



exploring

the

costs

of

accountability



feature

Claims that the
No Child Left Behind Act
represents an
“unfunded mandate”
wilt under close scrutiny

by **JAMES PEYSER & ROBERT COSTRELL**

How

much will the federal No Child Left Behind Act (NCLB) cost? Critics argue that NCLB's requirement that states bring all students up to academic proficiency by the year 2014 represents a massive unfunded mandate. William J. Mathis, for example, claims in a recent *Phi Delta Kappan* article that public K–12 spending needs to rise by at least 20 to 35 percent to meet the goals of NCLB—an increase of \$85 to \$150 billion a year. These critics reason that unless the feds put a lot more money on the table, financially strapped state and local governments will be forced to raise taxes sharply. Otherwise, the entire reform effort could collapse of its own weight.

The funding issue has three components. First is the cost of designing and implementing a statewide testing system. Second is the cost of establishing a state-level system for evaluating schools and districts and for intervening in those schools that continuously underperform. Third, and most controversial, is the cost of ensuring that schools have enough resources to provide the high-quality

PHOTOGRAPH BY CORBIS

**Even before
No Child
Left Behind,
most states
had already
committed
themselves to
a standards-
based reform
strategy.**

educational opportunities that students need to meet the academic standards required by NCLB.

We have examined these three components in light of our own experiences in Massachusetts. Since Massachusetts embarked on a path similar to that mandated by NCLB well before the law's adoption and has proceeded further than most other states, its experience may be useful in evaluating what is required nationally.

Our analysis suggests that many critics greatly exaggerate the shortfall of federal resources. Specifically,

- ♦ Federal spending to develop and administer mandated assessments is adequate for now, but will need to increase over time. The needed dollar amounts are relatively small and could be met easily by reallocating funds from lower-priority programs.
- ♦ Federal support of school evaluation and technical assistance, required under NCLB, is underfunded. This gap is likely to grow significantly as more schools are found to be “in need of improvement.” Much of the gap can be filled, however, by allowing states to allocate more of their federal dollars to supporting turnaround efforts in low-performing districts.
- ♦ No one—neither critics nor supporters of NCLB—really has any idea what it would cost to bring all students to proficiency by 2014 (or even 95 percent of all students, given the exceptions already built into the law), or if it can be done at all. The only question we can reasonably answer is whether there is enough money in the system to allow well-run schools to meet their goals for improvement.

From this perspective, school spending may, in some states or districts, be below what is required to steadily improve student achievement in line with federal requirements for “adequate yearly progress.” However, the extent of the shortfall appears to be a small fraction of the figures put forth by NCLB's critics. Our calculations using a school improvement approach based on data from Massachusetts suggest a national gap of perhaps \$8 billion a year, concentrated in a few states. This is only 5 to 10 percent of the critics' estimates, which are based on far more speculative and problematic models.

One of Several Mandates

In approaching these questions, it is crucial to remember that the federal government has a long-established interest in the operation of public schools and the academic achievement of students, dating back at least to the National Defense Education Act of 1958. Even before NCLB, most states had already committed themselves to a standards-based reform strategy. In some states, such as Texas and North Carolina, this was the result of internal pressures for reform. Other states were responding to the mandates of

the 1994 reauthorization of the Elementary and Secondary Education Act, the predecessor to NCLB. Under the reauthorization, each state was supposed to develop comprehensive academic standards with curriculum-based tests that would be administered annually at three grade levels, in both reading and math. The problem was that these federal requirements lacked

teeth. By the time the 1994 reauthorization was superseded by NCLB in 2002, only 21 states were in compliance with its accountability provisions. At its most basic level, NCLB attempts to fulfill the promise of earlier efforts by putting in place specific implementation timelines for those states that wish to continue to receive federal aid.

In addition, a wave of school finance lawsuits has placed half the states under court orders that effectively mandate comprehensive reforms and increased spending. The central claim in the latest, and most successful, of these lawsuits is that schools need a minimum amount of per-pupil funding in order to provide an “adequate” education. The plaintiffs in these lawsuits often draw on the same or similar reports that critics of NCLB use to make their case that a large funding shortfall exists related to NCLB's mandates.

Later in this essay we will demonstrate that these methods of determining the spending necessary for an adequate education suffer from severe shortcomings, and we will present an alternative approach that generates more credible numbers. For now the point is simply that much of what is alleged to be an NCLB mandate is either not new or actually results from states' actions. Most states have already dramatically increased their spending on education and have poured considerable resources into testing programs—changes driven by earlier federal initiatives, state-level policy, and court decisions, not NCLB. That NCLB raises the bar (and the stakes) is not in question. In assessing the burden of implementing the law, however, analysts must take care to focus on the fiscal impact of changes at the margin, instead of the total cost of education reform.

Student Assessment

Under the 1994 reauthorization, each state was supposed to put in place criterion-referenced tests to be administered annually at three grade levels, in both reading and math.

By the 2005–06 school year, NCLB requires states to implement annual standards-based math and reading assessments in grades 3 through 8, plus at least one more round in grades 10 through 12. In addition, beginning in the 2007–08 school year, states must administer annual science tests at three grade levels (once each in grades 3–5, 6–9, and 10–12). In total, NCLB will require 17 tests per year, while the previous federal law required only six.

A number of states, however, had already committed themselves as a matter of law or policy to roll out many of these tests, with or without a federal mandate. Massachusetts, for example, has a statutory requirement to develop student assessments at three grade levels in five subject areas (English, math, history, science, and foreign languages). In addition, the commonwealth's board of education determined as a matter of policy to implement a somewhat broader assessment program, adding an extra year of testing in both English and math in 2001. Other states, such as California, Colorado, Florida, Iowa, North Carolina, and Texas, adopted assessment systems that exceeded the old federal requirement. Five states met the full NCLB testing mandate before it even went into effect.

A survey by the Government Accounting Office (GAO) reports that, on average, states will have to add eight or nine tests to their existing programs as a result of NCLB. This number probably overstates the added burden, however, since it includes tests that were required under the previous reauthorization but not implemented by the states. It is reasonable to assume that NCLB itself will actually account for fewer than half of the 17 tests it eventually requires.

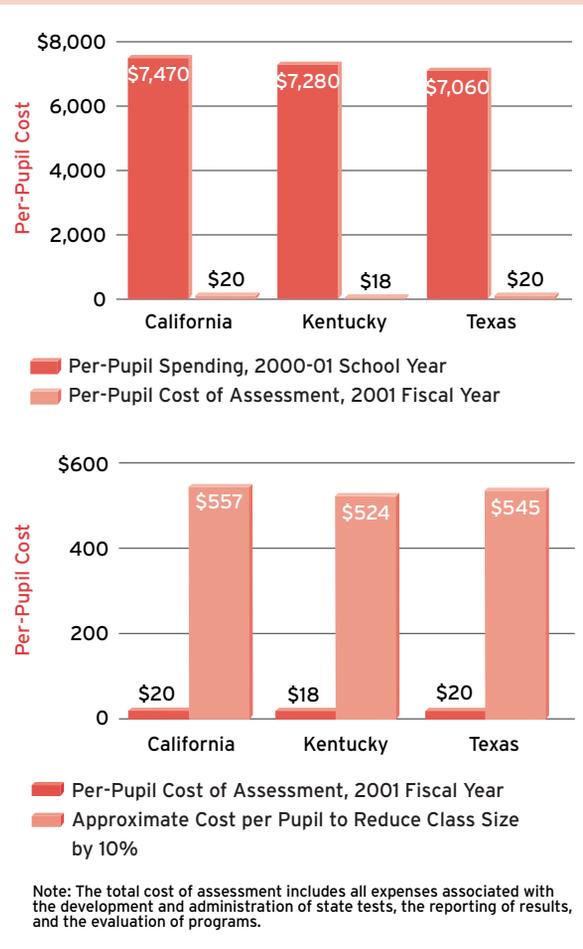
When fully implemented, the law's reading and math testing provisions will require states to administer approximately 45 to 50 million tests a year. The GAO's report estimates that the total cost to states of developing, administering, scoring, and reporting tests of the type currently administered would have been \$442 million in fiscal year 2003, or roughly \$9 per student tested.

Based on cost data from Massachusetts, we believe that the actual costs for tests of the quality necessary for an effective accountability program may run as much as twice the GAO's estimate. However, many states do not have all these tests in place for this school year. Moreover, many states have laws and policies requiring administration of such tests—with or without NCLB. All in all, it looks as if the incremental cost for implementing the new English and math tests soon to be required is about \$400 million.

In short, the Department of Education's current level of appropriation for state assessments—\$391 million—appears to be enough to cover the full marginal cost of federally mandated tests that states actually administered. This level of funding may not be sufficient going forward, however, as

Putting the Costs of Accountability in Perspective

In states like California, Kentucky, and Texas, whose accountability systems are among the most elaborate and expensive in the nation, the cost of testing each student amounts to less than 0.3 percent of per-pupil spending. Moreover, accountability programs are just a fraction of the cost of other popular school reforms, such as reducing class size.



SOURCE: : Caroline M. Hoxby, "The Cost of Accountability," in *School Accountability* (Hoover Institution Press, 2002)

more states roll out the full complement of required tests and enrich the tests with more open-response questions.

School Evaluation and Intervention

Besides measuring and tracking student achievement, NCLB requires districts and states to respond when schools

The professional judgment model assumes that educators already know, instinctively, what resources are necessary to meet higher standards.

repeatedly fail to meet improvement expectations. Such intervention was also required under the previous federal law, but without the explicit requirements and timetables of NCLB. As a result, little outside intervention in underperforming schools has occurred.

Last year, under the NCLB rules, more than 8,600 schools around the country were listed as “in need of improvement” by state departments of education—less than 10 percent of all public schools in the country. As the “adequate yearly progress” aspect of the law results in increasingly heightened performance expectations, this number will probably rise, too, even though many schools will “graduate” off the list due to improving (or at least fluctuating) test scores.

Under NCLB rules, schools that fail to meet their adequate yearly progress targets are subject to an escalating series of interventions. The mandated roles for state departments of education in this process are scorekeeper—determining which schools and districts are missing their targets—and provider of technical assistance. Many states have enacted legislation that provides considerably greater authority for intervention, up to and including takeover powers in cases of persistent underperformance.

NCLB requires states to set aside about \$230 million of their federal funds for grants to schools in need of improvement. Individual grants are supposed to be at least \$50,000, with a ceiling of \$500,000. Given 8,600 underachieving schools nationwide, however, the \$230 million set-aside can support an average grant amount of only about \$25,000, assuming each qualifying school receives a grant. As a practical matter, grants of this size will likely be focused on planning and professional development for staff.

The number of schools needing improvement under NCLB may increase substantially. In fact, this seems likely to occur once the requirement that all subgroups of students within a school make adequate yearly progress comes into effect. In this case, much more money will have to be appropriated or reallocated to school improvement grants in order to fund the minimum mandated interventions. If, for example, one-third of all schools found themselves “in need of improvement,” then the minimum amount of federal support required to fund grants of \$50,000 per school would be \$1.6 billion.

Although the specific allocation for school improvement grants appears to fall well short of the minimum amount required by federal regulations, other sources of federal funds could more than close the gap, if they were directed

to low-performing schools. For example, grants to states for “innovative programs” total more than \$380 million. These resources, however, are allocated to districts on the basis of an enrollment-driven formula. Grants for improving teacher quality are much larger, exceeding \$2.9 billion in 2003. Once again, these funds are allocated to districts by formula. Both of these grant programs provide relatively more money to districts with higher proportions of low-income students, which also tend to be lower-performing districts. Nevertheless, a more flexible allocation method would free up considerable resources to support the specific school improvement activities mandated under NCLB.

Rarely mentioned in the discussion of intervention programs is the cost of developing and sustaining a state evaluation infrastructure capable of conducting in-depth analysis and diagnosis of struggling schools and districts. Even with just 8,600 schools in need of improvement, it is not possible for state education agencies to develop and implement effective turnaround strategies, no matter how much money is available. Some form of triage must be used to identify those schools and districts most in need of help and least able to help themselves. Scanning the test results is not enough to determine which schools and districts require acute care and to prescribe the appropriate treatment. States must conduct in-depth evaluations of the facts on the ground, using a reliable protocol and experienced educators.

Massachusetts has been pioneering such a process for the past several years. To date, 18 schools have been declared underperforming, and each one has been required to develop an improvement plan, in collaboration with the state department of education, based on the findings of a diagnostic evaluation. A similar district-level procedure recently resulted in two districts’ being identified as underperforming. The annual budget for this in-depth evaluation work is \$2.4 million, and it is still in its launch phase. When fully implemented, the budget could easily double. Virtually none of these costs is covered by federal grants. If a similar evaluation infrastructure were put in place in every state, the total cost might reach \$250 million a year.

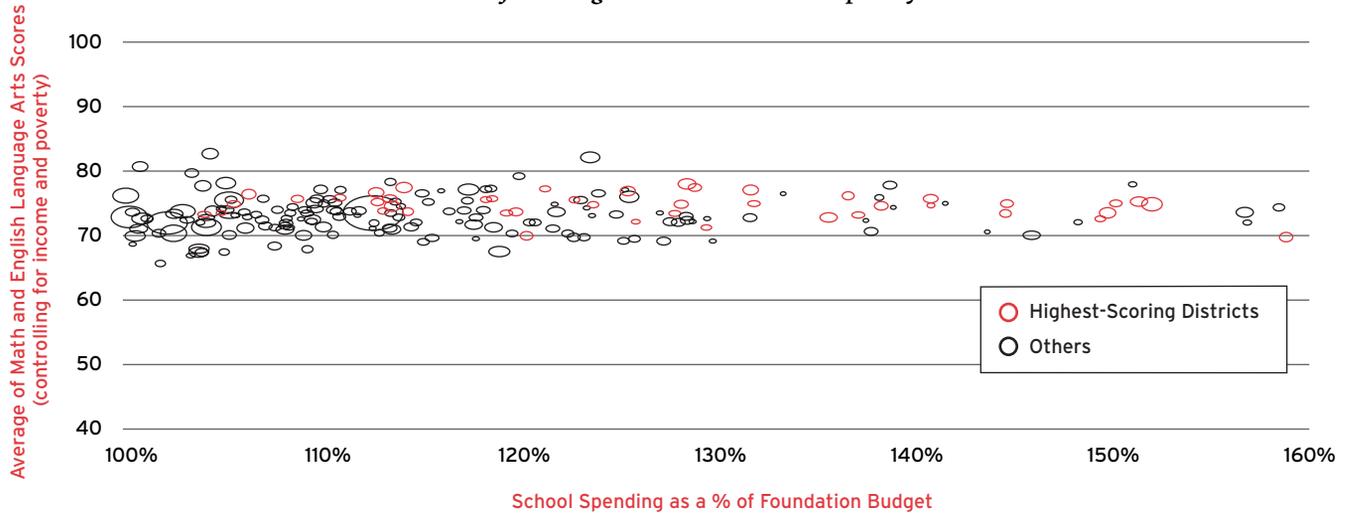
The Cost of an “Adequate” Education

These discrete funding issues are important, but they pale in comparison with the claim that schools need to spend at least 20 to 35 percent more (\$85 to \$150 billion) to meet NCLB’s performance goals. The source of this claim is a series of recent consultant reports commissioned by teacher unions, school board associations, legislative bodies, and others, often for use in school finance cases.

A fundamental problem with these reports is their tendency to rely on the outdated notion that education can be reduced to a simple production function between input and

Spending vs. Performance (Figure 1)

No relationship between school districts' spending levels and their performance on standardized tests was observed in Massachusetts after taking into account income and poverty levels.



Note: Size of bubble is proportionate to district enrollment. Data are for all K-12 districts in Massachusetts during the 2001-02 school year. Test scores are adjusted based on a regression of the average of math and English language arts performance on median income, the percentage of students eligible for a reduced-price lunch, and spending.

SOURCE: Authors

output. Of course, the amount of spending can matter a great deal if it is raised from very low levels. But over the range of spending commonly observed among school systems in the United States, the effect on student achievement is often swamped by how wisely the money is spent, by bureaucratic and contract rigidities, and by a host of important policies and decisions that have nothing at all to do with money. The fact is that most research finds, after controlling for demographic factors, no consistent causal relationship between expenditures and achievement over the current range of spending levels.

Consider, for example, data from the school finance case currently being litigated in Massachusetts. The plaintiffs point out that high-performing districts often spend considerably in excess of the foundation budget, the state's measure of what is necessary to provide an adequate education. But the association between performance on state tests and spending as a percentage of the foundation budget—the plaintiffs' preferred measure of spending—vanishes after applying even the most rudimentary demographic controls (see Figure 1).

Nonetheless, the claims that schools are underfunded rest on models that purport to quantify the level of expenditure necessary to meet higher performance standards. Two approaches are used most frequently: the "professional judgment" and "successful schools" models. The professional judgment approach asks educators to build their ideal school budget from the bottom up, by answering questions such as:

What is the optimal class size? How many teacher aides, computers, and professional development days should there be? Their instructions typically encourage them to "be creative and innovative," to create new programs or services, and to assume there are no revenue constraints. By contrast, the successful schools approach uses observed spending levels in the highest-performing schools as models from which to calculate necessary spending in other, lower-performing schools.

There are several methodological problems with the typical implementation of both of these approaches. The principal problem with the professional judgment model—the model drawn on most heavily by Mathis—is that there is no attempt to tie observed spending levels to actual student outcomes. It assumes that educators already know, instinctively or by dint of personal experience, what resources are necessary to meet higher standards, so the analysis of data is considered superfluous. This raises questions of possible subjectivity.

Take the Massachusetts case, for example. The plaintiffs' professional judgment study, performed by University of Virginia professor Deborah Versteegen, relied on employees of the school systems where the plaintiff students are enrolled to determine the ideal level of school spending. Versteegen's panelists included the superintendents of seven plaintiff districts. As the judge pointed out, one of the panelists was the mother of the named plaintiff in the original 1993 case. The study's findings implied that almost every district in the state—even the wealthiest—was underfunded,

If the increase in federal spending does not fully cover the fiscal gap, it comes pretty close.

with an average shortfall of 66 percent. Ironically, the only sizable district judged to be spending enough was Cambridge, where student performance has been persistently low.

While in the past there may have been few alternatives to the professional judgment model's input-based approach, today it seems out of step with the spirit of standards-based reform. The successful schools approach, to its credit, does focus on measured performance in a standards-based system. However, in practice, the approach is typically flawed by its method for selecting high-performing schools. As has been long established, family background is the strongest predictor of academic success. Children from wealthier, better-educated families also tend to live in communities where property-tax revenues and school budgets are high. The successful schools model typically assumes that a school's high test scores are primarily a function of its budget, rather than a student's family background. As a result, high-spending schools may be identified as "successful," not because they add more educational value, but because they enroll children from high-income families. The red bubbles in Figure 1, which indicate the districts identified as successful by the plaintiffs in the Massachusetts litigation, illustrate a concrete example.

As typically applied, the successful schools approach also has a number of technical flaws. Most important is its use of *average* spending among high-performing schools as the *minimum* level necessary for fiscal adequacy. By using this method, half or more of the successful schools themselves are inevitably found to have insufficient funding to be successful. While it may be true in Lake Wobegon that all districts can be at or above average, the laws of mathematics rule that out for the rest of us.

In Massachusetts, the plaintiffs commissioned a successful schools study by John Myers. Myers selected the top 75 districts and estimated their average level of per-pupil expenditures for regular education students. Because the selection criterion was the level of state test scores (with no attempt to gauge value added), the 75 districts were disproportionately high income—with incomes 50 percent above the state average—and the share of their students who were eligible for free or reduced-price lunches was only 4 percent, compared with 24 percent statewide. Myers' practice of using the average spending from this high-spending group (as well as some other flawed procedures) implied that 90 percent of Massachusetts students are in under-funded districts, including 70 percent of those in the "successful" dis-

tricts. On average, Massachusetts spending was judged to be 20 percent too low.

The studies used to claim that NCLB requires a 20 to 35 percent increase in school funding are similar to those brought to court in Massachusetts. Many are by the same author. Lacking scientific grounding, they are too flawed and too sensitive to changes in assumptions to sustain a compelling critique.

Improvement Model

Although any approach to calculating fiscal adequacy is bound to have limitations and flaws, state laws and courts increasingly require that some rational method be developed to approximate the price of an adequate education. Given that mandate, we suggest an "improvement" version of the successful schools approach, based at least in part on the rate of growth in measured academic achievement rather than a simple snapshot of average performance. This is not a full-blown value-added model, which would track gains in individual student performance from one year to the next. But it at least roughly controls for a district's previous performance, rather than effectively selecting districts with favorable demographics.

The improvement approach would first identify a set of K–12 districts that have realized the greatest gains in measured student achievement over a multi-year period. The next step is to identify a spending level somewhat lower than the group average to capture the minimum necessary expenditure as demonstrated by successful districts. In identifying that level, one should keep in mind that some of the lowest spending districts may be outliers whose success would be difficult to replicate. In general, one to two standard deviations below the average among improving districts might be a reasonable benchmark.

Let's again use Massachusetts as an example to see how this might work in practice. With the state's index for determining overall progress across grades and subjects as the criterion, 183 of the state's 207 K–12 districts met their adequate yearly progress targets in 2001–02. Among these districts, 114 also achieved a performance rating of at least "moderate" in both English and math. Of these districts, 46 had improvement rates that were above average. These two sets of improving districts (114 and 46) are still above average in terms of income, but less so than the set of K–12 districts selected by Myers solely on the basis of their performance levels. Incomes in the groups of 114 and 46 districts are 26 percent and 8 percent, respectively, above the state average, while Myers' districts were 50 percent above the average.

The average per-pupil spending figures for these two sets of improving districts were \$7,499 and \$7,105, respectively, in 2001–02 (compared with \$7,840 for all K–12 dis-

tricts). One standard deviation below the mean of the improving districts was \$6,320 and \$6,354. This would imply that the necessary spending level for adequate progress is approximately 80 percent of the K-12 average spending level in Massachusetts. By contrast, the necessary spending levels implied by the plaintiff models discussed above, which are similar to the models used by NCLB's critics, were well above the average spending level.

Whether districts can continue to meet their adequate yearly progress targets with the current level of spending, right on through NCLB's ambitious goals for 2014, remains to be seen. It will demand close monitoring at both the state and federal levels. But for now, it would appear that spending in Massachusetts is adequate to achieve the NCLB student achievement mandate.

The National Picture

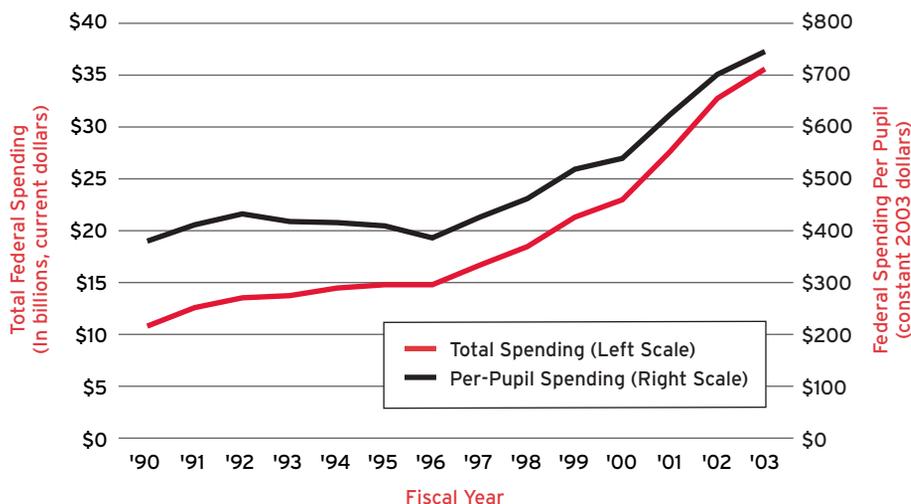
To illustrate what this analysis might mean nationally, consider *Education Week's* calculation of spending per student, adjusted (albeit imperfectly) for regional cost differences. By *Education Week's* measure, only 11 states have average spending levels below our benchmark for adequacy (in other words, 80 percent of Massachusetts' per-pupil average). For each of these states, a total fiscal gap can be estimated by multiplying the per-pupil gap by enrollment. Adding these up for the 11 states results in an estimated national fiscal gap of approximately \$8 billion (almost half of which is in California). While this \$8 billion estimate is not trivial, it is only 5 to 10 percent of the projections claimed by the law's critics.

To be sure, this estimate carries with it certain caveats. For example, if a state's average per pupil spending exceeds the adequacy measure, but some of its districts do not, the estimated fiscal gap of zero for that state assumes it will redistribute some of its spending. Progress in Massachusetts is also no doubt attributable in part to the state's strong system of student accountability, including a universal graduation requirement pegged to the 10th grade statewide test—a provision missing from the NCLB mandate. Other states that spend at levels similar to those in Massachusetts, but without student accountability, may fall short of NCLB goals.

Although NCLB's critics charge that Washington has not

Increasing Its Share (Figure 2)

Federal aid to K-12 education has continued to grow since the passage of the No Child Left Behind Act.



Note: Spending data include all U.S. Department of Education spending for K-12 education.

SOURCE: U.S. Department of Education

been doing its fair share, especially when compared with the original spending plan authorized by the act, total federal Department of Education appropriations for K-12 education grew by more than 25 percent—almost \$8 billion—from 2001 to 2003, despite a sharp drop in federal revenues (see Figure 2). Title I appropriations alone grew by more than \$2.9 billion over the same period—a 33 percent increase—and as of this writing are set to grow another \$0.7 billion in fiscal year 2004, for a total of a 41 percent increase since NCLB's enactment.

If this spending increase does not fully cover the fiscal gap, it would appear to come pretty close—especially when combined with state-level spending increases already required under various state laws and court decisions. Given that many states have been slow to implement the statewide assessment and accountability systems required by NCLB, one might even argue that in some instances federal spending growth has overshot the target.

—James Peyser is chairman of the Massachusetts Board of Education. Robert Costrell is a professor of economics at the University of Massachusetts at Amherst (on leave) and currently serves as chief economist in the Massachusetts Executive Office for Administration and Finance. Peyser is a named defendant in the Massachusetts school finance case, and both Peyser and Costrell testified for the defense. An unabridged version of this article is available at www.educationnext.org.