of the teachers as measured by their impact on the achievement gains of individual students. Nor is there *any* research that shows that teacher salaries are related to the ability to raise student achievement. So this becomes a whimsical adjustment based on the consultant's own sense of whether or not average salaries are high enough (for some unspecified quality level). If they want to improve teacher quality, they simply increase the average salary by some arbitrary percentage.

Achievement versus “Opportunity”

As previously noted, virtually none of the reports actually says that it has calculated the level of resources that will yield the outcomes that they are striving to obtain. When it comes time to write the reports – and to produce a document for which by which the consultants might be judged – the language generally changes to providing an “opportunity” to achieve the standard, not actually achieving any standard.

The motivation for the underlying costing out analyses is that children are not learning at a putative constitutional level (or an NCLB level or a state standards level), but the reports essentially never say explicitly that the resources identified in the study are either necessary or sufficient to achieve these levels. Instead, they say that the resources will provide an opportunity to reach the standards.

This change of language means that the consultants are not predicting any level of achievement if the stated resources are provided. In fact, none of the reports states that the added resources will yield achievement that is any higher than currently observed. The reports provide no predictions about outcomes, and thus they are *completely unverifiable*. Said differently, there is no scientific basis for deciding among alternative estimates, because data on student outcomes are not informative.

This situation is pervades all of the methods and all of the currently available reports. The possible exception is some of the successful student or expenditure projection studies, where the authors might suggest that a given school could achieve a given level of performance *if it could figure out why some other school achieved that level and if it could reproduce it in another setting*. Yet no guidance on either the source of achievement or the way to reproduce it is ever given.

If the costing out studies do not provide any clear view of the outcome that would be expected, they become just the whim of the consultant – even when based on a methodology that has previously been applied or has a “scientific” air to it. There is no way to judge among alternative spending projections based upon any evidence about outcomes that will become available, thus putting each in the category of personal opinion and not science. There is no obvious reason for giving deference to the personal opinion of consultants hired by interested parties in the debates.

This work also does not help the political and legislative debate on school finance. The studies are designed to give a spending number. They do not indicate how achievement is likely to be different from the current level if such an amount is spent. Neither do they suggest how achievement (or even opportunity) would differ if a state spent 25 percent more or 25 percent less than the consultants personal opinions about what should be spent.

Returning to the court dilemma, the terms of the ‘does money matter?’ debate are central. Simply stating that money can be effective if it is spent in the right way is tautological. Without a proven strategy for using money wisely, the existing evidence overwhelmingly indicates just adding money is likely to be broadly ineffective. The historical record indicates that it has been exceedingly difficult to ensure that added money is spent wisely, because districts have not been using added money in a consistently effective manner. Moreover, with the possible exception of the consultant’s choice model (which, as described above is not credible), none of the approaches even attempts to offer any guidance about effective programs or policies that would provide for enhanced achievement when broadly employed.

Conclusions

Early school funding lawsuits centered on equity, defined simply as equal per pupil funding across school districts. This has given way to an emphasis on adequacy, as measured by student performance and other educational outcomes, moving the courts into areas in which they are completely unprepared. They cannot simply mandate a given level of student achievement. Instead, the courts must define their remedy in terms of instruments that are expected to lead to desired outcomes, instruments that can be monitored by the court. The easiest thing to monitor is the amount that states are spending, which has led to an inevitably focus on the financial resources committed to education. But how much money is needed to achieve desired schooling outcomes? To answer that question, the courts have come to rely on outside consultants (frequently hired by interested parties). These consultants, and the people who hire them, suggest that “costing out” exercises provide a scientific answer to a simple question, How much does it cost to provide an adequate education?

The methodologies that have been developed lack all semblance to a scientific determination of what the court needs to know—how much is needed to reach desired levels of proficiency. They do not provide reliable and unbiased estimates of the necessary costs to achieve desired goals. Nor do they provide any reason to expect that once the financial remedy is ordered, the desired educational goal will be achieved. In many studies, most especially those that use the popular professional judgment model, the results employ a methodology that cannot be replicated by others. And they obfuscate

the fact that they are unlikely to provide a path to the desired outcomes. Even the consultants themselves admit the weakness of the studies' underlying premise:

The effort to develop these approaches stems from the fact that no existing research demonstrates a straightforward relationship between how much is spent to provide education services and performance, whether of student, school, or school district.⁴³

All of the methods rely crucially on existing educational approaches, existing incentive structures, and existing hiring and retention policies for teachers. Each calls for doing more of the same—reducing pupil-teacher ratios, paying existing teachers more, retaining the same administrative structure and expense. Thus, they reinforce and solidify the existing structure, which is arguably incapable of bringing about the kinds of improvements that they purport to cost out.

As they typically have no relevant expertise in the funding, institutions, and incentives of schools, courts are generally quite eager to have somebody tell them the answer; they jump on “the number,” even while recognizing it may not be correct. Costing out studies do not and cannot support such judicial decision making.

Those people who wish the courts to be more deeply involved in the appropriations process – faced with evidence that the existing costing out methods lack credibility – frequently push for an alternative. After all, they note, we need to have some method of determining how much should be spent on the schools. But, in fact, we have historically had a method. Duly elected legislatures, local school boards, and other officials are charged with resolving differences of opinion, including those regarding education funding. Certainly, the outcome of these democratic processes will not satisfy everyone, but, as currently conducted, costing out studies do not provide a scientific alternative.

We simply do not have any reliable, objective, and scientific method to answering the question of how much it would cost to obtain achievement that is noticeably different from today's level. The courts can judge the constitutionality of the schools, but they should show some humility in attempts to change the outcomes radically. Their chosen instrument – the level of funding for schools – simply is, by the historical record, not the key to solving the current achievement problem.

Footnotes

¹ This analysis benefited from the research assistance of Brent Faville, and the editorial acumen of the editors of the forthcoming Brookings book, *School Money Trials: The Legal Pursuit of Educational Adequacy* (Martin R. West and Paul E. Peterson, ed.).

² An early suit in federal court, *Rodriguez v. San Antonio*, was brought under the 14th Amendment to the U.S. Constitution, but the U.S. Supreme Court ruled in 1973 that state funding arrangements were not a federal constitutional violation.

³ Sheila E. Murray, William N. Evans, and Robert M. Schwab, "Education-finance reform and the distribution of education resources," *American Economic Review*, 88, no.4 (September 1998), pp. 789-812.

⁴ See, for example, Thomas A. Downes, "Evaluating the Impact of School Finance Reform on the Provision of Public Education: The California Case," *National Tax Journal*, 45, no.4 (December 1992), pp. 405-419; Eric A. Hanushek and Julie A. Somers, "Schooling, Inequality, and the Impact of Government," in Finis Welch (ed.) *The Causes and Consequences of Increasing Inequality* (University of Chicago Press, 2001), pp. 169-199; and William Duncombe and Jocelyn M. Johnston, "The Impacts of School Finance Reform in Kansas: Equity Is in the Eye of the Beholder," Ann E. Flanagan and Sheila E. Murray, "A Decade of Reform: The Impact of School Reform in Kentucky," Julie B. Cullen and Susanna Loeb, "School Finance Reform in Michigan: Evaluating Proposal A," and Thomas A. Downes, "School Finance Reform and School Quality: Lessons from Vermont," in John Yinger (ed.), *Helping Children Left Behind: State Aid and the Pursuit of Educational Equity* (MIT Press, 2004), pp. 148-193, 195-214, 215-249, 284-314.

⁵ A review of past costing out studies can be found in "Quality Counts 2005: No Small Change--Targeting Money Toward Student Performance," *Education Week*, January 6, 2005. See also the ACCESS Project website (www.accessednetwork.org), a project of the Campaign for Fiscal Equity (CFE), the plaintiffs in the New York City adequacy case, *Campaign for Fiscal Equity v State of New York*. CFE states that its primary mission is to "promote better education by conducting research, developing effective strategies for litigation and remedies (including cost studies), and providing tools for public engagement." The count of prior costing out studies comes from (www.schoolfunding.info/index.php3), accessed on October 7, 2005.

⁶ This explains why the websites for advocacy organizations give top-billing to costing-out studies. For example, see the ACCESS Project at www.schoolfunding.info.

⁷ See, for example, *Campaign for Fiscal Equity v State of New York* and *Montoy. v. State of Kansas*.

⁸ Eric A. Hanushek, "The Failure of Input-based Schooling Policies," *Economic Journal* 113, no.485 (February 2003), pp. F64-F98.

⁹ Outside of the courtroom, most discussion of the 'money never matters' debate, a controversy of a decade ago, has subsided. For the historical framing of the question, see the following exchange: Larry V. Hedges, Larry V., Richard D. Laine, and Rob Greenwald, "Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes." *Educational Researcher* 23,no.3 (April 1994):5-14.

Eric A. Hanushek, Eric A., "Money might matter somewhere: A response to Hedges, Laine, and Greenwald." *Educational Researcher* 23,no.4 (May 1994):5-8.

¹⁰ Examples of this include Augenblick & Myers, Inc., *Calculation of the Cost of an Adequate Education in Indiana in 2001-2002 Using the Professional Judgment Approach, prepared for Indiana State Teachers Association, 2002*; John Augenblick, John Myers, Justin Silverstein, and Anne Barkis, , *Calculation of the Cost of a Suitable Education in Kansas in 2000-2001 Using Two Different Analytical Approaches, prepared for Legislative Coordinating Council, Augenblick & Myers, Inc., 2002*; Augenblick, Palaich and Associates, Inc., *Calculation of the Cost of an Adequate Education in North Dakota in 2002-2003 Using the Professional Judgement Approach* (North Dakota Department of Public Instruction, 2003); American Institutes for Research and Management Analysis and Planning (AIR/MAP), *The New York Adequacy Study: Determining the Cost of Providing All Children in New York an Adequate Education, 2004*; Lawrence O. Picus, Allan Odden, and Mark Fermanich, *A Professional Judgment Approach to School Finance Adequacy in Kentucky*, Lawrence O. Picus and Associates, 2003; and Versteegen and Associates, *Calculation of the Cost of an Adequate Education in Kentucky, prepared for the Council for Better Education, 2003*.

¹¹ Augenblick, Palaich and Associates, *Calculation of the Cost of an Adequate Education in North Dakota in 2002-2003 Using the Professional Judgment Approach*.

¹² AIR/MAP, *The New York Adequacy Study*.

¹³ Augenblick, Palaich and Associates, *Calculation of the Cost of an Adequate Education in North Dakota in 2002-2003 Using the Professional Judgment Approach*

¹⁴ Augenblick, Palaich and Associates, *Calculation of the Cost of an Adequate Education in North Dakota in 2002-2003 Using the Professional Judgment Approach*, p. II-3. [italics added]

¹⁵ See Allan Odden, Mark Fermanich, and Lawrence O. Picus, *A State-of-the-Art Approach to School Finance Adequacy in Kentucky*, Lawrence O. Picus and Associates, 2003.

¹⁶ Ibid.

¹⁷ Eric A. Hanushek, "The Evidence on Class Size," in Susan E. Mayer and Paul E. Peterson (eds.), *Earning and Learning: How Schools Matter* (Brookings Institution, 1999), pp. 131-168.

¹⁸ See, for example, Augenblick & Myers, Inc., *Recommendations for a Base Figure and Pupil-Weighted Adjustments to the Base Figure for Use in a New School Finance System in Ohio, 1997*; John L. Myers and Justin Silverstein, *Successful School Districts Study for North Dakota*, Augenblick, Palaich and Associates, Inc., 2005; and Standard & Poor's School Evaluation Service, *Resource Adequacy Study for the New York State Commission on Education Reform, 2004*.

¹⁹ A second extrapolation problem frequently occurs. Schools identified as successful just by proficiency levels on state tests tend to be higher-SES schools where the parents have provided considerable education to the students. The methodology concentrates on base spending for a typical successful school but then must indicate how much remedial spending would be necessary to bring schools with students of lower-SES backgrounds

up to the proficiency of the higher-SES schools. The appropriate way to do this is entirely unclear, because again the situation is largely outside of the observations going into the successful schools analysis.

²⁰

The classic misstatement of efficiency in education is found in Raymond E. Callahan, *Education and the Cult of Efficiency* (University of Chicago Press, 1962). Callahan failed to hold outcomes constant but instead looked at pure minimization of spending.

²¹

John D. Feerick, E. Leo Milonas, and William C. Thompson, *Report and Recommendations of the Judicial Referees* (Supreme Court of the State of New York, 2004).

²²

Gronberg et al. explicitly analyzed the efficiency of districts, but this analysis was not well-received in the courtroom; see the decision of Judge John Dietz in *West Orange-Cove Consolidated Independent School District et al. v Neeley et al.*, November 30, 2004. Timothy J. Gronberg, Dennis W. Jansen, Lori L. Taylor, and Kevin Booker, *School Outcomes and School Costs: The Cost Function Approach* (Texas A&M University, 2004).

²³

Eric A. Hanushek, "The Failure of Input-based Schooling Policies."

²⁴

Note that these estimates bear little relationship to classic cost functions in microeconomic theory that would use an underlying assumption of optimal firm behavior to translate the production function (achievement as related to various inputs) into a cost function that describes how cost relates to the prices of inputs. None of the work in education observes any variations in input prices (e.g., teacher wages, textbook costs, and the like). The empirical work in education described here relates spending to outputs and inputs such as the number or type of teachers, the poverty rate, and so forth.

²⁵

Some approaches to cost estimation are not done in this way but instead use various optimization methods to obtain the minimum cost of achieving some outcomes. They are nonetheless subject to the same interpretative issues about causation.

²⁶

There are some serious statistical complications in this work. The econometric methodology places requirements on the modeling that are almost certainly violated in this estimation. The cost function estimation essentially assumes that districts first specify the outputs they will obtain and that this chosen achievement level and the characteristics of the student body determine the spending that would be required (i.e., achievement is exogenous in statistical parlance). This approach, while summarizing the average spending patterns of different districts, is inconsistent with the interpretation that the level of resources available to a district determines student outcomes. The specific data and modeling are also very important. As Gronberg, Jansen, Taylor, and Booker state, "The measurement of efficiency in producing a set of outcomes is directly linked to the particular set of performance measures that are included in the cost model and the particular set of input measures." Gronberg et al., *School Outcomes and School Costs: The Cost Function Approach*.

²⁷

Other techniques found in the scholarly literature have been developed to consider cost minimization. See Eric A. Hanushek, "Publicly Provided Education," in Alan J. Auerbach and Martin Feldstein (eds.), *Handbook of Public Economics* (Amsterdam: Elsevier, 2002), pp. 2045-2141. Even when considered, it is generally impossible to

describe how efficiency is achieved. See Gronberg et al., *School Outcomes and School Costs: The Cost Function Approach*.

²⁸ E.g., see Eric A. Hanushek, "Pseudo-science and a Sound Basic Education: Voodoo Statistics in New York," *Education Next* 5, no.4 (Fall 2005), pp. 67-73.

²⁹ "Quality Counts 2005."

³⁰ For example, Thomas Decker describes the choice of professional judgment model for the costing out study to be commissioned by the North Dakota Department of Public Instruction: "The professional judgment approach we were aware would probably produce a higher cost estimate for achieving adequacy than successful schools."

Transcript of Deposition of Thomas G. Decker, August 17-18, 2005, p. 312.

³¹ "Quality Counts 2005."

³² New York State traditionally had two different diplomas with varying requirements. In 1996, the New York Regents determined that all students would have to qualify for a Regents Diploma (the previously optional high standard undertaken by roughly half of the students in New York State). This requirement has had a long phase-in period with altered testing requirements.

³³ Details of the costing out exercises in the CFE case can be found in Hanushek, "Pseudo-science and a Sound Basic Education: Voodoo Statistics in New York."

³⁴ Ibid.

³⁵ Feerick, Milonas, and Thompson, *Report and Recommendations of the Judicial Referees*.

³⁶ Instructions of what is needed were given to the panelists: Sufficient oral and written communication skills to enable students to function in a complex and rapidly changing civilization; Sufficient knowledge of economic, social and political systems to enable the student to make informed choices; Sufficient understanding of governmental processes to enable the student to understand the issues that affect his or her community, state, and nation; Sufficient self-knowledge and knowledge of his or her mental and physical wellness; Sufficient grounding in the arts to enable each student to appreciate his or her cultural and historical heritage; 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 Sufficient levels 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. Picus, Odden, and Fermanich, *A Professional Judgment Approach to School Finance Adequacy in Kentucky*.

³⁷ Verstegen and Associates, *Calculation of the Cost of an Adequate Education in Kentucky*.

³⁸ Augenblick, Myers, Silverstein, and Barkis, *Calculation of the Cost of a Suitable Education in Kansas in 2000-2001 Using Two Different Analytical Approaches*.

³⁹ Augenblick, Palaich and Associates, *Calculation of the Cost of an Adequate Education in North Dakota in 2002-2003 Using the Professional Judgement Approach*.

⁴⁰ William Duncombe, Anna Lukemeyer, and John Yinger, “Education Finance Reform in New York: Calculating the Cost of a ‘Sound Basic Education’ in New York City,” CPR Policy Brief, No. 28/2004, Center for Policy Research, Syracuse University, 2004.

⁴¹ See Odden, Fermanich, and Picus, *A State-of-the-Art Approach to School Finance Adequacy in Kentucky*.

⁴² See Augenblick, Palaich and Associates, *Calculation of the Cost of an Adequate Education in North Dakota in 2002-2003 Using the Professional Judgement Approach* and Odden, Fermanich, and Picus, *A State-of-the-Art Approach to School Finance Adequacy in Kentucky*.

⁴³ Augenblick & Myers, Inc., *Calculation of the Cost of an Adequate Education in Indiana in 2001-2002 Using the Professional Judgment Approach*.