

Pseudo-Science and a Sound Basic Education

Voodoo Statistics in New York

Checked:

“The New York Adequacy Study: Determining the Cost of Providing All Children in New York an Adequate Education,” American Institutes for Research and Management Analysis and Planning (March 2004).

“Resource Adequacy Study for the New York State Commission on Education Reform,” Standard & Poor’s School Evaluation Service (March 2004).

“Report and Recommendations of the Judicial Referees,” in Campaign for Fiscal Equity, Inc., et al., Plaintiffs, against The State of New York, et al., Defendants (November 2004).

Checked by Eric A. Hanushek

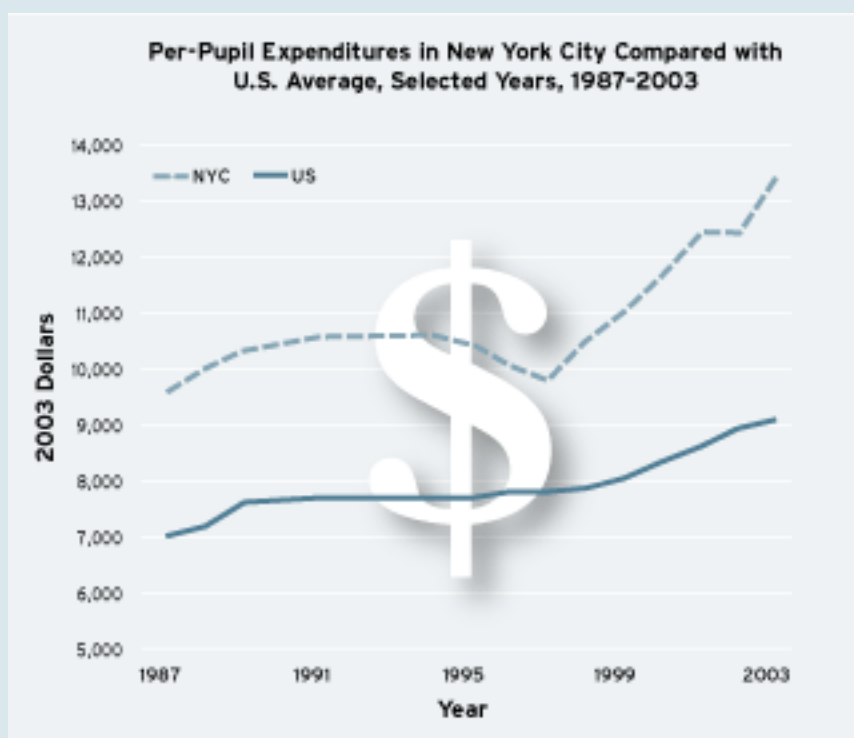
Most people who read the headlines last February were stunned to learn that New York City schools were being short-changed by \$5.6 billion per year, or more than \$5,000 per student. The 43 percent court-ordered budget increase, from around \$13 billion in operating expenditures to something approaching \$19 billion (not including some \$9 billion over five years for building improvements), is the largest school finance “adequacy” judgment ever awarded.

Of course, most people do not have a good grasp on either the economics or the performance of New York City schools. If they did, they would be even more stunned by the declared shortfall.

Figure 1 shows the recent history of spending in New York City, now nearly \$13,000 per student per year, which is

Infinity and Beyond (Figure 1)

Per-pupil expenditures have skyrocketed in New York City since the late 1990s, both in absolute terms and relative to the rest of the country.



SOURCE: New York State Education Department; National Center for Education Statistics

more than 50 percent above the national average and pulling away.

The city does, by any standard, face huge education problems. Indeed, despite a drastic restructuring of the school bureaucracy, implemented by Mayor Michael Bloomberg beginning in 2002 (see *Forum*, p. 11), and despite the heavy infusions of cash shown described in Figure 1, Gotham’s academic outcomes remain poor. On the 2003 National Assessment of Educational

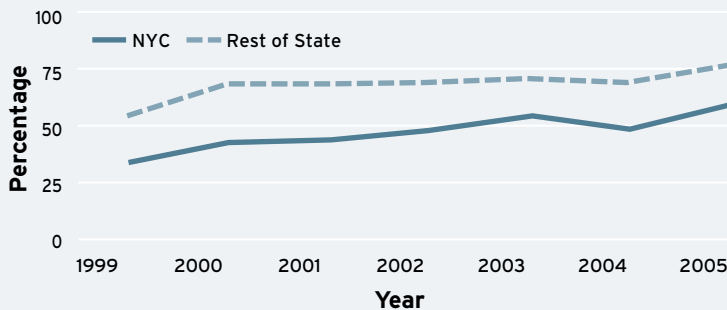
Progress (NAEP) tests, 46 percent of the city’s students scored “below basic” in mathematics, and 38 percent were below that low threshold in reading (compared with 33 and 28 percent for the nation, respectively). On the state exams that can be tracked over time, New York City has had mixed results—improvement in some areas but declines elsewhere.

But the discrepancy between years of budget increases and years of mediocre

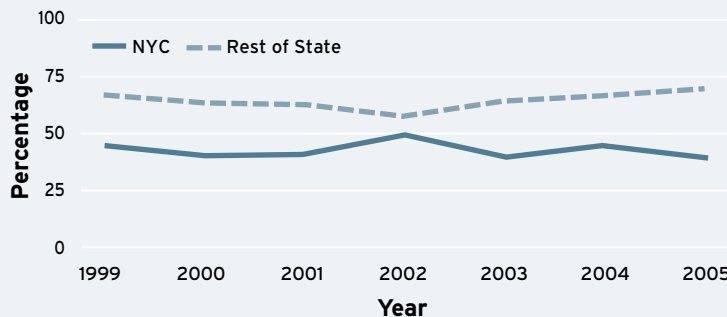
A Reading Frenzy? (Figures 2a and 2b)

Though widely celebrated as an indication of real improvement, the reading test scores for New York City 4th graders trailed those in the rest of the state and have only slightly closed the gap in recent years. For 8th graders, the gap has widened.

a) Fourth-Grade Students Passing the New York State Reading Test



b) Eighth-Grade Students Passing the New York State Reading Test



SOURCE: New York State Education Department

academic outcomes did not deter New York State Supreme Court Judge Leland DeGrasse from deciding that the problem could be solved by an annual addition of \$5.6 billion.

The very process of budget determination implicit in such judicial appropriations gives the first indication that something is fundamentally haywire. Ordinarily, courts have nothing to do with expenditures. That is a matter for the political branches, not the courts, to decide—a constitutional arrangement that led that great New

Yorker, Alexander Hamilton, to declare the judiciary the weakest branch. In New York, as in all other states, the normal appropriations process begins with the governor's creating the budget recommendations for education and other state services. The legislature, subject to gubernatorial veto, appropriates the funds. But such constitutional proprieties were set aside when Judge DeGrasse—with no previous education expertise and no relevant staff support and without considering the impact on other areas of

expenditure—intervened to establish the level of education appropriations for New York City. Suddenly the weakest branch had declared itself the boss.

Given the fundamental constitutional conflict involved, this judicial decision will probably be in and out of the courts and legislature for some time. To get some hint of the future, one may look no farther than neighboring New Jersey, where the courts have retained control over the financing of several city school districts for decades.

Nonetheless, it is informative to investigate what is behind the DeGrasse appropriations, because New York is only the leading edge of a national movement. In more than two-thirds of the states, teacher unions, school districts, and other interested parties have filed similar lawsuits that seek judgments resembling the stunning result handed down in New York.

The DeGrasse judgment is the result of a decade-long political and legal struggle (described by *New York Daily News* reporter Joe Williams in this journal earlier this year: "Legal Cash Machine," *Education Next*, Summer 2005). Several groups, led by the Campaign for Fiscal Equity (CFE), a nonprofit legal advocacy organization, filed suit in 1993 claiming that New York State was depriving New York City public school students of their constitutional rights to a "sound basic education," a standard that had been prescribed in 1982 by the state's highest court (in New York, the Court of Appeals). Despite its name, the lead plaintiff in the 1993 complaint, CFE, did not argue that the state's financing arrangements were inequitable, but that the funds given to New York City were not "adequate" for a sound basic education. From his Manhattan courtroom, Judge DeGrasse sided with the plaintiffs. The decision was ultimately upheld by the Court of Appeals, which remanded the case to DeGrasse to ensure that the Constitution was served; hence his appropriations figure.

But the interesting question, ignored in all the righteous hoopla over the court decision, is: Where did Judge DeGrasse get that \$5.6 billion figure? Why not \$10 billion? Or just \$1 billion? How much does a sound basic education cost?

The Inexact Science of Costing Out

The paternity of the \$5.6 billion figure is easily traced to the plaintiffs in the case, whose expertise was treated as authoritative, despite their obvious vested interest in the outcome. The Campaign for Fiscal Equity had commissioned a costing-out study by a consortium of two consulting firms, the American Institutes for Research (AIR) and Management Analysis and Planning, Inc. (MAP). Both firms claimed to have the analytical capacity to determine objectively the funding schools need to perform adequately. The consortium, known as AIR/MAP, made the extraordinary claim in its November 2002 proposal that its study 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?”

The following year AIR/MAP submitted its final costing-out analysis to its client, the plaintiff, who then submitted the document to the court by way of the panel of three referees appointed by Judge DeGrasse to assist in fashioning an appropriate remedy. These referees were a Fordham Law School dean and two retired New York judges, none with any particular expertise in school finance. After an intensive and expensive period (the three referees submitted combined bills in excess of \$350,000 for their part-time work over the course of four months), they issued a 57-page

report accepting the essential elements of the AIR/MAP document that CFE had submitted to the court. The referees recommended that funding of New York City schools be ramped up an additional \$5.6 billion a year within four years; that new studies be undertaken every four years to find out how much, if any, additional funding would be required; that \$9.2 billion be spent for capital projects spread over the following five years; and that another study be con-

The range of estimates underscores the arbitrary nature of any number the court would order the legislature to spend.

ducted after five years to see if additional spending was required.

Both Judge DeGrasse and the mainstream New York City media covering the story treated the referees’ report as authoritative. Little attention was given to the other studies reviewed by the referees that recommended quite different levels of expenditure. One might have thought the referees would give at least equal consideration to the report submitted by the New York State Commission on Education Reform appointed by Governor George Pataki. Known as the Zarb Commission, after its chairman, Frank G. Zarb, a former chairman of NASDAQ, the commission estimated that the city needed \$1.9 billion to provide an adequate education. Meanwhile, the City of New York, eager to get as much state money as possible, proposed additional spending of \$5.4 billion, an amount that resembled the AIR/MAP recommendation. It added

the caveat that none of this funding should come from the city. Not to be left out, the New York State Board of Regents calculated its own figure, \$3.8 billion. Even Standard & Poor’s jumped in, with an independent study that included 16 different estimates for the resource gap, ranging from as high as \$7.3 billion to as low as \$1.9 billion, depending on achievement targets, regional cost adjustments, and cost effectiveness of districts. The Zarb Commission, in fact, used the lowest of S&P’s estimates as the basis for its own recommendation.

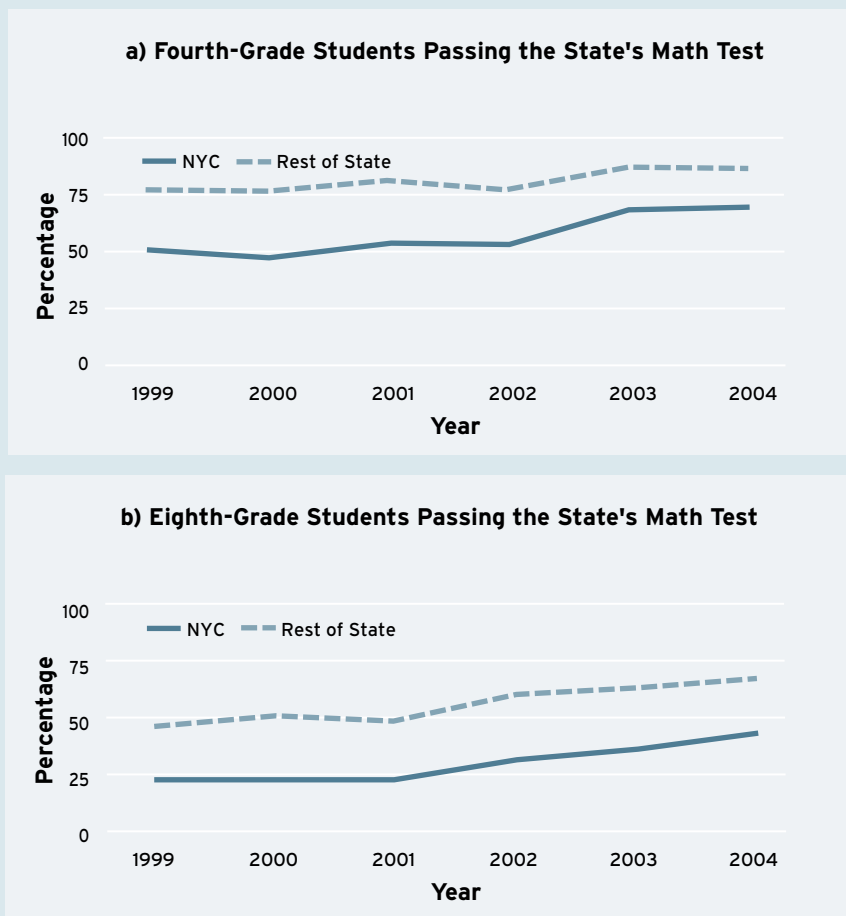
This range of estimates underscores the arbitrary nature of any number the court would order the legislature to spend. Even the plaintiff’s own consultant, AIR/MAP, admitted that its “costing out” methods are not based on an exact science.” Far from being an exact science, the method they chose, as we shall see, was profoundly subjective, a matter of judgment by and for self-interested parties.

Aligning Professional Judgment and Self-Interest

The AIR/MAP study relied on the “professional judgment” method in its costing-out analysis. The consultants brought together multiple panels of school personnel and asked them to design a program that would ensure that all New York City students could get a sound basic education and determine the resources needed to deliver the program. But these program designers, 56 in all, were also service providers whose pay, working conditions, and other funds were directly dependent on the resources put into the system. Such a procedure is akin to asking Martha Stewart how much you should pay for her to decorate her own house. When someone else is to pay,

Go With the Flow (Figures 3a and 3b)

From 1999 to 2004, the math scores of New York City 4th and 8th graders continued to trail those of students in the rest of the state.



SOURCE: New York State Education Department

and Martha is to enjoy, one can only expect the sky to be the limit.

Admittedly, not all 56 panelists worked within the New York City school system. But all except one, a retired employee, were currently working somewhere within the New York State school system. Since the panelists were asked to cost out programs statewide (presumably in anticipation that any financing changes would spill over to districts outside New York City), the conflict of interest could hardly be more direct, unless

the panelists had been paid for their labors in proportion to the amount they recommended.

These arguments are not against professional judgment per se, but against its misuse in this case. There is a big difference between asking professional educators to make education decisions and resource allocations within the constraints of a fixed budget and asking them to determine what that budget should be. The former endeavor is what they traditionally do, exactly where the professional

judgment of an administrator might be helpful, just as it would be useful to have Martha Stewart's decorating opinion. But that opinion is solicited after a fixed budget has been set. Asking the professional educators to determine the budget only guarantees solutions that retain the basic organization of the current system, including the existing incentive structure. After all, it is a structure that the participants have accepted and to which they have grown accustomed.

Notably, the AIR/MAP approach did not consider any ways of reconfiguring the education system so as to make it more efficient. Instead, it assumed that existing arrangements were fixed and made their best guess as to how much more money that system might need to get the job done. Not surprisingly, the professionals' recommendations included such nostrums as paying employees (themselves) more and giving them less work to do (reducing class size). The notion that the city's current stable of teachers should be paid more is particularly ironic, given that much of the plaintiffs' evidence at trial was devoted to documenting their shortcomings. Moreover, research has shown that any of these steps would cost a fortune, far beyond any reasonable expectations of achieving adequate performance levels. The professional judgment panels paid such research no attention whatsoever.

Substituting Self-Interested Judgment for Data

The AIR/MAP analytic approach ignores ample evidence from New York indicating the absence of a clear connection between performance and expenditure. Take, for example, the percentage of students in a district who obtain a Regents' diploma, a key measure of education quality in New York. Districts that are higher performing by this indicator actually

check the facts

PSEUDO-SCIENCE HANUSHEK

spend, on average, no more than the lower performing districts (after adjustment for differences in family income, special-education placements, and the percentage of students who are of limited English proficiency). Thus the normal operations of districts in the state give no indication that increasing expenditure alone would necessarily enhance student achievement.

Now consider New York City itself. The judicial referees call for a 43 percent increase in spending. Between 1998 and 2003, as Figure 1 shows, expenditures in New York City increased by almost exactly that amount, 44 percent, an increase that surpassed the rate of increase for the state as a whole and for the nation. If money is the answer, this history should help foretell the results of the next infusion. But as Figures 2a through 3b demonstrate, student passing rates in reading and math for New York City students have remained barely above 50 percent—in fact, have worsened in 8th-grade reading. Whatever small gains have occurred, they hardly support the conclusion that spending increases constitute the solution to the city's inadequate schools. Perhaps these numbers led AIR/MAP to qualify their findings so dramatically as to undermine the validity of their study:

The success of schools also depends on other individuals and institutions to provide the health, intellectual stimulus, and family support on which public school systems can build. Schools cannot and do not perform their role in a vacuum. Furthermore, schools' success depends on effective allocation of resources and implementation of programs in school districts.

If more resources are not sufficient, what is the evidence that they are necessary? Are there reasons to believe that the next 40-plus-percent spending increase will have a greater impact than the last? Or should we expect the next quadrennial costing-out study to call for yet another 40-plus-percent increase in spending to meet the achievement goals?

Program designers, 56 in all, were also service providers whose pay, working conditions, and other funds were directly dependent on the resources put into the system. Such a procedure is akin to asking Martha Stewart how much you should pay for her to decorate her own house.

S&P's Successful Schools Model

The Standard & Poor's study relied on the "successful schools" method, focusing on observed costs for a set of New York districts that obtain good student outcomes. Even after allowing for the cost of educating students with special needs, S&P's analysis showed a wide dispersion across school districts in the spending observed to achieve equivalent outcomes. The lower-spending half of successful districts spent 50 percent less than the higher-spending districts, proving that many good schools do quite well with much less than other schools. Recognizing this, the Zarb Commission

went with the average expenditures of the lower-spending half of the successful districts.

The definition of success is particularly relevant to understanding the synthesis of the different approaches, since, as noted, the full S&P analysis considered a variety of possible definitions of "successful schools." The Zarb Commission relied on the set of school districts meeting the Regents' operational definition of an adequate education: 80 percent of their 4th graders passed the math and English exams and passed five of the high-school graduation tests. This definition of the objective of an adequate education was consistent with the court's decision on how to interpret the requirement of a sound basic education.

Curiously, however, AIR/MAP defined a sound basic education quite differently. It determined that a successful school district was one in which all students meet the full Regents Learning Standards, a much higher bar that moved the 80 percent pass rate to 100 percent. That measure was explicitly rejected in the New York Court of Appeals decision, which the referees were being asked to implement.

Meeting more stringent standards should clearly cost more than meeting the lesser standards. Yet the referees, by carefully selecting and modifying components of the S&P study, were pleased that they could extract similar estimates of adequate funding requirements from the various studies. They state, "This relative convergence of costing-out results derived from three different methods—the successful school district method used in the State's costing-out analysis, the professional judgment method used in plaintiffs' costing-out analysis, and the City's detailed planning method—provides comfort that our \$5.63 billion costing-out recommendation to the Court is indeed sound."

Meeting more stringent standards should clearly cost more than meeting the lesser standards. Yet the referees, by carefully selecting and modifying components of the S&P study, were pleased that they could extract similar estimates of adequate funding requirements from the various studies. They state, "This relative convergence of costing-out results derived from three different methods—the successful school district method used in the State's costing-out analysis, the professional judgment method used in plaintiffs' costing-out analysis, and the City's detailed planning method—provides comfort that our \$5.63 billion costing-out recommendation to the Court is indeed sound."

If the costing-out studies have any validity, the cost of achieving very different outcomes should not be the same.

Why Worry about Efficiency?

The most basic problem is the absence of a scientific method in the application of the costing-out models. The reasonable scientific question is, “What level of funding would be *required* to achieve a given level of student performance?” In fact, there is no evidence to suggest that the methodology used in *any* of the existing costing-out approaches, including the two considered here, is capable of answering that question.

The existing analyses never consider the minimum cost, or efficient level of spending, needed to achieve the desired outcome. Instead, they are fixated on identifying any policies that might lead to an improvement in performance, almost without regard to the magnitude of gains or cost. The focus on minimal required spending is a necessary ingredient, because without this restriction the question of cost is completely arbitrary (and thus beyond science). Actual spending to achieve an outcome can obviously range anywhere from the efficient level to infinity. But none of the available methodologies focuses on the efficient spending required for any given performance level.

Moreover, locking in the current technology (through professional judgment or successful schools) can at best produce marginal changes in outcomes. Overcoming the deficits illustrated in Figures 2a through 3b will require more dramatic improvements.

Consider again the AIR/MAP analysis. There is, first, no demonstration that the schools that employ the panel members are using their funds in a particularly effective manner or that their experiences indicate they have the data to answer the “level of funding” question. Second, there is

no way to replicate the wish lists of the specific panels, because they are based solely on personal opinions of the selected panelists and not on any data about school operations.

More important, the specific approach of AIR/MAP for combining the judgments of the separate professional judgment panels led directly to costing out the *maximum*, not minimum, recommended resource use to achieve the Regents Learning Standards. Thus, ignoring whether the choices would conceivably lead to the desired outcomes, the methodology necessarily produced a biased answer, albeit one that suited the interests of the clients.

The referees seemed unconstrained by any of this logic, however. The state, using S&P’s estimates, had suggested that it was reasonable to concentrate on the spending patterns of the most efficient of the successful schools, those that did well with lower expenditure, and thus excluded the top half of the spending distribution in its calculations. But when the referees attempted to reconcile the state’s recommendation of \$1.9 billion with the AIR/MAP estimates of more than five billion dollars, they insisted on adding in all the high-spending districts, even when such districts did not produce better academic outcomes. Thus they forced on S&P an inefficiency standard that, on its face, violates the premise of the successful schools model. After all, the referees reasoned, “there was no evidence whatsoever indicating that the higher spending districts ... were in fact inefficient.” 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?

Perhaps, however, the top-spending districts are using the money for some unmeasured reason. If so, this would only magnify the analytical problem, for if the top-spending districts are not comparable, then their spending level does

not indicate what would happen if funds were added to a typical district. It would not reflect the causal effect of added funds on student outcomes, but rather the effects of unknown underlying differences between the districts. But, again, neither AIR/MAP nor the referees made any use of historical data, so no consideration of variations in spending across districts entered their deliberations.

Furthermore, in neither the successful schools nor the professional judgment methodologies is there a sense that the results of the successful districts could be reproduced without instituting a host of reforms (unmentioned by the referees) to ensure that the extra money led to better schools. In fact, the multiplicity of high-spending/low-achievement districts would seem to indicate that money is decidedly not the measure of a good school, that the approach fails on fundamental grounds of science.

To avoid the dead end that both logic and the facts create for costing-out proponents, the referees use a clever bit of language throughout their report. They calculate the amount of annual funding required “to provide all New York City school children the *opportunity* for a sound basic education” (emphasis added). They never say that the spending they propose *will achieve* the desired results. Such a statement, or rather, such an omission, clearly suggests that the referees and Judge DeGrasse are not interested in improving student outcomes as much as they are in equalizing opportunities for inefficiency. Unfortunately, doubling the dosage of an ineffective pill seldom provides an effective cure.

Just Send the Money

The courts, of course, do not condone wasting funds. In fact, court judgments about school finance frequently contain explicit notes cautioning that the funds will lead to improvements only if they

check the facts

PSEUDO-SCIENCE HANUSHEK

are used effectively. Such tautological statements seldom recognize that New York City (and other states under judgments) have no history of spending funds effectively.

At the same time, the objective is a serious one. The education problems in New York City (and a number of other jurisdictions that face court financing challenges) are real and important. Many people would indeed be willing to put more money into New York City schools (or any poorly performing school for that matter) if they had any reason to believe that students' achievement would improve significantly.

Unfortunately, addressing these problems by simply augmenting the current system, which has virtually nonexistent performance incentives, will not solve the problems. At such a critical juncture, students and taxpayers

alike deserve an approach that embraces the best of what we already know about investments in public schooling that work. This is not ensured by any of the legal proceedings to date.

In the end, the big difficulty with the costing-out exercise is that it purports to provide something that cannot currently be provided: a scientific assessment of what spending is needed to bring about dramatic improvements in student performance. By their very nature such studies provide little information about the costs of achieving improvements efficiently. They contain nary a word about changing the reward structure for teachers (other than paying everybody more). They avoid any consideration of accountability systems based on student outcomes. And they lack any appropriate empirical basis.

Asking the courts or, more precisely, outside consultants to provide

a scientific answer to the question of how much should be spent on schools is irresponsible. Decisions on how much to spend on education are not scientific questions, and they cannot be answered with methods that effectively rule out all discussion of reforms that might make the school system more efficient.

Even the weak statement from the New York Court of Appeals that new accountability should accompany added funding was met with indifference by the judicial referees, who accepted the thrust of Mayor Bloomberg's testimony when he appeared before them: he is already accountable through the electoral system, so just send the money.

Eric A. Hanushek is a senior fellow at the Hoover Institution, Stanford University, and a member of its Koret Task Force on K-12 Education.

"Highly recommended"

—*Library Journal*

Ask for *The Common Review* at your favorite library, newsstand, or bookstore. Subscribe today to *The Common Review*, the magazine of the Great Books Foundation (\$12 US/\$20 Int.), and become part of America's most thoughtful reading community. Visit us online at www.thecommonreview.org for selected content and more, or call **1-800-222-5870** to subscribe today.

