

# Strategies For Improving Student Achievement

by David W. Grissmer

Researchers attempting to assess the effectiveness of different educational strategies have had a rich database from which to draw. Since the early 1980s individual states have leveraged the reform powers found in funding formulas to create varying policies that influence who teaches and what is taught. Additionally, state courts have played a role in deciding whether educational funds are adequate and fairly distributed. This combination of factors has created a widely diverse set of state educational systems.

If research and evaluation can identify the successful and unsuccessful approaches in this variety of systems, they can provide valuable information for states to use in the ongoing process of refining and adapting successful policies. Evaluating the effects of different levels and uses of resources and changing state policies, then becomes critical to improving schools and student outcomes.

## Assessing the Effect of Resources

The question of whether additional educational resources affect educational outcomes has not been definitively answered through empirical nonexperimental research. Experimental research, in combination with new reviews and interpretations of the empirical literature, is pointing to a hypothesis that additional resources primarily affect disadvantaged students. Because of wide state variances in the proportions of disadvantaged students and per-pupil expenditures, an analysis of state achievement scores can help test this hypothesis.

Since resources are spent differently across states, estimates of the effectiveness of the different uses can be made. More importantly, the different ways in which resources are used can provide measures of both the marginal cost and marginal achievement benefit of changing resource usage, allowing cost-effectiveness comparisons. These measures can help answer two important questions:

- What uses of resources are most cost-effective in boosting student achievement?
- To what extent do resources affect achievement for disadvantaged students?

Until 1990, when the Department of Education (DOE) began to use the National Assessment of Educational Progress (NAEP) test, no test gave representative samples of students in each state the same test, and achievement could not be validly compared across states. The DOE used the NAEP to test representative samples of students in participating states, testing them in reading and math at the 4th- and 8th-grade levels for seven years from 1990 to 1996. It is probably too early to use this period as a definitive test of whether reforms are successful, since reform initiatives are expected to take years to be fully reflected in achievement outcomes. Evidence of no achievement gains, however, would certainly challenge current reform directions.

The RAND report (Grissmer et. al, 2000) uses data from the NAEP to estimate score gains both nationally and by state, to estimate the effects of varying levels and uses of per-pupil expenditures on student achievement, and to estimate the cost-effectiveness of the major alternatives for utilizing educational resources.

## RAND Study Objectives

This study had several specific objectives:

- Compare raw achievement scores across states and determine which states have statistically significant improvements, taking account of all NAEP tests between 1990 and 1996.
- Estimate NAEP scores for students with similar family characteristics across states to develop a better measure for the overall effects of educational policies and environments.
- Determine whether trends and differences in scores across states for students from similar family backgrounds can be statistically linked to differences in state educational system characteristics that are resource intensive. (These characteristics include per-pupil expenditures, pupil-teacher ratios, public prekindergarten participation rates, teacher-reported adequacy of resources for teaching, teacher salary levels, teacher education, and teacher experience.)
- Determine whether significant trends exist (unaccounted for by these resource-intensive variables) that might suggest effects from unobserved variables linked to reform efforts.
- Estimate the costs of changing these resource-intensive policies and characteristics, and compare their cost-effectiveness in improving scores.
- Propose a broader explanation for the pattern of achievement results (in the NAEP study and in the empirical literature) that incorporates new experimental class-size results and the historical pattern of spending and achievement in the nation.

Given the RAND results, the study authors propose a broader explanation of the effectiveness of resources in the public school system as follows:

*Additional resources provided to public schools mainly affect minority and less-advantaged students; these effects can be large and significant if properly allocated and targeted. Additional resources deployed in historical ways have had much less, if any, effect on more-advantaged students.*

## Methodology

Comparative state analysis became possible when the DOE gave the NAEP tests to representative samples of students across a voluntary sample of states in 1990, 1992, 1994, and 1996. Seven tests were given in reading and mathematics at either the 4th- or 8th-grade level; each test was administered to approximately 2,500 students in 44 states.

### Barriers to Analysis

Although these tests represented the first valid, comparable measures of achievement for representative samples of children in various states, there remained significant barriers to carrying out analysis and obtaining the kind of reliable results policymakers need.

1. Because of the wide variation in state demographic composition and family characteristics, previous research suggests that family variables could account for a substantial part of the variation of scores across states. Family variables collected by NAEP were limited; those collected were reported by 4th- and 8th-graders, making their quality problematic.
2. The sample was small. State scores lacked independence across tests, and states participated in an unequal number of tests.
3. The credibility of results derived from models aggregated across states is at issue. Previous studies using state-level data have shown that educational resources have consistent positive, statistically significant effects on educational outcomes, dissimilar from the generally null effects found at low levels of aggregation.
4. Models using nonexperimental data are deemed more credible if they agree with results using experimental data.

*... these tests represented the first valid, comparable measures of achievement for representative samples of children in various states...*

### Addressing the Barriers to Analysis

Instead of relying on NAEP-reported family variables, RAND used Census data and data from the National Educational Longitudinal Survey (the largest survey that collected both achievement scores and parent-reported family characteristics) to develop three sets of family variables that use different sources of data and methods of weighting the influence of family characteristics.

A Tennessee Student-Teacher Achievement Ratio (STAR) class size experiment showed that reducing class size in K-3 had positive and statistically significant effects through 8th grade; a more recent quasi-experiment in Wisconsin showed initial results similar to Tennessee's. This analysis used a model specification consistent with the results from the Tennessee class-size

experiment, compared the results to the experimentally determined results from Tennessee, and showed agreement.

## Main Findings

### Highlights

The results paint a more positive picture of public education in America than is commonly portrayed, especially with respect to effective allocation of resources. Following are some highlights:

- Public elementary students across states in this sample showed statistically significant gains (about 1 percentile point) in mathematics between 1990 and 1996 (The reading data are insufficient for analysis until the 1998 state NAEP reading data are included.)
- There is a disparity in progress made by states: The math gains across states showed that a few made gains of around 2 percentile points a year, while other had almost no gains.
- The highest average achievement scores were found in the more-rural northern states; southern states were usually among the lowest. The more-urban northern states generally fell in the middle of the score distribution. This distribution is explained primarily by family rather than school characteristics.
- Statistically significant differences—as large as 11 to 12 percentile points—were found among students with similar family characteristics across states, with all regions of the country having states with both higher and lower student scores from similar families.
- Both the level of expenditure per pupil and allocation affected student achievement, particularly for states with disproportionately higher numbers of minority and less-advantaged students.
- Some educational expenditures were much more cost-effective, with the difference depending on how the expenditures were directed. Cost-effectiveness also varied markedly depending on the SES level of the state, the current allocation of expenditures, and the grades targeted.

### Evidence for the Effects of Reform

*Math scores increased from 1990 through 1996 in most states for public school students. Reform efforts are the leading candidates to explain the gains.*

Controlling for population changes and participation rates, this analysis provides strong evidence that math scores from 1990 through 1996 increased in most states for public school students by statistically significant amounts. Small changes in resource-intensive variables during this period do not explain this improvement, suggesting reform efforts as the leading candidates to explain the gains.

Additional research is necessary, however, to adequately test whether and which reform efforts are linked to achievement gains.

States varied in their estimated math gains, with some gaining 2 percentile points and others showing little gain. Texas and North Carolina were among the states making large, statistically significant gains; state administered tests during this period also showed large gains. Resource-intensive variables included in the analysis did not explain much of these gains over time. Thus, reform efforts emerge as the leading candidates.

Trends in reading scores cannot be assessed with the current data, with only two reading tests given two years apart currently available.

### **Scores for Students from Similar Backgrounds**

Scores of students with similar family and demographic characteristics varied as much as 12 percentile points. This analysis distinguished three groups of states: those whose scores for students from similar families are significantly above the median state, those whose scores are below, and a broad middle group. Statistically significant differences for students with similar family characteristics are found in adjoining and other states.

These score differences can be traced, in part, to several systemic features:

- lower pupil-teacher ratios
- higher public prekindergarten participation
- lower teacher turnover
- higher levels of teacher-reported adequacy of resources for teaching

Scores for students from similar families placed Texas in the highest group of states and California in the lowest. Contributing to the higher Texas scores are lower pupil-teacher ratios, a larger percentage of children in public pre-kindergarten, and teachers with greater resources. Using these measures as a guide, an analysis can reveal what creates the differences.

### **Effects and Cost-effectiveness of Educational Resource Allocation**

NAEP scores are higher in states that have:

- higher per-pupil expenditures
- lower pupil-teacher ratios in lower grades
- higher levels of teacher-reported adequacy of resources for teaching
- higher public prekindergarten participation
- lower teacher turnover

States with higher teacher salaries or greater percentage of teachers with master's degrees did not have higher scores. Further research is needed to identify the reason for a lack of effect from direct investment in salaries. Possible explanations include:

- Interstate differences in salary may be less sensitive to student achievement than are intrastate salary differences.

- Teacher salary is a variable that correlates highly with family SES variables; it may be difficult to separate salary and social-capital effects.
- These measurements occurred during a period of adequate teacher supply, and lower salary sensitivity is expected when supply is more readily available. Labor market conditions are changing markedly, however, because of demand increases due to retirements, lower class sizes and attrition rates.
- The results could reflect the inefficient structure of the current teacher-compensation system that rewards experience and education, neither of which is strongly related to producing higher achievement. If higher compensation could be provided to higher-quality teachers and those who are effective with lower-scoring students, then one could expect compensation to be more effective.

An examination of the effects of factors that influence achievement must take into account the type of students targeted and current program funding. Lowering pupil-teacher ratios in states with high SES levels that already have ratios below the national average appears to have little effect. Conversely, lowering pupil-teacher ratios for students in lower grades in states with low SES and higher than average ratios has large predicted effects. Prekindergarten has stronger effects in states with lower SES. The adequacy of teacher resources, however, appears to have significant effects regardless of family characteristics.

The cost-effectiveness of resource expenditures could change by more than a factor of 25, depending on the program or policy, which types of students are grades are targeted, and the current program levels. This analysis predicted the most cost-effective policies to be:

- Provide teachers with greater discretionary resources in all states.
- Lower pupil-teacher ratios in the lower grades to below the national average, expand public prekindergarten, and provide additional teaching resources in states with a disproportionate percentage of lower-SES students.
- Lower pupil-teacher ratios in the lower grades to equal the national average in states with average SES characteristics.

This analysis also estimates that the use of in-classroom teacher aides is far less cost-effective than the above recommendations.

*In summary, investing in better working conditions to make teachers more productive can produce significant gains in achievement. Although increasing the quality of teachers is important in the long run, this analysis suggests that significant productivity gains can be achieved now with the current teaching force if working conditions are improved.*

## **The Bigger Picture: Understanding Effects of Investment in Public Schools**

Any general theory about the effects of public-school expenditures must account for the following:

- The pattern of results in previous nonexperimental measurements
- The results of the Tennessee experiment and the Wisconsin quasi-experiment
- The pattern of national score gains and expenditure growth from 1970 through 1996

One frequently advanced explanation holding that public schools lack a consistent ability to utilize additional resources to improve outcomes depends on the inconsistency in nonexperimental measurements at levels of aggregation below the state level. This explanation assumes that the inconsistency in measurements is a result of inconsistency in the utilization of resources, but overlooks the possibility of inconsistency in the measurement process itself. This explanation is not consistent with the experimental results from Tennessee and Wisconsin,—where the 1970s and 1980s brought large score gains for minority and disadvantaged students—and with positive and consistent nonexperimental results at the state level of aggregation.

RAND researchers propose a different explanation—consistent with the current experimental and nonexperimental evidence, and historical expenditure and achievement trends—suggesting that additional resources are effective for minority and disadvantaged students, but that resources directed toward more-advantaged students have only small, if any, effects. This is consistent with the pattern of national score gains and expenditures from 1970 through 1996: Minority and lower-SES white students made significant gains, but more-advantaged students made much smaller, if any, gains.

The Tennessee experiment and Wisconsin quasi-experiment results show positive, statistically significant long-term effects on achievement, but were based on samples that were disproportionately drawn from minority and disadvantaged student populations. RAND's state-level results also produced estimates for pupil-teacher ratio consistent with the size of effects measured in the Tennessee experiment, and produced a similar pattern of larger effects for minority and lower-SES students, suggesting that aggregate-level measurements may provide more unbiased effects than less-aggregate models.

This analysis does not account for the lower, and inconsistent, pattern of previous measurements at levels of aggregation below the state level. Most independent literature reviews conclude that previous nonexperimental results show the effects of additional resources on educational outcomes as generally positive. These reviews, however, have not yet explained the wide variance in previous results, nor why more-aggregate measurements show more positive and consistent effects than measures at lower levels of aggregation. RAND researchers hypothesize that the inconsistency reflects the measurement process itself rather than an inconsistency in the use of resources.

Inconsistencies in previous measurements used may be accounted for by widely different specifications and assumptions. Previous measurements did not measure separate effects for high- and low-SES students, and most measurements contained typical student populations with large proportions of more-advantaged students. Smaller effects might be expected in such samples, and effects would be "inconsistent" across studies if student characteristics changed. Effects could also differ across grade levels and lead to "inconsistent" results across studies that focus on measuring different grade levels.

## Implications for Policy: Improving American Education

As noted, one interpretation of the empirical evidence implies that additional resources for public education are not the answer to improving schools if there remains an absence of fundamental reforms in incentives and organizational culture. Underlying this view is the idea that it is necessary to create either alternatives outside the current system or increased choice within the system to foster greater competition for public schools.

RAND's results show that resources can make significant differences for minority and lower-SES students. Between-state differences in resources are the main reason for inequitable resource levels for these students, and can only be addressed with federal programs. Results also suggest, however, that significant gains are occurring in math scores across many states—gains that cannot be traced to changing resources.

Much research is required to attribute these gains to specific reforms, but a plausible explanation suggests that ongoing systemic structural reform within public education might be responsible, certainly challenging the traditional view of public education as "unreformable."

Significant reform may be achieved in public education if the output of its separate and diverse units can be measured and compared, leading to the identification and diffusion of successful initiatives. Caution is warranted, however, until student gains in elementary schools result in longer-term gains in secondary schools, leading to completion of more years of education and greater success in the labor market.

*RAND's results show that resources can make significant differences for minority and lower-SES students. Between-state differences in resources are the main reason for inequitable resource levels for these students, and can only be addressed with federal programs. Results also suggest, however, that significant gains are occurring in math scores across many states—gains that cannot be traced to changing resources.*



There are reasons to believe that improvements in achievement will continue:

- The full effect of structural reform initiatives is not reflected in current achievement
- The identification of successful initiatives may result in diffusion across states
- Better allocation of future resources can also raise achievement

## Implications for Research

### Experimentation and Improving Nonexperimental Analysis

Expanded experimentation in education is critical to understanding educational processes and helping to determine the application of appropriate assumptions and specifications to nonexperimental data. Experimentation should be directed toward *measuring the effects* of major resource variables and *the critical assumptions* used in nonexperimental analysis. In addition, both experimental and nonexperimental research must seek an understanding of how resources impact both student development and what occurs in the classroom. It is unlikely that research consensus will emerge until we can answer some critical questions:

- What causes differences in experimental and nonexperimental measurements and the differences among nonexperimental measurements?
- What theories explain how changing resource levels affect parent, teacher and student behavior in the classroom and families?
- How do these changes affect long-term student development in ways that result in higher long-term achievement?

Two hypotheses that arose from the RAND analysis also need much more study.

### Hypothesis #1: The Dynamic Effect of Schooling Variables

The first is the dynamic nature of achievement effects across grades suggested in the Tennessee experiment. Schooling variables in one grade appear to influence achievement at *all* later grades, so conditions during all previous years of schooling need to be specified. Pretest scores may not adequately control for previous schooling characteristics. *The Tennessee experiment results suggest that two students can have similar pretest scores and similar schooling conditions during a grade and still emerge with different posttest grades that have been influenced by different earlier schooling conditions.*

For example, despite having similar schooling conditions in grades 4 through 8, relative changes in achievement occurred in grades 4 through 8 for those students who had one to two, versus three to four, years in small K-3 class sizes. Whether or not a smaller class size in 2nd grade had an effect cannot be known until later grades, and even then the answer will depend on what class sizes were experienced in both previous and higher grades.

Conceptually, the effect of class-size reductions resembles a human "capital" input that can change outputs over all future periods. Thus, models that specify the effects of capital investments may be more appropriate. These results are consistent with the concepts of risk and resiliency in children from the standpoint of child development: Different levels of risk and resiliency in children appear to interact with schooling conditions to produce gains or losses.

## **Hypothesis #2: Resource Substitutions Affect Achievement and Measurement**

A second key hypothesis underlying the RAND analysis is that resource substitutions can affect student achievement. High family resources can substitute for and supplement school resources in indirect and unmeasured ways that affect accurate measurement of policy variables. Families who are able may apply more of their own resources when school resources are lower, and less when schools are devoting more resources. Students with lower levels of family resources may be affected more by changing school resources, and show the most sensitivity to levels of school resources. Taken at face value this would imply that more school resources can substitute for lower family resources; these substitutions need to be the focus of more research.

## **Assumptions and Caveats for Interpreting the Study Results**

### **Achievement is only one of many desirable outcomes expected from schools.**

Test scores will continue to receive a disproportionate share of attention until other comparable measures of outcomes are available. It is possible to overemphasize achievement at the expense of other outcomes; it is also possible to have good schools that satisfy parents even though they are not among the highest achieving. While achievement is certainly a very important outcome expected of schools (and we should continue to try to understand the policies that contribute cost-effectively to increasing achievement), we must also begin collecting a broader range of measures of school outcomes to achieve balance.

### **No test is a perfect indicator of what students have learned.**

Achievement scores reflect particular test items that can emphasize more basic skills than critical-thinking skills. Further, scores can reflect the timing of when students learn skills: Students in different states do not learn certain skills in the same sequence or at the same grade level because of differences in curricula. Finally, different state standards and assessment systems may not be aligned with NAEP test items; states having systems that reflect NAEP might be expected to score higher.

### **Measured effects should be seen primarily as long-term effects of differences in policies.**

States will not see the full effects measured in this analysis in the first few years. State differences have existed over long periods of time, allowing students, teachers, parents and curricula to make longer-term adjustments.

### **A variety of factors are reflected in the estimated differences in scores for students from similar families.**

Several factors related to characteristics of the state education system have been identified, and account for part of the differences. Less than one-half of the differences are accounted for, however, with the remaining variance arising from:

- unmeasured family characteristics
- unmeasured characteristics of the educational system
- characteristics of other social support systems for families and children
- particular factors (such as foundations) creating social capital in states

**Effects and rankings all have ranges of uncertainty.**

Use of these results for policy guidance must take into account the ranges of uncertainty associated with the effects and rankings. The effectiveness of certain policies may hide the presence of context-sensitive factors that make the policy more or less effective. Further, the particular predicted effects may vary within state or local contexts.

**These results identify effective policies and states where students from *similar backgrounds* are performing at *different levels*.**

This is a first step toward identifying policies and practices that contribute to higher achievement, and toward understanding constraints upon broader implementation.

**The tendency to blame or credit policymakers for achievement results must be tempered by three factors.**

1. Achievement results from 1990-96 can reflect policies and practices from the early 1980s through 1996. Eighth-graders tested in 1990 entered school in 1992; their scores reflect the quality of education throughout their schooling. Fourth-graders tested in 1996 have scores that reflect more-recent policies.
2. Many reforms initiated since the mid-1980s require significant organizational adjustment; their effect on schools, teachers and students occurs gradually and is not necessarily reflected in current scores.
3. The research and development community in education has been unable to provide consensus results or pilot-tested policies and procedures to guide policymakers and educators in adopting more effective practices. Without good research and development, policymakers lack the key process required to improve the system of education; progress in education reform will thus be slow, uncertain and inefficient.

**Final Note: The Importance of Linking Educational Reform to Social Services for At-risk Students**

Policy decisions need to include a broader mix of school, family and community programs to improve educational outcomes. The narrow focus of educational research is a function of using easily measurable and available objectives. "Achievement" or "high school completion" or "total years of education" are the common measures used to evaluate the effectiveness of schooling expenditures. These narrower objectives do not provide direction for the ultimate objective: connecting both family and school expenditures to their effect on longer-term contributions (taxes raised) and disbursements (welfare, criminal justice, Medicaid, etc.). Better educational outcomes presumably contribute to higher wages and more taxes, and to reduced social welfare, health and criminal justice expenditures.