TECHNICAL REPORT

Improving the Achievement of Tennessee Students

Analysis of the National Assessment of Educational Progress

David W. Grissmer, Ann Flanagan

Supported by the Tennessee Advisory Commission on Intergovernmental Relations

RAND EDUCATION

The research described in this report was carried out by the RAND Corporation under the auspices of RAND Education and supported by the Tennessee Advisory Commission on Intergovernmental Relations.

Published 2006 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213
RAND URL: http://www.rand.org/
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Preface

This study focuses on the state of Tennessee to help state and local policymakers make decisions about educational spending and policies. Like their contemporaries in every state, educational policymakers in Tennessee want answers to a set of complex questions:

- How are Tennessee's students performing?
- What factors explain differences in the performance of Tennessee's students relative to that of students in other states?
- How can policies be improved and spending be made more effective and efficient?

To address these questions, this study compares the performance of Tennessee's students with that of students in other states. The study uses data primarily from the National Assessment of Educational Progress (NAEP) from 1990 to 2003. The NAEP tests are the only readily available achievement tests that permit valid comparisons of student performance across states. Data from surveys of teachers administered as part of the NAEP are examined as well. In analyzing the NAEP data, we compare Tennessee's performance with that of a set of comparison states in the south with similar family characteristics. The study also employs statistical models using state-level data to investigate the role of family background and educational resource measures in explaining differences in student achievement across states and over time.

While the study is directed primarily to Tennessee policymakers, policymakers from other states may find this study of interest for its approach and its comparisons of achievement and other measures across states. However, the policy recommendations made in this study cannot be applied readily to other states, both because each state has different policies and resources and because it is also important to consider the uncertainties that are specific to each state.

This study was supported by the Tennessee Advisory Commission on Intergovernmental Relations through an appropriation from the Tennessee legislature. The RAND Corporation carried out the research under the auspices of the RAND Education program. The mission of RAND Education is to bring accurate data and careful, objective analysis to the national debate on education policy.

Contents

Tables

Summary

Like their contemporaries in every state, educational policymakers in Tennessee want answers to a set of thorny—and pressing—questions:

- How are Tennessee's students performing?
- What factors explain differences in the performance of Tennessee's students relative to that of students in other states?
- How can policies be improved and spending be made more effective and efficient?

As in most southern states, the scores of Tennessee students on standard achievement tests are below the national average. But the explanations commonly cited for this underperformance generally are not based on sound empirical evidence. This study was designed to provide the empirical evidence that would lead to a more accurate understanding of what factors are linked to the differences in scores between states. How much of the variation between scores in Tennessee and those in other states can be attributed to different family characteristics, for example? How much can be traced to educational policies, such as the way spending is distributed among low pupil-teacher ratios, teacher salaries, teacher resources, and public pre-kindergarten?

Because of data limitations, none of these questions offers an easy or certain response. To increase the confidence we attach to our findings, we use four sources of evidence in this study. These sources are (1) literature from research based on experimental designs; (2) literature from research based on nonexperimental designs; (3) a regression analysis linking differences across states on 17 achievement tests administered through the National Assessment for Educational Progress (NAEP) in 4th and 8th grade math and reading from 1990 to 2003 with state differences in family background and educational resource policies; and (4) responses from surveys given to teachers during NAEP testing on questions about their credentials and training, their pedagogical practices, the adequacy of training and resources, the school climate, and their attitudes toward the state standards-based accountability. When possible, we use the results of the literature to confirm the results of our regression models. In particular, we have the most confidence in results that are found in both the experimental literature and other sources.

We begin in the next section by presenting a rationale for using NAEP scores for comparison across states rather than scores on the SAT®, an approach taken in other studies. We then summarize our findings with respect to the performance of Tennessee's students using the 2003 NAEP 4th and 8th grade reading and math tests and earlier writing and science tests. We also compare the performance of Tennessee students with that of students in other southern states with similar family characteristics. The subsequent sections focus on possible explanations for Tennessee's performance. As part of that discussion, we summarize key findings from the nonexperimental and experimental empirical literature relevant to the effects of major expenditure variables on achievement. We then highlight our results linking state achievement scores to each state's family characteristics, as well to the pattern of educational spending in the state, identifying the estimated quantitative differences in the effects of different types of resources. This model helps identify what factors explain Tennessee's relative performance. We also discuss the potential role of other factors not directly examined in the quantitative analysis. We conclude with a discussion of the implications of the findings for future directions in Tennessee educational policy.

NAEP Tests Provide a Basis for Comparing States

For this study, we used the NAEP tests given across states from 1990 to 2003 as our primary measure of student achievement for comparing states. The NAEP tests have been given in reading and math from the early 1970s until today using representative samples of U.S. students at ages 9, 13, and 17. NAEP are the only achievement tests with which one can readily make valid comparisons of student performance across the nation because, unlike the other tests, they offer a broad, representative sample of students across time and states.

From its inception until 1990, NAEP provided only estimates of overall national performance. However, starting in 1990, NAEP expanded to provide state estimates for individual states that volunteered to participate. Consequently, starting in 1990 the NAEP included large, random samples of students in a subset of states. Before 2003, participation was voluntary, with between 35 and 44 states participating in any given test. Since 2003, because of the No Child Left Behind (NCLB) legislation, all states have been required to participate. The four tests in 4th and 8th grade math and reading given in 2003 were the first that included all 48 contiguous states.

Recognizing the advantages of the NAEP scores, we used them as the basis for our analysis. We utilized the 2003 scores with all states participating to compare Tennessee's performance with that of other states. The 2003 data are used to rank each state's performance on a given test for all students and subgroups of students. We also analyzed 700 state scores associated with 17 NAEP tests given in participating states to 4th and 8th grade students in math and reading from 1990 to 2003. These data are used in a regression analysis to assess the importance of several specific educational resources on student achievement. The models consider the contribution of specific educational resources, controlling for differences in family characteristics across states. We also draw on responses from teachers surveyed as part of the NAEP test administration to consider other possible factors that are not always included in our models but might explain test score differences across states.

In the past, some comparisons of educational performance across states have used scores from the SAT. However, SAT scores have several important flaws in such applications that NAEP scores do not possess. The primary limitation of SAT scores is that only high school students applying to college complete the SAT. This selective subset of all U.S. students excludes younger students and precludes a nonrandom sample of high school students. Moreover, the SAT sample has changed as the characteristics of students applying to and attending college has changed dramatically over the last 40 years. The sample today includes a larger proportion of all high school students than it did in 1967, as well as a greater proportion of minority and female students. Thus, changes in the SAT can reflect changes both in the achievement of students and in the population of test takers. Similarly, any differences among states in the populations of students taking the SAT will confound cross-state comparisons. Finally, the SAT provides no information on students with a low propensity to attend college, and other data sources suggest these students have been the focus of many educational reforms and have made significant achievement gains over the last 30 years.

How Are Tennessee's Students Doing?

To address the first question, we compared the performance of Tennessee students with that of students in other states on recent NAEP tests, first for students in aggregate and then for subgroups of students. We used the 2003 4th and 8th grade math and reading tests because these tests are the first to include all 48 contiguous states. We also examined the most recent results for 4th and 8th grade writing and science tests across all participating states. In addition, we focused on the 2003 results of the 8th grade math test to compare the performance of black and white students, and among students in central cities and suburban and rural areas.

As part of this analysis, we compare the performance of Tennessee students to that of students in a comparison group of eight southern states with similar family characteristics. These states are Alabama, Arkansas, Georgia, Kentucky, North Carolina, South Carolina, Virginia, and West Virginia. Among the states in the southern region of the United States, these are the eight in which the family characteristics that predict achievement are most similar to those of Tennessee. The similarity of these family characteristics suggests that any differences in achievement in these states are more likely linked to differences in characteristics of the K–12 education systems.

Our analysis of the NAEP data generated the following key findings with respect to Tennessee's performance:

- Tennessee consistently ranked in the bottom fifth of states on 4th and 8th grade reading and math scores (as low as 42nd out of 48 states on the 4th grade reading test and as high as 38th out of 48 states on the 8th grade reading test).
- The NAEP also tested writing and science at the 4th and 8th grade level in 2002 and 2000, respectively, with about 35–40 states participating. Had all 48 states participated, Tennessee's estimated ranking on these tests would have been between 33 and 37 out of 48 states.
- On the 2003 tests, three states have consistently higher scores than Tennessee: North Carolina, Virginia, and Kentucky. Tennessee has consistently higher scores than Alabama. South Carolina, Georgia, West Virginia, and Arkansas generally have scores similar to those of Tennessee.
- Tennessee also made slower gains in scores between the early 1990s through 2003 than some of the comparison states. Between 1990 and 2003, the average annual score gain for Tennessee across all tests was 0.5 percentile points, below the national average and significantly below North Carolina and South Carolina, the highest-performing comparison states. In the early 1990s, North and South Carolina and Tennessee generally had similar scores, but by 2003, both of those states had significantly higher scores than Tennessee.
- Black students and students in central cities in Tennessee fare worse in terms of 2003 NAEP scores in 8th grade math when compared with the eight comparison states, whereas white students and students in rural and suburban areas are more comparable to their counterparts in the comparison states.

Explaining Tennessee’s Performance on NAEP

In order to address the second question, concerning the factors that explain difference in the performance of Tennessee’s students relative to that of students in other states, our analysis used a methodology employed in an earlier RAND report (Grissmer, Flanagan, et al., 2000) that presented results for NAEP tests from 1990–1996 but included only about 300 state observations. The results presented here are based on a larger sample over a longer time period and are consistent with those of the earlier study. Before summarizing the key findings of our empirical analysis, we briefly review findings from the experimental and nonexperimental literature that considers the effects of specific policies or interventions on student achievement.

Shifting Paradigms in the Research Literature

No issue in educational research or policymaking has received as much attention as the role of resources and their effect on achievement and other educational outcomes. Until about 1993, a dominant view based on reviews of the nonexperimental literature was that additional resources put into public education would not improve outcomes. Underlying this view was a theory of the efficiency of markets and the inefficiency that normally has been associated with public-sector activities. Public schools were viewed as public bureaucracies that had few internal incentives to improve or use resources efficiently.

The view that additional resources cannot improve outcomes in the current public education system has been challenged by recent literature reviews and results from experimental research. Since 1993, literature reviews have supported the hypothesis that more resources can improve educational outcomes, but these reviews have been unable to identify consistently which use of resources is most effective or efficient. However, research based on experimental design involving specific programs or specific uses of resources has provided stronger evidence that some targeted uses of resources can improve achievement. This view was also supported by the long-term gains in NAEP scores from 1970 to 1990 that occurred only among minority and disadvantaged students during a time when significant additional resources were targeted to programs expected to benefit such students.

The research evidence using experimental data also tends to converge on a hypothesis that specific programs can boost the achievement of minority and disadvantaged students, but there have not been many adequate experimental evaluations. A major experiment on different class sizes and teacher aides suggested that lower class sizes at the K–3 level could increase achievement not only in these lower grade levels: A significant part of the gain extended through high school. The research suggested that the achievement gains were larger for minority and disadvantaged students, and that 3-4 years of class size reductions were necessary for sustained gains.

There have also been experimental evaluations of many early childhood interventions involving preschool, kindergarten, and other early interventions before school enrollment begins. This research suggests that such interventions have a variety of educational and noneducational benefits, and that these benefits can markedly exceed their costs. The educational effects can include higher test scores and school attainment (e.g., high school graduation) and reductions in grade retention and special education placement. Research also suggests that such effects are greater when targeted to minority and disadvantaged students. Finally, this research has also suggested that interventions earlier in life are likely to be more efficient than later interventions.

The more recent evidence from both experimental and nonexperimental studies tends to support an emerging educational reform strategy that has two components. The first component is to provide more resources to education and to target these resources efficiently to programs and students on the basis of the best research evidence. The second component is to introduce standards-based accountability systems to provide better information management and incentives to improve the efficiency of the system. Such systems would develop specific standards for student knowledge and test students to provide evidence of whether they are meeting the standards. Students’ test results are to be used to assess progress; diagnose why results are different across students, schools, and school districts; and provide the basis for incentives to schools and teachers meeting certain progress criteria. The standards are to be used to align curriculum, teachers’ professional development, and other resources in a focused way. Almost all states have developed such systems, but have given them widely varying characteristics, making it important to attempt to measure their differential effects across states.

There is no consensus in the research community about the effects or efficiency of resource or reform policies. The evidence from well-designed experiments is considered the most reliable when such evidence is available. However, not many reliable experiments have been undertaken, and the results of experiments can make reliable predictions only for the conditions present in the experiment. Predictions of effects in different contexts are less reliable.

There is also no consensus about why results from research based on nonexperimental data have such wide variance. The vulnerability of nonexperimental data in education arises for at least two reasons. First, many variables that can affect educational outcomes are often not present in models. Such missing variables can bias the effects of variables included in the models through correlations with these variables. Second, the effects of expenditures and schooling variables account for a relatively small part of the explained variance, with family and community variables accounting for the larger share. Variables with small effects require large variations in the sample and/or large sample sizes for effect measurements to show the desired range of variation. Many research studies in education have limited ranges of variation of key variables and/or small sample sizes that can increase the range of variation of the results.

The literature has tried to find circumstances in which more reliable results have emerged. There is some evidence that measurements using state-level data may be more consistent than school district, school classroom, or student-level analysis. The argument for why aggregate state data may be more consistent is that the range of variation in key variables is larger across states than school districts or schools and that certain forms of bias can cancel at higher levels. However, there are also hypotheses that suggest that more aggregated analysis might be more biased.

In the current environment, no single analysis—especially using nonexperimental data—will be definitive. Every analysis can be vulnerable to bias. Rather, the results of each model must be evaluated with respect to the set of assumptions made and with respect to its agreement with the more reliable experimental data and with previous nonexperimental research. It is the triangulation of results from different empirical methods and from different periods that can help provide more reliability to policy suggestions.

Family Characteristics Contribute to Score Differences

In our estimated models using the NAEP data, family characteristics and characteristics of the state educational system both predict how states rank on NAEP scores, but family characteristics have much larger effects. States that score higher on these achievement tests have higher levels of parent education and income, lower proportions of single-parent families and births to teen mothers, and lower proportions of minority and disadvantaged students. Family characteristics such as these appear to either place children at educational risk or provide an educational advantage. These factors tend to cluster within families, creating multiple risks or advantages. For example, families with lower parental education are also more likely to have a lower income, be headed by a single parent, and have a mother who was a teen at the time of the birth of one or more of her children.

Tennessee's family characteristics rank about 36th out of 48 states on a combined measure of relative educational risk. In general, its families have a higher level of combined risk factors than families in many other states. This is consistent with its relatively lower ranking on NAEP scores.

Across the United States, families with higher levels of risk factors tend to be clustered in the southeastern and southwestern states. NAEP scores in these states tend to be among the lowest in the country. In contrast, families with higher levels of advantage tend to be clustered in northern rural states where NAEP scores are also highest. Northern urban states tend to have NAEP scores near the average because they have a mix of families: those with higher advantage in rural and suburban areas and those with higher risk in central cities. This clustering of families with different characteristics is a significant factor in variations in scores across regions and states.

Educational Resources Also Affect Test Scores

After controlling for differences in family background across states, the results from our models of 4th and 8th grade achievement scores indicate that educational resources matter as well. We find that a lower pupil-teacher ratio in grades 1 to 4 and higher participation rate in public pre-kindergarten programs positively affect achievement. Both of these results are consistent with experimental studies that evaluate the effects of reducing class size and providing high-quality preschool. We also find positive effects on student achievement in raising teacher salaries and teacher resources. The finding regarding teacher salaries is confirmed in other nonexperimental studies but has not been the subject of experimental evaluations. No previous studies are available on the effect of the adequacy of teacher resources as measured in this study.

When focusing on 4th grade achievement scores, our models indicate that the effect on achievement of lowering pupil-teacher ratios in the early grades and of raising pre-kindergarten participation rates is strongest in states with a higher proportion of disadvantaged families. Again, this is consistent with findings from experimental evaluations of class size reduction and preschool interventions, where the effects have been found to be strongest for children at risk of poor educational performance.

Given the composition of families in Tennessee and the state's allocation of educational resources, our model can be used to explain Tennessee's performance vis-à-vis the eight comparison states we identified, as well as all 48 states included in our analysis. First, as noted above, Tennessee has a relatively higher-risk population, ranking 36th out of 48 states on our composition measure of family background. Second, the dollar resources devoted to education measured by per pupil spending are among the lowest in the country, ranking 42nd out of 48 states. Considering the specific resource measures included in our model, with the exception of the pupil-teacher ratio, Tennessee ranks in the bottom half among all states on the resource measures that raise student achievement. On the measure of teacher resources, Tennessee ranks well below the eight comparison states. Teacher salaries and participation in public pre-kindergarten are near the average for the comparison states. In contrast, Tennessee's pupil-teacher ratio is lower than that of five of the other comparison states and lower than that of 27 of the 48 states we examine.

Other Factors Deserve Consideration

While family background and the specific educational resources we considered can explain some of the variation in achievement scores across states, some unexplained variation remains. When we use the model to explain the average annual score gains between 1990 and 2003, there is little unexplained variation for Tennessee. However, there are large score gains in other states, including comparison states such as North Carolina and South Carolina, that are not explained by the family background and educational resource measures included in the model. These and other states made large gains in achievement scores beyond what would have been expected from the resources we analyze.

Two other resource measures were included in our models. The results indicate that a higher fraction of inexperienced teachers is associated with lower achievement. Tennessee has higher levels of inexperienced teachers than all comparison states except North Carolina. This may be the result of reductions in pupil-teacher ratios in the late 1990s. We found no effect on test scores for a measure of the proportion of teachers with advanced degrees. Although other research has also found that, in general, advanced degrees among teachers do not correlate with higher achievement, there is some evidence that subject-specific degrees can raise achievement scores.

Beyond these two measures, our analysis looked to other sources, including teacher responses on surveys administered with the NAEP tests, to identify other possible explanations for the residual score gains. These surveys suggest that

Tennessee teachers report significantly lower positive connections to their accountability systems than teachers in most of the comparison states. They report that standards are less clear and less useful for planning curriculum, and that there are fewer resources for training and implementing the system than in most comparison states and in particular comparison states with the largest unexplained gains in NAEP scores. Tennessee English/reading teachers also report using less advanced pedagogical techniques, and Tennessee math teachers seem to have fewer credentials and less of the knowledge required to teach more advanced courses. Tennessee students also seem to less frequently take algebra for high school credit.

The data we have presented are not definitive, but only suggestive that Tennessee's accountability system and its teachers are not tightly linked in a way that might drive curriculum, pedagogy, and credentials to higher levels. Moreover, researchers have yet to demonstrate the benefits in terms of achievement scores from particular features of a standards-based accountability system. A much more comprehensive study is required that would focus on the relationships between the differences in the standards-based accountability systems, the structure of state tests, the training of teachers and their pedagogical approach, and the adequacy of resources provided to teachers.

Strengths and Weakness of the NAEP Score Analysis

The analysis undertaken here using NAEP data across states aims to link differences in state achievement with family characteristics and educational spending and policies during an important period in American education. In this period, each state and its school districts made effort to reform and improve its education system. The NAEP data from 1990 to 2003 are the only achievement data that can be used readily to validly compare state performance and to explain differences at this important time in American education. Thus, NAEP data from this period must be analyzed to try and understand what might explain differences across states and whether reforms are working. But such results also need to be assessed with respect to the wider experimental and nonexperimental results, and they must take account of the strengths and weaknesses of the analysis.

The strengths of this analysis include the following: (1) the model is based on 17 separate tests in two subjects and two grades over a 13-year period and provides over 700 observations of state achievement; (2) the NAEP evaluates not only lower-level skills through multiple-choice items, but also higher-level critical-thinking skills through open-ended items; (3) variation across states in almost all dependent variables is quite large compared to within-state district or school variation; (4) the analysis uses both random- and fixed-effects models that incorporate different statistical assumptions; (5) the model is consistent with the experimental effects of class size reductions in lower grades and pre-kindergarten programs; (6) these results also show consistency with the historical trends in achievement and spending that suggested that large achievement gains among minority and disadvantaged students occurred at the time when additional spending was directed to programs that would primarily benefit minority and disadvantaged students; and (7) none of the effects measured are inconsistent with the results of the nonexperimental literature, although because of the wide range of such measurements, this standard is not hard to meet.

The weaknesses of the model include the following: (1) possible bias in the results from several sources, including missing variables, selectivity, and non-linearities; (2) bias resulting from the inability to incorporate district- and school-level information in the analysis (also known as the ecological fallacy); (3) the limited data on family variables available directly from the NAEP, necessitating the use of U.S. Census data and a weighting procedure for family variables using an alternative achievement test; (4) the absence of several family variables that other research has shown to be linked to achievement, but which can be collected only through parental surveys; (5) a lack of data on within-race/ethnicity changes in family characteristics across states; and (6) inconsistency in the participation of states so that data are not available for all 48 contiguous states for all 17 tests.

Implications for Tennessee’s Education Policy

The findings of this study have several implications for Tennessee’s future educational policy. The research evidence suggests that Tennessee is justified in devoting substantial resources to lowering class sizes in the elementary grades and raising the proportion of children in public pre-kindergarten programs. While our findings do not indicate the optimal level of spending in these areas, our estimates suggest that these are areas that have generated the largest returns in the past in terms of test score increases for a given dollar of investment. In its effort to expand public pre-kindergarten programs, Tennessee should continue to maintain the research-based standards associated with high-quality programs.

Given that Tennessee lags other states in how teachers assess the adequacy of resources—another factor associated with higher achievement—the state should examine potential deficiencies in this area and consider ways to reallocate other spending toward efficient forms of teacher resources. On the other hand, while higher teacher salaries were shown to raise achievement, they do so at a relative higher cost. Given that Tennessee has salaries close to the national average, there may be less justification for using this policy lever to raise educational attainment. Since teacher salaries are the largest expenditures in education budgets, modest restraints in future salary increases may provide a source for channeling more funds into teachers' resources.

Finally, although the research base needed to guide decisionmaking is weak, Tennessee should assess the need for reforms in other areas that may be linked to improved school performance. This includes the state's standards-based accountability system, as well as its approach to teacher compensation, teacher training, and pedagogy in the classroom. For example, a useful next step would be to investigate the current standards-based accountability system in Tennessee and selected other states with the goal of discovering possible differences that might explain Tennessee's slower NAEP score improvements between 1990 and 2003, and addressing the issues that teachers have with the current system. A suggested set of objectives for such a study would include the following:

- Determine the differences in improvement, particularly between Tennessee and North and South Carolina
- Assess the link between standards and curriculum to determine why teachers in Tennessee report that standards are less clear and useful for curriculum planning
- Assess whether Tennessee's standards and testing program have the appropriate balance of emphasis on basic and critical-thinking skills
- Ensure that teachers are provided appropriate training and resources to understand and support the system.

Acknowledgments

This project was made possible through the initiative, cooperation, and ongoing advice and review of many individuals in the legislative and executive branches of government in Tennessee. The project received support from the Tennessee Advisory Commission on Intergovernmental Relations (TACIR), its chairman, State Representative Randy Rinks; vice-chairman, Tom Rowland; and executive director, Harry A. Green, as well as Lynnisse Roehrich-Patrick, Director of Special Projects for TACIR; John G. Morgan, Comptroller of the Treasury; and Ethel R. Detch, Director of the Offices of Research and Education Accountability. In addition, former TACIR chairman Robert Rochelle played an important role in initiating the project.

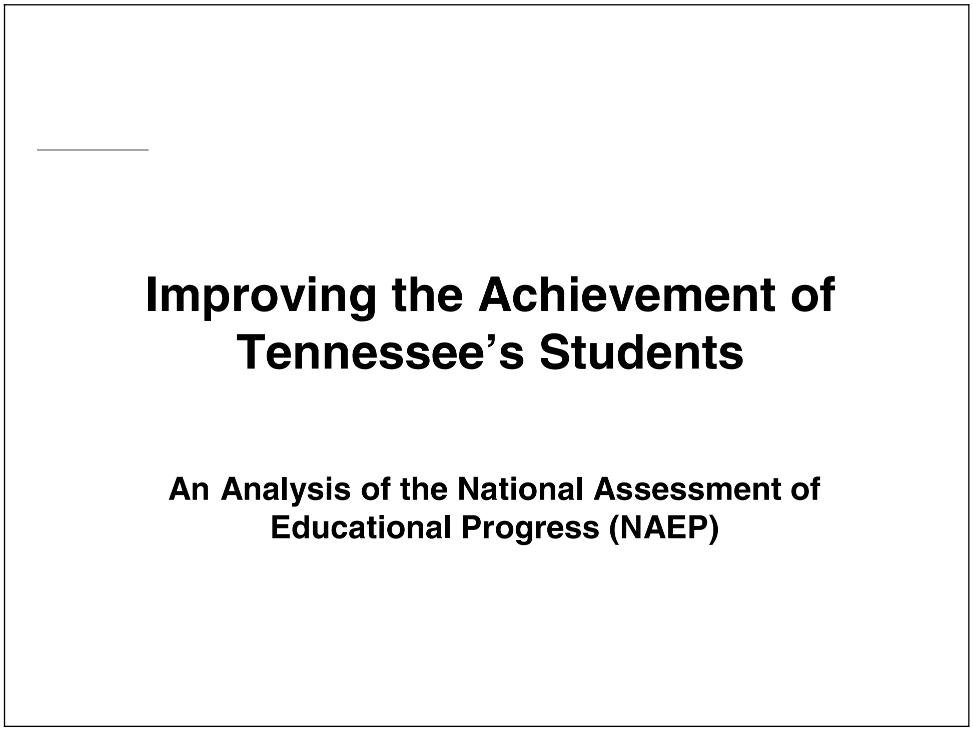
The project required continuous advice and review by individuals knowledgeable about educational budgeting and funding and educational conditions in Tennessee in order to interpret and develop explanations for the statistical results. We were fortunate to get such ongoing advice and review from Lynnisse Roehrich-Patrick of TACIR, who provided many insights that helped link the results of the analysis to educational conditions in Tennessee. The report also benefited from data provided by the Tennessee Department of Education under the direction of Lana C. Seivers, Commissioner of Education, and from reviews and comments by her staff and staff of the Office of Education Accountability.

This report benefited immensely from technical reviews by Lynn Karoly and Dan McCaffrey of RAND.

Acronyms

COL	cost of living
ECLS-K	Early Childhood Longitudinal Survey of the kindergarten class of 1998–1999
IEP/DS	Individualized Education Plan/Disabled
LEP	Limited English Proficiency
NAEP	National Assessment of Educational Progress
NCLB	No Child Left Behind
NCTM	National Council of Teachers of Mathematics
NELS:88	National Education Longitudinal Study of 1988
NIEER	National Institute for Early Education Research
NSF	National Science Foundation
SES	socioeconomic status
SES-FE	SES fixed effects
TACIR	Tennessee Advisory Commission on Intergovernmental Relations

Presentation of Results

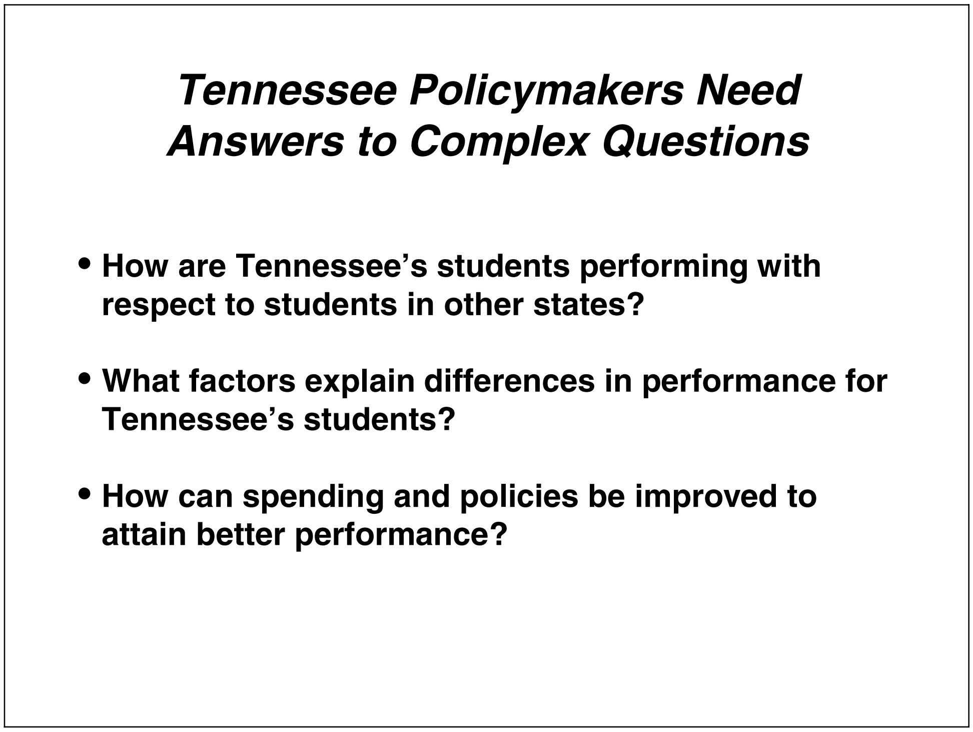


Improving the Achievement of Tennessee's Students

An Analysis of the National Assessment of Educational Progress (NAEP)

This study is focused on the state of Tennessee to help state and local policymakers make decisions about educational spending and policies. As is true of most southern states, the scores of Tennessee students on standard achievement tests are below the national average. But the explanations commonly cited for this underperformance generally are not based on sound empirical evidence. This study was designed to provide improved empirical evidence that would lead to a more accurate understanding of what factors are linked to the differences in scores between states. These results are then used to assess current educational spending and policies in Tennessee and suggest directions for improvement.

The study uses data primarily from the National Assessment of Educational Progress (NAEP). The NAEP tests are the only readily available achievement tests that permit valid comparisons of student performance across states. Unlike the SAT®, the NAEP assessments select representative samples of students, making valid state comparisons possible. Seventeen tests have been administered in participating states in 4th and 8th grade math and reading from 1990 to 2003. In conjunction with these tests, surveys of teacher characteristics and attitudes, pedagogical practices, and teaching environments have been given in various NAEP test years to representative samples of teachers.



Tennessee Policymakers Need Answers to Complex Questions

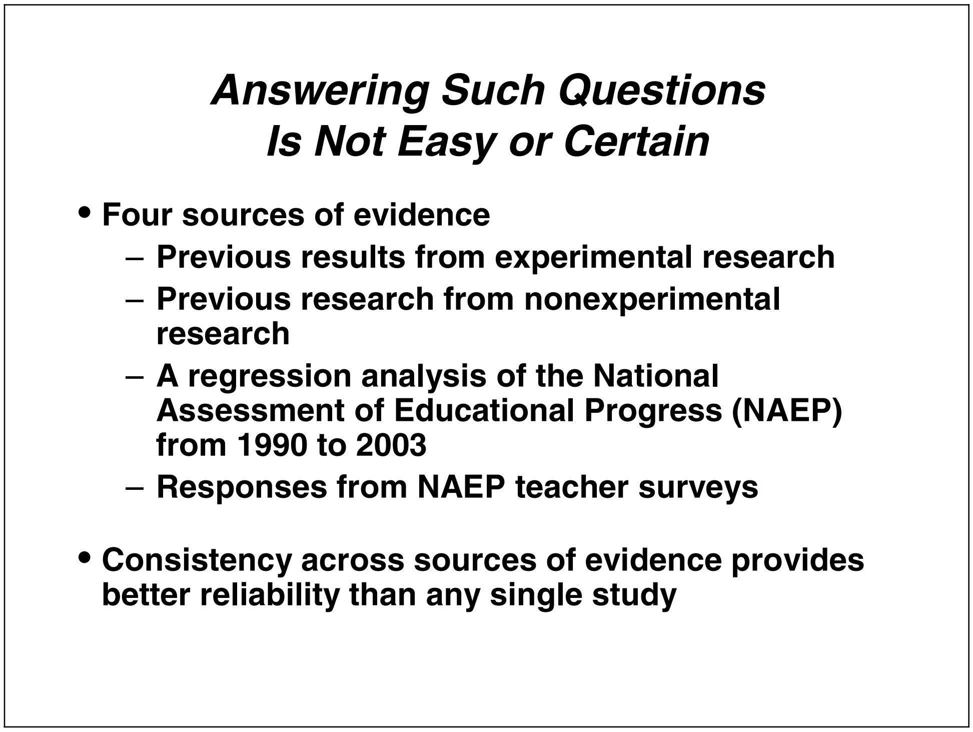
- **How are Tennessee’s students performing with respect to students in other states?**
- **What factors explain differences in performance for Tennessee’s students?**
- **How can spending and policies be improved to attain better performance?**

Like their contemporaries in every state, educational policymakers in Tennessee want answers to a set of thorny—and pressing—questions. The three questions listed here pertain to the relative performance of Tennessee’s students, the factors that can explain differences in performance, and policy changes that would lead to better student outcomes.

To address the first question, this study compares the performance of Tennessee’s students with that of students in other states on the most recent 4th and 8th grade math, reading, writing, and science NAEP tests. In addition, we compare Tennessee’s performance with that of a set of comparison states in the south with similar family characteristics. For these comparison states, we also examine differences in 8th grade math scores for subgroups of students defined by race/ethnicity and urban/rural status.

To address the second question, we briefly review the research literature that examines the effect of specific policies or interventions on student achievement. We then assess the importance of several specific educational resources relative to student achievement using a regression analysis of 17 NAEP reading and math tests given to 4th and 8th graders from 1990 to 2003. The models consider the contribution of educational resources, controlling for differences in family characteristics across states. We also draw on responses from teachers surveyed as part of the NAEP test administration to consider other possible factors that are not always included in our models but might explain test score differences across states.

Finally, our analysis of NAEP achievement scores and teacher responses, together with results from the research literature, then serves as a basis for suggesting future directions for Tennessee state and local policymakers.



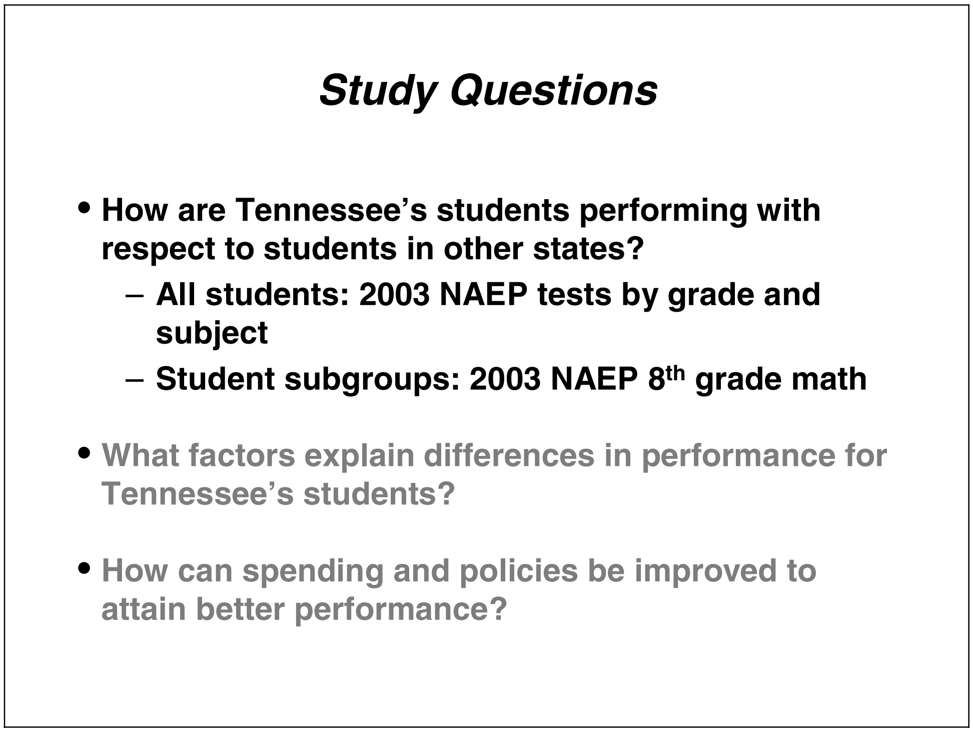
Answering Such Questions Is Not Easy or Certain

- **Four sources of evidence**
  - **Previous results from experimental research**
  - **Previous research from nonexperimental research**
  - **A regression analysis of the National Assessment of Educational Progress (NAEP) from 1990 to 2003**
  - **Responses from NAEP teacher surveys**
- **Consistency across sources of evidence provides better reliability than any single study**

Because of data limitations, none of the study questions offers an easy or certain response. In order to increase the confidence we attach to our findings, we use four sources of evidence in this study. These sources are (1) literature from research based on experimental designs; (2) literature from research based on nonexperimental designs; (3) a regression analysis linking differences across states on 17 NAEP achievement tests in 4th and 8th grade math and reading from 1990 to 2003 with state differences in family background and educational resource policies; and (4) responses from surveys given to teachers during NAEP testing on questions about their credentials and training, their pedagogical practices, the adequacy of training and resources, the school climate, and their attitudes toward the state standards-based accountability system.

There is general agreement between the results of the NAEP regression analysis and previous experimental results when such results are available. Agreement with experimental results provides the most reliable validation of the models we estimate, but experimental results are available only for a limited set of variables. For variables that have no experimental results for comparison, we also utilize the nonexperimental literature. But since this literature often shows inconsistency, this agreement does not provide a very strong validation. Moreover, some educational policy reforms have yet to be as carefully studied using either experimental or nonexperimental methods, so there is little guidance from the evaluation literature. Thus, we have more confidence in suggesting future directions for educational policy in Tennessee when the results of our NAEP analysis are consistent with results arising from the experimental and the nonexperimental literature, collectively. In other cases where the evidence is more limited, our conclusions are naturally more tentative.

Our NAEP analysis is based on nonexperimental student data aggregated to the state level. Such data require several assumptions to produce unbiased estimates. There is no consensus in the research community about what kinds of nonexperimental analyses produce the most unbiased results. Rather, the results of each model must be evaluated with respect to the set of assumptions made and with the results' agreement with the more reliable experimental data and with previous nonexperimental research. We discuss some of the limitations associated with our approach when we discuss our findings.

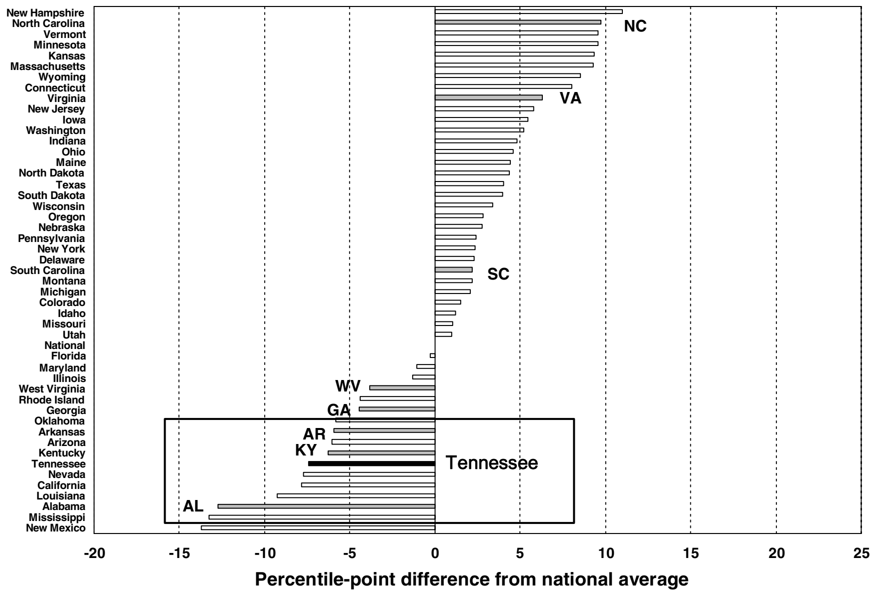


Study Questions

- **How are Tennessee's students performing with respect to students in other states?**
  - **All students: 2003 NAEP tests by grade and subject**
  - **Student subgroups: 2003 NAEP 8th grade math**
- **What factors explain differences in performance for Tennessee's students?**
- **How can spending and policies be improved to attain better performance?**

In order to address the first question, we begin by providing comparisons of the performance of Tennessee students with students in all other available states on recent NAEP tests, first for students in aggregate and then for subgroups of students. We use the 2003 4th and 8th grade math and reading tests because these tests are the first to include all 48 contiguous states. We also present the most recent results for 4th and 8th grade writing and science tests across all participating states. We then focus on the 2003 results of the 8th grade math test to compare the performance of black and white students and students in central cities and suburban and rural areas. Throughout the discussion, we compare the performance of Tennessee students to that of students in other southern states with similar family characteristics that also have large minority and central city populations.

Average 4th Grade Math NAEP Scores in 2003 Show Tennessee to Rank 42nd out of 48 States



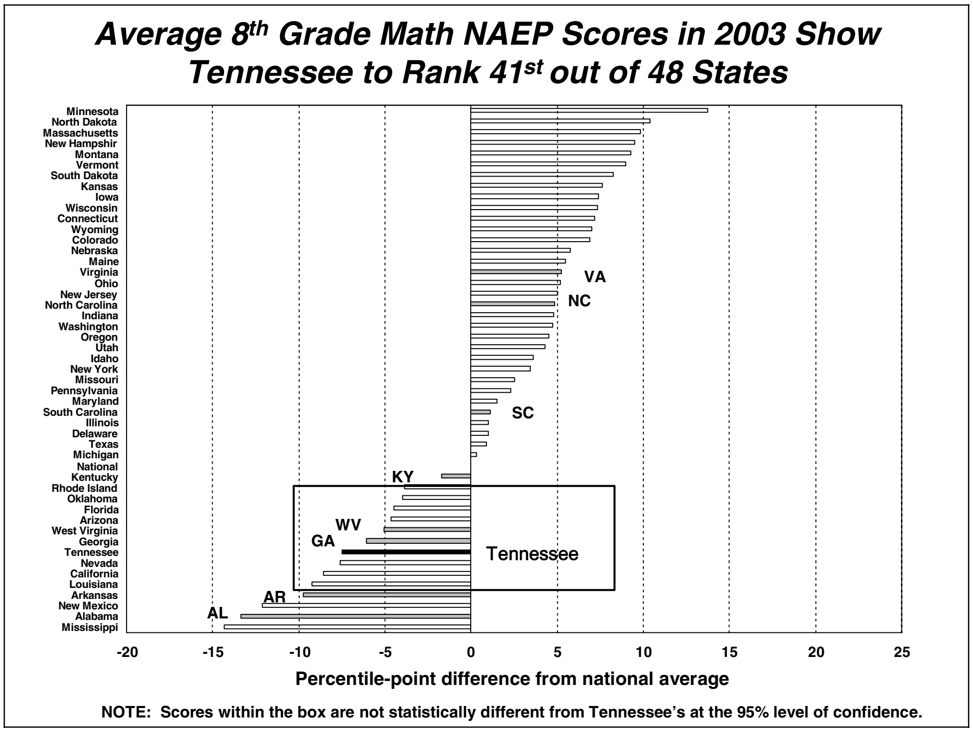
NOTE: Scores within the box are not statistically different from Tennessee's at the 95% level of confidence.

We utilize the 2003 NAEP tests to characterize the recent performance of Tennessee with respect to other states. The 2003 tests are the first to have included all 48 contiguous states. This chart shows that Tennessee's scores ranked 42nd out of the 48 states on the 2003 NAEP 4th grade math test. The achievement for each state is in percentile points above or below the national average score.

The data show that northern rural states tend to score above the average and usually closer to the top of the rankings. Northern urban states tend to score closer to the middle of the rankings, while southern states tend to be the lowest-scoring.

Tennessee students scored about 8 percentile points below the national average, indicating that the average Tennessee student scored above approximately 42 percent of the nation's students. The states within the box are states whose scores show no statistically significant difference from those of Tennessee at a 95-percent level of confidence. The difference between Tennessee's average score and the national average is statistically significant.

We chose eight comparison states in the same region as Tennessee that have similar family characteristics. These states are Alabama, Arkansas, Georgia, Kentucky, North Carolina, South Carolina, Virginia, and West Virginia. The neighboring comparison states show a wide variance in results, with North Carolina scoring the second highest in the nation and Virginia scoring ninth in the nation. South Carolina is also ranked above the national average. Four of the comparison states show statistically significant higher scores: North Carolina, Virginia, South Carolina, and West Virginia. The remaining comparison states do not have statistically significant scores different from those of Tennessee.



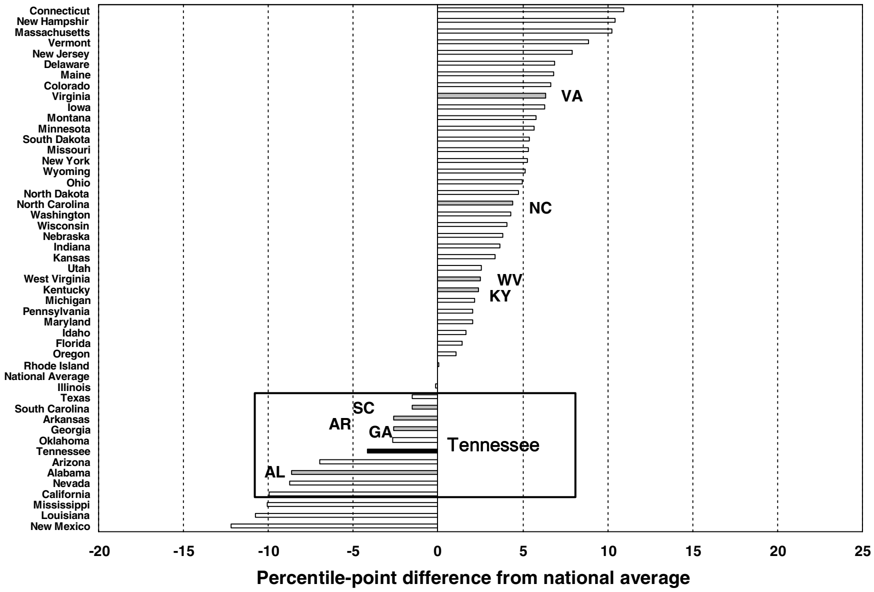
Average 8th Grade Math NAEP Scores in 2003 Show Tennessee to Rank 41st out of 48 States

NOTE: Scores within the box are not statistically different from Tennessee's at the 95% level of confidence.

Tennessee's students ranked 41st out of 48 states on the NAEP 8th grade math test in 2003—a ranking similar to its 4th grade math results. The pattern of 8th grade math scores across states shows a generally similar pattern to the 4th grade scores, with some exceptions.

Tennessee's average score is about 8 percentile points below the national average—a statistically significant difference. Virginia, North Carolina, South Carolina, and Kentucky have statistically significant higher scores than Tennessee. Tennessee has a statistically significant higher score than Alabama and Arkansas, but similar scores to West Virginia and Georgia.

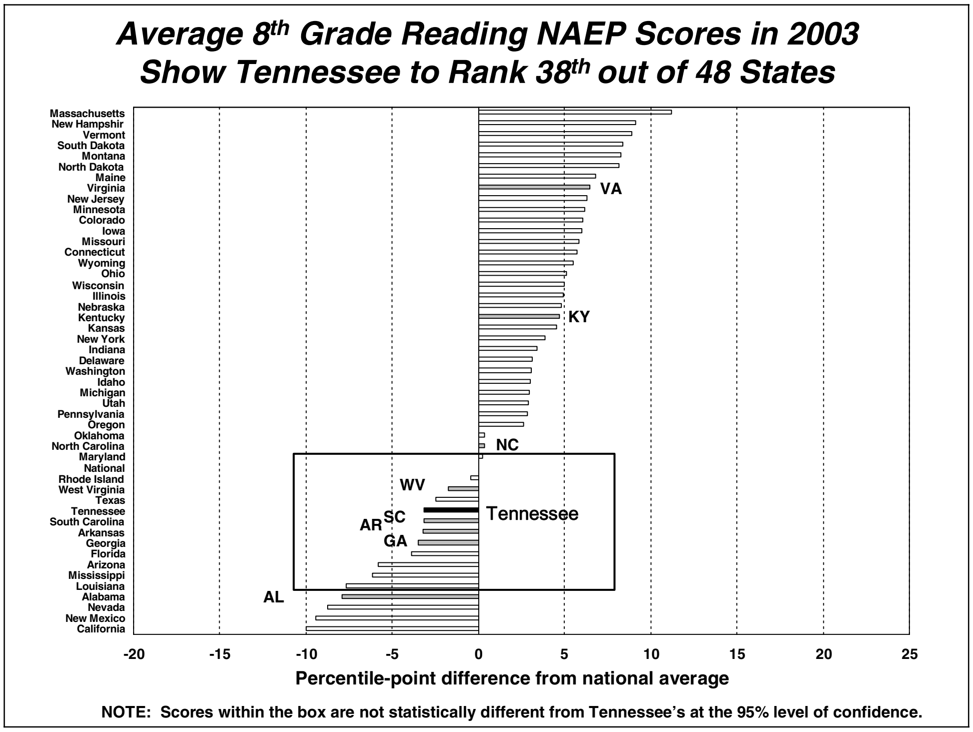
Average 4th Grade Reading NAEP Scores in 2003 Show Tennessee to Rank 41st out of 48 States



NOTE: Scores within the box are not statistically different from Tennessee's at the 95% level of confidence.

Tennessee's 4th grade NAEP reading scores have a ranking similar to its 4th and 8th grade math scores: 41st out of 48 states. The pattern of 4th grade reading scores is generally similar to that of 4th grade math scores, with some exceptions. Delaware, Colorado, and Missouri rank much higher in reading than in math, while South Carolina and Texas have much higher rankings in math than in reading.

Tennessee's scores are about 4 percentile points below the national average in 4th grade reading; that difference is barely statistically significant. Four comparison states have statistically significant higher scores than Tennessee: Virginia, North Carolina, West Virginia, and Kentucky. The remaining comparison states show no statistically significant results different from Tennessee.



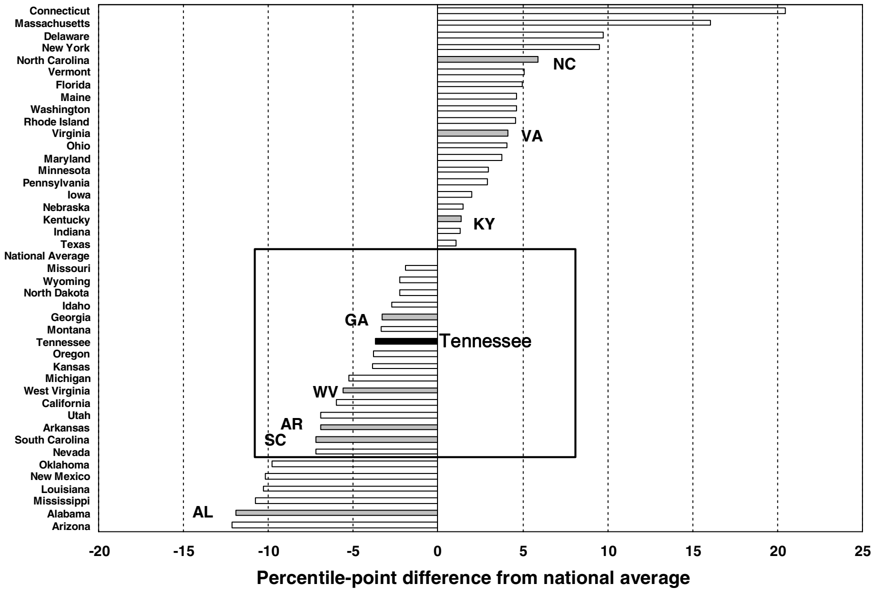
Average 8th Grade Reading NAEP Scores in 2003 Show Tennessee to Rank 38th out of 48 States

NOTE: Scores within the box are not statistically different from Tennessee's at the 95% level of confidence.

Tennessee's 8th grade NAEP reading tests ranked 38th out of 48 states—the highest ranking in math or reading. The general pattern is similar to the other reading and math scores, with some exceptions. Missouri and Illinois ranked much higher on the 8th grade reading tests than on the 8th grade math tests, while Kansas and North Carolina ranked much lower.

Tennessee's average score was about 3 percentile points below the national average, but that difference was not statistically significant. Virginia, Kentucky, and North Carolina had statistically significant higher scores than Tennessee. Alabama had statistically significant lower scores, while West Virginia, South Carolina, Arkansas, and Georgia had statistically similar scores.

Average 4th Grade Writing NAEP Scores in 2003 Show Tennessee to Rank 28th out of 42 States

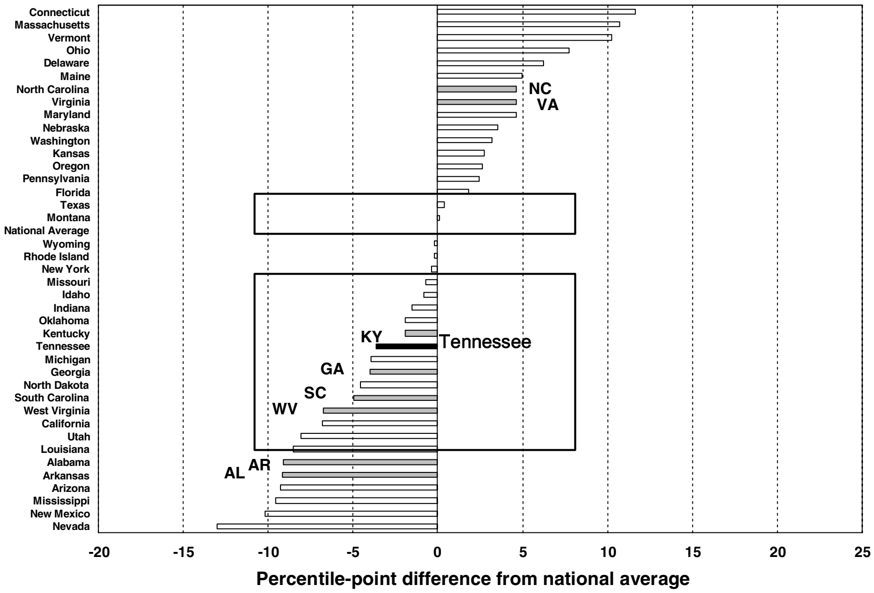


NOTE: Scores within the box are not statistically different from Tennessee's at the 95% level of confidence.

Of the four subjects tested by the NAEP—math, reading, writing, and science—Tennessee students did best with respect to other states in writing. Writing tests were given to 4th grade students in 2002 and 42 states participated. Tennessee ranked 28th out of the 42 states on the 4th grade writing test. Five of the six states that did not participate in the 4th grade writing test scored much higher than Tennessee on all the reading and math tests and likely would have scored higher than Tennessee if they had participated in the writing tests as well. As a rough approximation, we estimate that Tennessee's ranking among 48 states would have been around 33rd or 34th. But this is still a much better performance than the state achieved in reading and math.

Tennessee scored about 4 percentile points below the national average in writing—a difference that is not statistically significant. Among comparison states, North Carolina, Virginia, and Kentucky had statistically significant higher scores, while Alabama had a statistically significant lower score. South Carolina, Georgia, Arkansas, and West Virginia had similar scores.

Average 8th Grade Writing NAEP Scores in 2003
Show Tennessee to Rank 27th out of 40 States

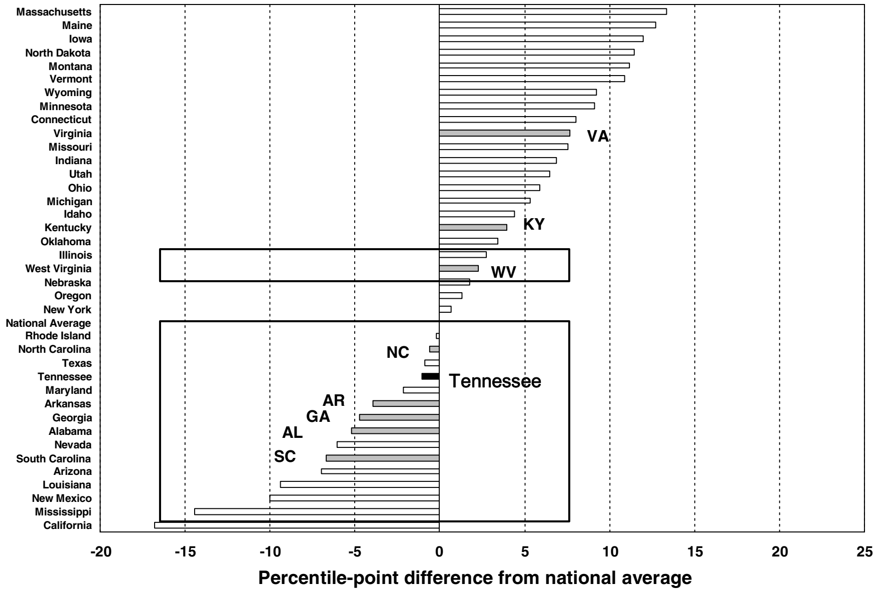


NOTE: Scores within the box are not statistically different from Tennessee's at the 95% level of confidence.

Tennessee students did as well on the 8th grade NAEP writing test as on the 4th grade writing test, ranking 27th out of 40 states. Seven of the eight states that did not take the 8th grade writing test had much higher scores than Tennessee in reading and math. Again, as a rough approximation, Tennessee's likely rank if all 48 states had taken the test would have been 34th or 35th. Tennessee was about 3 percentile points below the national average—a statistically significant difference.

Of the comparison states, North Carolina and Virginia had statistically significant higher scores than Tennessee, while Alabama and Arkansas had statistically significant lower scores. Kentucky, West Virginia, Georgia, and South Carolina had similar scores. Because of variations in sample size and within-state sample designs, p-values are not monotonically decreasing with size of the difference between the mean for Tennessee and the mean for other states. Hence, there are two boxes in the figure to capture all states not statistically significantly different from Tennessee.

Average 4th Grade Science NAEP Scores in 2003 Show Tennessee to Rank 27th out of 38 States

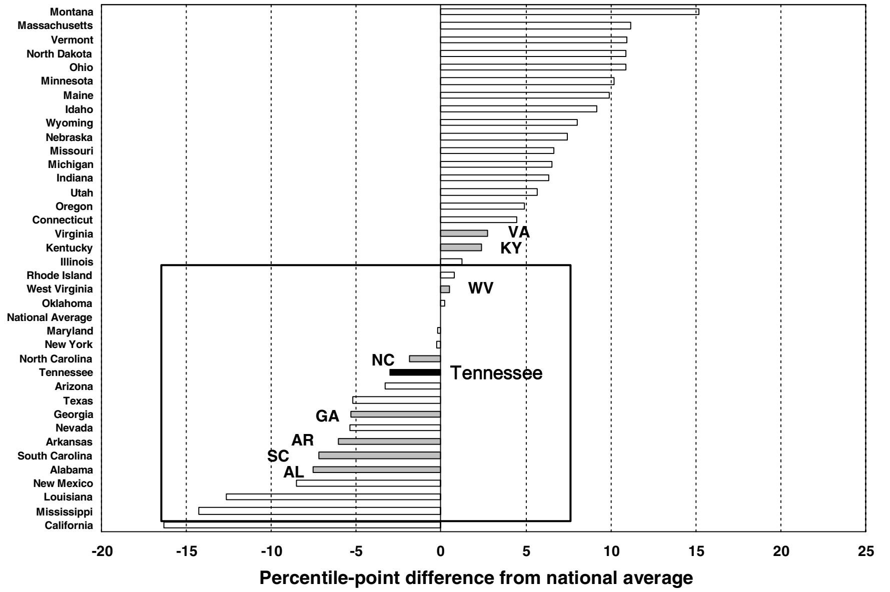


NOTE: Scores within the box are not statistically different from Tennessees at the 95% level of confidence.

NAEP 4th and 8th grade science tests were given in 2000. Only 38 states participated in the 4th grade science test. The pattern of science scores is more similar to patterns on reading and math tests than to the pattern of writing scores. Tennessee ranked 27th out of 38 states on the 4th grade science test. Nine of the 10 states that did not take the tests had statistically higher scores than Tennessee on all reading and math tests and would likely have scored higher on science as well. So Tennessee's likely ranking, had all 48 states taken the test, would have been about 36th or 37th.

Tennessee scored 1 point below the national average on this test—a difference that was not statistically significant. The state scored closer to the national average on the 4th grade science test than on any other test. Two comparison states had statistically significant higher 4th grade science scores: Virginia and Kentucky. The remaining comparison states had no statistically significant differences from Tennessee. As with the previous figure, there are two boxes in this figure because variations in sample size and within-state sample designs make p-values non-monotonic.

Average 8th Grade Science NAEP Scores in 2003 Show Tennessee to Rank 26th out of 37 States

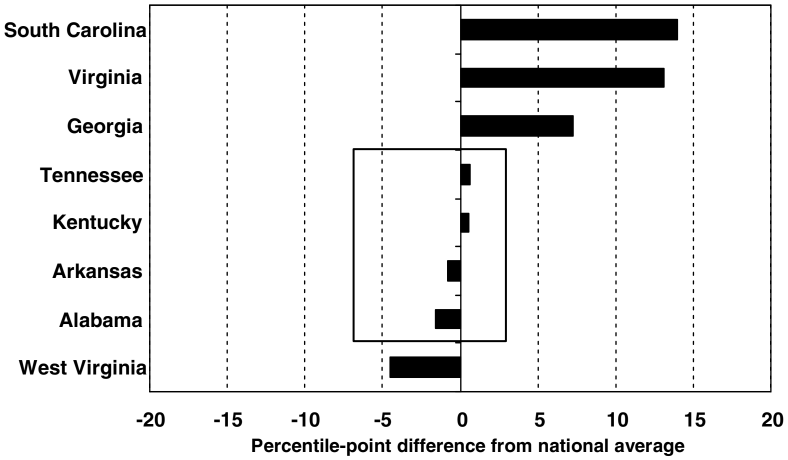


NOTE: Scores within the box are not statistically different from Tennessee's at the 95% level of confidence.

Tennessee ranked 26th out of 37 states taking the NAEP 8th grade science test in 2000. The general pattern across states is similar to that of the 4th grade science test. Nine of the eleven states that did not take the test scored statistically significantly higher than Tennessee on the math and reading tests. So Tennessee's ranking if all 48 states had participated would likely have been 35th–37th.

Tennessee's score is 3 points below the national average—a difference that is not statistically significant. Two states, Virginia and Kentucky, have statistically significant higher scores than Tennessee. The remaining six states are not statistically different from Tennessee.

Tennessee's White Students Score at the National Average, but Below Three Comparison States in 8th Grade Math Scores in 2003



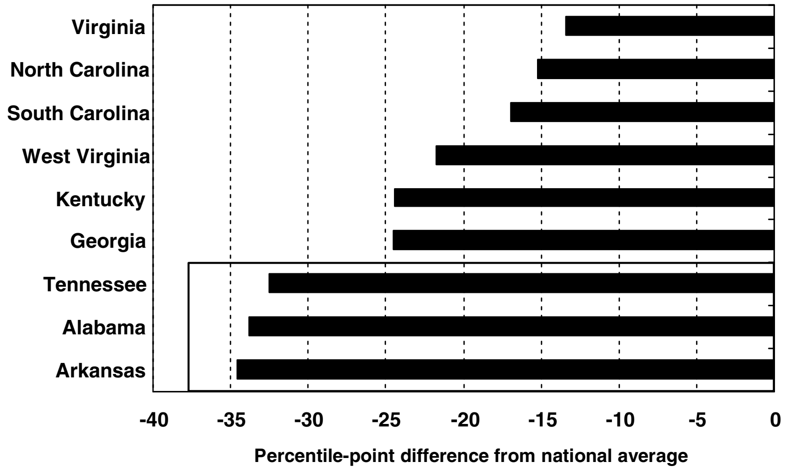
NOTE: Scores within the box are not statistically different from Tennessee's at the 95% level of confidence.

We now provide comparisons of the performance of black and white students in Tennessee as well as students living in central city, suburban, and rural areas. We make comparisons to similar students in the comparison states. We present only the results for 8th grade math. The results for the 4th and 8th grade reading and 4th grade math show somewhat similar patterns.

About 75 percent of students in Tennessee are white. These students scored at the national average for all students on the 8th grade math test, but 10 percentile points below the national average for white students (not shown).

Three comparison states have statistically higher scores among white students: South Carolina, Virginia, and Georgia. The scores of white students in Tennessee are statistically significantly higher than the scores of white students in West Virginia and show no statistically significant differences from those of Kentucky, Arkansas, and Alabama.

The Scores of Black Students in Tennessee in 8th Grade Math Are Among the Lowest of All Comparison States

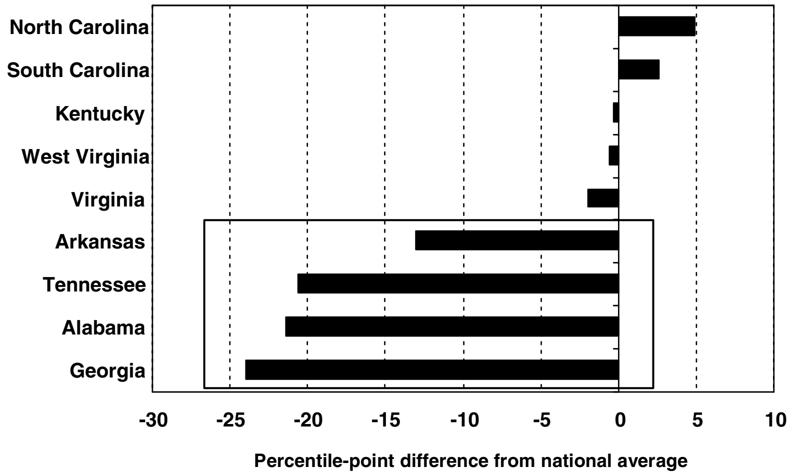


NOTE: Scores within the box are not statistically different from Tennessee's at the 95% level of confidence.

About 22 percent of K–12 students in Tennessee are black. These students scored about 33 percentile points below the national average for all students on the 2003 8th grade math test. This score was 10 percentile points below the average score of black students nationally (not shown).

Compared with black students in the comparison states, Tennessee's black students have among the lowest scores. There is no statistically significant difference between Tennessee's scores and those of Alabama and Arkansas. The scores of black students in the six other comparison states are statistically significantly higher than the scores of black Tennessee students.

The Scores of Students in Tennessee's Central Cities in 8th Grade Math in 2003 are Among the Lowest of Comparison States

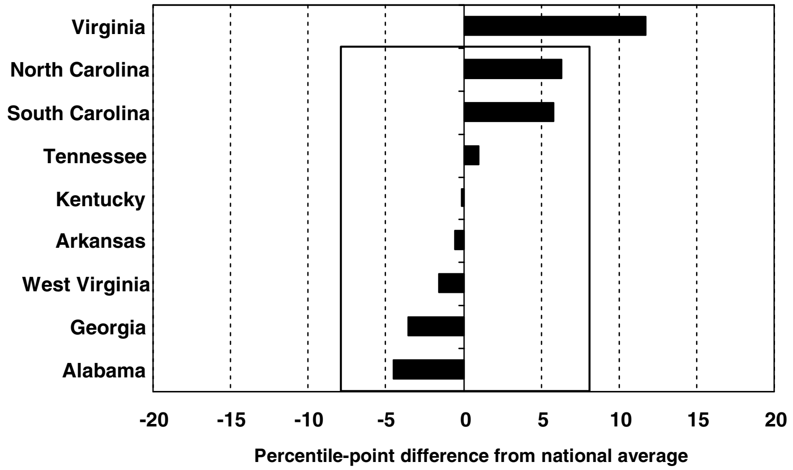


NOTE: Scores within the box are not statistically different from Tennessee's at the 95% level of confidence.

About 27 percent of Tennessee's students live in central cities, compared with 30 percent nationally. These students scored about 21 percentile points below the average for all students nationally on the 2003 8th grade math test and 9 percentile points below the average score of students in central cities nationally (not shown).

Tennessee's central city students score among the lowest within comparison states, with no statistically significant differences between Tennessee, Georgia, Alabama, and Arkansas. Five comparison states have statistically significantly higher scores than Tennessee.

The Scores of Tennessee's Suburban Students in 8th Grade Math in 2003 Compare Favorably with Comparison States

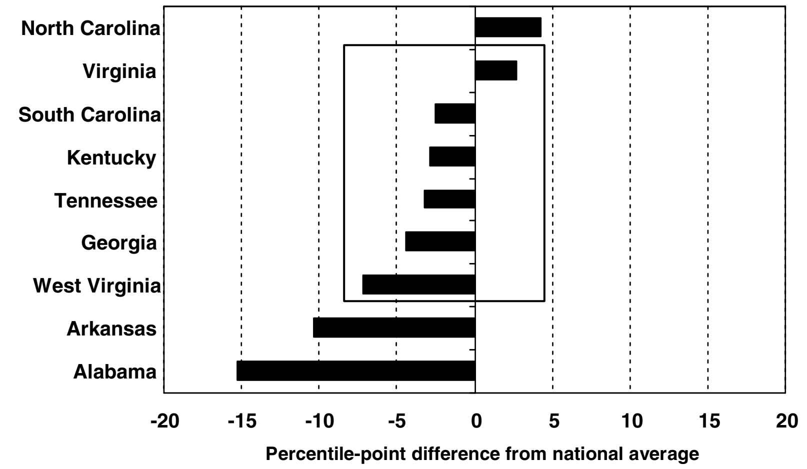


NOTE: Scores within the box are not statistically different from Tennessee's at the 95% level of confidence.

About 26 percent of K–12 students in Tennessee live in suburban areas, compared with 31 percent nationally. These students scored about 2 percentile points above the average for all students nationally on the 2003 8th grade math test, and only 3 points below the average scores of suburban students nationally (not shown).

The scores of Tennessee's suburban students are not statistically different from scores in seven comparison states. Only Virginia's suburban students have statistically significantly higher scores.

The Scores of Tennessee's Rural Students in 8th Grade Math in 2003 Also Compare Favorably with Comparison States



NOTE: Scores within the box are not statistically different from Tennessee's at the 95% level of confidence.

About 47 percent of Tennessee's students live in rural areas, compared with 39 percent nationally. These students scored about 3 points below the national average for all students on the 2003 8th grade math test and 6 percentile points below the national average for rural students (not shown).

Only one comparison state, North Carolina, had rural students who scored statistically significantly higher than students in Tennessee. Tennessee's scores are statistically significantly higher than Arkansas's and Alabama's.

Tennessee’s Performance is Fairly Similar Across Tests with Somewhat Better Performance in Writing and Science

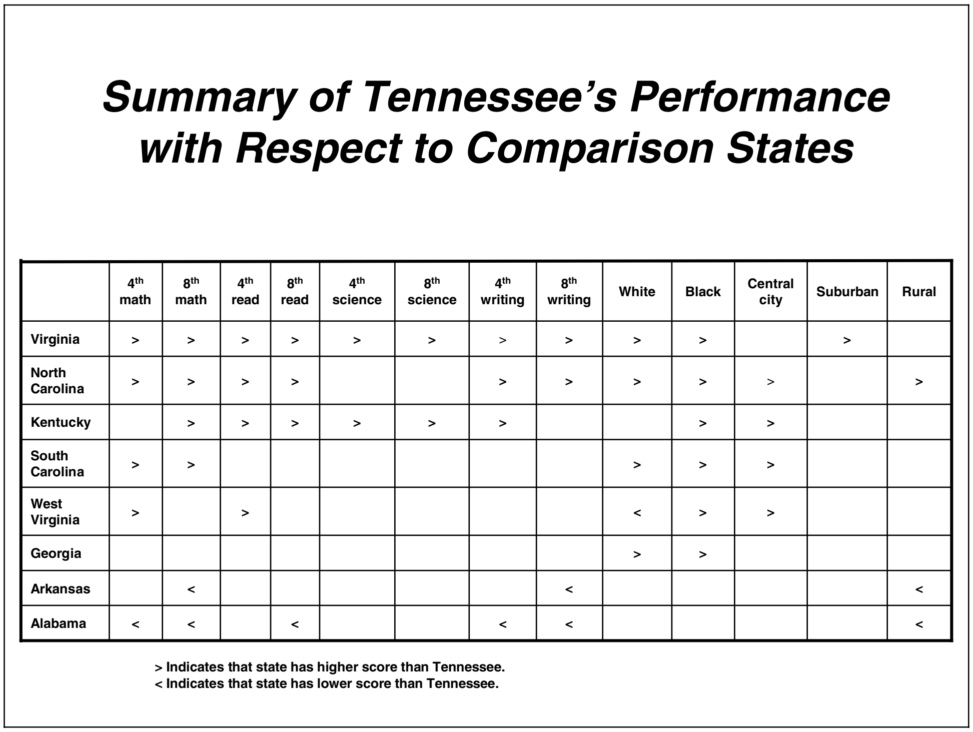
	Rank out of 48 states (*=estimated)	Percentile-point difference from national average (*=statis. significant)	Number of comparison states with statistically significant higher scores	Number of comparison states with statistically significant lower scores
4th math	42	7.5*	4	1
8th math	41	7.5*	4	2
4th reading	41	4.1*	4	0
8th reading	38	3.1*	3	1
4th writing	33–34*	3.7*	3	1
8th writing	34–35*	3.6*	2	2
4th science	36–37*	1.1	3	0
8th science	36–37*	3.0	2	0

We use four measures to summarize Tennessee's performance on the most recent NAEP tests in four subjects and two grades. The measures are (1) Tennessee’s rank, (a rough approximation for writing and science tests assuming that 48 states took all tests), (2) the percentile-point difference in Tennessee’s score from the national average, (3) the number of comparison states that scored statistically higher than Tennessee, and (4) the number of comparison states scoring statistically significantly lower.

Overall, there is much more similarity than difference in performance across grades and tests. Tennessee tends to rank between 33rd and 42nd among states across subjects and grades. It also has scores that are usually below the national averages and usually statistically significant, has statistically significant and lower scores than two to four of the comparison states, and only has statistically significant and higher scores than two or fewer comparison states.

The data would suggest that math performance might be somewhat weaker than reading performance at both the 4th and 8th grade levels, and that performance in writing and science tends to be stronger than performance in reading. To the extent that differences do exist, they tend to be between subjects rather than grades.

One hypothesis for Tennessee’s relatively better performance in science and writing is that the state has tested students in those subjects through state-administered tests at least since the early 1990s. Few other states have tested writing or science skills in elementary or middle school during that period. Perhaps in those states with no writing or science tests, those subjects were not emphasized as much as reading and math, while Tennessee emphasized them more, perhaps with less emphasis on reading and math. Some evidence exists that teachers emphasize and spend more time on subjects that are tested statewide (Stecher, 2002; Hamilton, 2003).

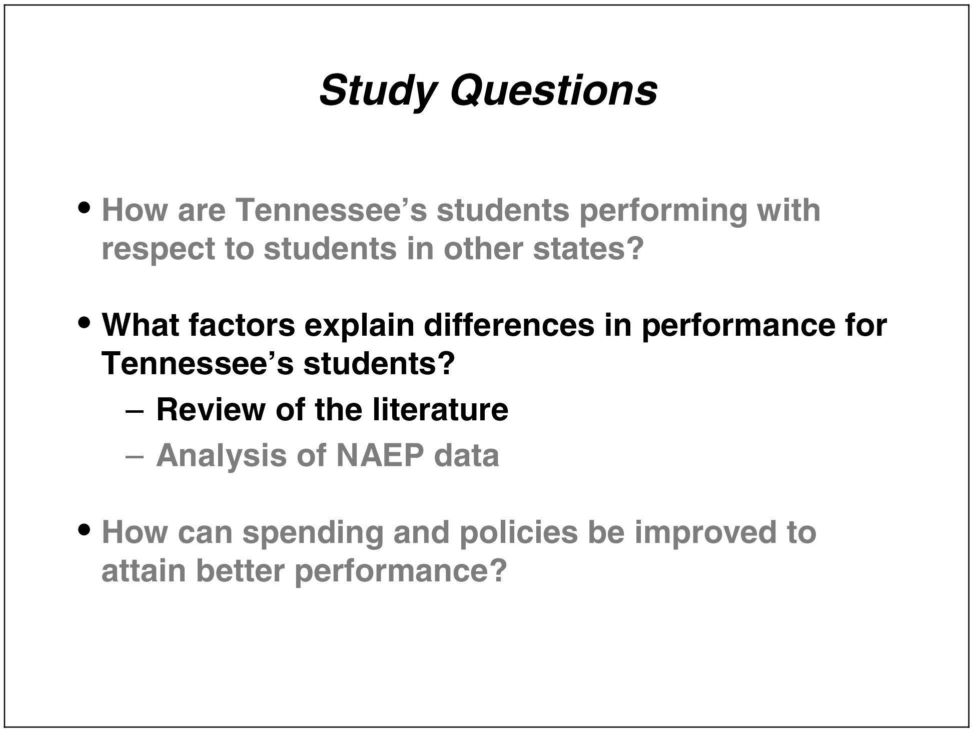


Summary of Tennessee's Performance with Respect to Comparison States

	4th math	8th math	4th read	8th read	4th science	8th science	4th writing	8th writing	White	Black	Central city	Suburban	Rural
Virginia	>	>	>	>	>	>	>	>	>	>		>	
North Carolina	>	>	>	>			>	>	>	>	>		>
Kentucky		>	>	>	>	>	>			>	>		
South Carolina	>	>							>	>	>		
West Virginia	>		>						<	>	>		
Georgia									>	>			
Arkansas		<						<					<
Alabama	<	<		<			<	<					<

> Indicates that state has higher score than Tennessee.
< Indicates that state has lower score than Tennessee.

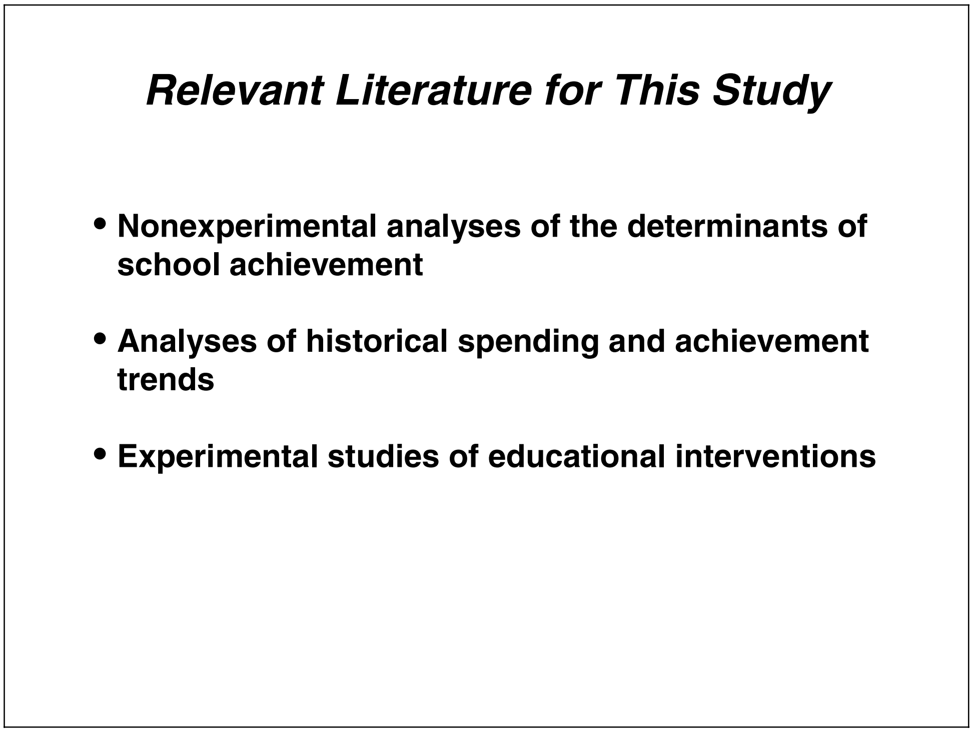
With respect to performance among the comparison states, Virginia, North Carolina, and Kentucky have statistically significant higher scores than Tennessee across most tests and population subgroups. Tennessee only has statistically significant higher scores than one other state, Alabama, across most tests and population subgroups. The remaining states, South Carolina, West Virginia, Georgia, and Arkansas, usually have no consistent statistical differences from Tennessee.



Study Questions

- How are Tennessee's students performing with respect to students in other states?
- **What factors explain differences in performance for Tennessee's students?**
  - **Review of the literature**
  - Analysis of NAEP data
- How can spending and policies be improved to attain better performance?

We now turn to developing and assessing the evidence for possible explanations for the relative performance of Tennessee students on the NAEP tests. We first provide a brief review of the status of the experimental and nonexperimental literature that considers the effect of specific policies or interventions on student achievement. This literature motivates our empirical analysis of the NAEP and provides a basis for comparison with the results of our statistical model.



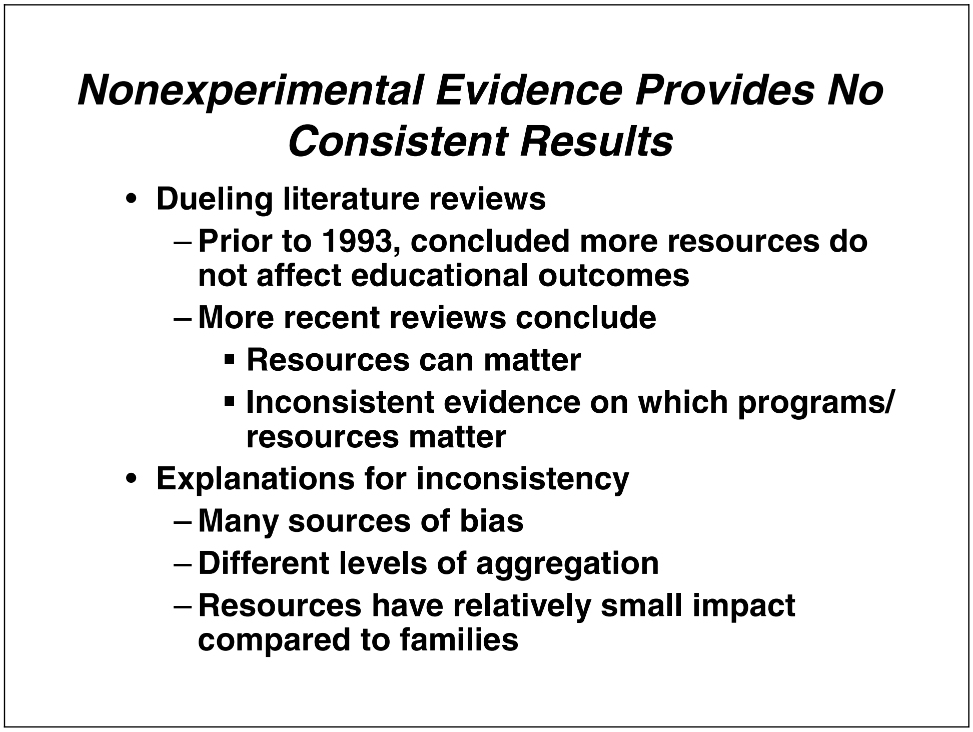
Relevant Literature for This Study

- **Nonexperimental analyses of the determinants of school achievement**
- **Analyses of historical spending and achievement trends**
- **Experimental studies of educational interventions**

To set the stage for our empirical analysis, we briefly review several relevant strands of the research literature on educational inputs and student outcomes. First, we discuss findings from nonexperimental studies mostly conducted before the early 1990s. Several published syntheses of this literature conclude that nonexperimental studies demonstrate that additional resources do not improve educational outcomes. Although the negative conclusion is often cited, other studies review the same literature using different synthesis methods and come to the conclusion that resources do indeed affect outcomes.

The second strand of research we review is the literature from studies of trends in resources and trends in SAT scores. These studies showed that from about 1967 to 1990, SAT scores fell while expenditures rose sharply. Again, this was cited as evidence that resources do not matter. However, recent re-analyses of this data also find that the early conclusions are not readily supported.

We then turn to studies that altered resources under experimental conditions. These studies address limitations of nonexperimental studies and provide compelling evidence that some expenditures can have a significant effect on student outcomes. They also suggest that effects are largest among minority and low-income students.



Nonexperimental Evidence Provides No Consistent Results

- **Dueling literature reviews**
  - **Prior to 1993, concluded more resources do not affect educational outcomes**
  - **More recent reviews conclude**
    - **Resources can matter**
    - **Inconsistent evidence on which programs/ resources matter**
- **Explanations for inconsistency**
  - **Many sources of bias**
  - **Different levels of aggregation**
  - **Resources have relatively small impact compared to families**

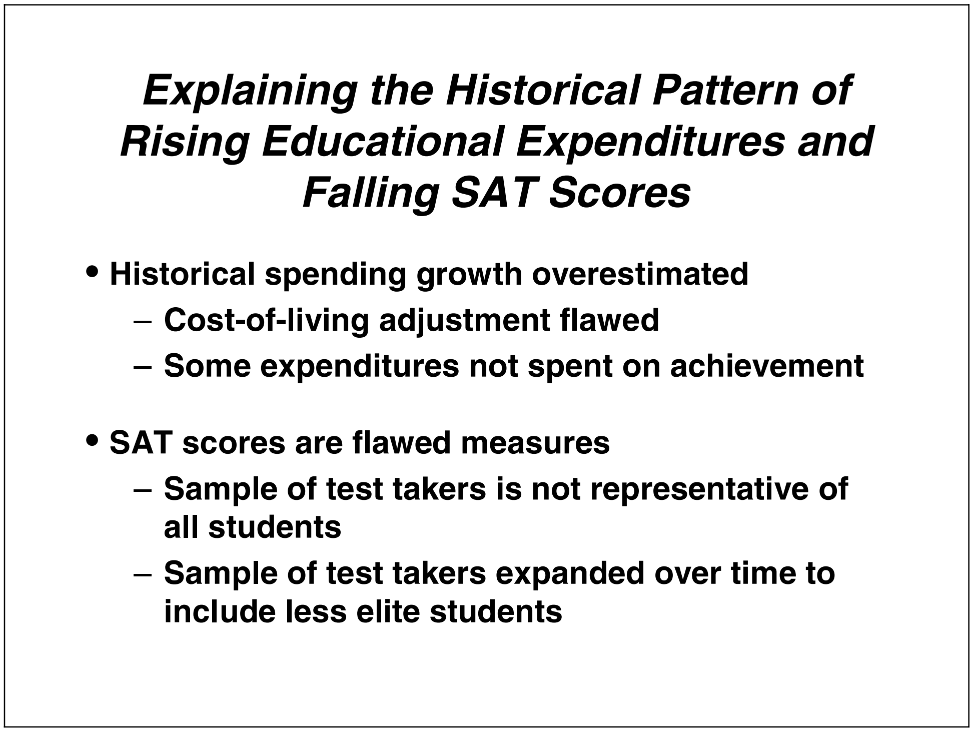
There is an ongoing debate in the academic literature about what conclusions, if any, can be drawn from the previous nonexperimental research that aims to measure the effect of various state and local policies on student achievement. Dueling literature reviews have provided ammunition for this debate. However, the heart of the debate is really about the quality and validity of the assumptions underlying the previous nonexperimental research. One focus of the argument has been whether such nonexperimental research is flawed, whereas experimental evidence can provide more consistent results (Thomas Cook, 2002).

The dominant literature reviews done before 1990 suggested that additional funding does not improve educational outcomes in K–12 education (Hanushek, 1989). Several updated literature reviews by the same author using additional studies, but the same methods for choosing and weighting previous studies, came to similar conclusions (Hanushek, 1994; 1996).

Other literature reviews by other authors since 1992 have concluded that the evidence supports a positive relationship between resources and educational outcomes (Hedges, Laine, and Greenwald, 1994; Greenwald, Hedges, and Laine, 1996; Krueger, 2000). These reviews used more refined statistical tests and different weighting procedures; they were also more discriminating in choosing measurements to include in the review. However, all reviews showed a large variance in outcomes, suggesting that the nonexperimental research was flawed and could not reliably inform policymakers about what specific uses of resources would be effective and efficient in improving student performance.

There are several possible explanations for such inconsistency: bias from missing variables and selectivity, use of different levels of aggregation in analyses, and the often low power of the samples considering the relatively small effect sizes of most educational interventions. No systematic analysis has been done to attempt to explain such inconsistency. One study did suggest that more consistent measurements emerged from analyses of data aggregated at the state level, as opposed to data measured at the school district, school, or student level (Hanushek, Rivkin, and Taylor, 1996). Grissmer, Flanagan, et al. (2000) suggested that a possible reason was the canceling of zero sum selectivity effects at higher levels of aggregation. But there is no consensus on any of these questions in the research community.

However, well-designed experiments could potentially eliminate most of these problems. In recent years, the experimental literature has been given much greater weight than the nonexperimental literature due to its theoretically superior design. However, while experimental results may usually provide more reliable results within the context in which they are derived, their reliability is more problematical when these results are used to predict effects in different contexts.



Explaining the Historical Pattern of Rising Educational Expenditures and Falling SAT Scores

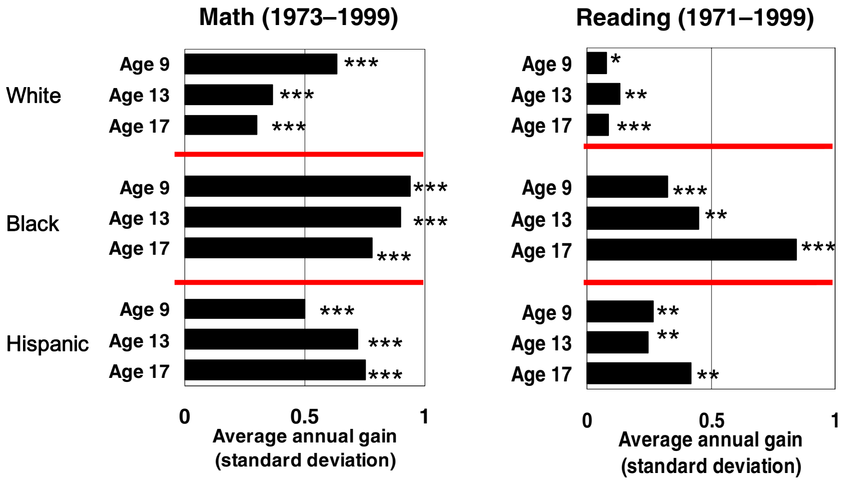
- **Historical spending growth overestimated**
  - **Cost-of-living adjustment flawed**
  - **Some expenditures not spent on achievement**
- **SAT scores are flawed measures**
  - **Sample of test takers is not representative of all students**
  - **Sample of test takers expanded over time to include less elite students**

Major support for the "money doesn't matter" hypothesis also came from another source: the historical evidence from the late 1960s to the early 1990s of falling SAT scores during periods of rapidly rising educational funding (Hanushek, 1994; Hanushek and Jorgenson, 1996). However, new research evidence suggests that this interpretation was flawed. This new evidence indicates that significantly less growth occurred in educational funding and that SAT scores were flawed measures of the trends in achievement in the nation.

Rothstein and Miles (1995) analyzed educational expenditures from 1967 to 1992. The authors suggested that the “real” increase in per pupil expenditure was overestimated for two reasons: a flawed cost-of-living adjustment and inclusion of expenditure growth not directed at achievement gains. The authors concluded that schools received about a 33 percent increase to improve the achievement of regular students rather than the 100 percent commonly cited. (Other spending was directed, for example, at special education students.) The authors also concluded that much of the growth in expenditures was the result of spending on programs expected to primarily benefit minority and disadvantaged students.

SAT tests are flawed measures of national achievement trends. Students taking SAT tests are not chosen randomly; the exams are taken only by students applying to college, a population that grew from 33 percent of high school seniors in 1967 to 43 percent as of 1994. This nonrepresentative sample changed significantly between those years as it broadened to encompass the changing composition of college-going students, including more women and minorities. These changes in the composition of test takers would be expected to have lowered average SAT scores. Moreover, as discussed on the next slide, the students who were in subgroups less likely to take the SAT actually made the largest score gains on tests representative of all students.

NAEP Scores Made Gains From 1971 to 1999 for Every Age and Racial/Ethnic Group



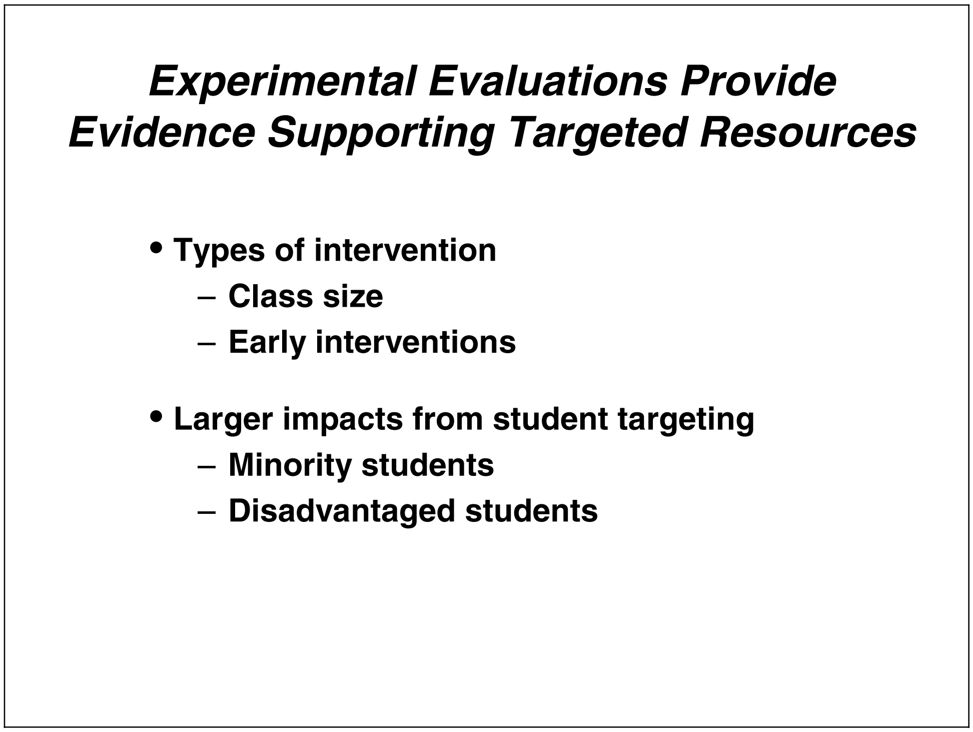
NOTE : Statistical significance: *** 1 percent; ** 5 percent; * 10 percent.

Fortunately, we do not have to depend on SAT scores to inform us about the trends in scholastic achievement in the United States. The NAEP tests have been given in reading and math since the early 1970s using representative samples of U.S. students at ages 9, 13, and 17. Such scores show different trends than the SAT scores.

This chart uses data from the NAEP scores given from the early 1970s to 1999 (see Campbell, Hombo, and Mazzeo, 2000). It fits simple trends to these data to derive annual score gains. The results show the annual average gain in scores for each age, subject, and racial/ethnic group. The data show that every group made statistically significant improvement in each subject during that time period. Math scores improved more than reading scores for nearly every age and racial/ethnic group.

The data also show that black and Hispanic students made much larger gains than white students in nearly every age group and subject. The gains among black students are especially large. Black 17-year-old students gained nearly 25 percentile points over the entire time interval in both reading and math.

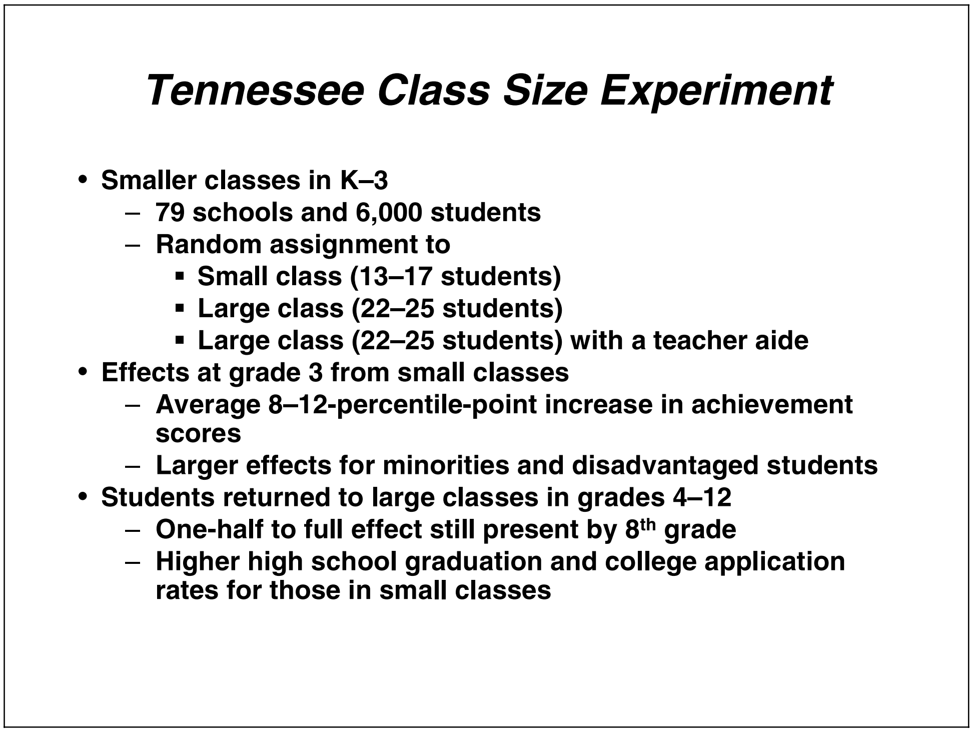
The large gains made by black students combined with the smaller gains by white students made the black-white gap in scores decline by between onequarter and one-half, depending on the age group and subject. Other research has shown that the gains among white students were larger for disadvantaged white students than advantaged white students (Hedges and Nowell, 1998). Thus, the smallest gains over this period were made by advantaged white students.



Experimental Evaluations Provide Evidence Supporting Targeted Resources

- **Types of intervention**
  - **Class size**
  - **Early interventions**
- **Larger impacts from student targeting**
  - **Minority students**
  - **Disadvantaged students**

Around 1995 and thereafter, the inconsistency of nonexperimental results led researchers to focus on the potentially more reliable results from experimental evaluations. However, few experimental results were available and they did not cover the full range of educational policies that might be of interest to policymakers. Experimental designs have been used to evaluate the effects of reducing class size, as well as the effects of several early childhood interventions. The results from these experiments showed large and significant positive effects on several different educational outcome measures as a result of reducing class size and providing services to children and families during early childhood. The results of these experiments also suggested that interventions targeted toward minority and disadvantaged students had greater effects compared with the effects of providing the same interventions to more advantaged children.



Tennessee Class Size Experiment

- **Smaller classes in K–3**
  - **79 schools and 6,000 students**
  - **Random assignment to**
    - **Small class (13–17 students)**
    - **Large class (22–25 students)**
    - **Large class (22–25 students) with a teacher aide**
- **Effects at grade 3 from small classes**
  - **Average 8–12-percentile-point increase in achievement scores**
  - **Larger effects for minorities and disadvantaged students**
- **Students returned to large classes in grades 4–12**
  - **One-half to full effect still present by 8th grade**
  - **Higher high school graduation and college application rates for those in small classes**

The Tennessee class size experiment assigned children randomly to small or large classes in early grades and provided evidence that long-term achievement could be improved through reduced class size in early grades (Finn and Achilles, 1999; Krueger, 1999; Nye, Hedges, and Konstantopoulos, 1999).

The experiment started at the kindergarten level, with students assigned randomly to small classes averaging 16 students per class, large classes averaging 24 students per class, and large classes with teacher aides. The class size experiment was maintained through 3rd grade. At that point, all pupils were returned to larger classes.

The experiment showed higher achievement in small classes at grade three by an average 8–12 percentile points for the students in small classes. However, minority and free-lunch students made larger gains than the typical student (Finn and Achilles, 1999; Krueger, 1999). Moreover, research also suggests that the higher achievement lasted through 8th grade, and that those in smaller classes had higher rates of high school graduation and college entrance (Nye, Hedges, and Konstantopoulos, 1999; Krueger and Whitmore, 2001). Additional evidence indicates that higher achievement at the 8th grade level only occurred among students who were in smaller classes for three or four years (Nye, Hedges, and Konstantopoulos, 1999). Molnar et al. (1999) also report the results of a pupil-teacher ratio quasi-experiment from Wisconsin that showed similar results. Grissmer (1999) provides a review of the class size literature and possible explanations for differences between experimental effects and nonexperimental effects.

The Tennessee experiment also included a random assignment of children to large classes with and without a teacher aide. The effect of teacher aides on achievement was generally positive, but not statistically significant. Grissmer, Flanagan, et al. (2000) carried out a cost-effectiveness analysis using the Tennessee results. This study showed that reducing class size was much more cost-effective in improving achievement than hiring teacher aides.

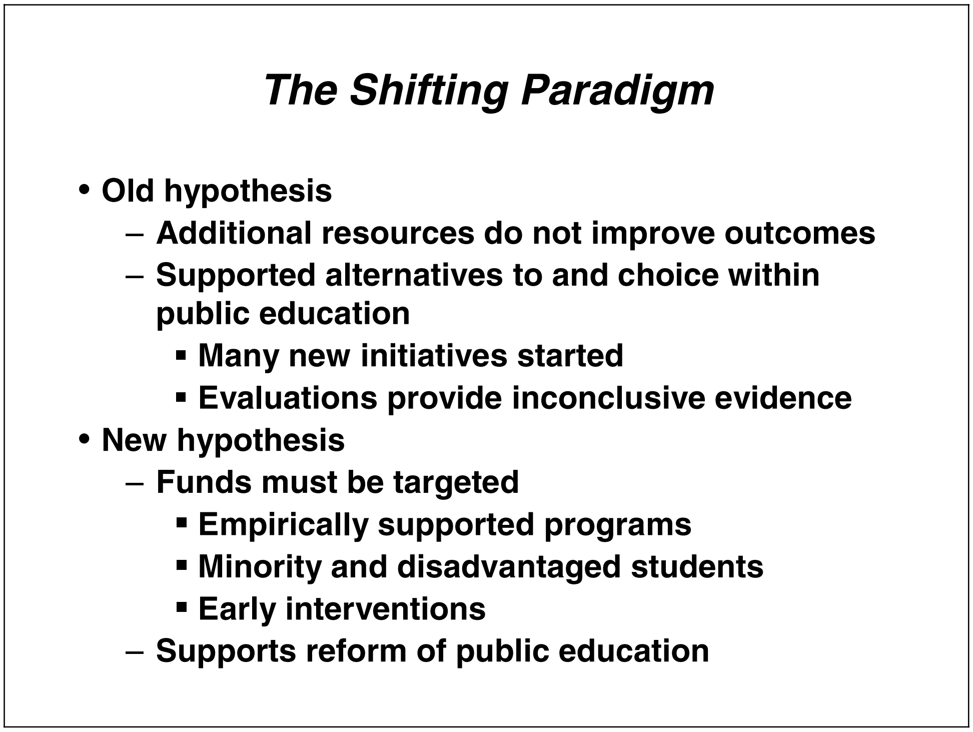
Early Childhood Interventions

- **Early childhood interventions with experimental or quasi-experimental evaluations**
  - **Nurse-Family Partnership**
  - **Abecedarian**
  - **Perry Preschool Project**
  - **Chicago Child-Parent Centers**
- **Results**
  - **Favorable effects on long-term outcomes**
    - **Higher educational achievement and schooling attainment**
    - **Lower special education placement and grade retention**
    - **Lower delinquency and crime**
    - **Lower welfare utilization**
    - **Higher employment and earnings**
  - **Estimated dollar benefits from targeted high-quality programs exceed costs**

This chart lists four early childhood interventions with experimental or quasi-experimental evaluations, ranging from a nurse home-visiting program in the first few years of life to center-based preschool programs one or two years before kindergarten entry. Because these studies included long-term follow-up of participants (between age 15 and age 40), they provide evidence of the potential for improving short- and long-term outcomes through intensive early interventions targeted at disadvantaged children (for descriptions and further references, see Karoly, Greenwood, et al., 1998; Masse and Barnett, 2002; Reynolds et al., 2002; Lynch, 2004; Schweinhart at al., 2005; and Karoly, Kilburn, and Cannon, 2005).

While the outcomes measured in these evaluations varied, at least one or more studies showed improvements for the program participants in the outcomes listed here. These include higher test scores and school attainment (e.g., high school graduation), reductions in grade retention and special education placement, lower involvement in the criminal justice system, lower welfare utilization, and higher employment and earnings. In addition, depending on the nature of the intervention, substantial benefits may also occur for the family and mother, including more employment, less welfare and food stamp utilization, higher education, and higher wages.

Benefit-cost analyses of these four interventions show that the present value of savings associated with the improved outcomes was typically more than $3 for each present-value dollar invested (Karoly, Kilburn, and Cannon, 2005). However, in at least one case, a program that served lower-risk families did not show benefits exceeding costs (Karoly, Greenwood, et al, 1998). Perhaps as importantly, the long–net term annual savings in public social and educational costs for a high-quality preschool program serving the 20 percent of children who were most disadvantaged was estimated at $31 billion by 2030 and $61 billion by 2050. This savings represents about one-fifth of the estimated deficit in the Social Security trust fund for those years (Lynch, 2004).



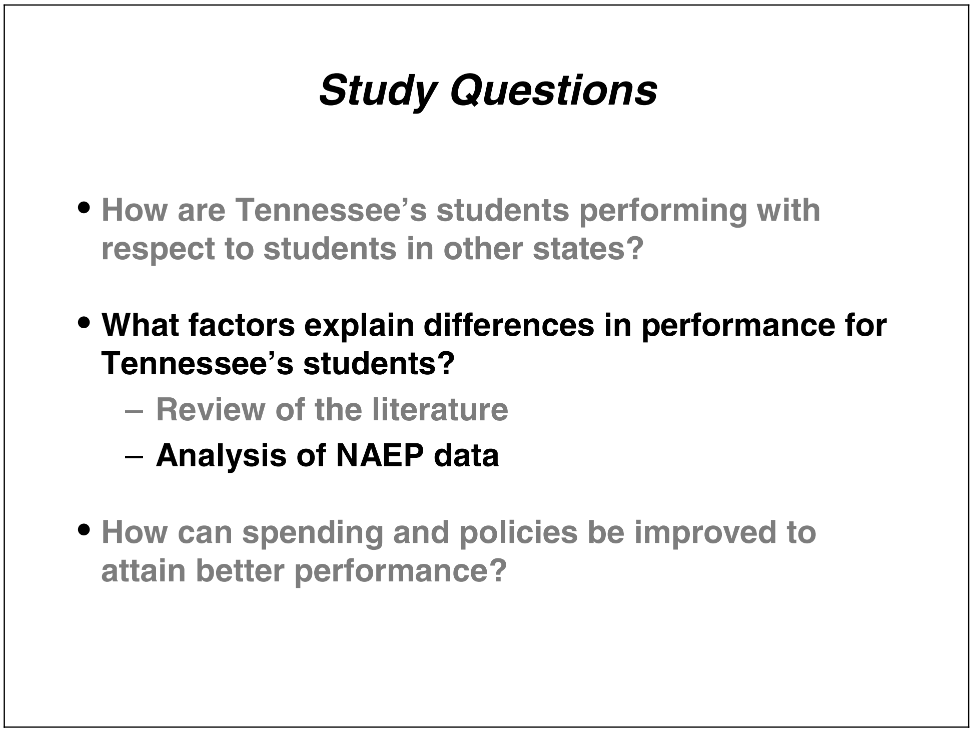
The Shifting Paradigm

- Old hypothesis
  - Additional resources do not improve outcomes
  - Supported alternatives to and choice within public education
    - Many new initiatives started
    - Evaluations provide inconclusive evidence
- New hypothesis
  - Funds must be targeted
    - Empirically supported programs
    - Minority and disadvantaged students
    - Early interventions
  - Supports reform of public education

In sum, the dominant view prior to about 1993 was that the empirical literature provided no consistent evidence that would support the hypothesis that more resources in public education would improve educational outcomes. Such a view supported a policy strategy that encouraged starting and evaluating alternatives that would introduce more competition into public education, as well as more alternatives to public education. These alternatives included charter schools, vouchers, school choice, and contracts to nonprofit and private-sector entities to deliver public education. By and large, such options had not been present previously in American education and a strategy developed to initiate such systems and provide research evidence for their effects.

Two issues are important in this context: whether alternative reforms provide higher performance for an equivalent cost and whether increasing competition raises performance in public schools. The evidence that has emerged so far from the trials has been mixed, generating little scientific consensus (see, for instance, Zimmer et al., 2003; Gill, Timpane, et al., 2001; and Gill, Hamilton, et al., 2005). However, such results may not yet measure long-term results and may not reflect the effects of larger-scale programs.

A more consistent story emerges from experimental evidence. This evidence suggests that resources can improve outcomes, but primarily when targeted to proven programs, minority and disadvantaged students, and early interventions (Heckman and Masterov, 2004). These hypotheses are also consistent with the re-interpretation of historical trends in expenditures and achievement.

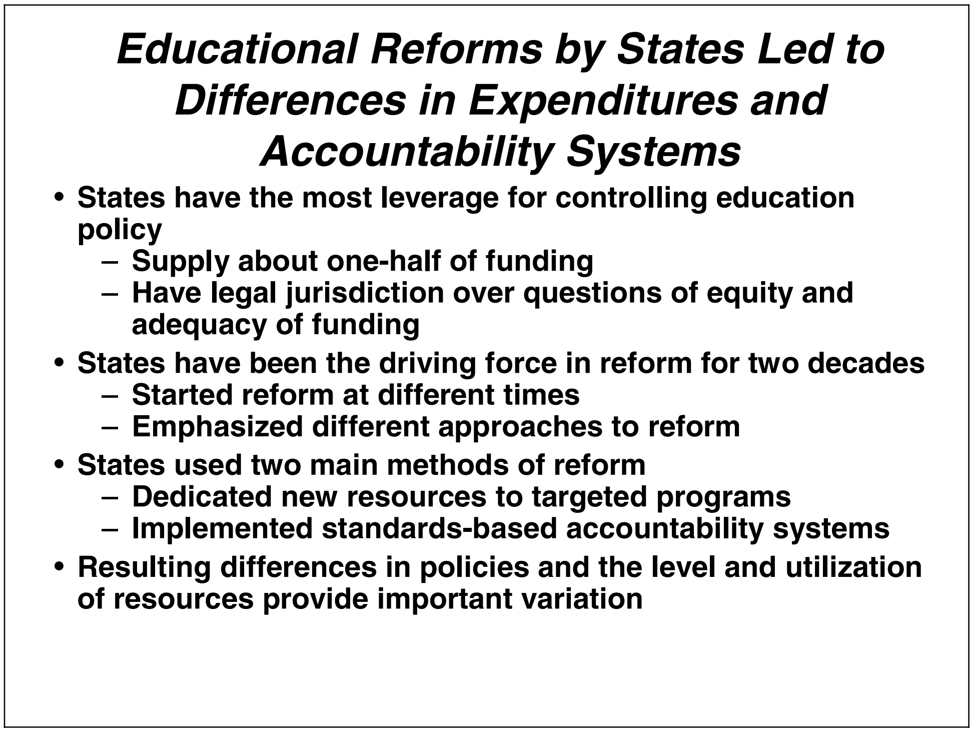


Study Questions

- How are Tennessee's students performing with respect to students in other states?
- **What factors explain differences in performance for Tennessee's students?**
  - Review of the literature
  - **Analysis of NAEP data**
- How can spending and policies be improved to attain better performance?

With this background, we now turn to our analysis of the NAEP tests in 4th and 8th grade reading and math given across states from 1990 to 2003, a period when intense effort was made across states to reform education. These data are used to explain the variation in achievement scores across states in terms of differences in students' characteristics and the resources allocated to education in the state, with a specific focus on the importance of these explanatory factors for outcomes in Tennessee.

This analysis updates an earlier study, which used a similar methodology but included only seven NAEP tests administered between 1990 and1996 (Grissmer, Flanagan, et al., 2000). Additional details on the methodology and regression results are presented in the Appendix.



Educational Reforms by States Led to Differences in Expenditures and Accountability Systems

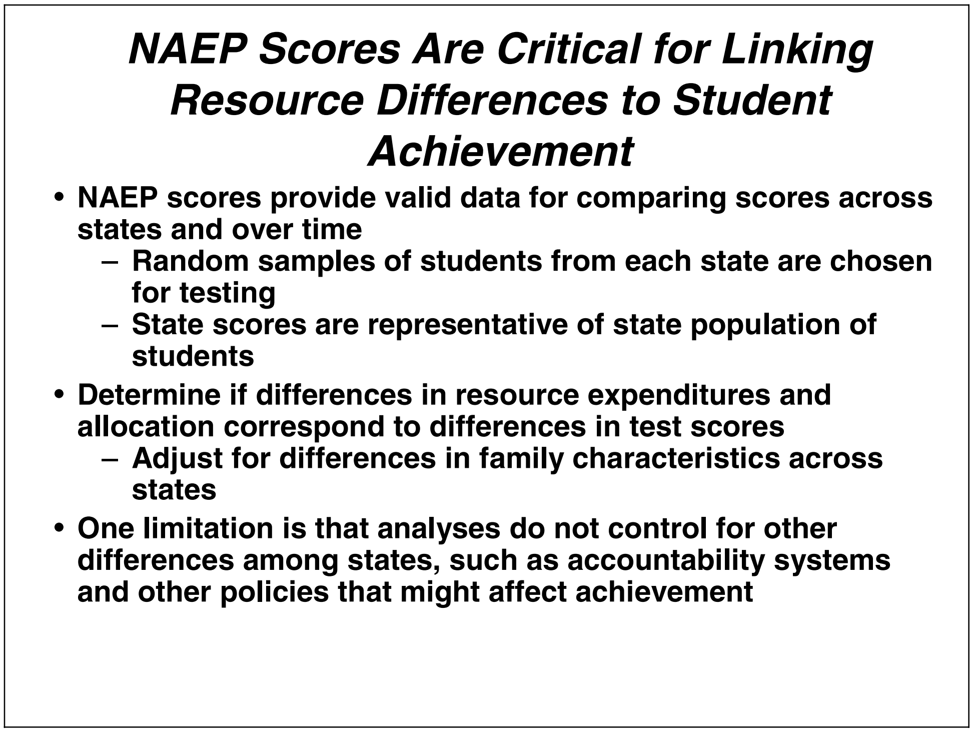
- **States have the most leverage for controlling education policy**
  - **Supply about one-half of funding**
  - **Have legal jurisdiction over questions of equity and adequacy of funding**
- **States have been the driving force in reform for two decades**
  - **Started reform at different times**
  - **Emphasized different approaches to reform**
- **States used two main methods of reform**
  - **Dedicated new resources to targeted programs**
  - **Implemented standards-based accountability systems**
- **Resulting differences in policies and the level and utilization of resources provide important variation**

States took the initiative in reforming education after the *Nation at Risk* report was published in 1983 (National Commission on Excellence in Education, 1983). States have the leverage to reform schools since states provide the largest share of resources—typically between 40 and 60 percent—and also have been given responsibility by the courts for creating equity and adequacy in educational funding. The federal government has historically had limited leverage in education policy because it provides less than 10 percent of K–12 funding, and the *San Antonio Independent School District v. Rodriguez* decision (Supreme Court, 1973) specified that questions of adequacy and equity in educational funding were not a matter for the federal courts. Thus, from 1983 to 2003, prior to the passage of the federal No Child Left Behind (NCLB) act, states led the effort to improve K–12 education, with variation across states and over time in the mix of policy changes and reforms implemented.

The states have used two major strategies to improve education. The first is resource-intensive and has involved increasing funding for education and/or reallocating funds to different programs. States have increased per pupil expenditures significantly between 1983–1984 and 2003–2004 and many have increased their relative allocations to poorer school districts. States have utilized their increased resources in a variety of ways—from reducing class size mainly at elementary levels to increasing kindergarten and pre-kindergarten participation, raising teacher pay, and providing teachers more resources and professional development.

The second strategy has been called "systemic standards-based reform" (Smith and O'Day, 1991; O'Day and Smith, 1993; Vinovskis, 1996). Standards-based reform refers to an approach to K–12 education that emphasizes the setting of measurable standards; assessments to measure progress toward standards; and alignment of curriculum, professional development, incentives, and teacher training. Nearly every state has established a standards-based accountability system, although there are many differences across states in such systems.

As a result of these initiatives, there are substantial differences across states and over time in their educational policies and the level and utilization of resources. These differences, which are the focus of our empirical analysis, are further highlighted below.



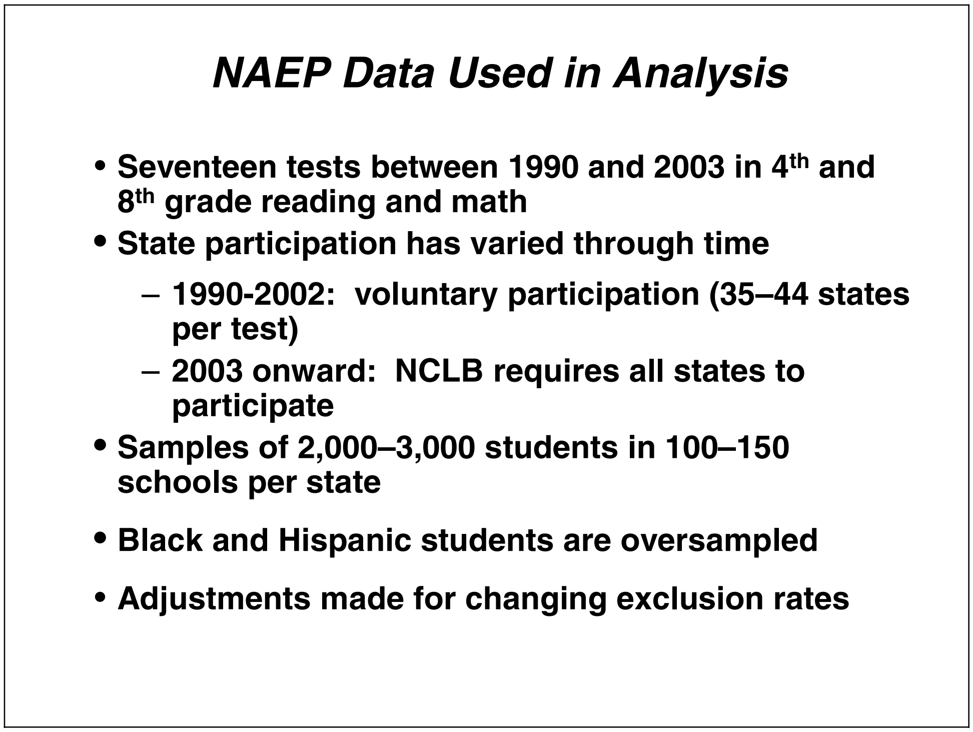
NAEP Scores Are Critical for Linking Resource Differences to Student Achievement

- **NAEP scores provide valid data for comparing scores across states and over time**
  - **Random samples of students from each state are chosen for testing**
  - **State scores are representative of state population of students**
- **Determine if differences in resource expenditures and allocation correspond to differences in test scores**
  - **Adjust for differences in family characteristics across states**
- **One limitation is that analyses do not control for other differences among states, such as accountability systems and other policies that might affect achievement**

If resources affect student outcomes, then we would expect states with different resource levels and allocations to have different levels of student achievement, if all else is equal. The NAEP tests allow us to explore this hypothesis by providing statistically valid data that can be used to compare scores across states. The NAEP selects random samples of schools and students from every state, so after controlling for differential exclusion rates, the data represent the achievement of each state's entire population of eligible students.

Using statistical models, we can then determine whether differences among states in the level and allocation of resources explain differences in NAEP scores. However, we need to do our best to make "all else equal" among states to confirm that the differences are due to resources and not some other factor. As described in the Appendix, we use the methods of our previous research on NAEP (Grissmer, Flanagan, et al., 2000) to control for differences in the family characteristics among state populations.

However, our analyses do not explicitly control for some other possible differences among states, including differences in state accountability systems and social or other policies that might also affect student achievement. To the extent that variations in these other factors are correlated with variation in resource allocations, our analyses might provide over- or underestimates of the effects of resources. Despite this limitation, the analyses provide important guidance about resources.



NAEP Data Used in Analysis

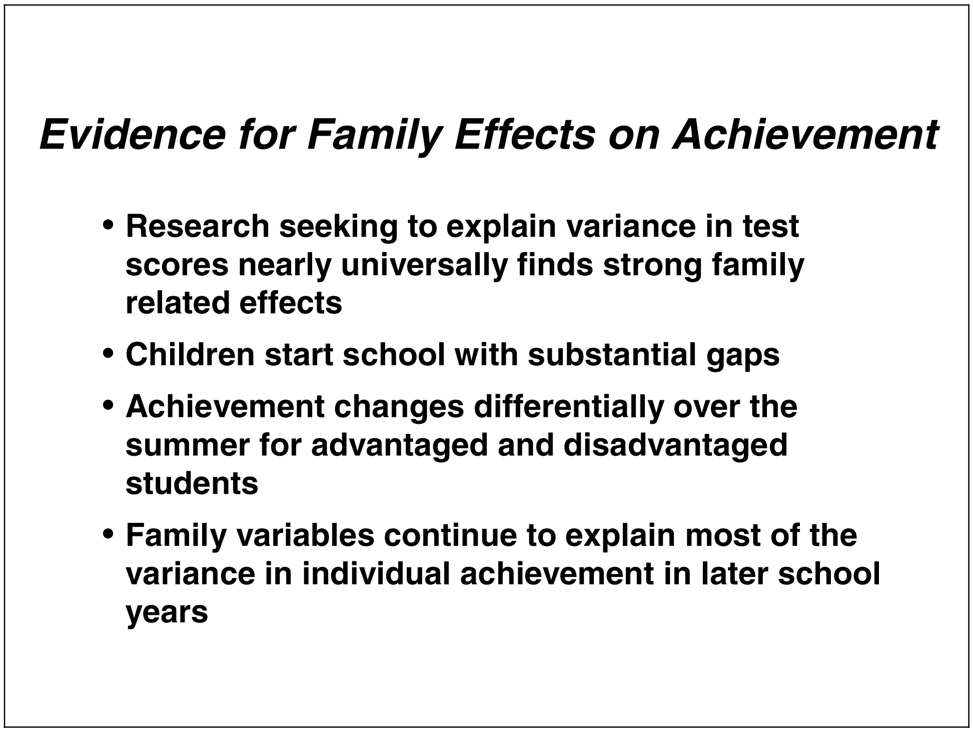
- **Seventeen tests between 1990 and 2003 in 4th and 8th grade reading and math**
- **State participation has varied through time**
  - **1990-2002: voluntary participation (35–44 states per test)**
  - **2003 onward: NCLB requires all states to participate**
- **Samples of 2,000–3,000 students in 100–150 schools per state**
- **Black and Hispanic students are oversampled**
- **Adjustments made for changing exclusion rates**

We rely on data from 17 NAEP tests implemented between 1990 and 2003. During that period, five NAEP 8th grade math tests, four 4th grade math tests, five 4th grade reading tests, and three 8th grade reading tests were administered across states. Before 2003, state participation was voluntary and between 35 and 44 states participated in a given test. Beginning in 2003, because of the NCLB legislation, all states were required to participate. (See Table A.1 in the Appendix for additional details.) A total of 700 state scores are analyzed in the models.

The state NAEP samples include for each state approximately 2,000–3,000 students from 100–150 schools in each state. Black and Hispanic students are over-sampled. Sample sizes were increased somewhat in 2003, and a pattern of testing 4th and 8th graders in math and reading every two years was initiated.

Our previous analysis utilized state NAEP scores for seven tests from 1990 to 1996 (Grissmer, Flanagan, et al., 2000). The current analysis uses results from 17 tests from 1990 to 2003. The methodology is similar to that used in the previous study, except that we have used scores adjusted for changing exclusion rates (see Appendix). Such adjustments are estimated by imputing missing scores using teacher-supplied information on excluded and included students (McLaughlin, 2000; 2001). The adjustments are relatively small for almost all states and do not affect the estimates of resource or family effects in any significant way.

We now present the results of the NAEP analysis. On prior slides, we presented raw data comparing the performance of Tennessee's students with respect to that of students in other states using the 2003 4th and 8th grade math and reading tests and the most recent writing and science tests. We now focus on differences in family traits across states and follow this analysis with equations to estimate the specific effects on achievement across states attributable to family characteristics, pupil-teacher ratios, teacher salaries, the adequacy of resources reported by teachers, and the proportion of students in public pre-kindergarten. We also transform the estimates from our achievement models to the scale of expected dollars per pupil required to increase student achievement by a given amount.

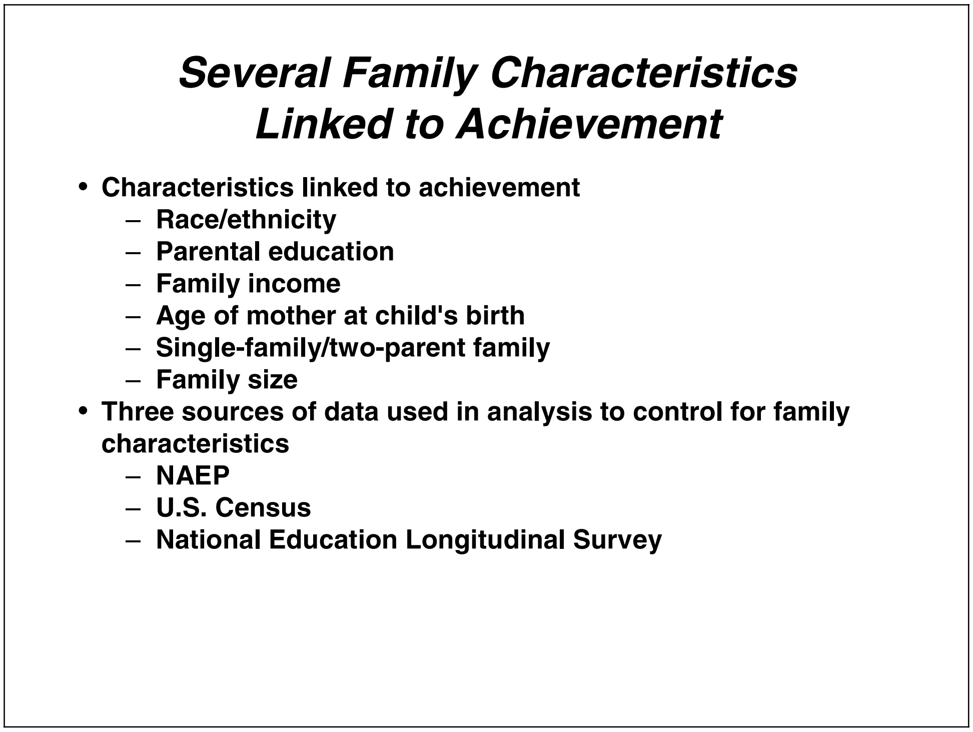


Evidence for Family Effects on Achievement

- **Research seeking to explain variance in test scores nearly universally finds strong family related effects**
- **Children start school with substantial gaps**
- **Achievement changes differentially over the summer for advantaged and disadvantaged students**
- **Family variables continue to explain most of the variance in individual achievement in later school years**

Virtually all research that tries to explain differences in individual test scores finds substantial predictive links to family characteristics. Links to characteristics related to school are generally found to be much weaker. For instance, the Early Childhood Longitudinal Survey of the kindergarten class of 1998–1999 (ECLS-K) shows that significant differences exist in early achievement tests before children begin school and that these differences are linked primarily to family characteristics (Lee and Burkam, 2002). (The ECLS-K is a longitudinal survey of a nationally representative sample of about 20,000 beginning-kindergarten students; see West, Denton, and Germino-Hausken [2000] for a description. Each student was given an extensive set of tests measuring readiness in math and reading at the start of the kindergarten year.)

Research evidence also suggests that achievement changes differentially for advantaged and disadvantaged children over the summer months, when families are the dominant influence (Alexander, Entwisle, and Olson, 2001). Family variables also continue to explain much more of the variance in individual score differences during later school years than variables related to school (see, for instance, Grissmer, Kirby, et al., 1994).



Several Family Characteristics Linked to Achievement

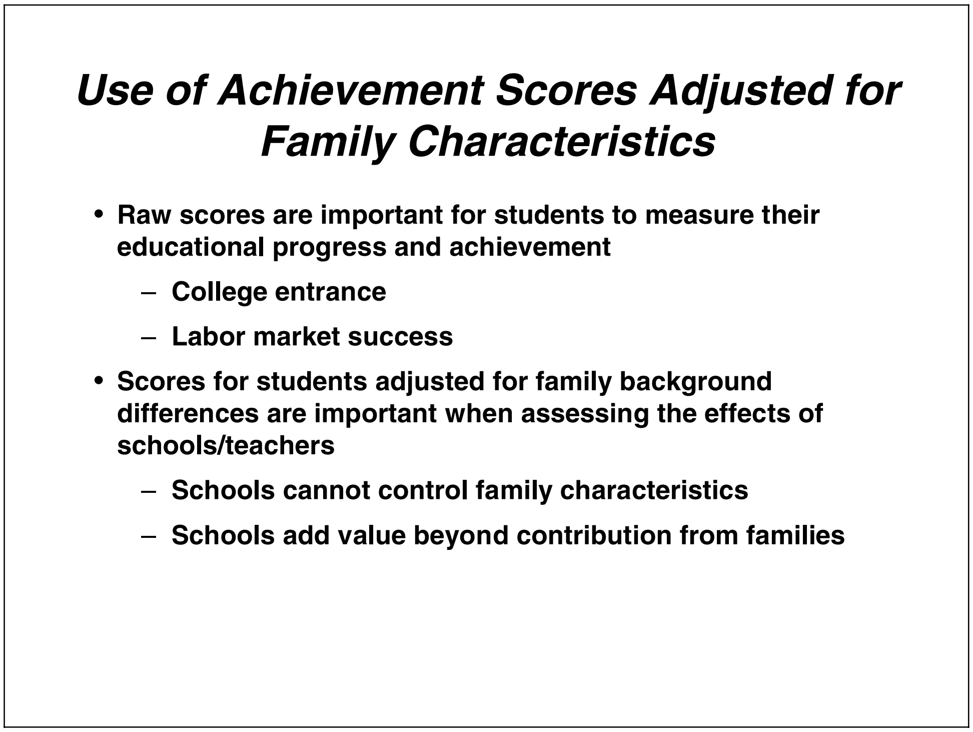
- **Characteristics linked to achievement**
  - **Race/ethnicity**
  - **Parental education**
  - **Family income**
  - **Age of mother at child's birth**
  - **Single-family/two-parent family**
  - **Family size**
- **Three sources of data used in analysis to control for family characteristics**
  - **NAEP**
  - **U.S. Census**
  - **National Education Longitudinal Survey**

What particular family characteristics show predictive links to achievement? Multiple studies of achievement have shown that the family variables consistently linked to achievement include race/ethnicity, education levels of both the mother and father, family income, age of the mother at the time of the child's birth, family structure (one- or two-parent), and number of siblings (see, for instance, Grissmer, Kirby, et al., 1994; Haveman and Wolfe, 1995; Hedges and Nowell, 1998; and Lee and Burkam, 2002).

If there are major differences in these family characteristics across states, these differences would be expected to account for part of the differences in state performance. It would be important to take into account such wide differences in family characteristics across states before making judgments about school performance because schools cannot control or influence the family characteristics of their students.

We use three sources of data in this study to characterize family characteristics and their predictive effects on achievement across states: U.S. Census data; student-reported data from NAEP; and parent-reported data from a large longitudinal study of 8th grade students, the National Education Longitudinal Survey.

We utilize these data to develop estimates of the characteristics of families with children in 4th and 8th grades. We also use these data to determine how much weight should be given to each family characteristic in accounting for score differences and to develop a composite measure of socioeconomic status (SES) as a measure of educational risk (see Grissmer, Flanagan, et al., 2000, for the detailed methodology of this index, which is defined there as the SES fixed effects [SES-FE] measure). The family SES variable in our model changes over time to the extent that the racial/ethnic characteristics of students taking the NAEP change over time. However, changes in characteristics with regard to race/ethnic groups are not tracked in the model.



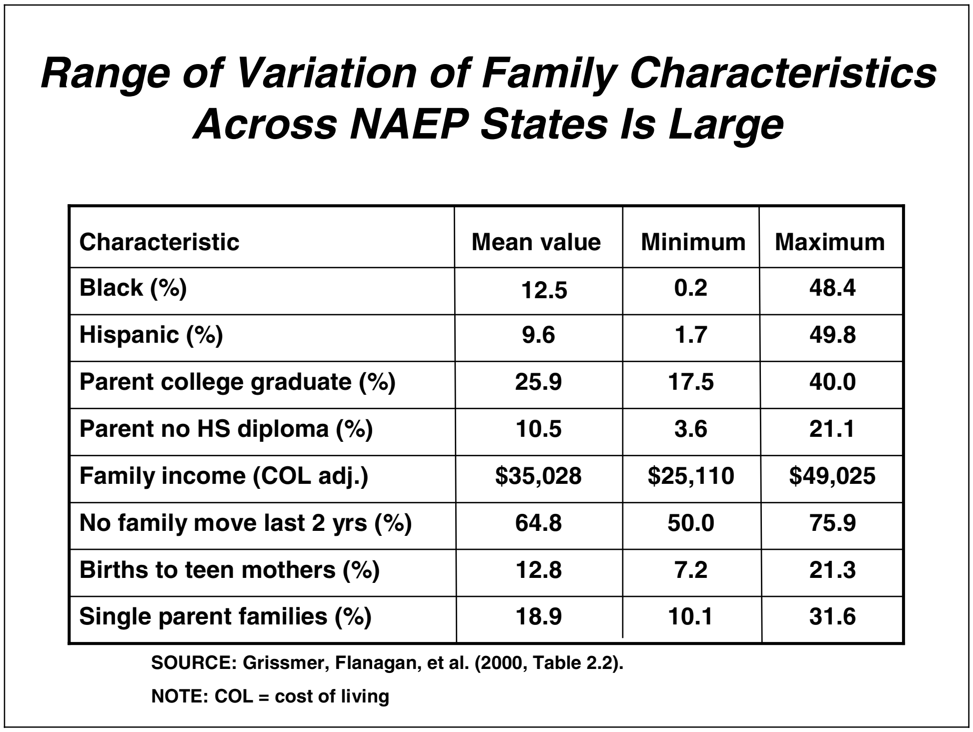
Use of Achievement Scores Adjusted for Family Characteristics

- **Raw scores are important for students to measure their educational progress and achievement**
  - **College entrance**
  - **Labor market success**
- **Scores for students adjusted for family background differences are important when assessing the effects of schools/teachers**
  - **Schools cannot control family characteristics**
  - **Schools add value beyond contribution from families**

It is important to distinguish between appropriate uses of different measures of student achievement. Raw achievement scores are the most important scores for students and families because these scores are a measure of students' educational progress and achievement with implications for subsequent outcomes, such as college admission and, ultimately, labor market success.

However, when we want to assess the effect of teachers, schools, and school districts on student achievement, a better measure is to compare the scores of students holding constant differences in family characteristics. Teachers, schools, and school districts cannot change the characteristics of their students, but given these characteristics, they are responsible for raising the scores of all students and erasing the gaps in scores between minorities and non-minorities and advantaged and disadvantaged students. Equivalent standards need to be set for all students.

In our analysis, the use of the state-level composite measure of educational risk is consistent with these objectives. Such measures can help to identify the characteristics of schools, teachers, and educational programs that are effective in raising scores for students after netting out the effects of average family characteristics at the state level.



Range of Variation of Family Characteristics Across NAEP States Is Large

Characteristic	Mean value	Minimum	Maximum
Black (%)	12.5	0.2	48.4
Hispanic (%)	9.6	1.7	49.8
Parent college graduate (%)	25.9	17.5	40.0
Parent no HS diploma (%)	10.5	3.6	21.1
Family income (COL adj.)	$35,028	$25,110	$49,025
No family move last 2 yrs (%)	64.8	50.0	75.9
Births to teen mothers (%)	12.8	7.2	21.3
Single parent families (%)	18.9	10.1	31.6

SOURCE: Grissmer, Flanagan, et al. (2000, Table 2.2).

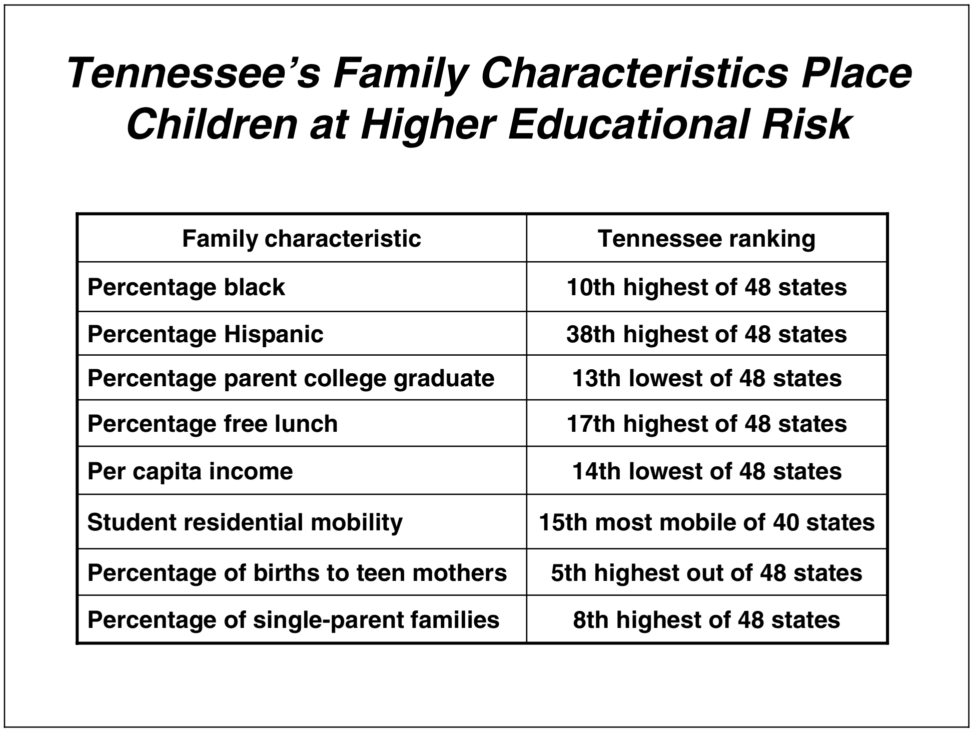
NOTE: COL = cost of living

The risk that differences in family characteristics among states will obscure the effects of differences in policies depends partly on how large the differences are in family characteristics and educational spending and policies across states. Family characteristics across states vary markedly, as shown here for indicators measured as of the early 1990s (see Grissmer, Flanagan, at al., 2000, Table 2.2, for sources for these indicators and for state values for these variables). For example, the racial/demographic composition of states varies dramatically; less than 1 percent of families being black in many small northern states, compared with almost 50 percent in Mississippi. The Hispanic population is less than 5 percent in most states, but almost 50 percent in New Mexico.

States also vary considerably in the average level of parental education. For instance, the percentage of parents who are college graduates varies from 17 to 20 percent in Mississippi, Arkansas, and West Virginia to 36 to 40 percent in Utah, New Hampshire, and Connecticut. The percentage of parents with no high school diploma varies from less than 5 percent in Utah, Nebraska, North Dakota, and Minnesota to about 20 percent in Texas, California, and Mississippi.

The cost-of-living-adjusted family income varies significantly, from less than $28,000 in Mississippi and Arkansas to more than $48,000 in Connecticut and New Jersey. Family mobility—the percentage of families that did not move in the last two years—varies from less than 55 percent in California, Texas, and Florida to more than 70 percent in Iowa, Maine, and New Hampshire.

States also differ in the percentage of births to teen mothers—a proxy for the mother's age at the time of the child's birth. In New Hampshire, Massachusetts and Minnesota, the percentage of births to teenaged mothers is 7–8 percent, while in West Virginia, Arkansas, Alabama and Mississippi, it reaches 18–21 percent. Finally, the percentage of single-parent families varies from 10 to 32 percent.



Tennessee’s Family Characteristics Place Children at Higher Educational Risk

Family characteristic	Tennessee ranking
Percentage black	10th highest of 48 states
Percentage Hispanic	38th highest of 48 states
Percentage parent college graduate	13th lowest of 48 states
Percentage free lunch	17th highest of 48 states
Per capita income	14th lowest of 48 states
Student residential mobility	15th most mobile of 40 states
Percentage of births to teen mothers	5th highest out of 48 states
Percentage of single-parent families	8th highest of 48 states

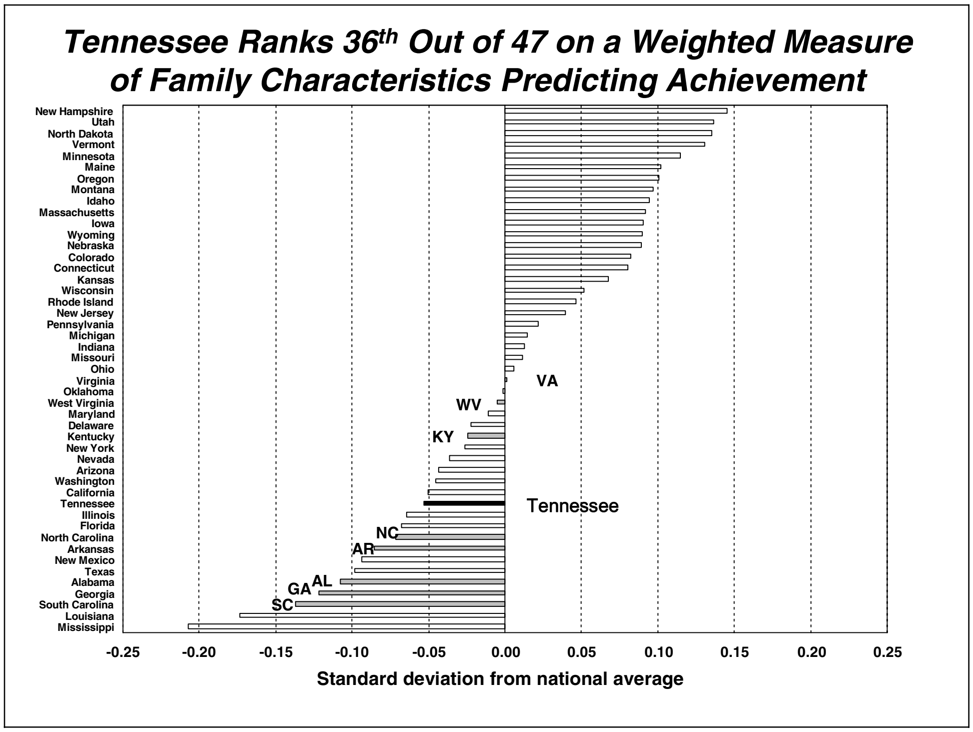
This chart summarizes the characteristics of Tennessee’s families in comparison with other states, again for indicators measured as of the early 1990s. Family characteristics that place a state’s children at higher educational risk include a greater proportion of black or Hispanic children, lower parental education, and lower family income, as well as a higher fraction of free-lunch children. Finally, higher residual mobility, a higher percentage of births to teen mothers, and a higher proportion of single-parent families also place a state’s children at higher risk.

About 27 percent of Tennessee’s K–12 student population is black, more than the typical state. Tennessee ranks 10th highest out of 48 states on this measure. In contrast, Tennessee has a relatively low percentage of Hispanics students (about 2 percent), ranking 38th highest out of 48 states.

As measured by the percentage of parents of 4th grade students who have a college degree, parental education places Tennessee 13th lowest out of 48 states. In terms of family income, Tennessee has the 17th highest percentage of free-lunch students out of 48 states and ranks 34th out of 48 states in per capita income.

Tennessee ranks as 15th highest in student mobility as measured by the percentage of students who report on the NAEP that they have moved in the last two years. Tennessee ranks 5th and 8th highest, respectively, among 48 states in the percent of births to teens and single-parent families.

Except for the Hispanic percentage, Tennessee is ranked about in the top third of states or much lower on every other risk factor. However, not all of these characteristics have equivalent strength when accounting for differences in student scores. As noted earlier, we develop an index, building on our prior research, that combines data from the NAEP, U.S. Census, and the National Education Longitudinal Survey to weigh eight characteristics (percentage of black and Hispanic students, education levels of the mother and father, family income, age of mother at the time of the child's birth, family structure, and number of siblings) to obtain an overall level of educational risk associated with each state's combination of family characteristics.



Tennessee Ranks 36th Out of 47 on a Weighted Measure of Family Characteristics Predicting Achievement

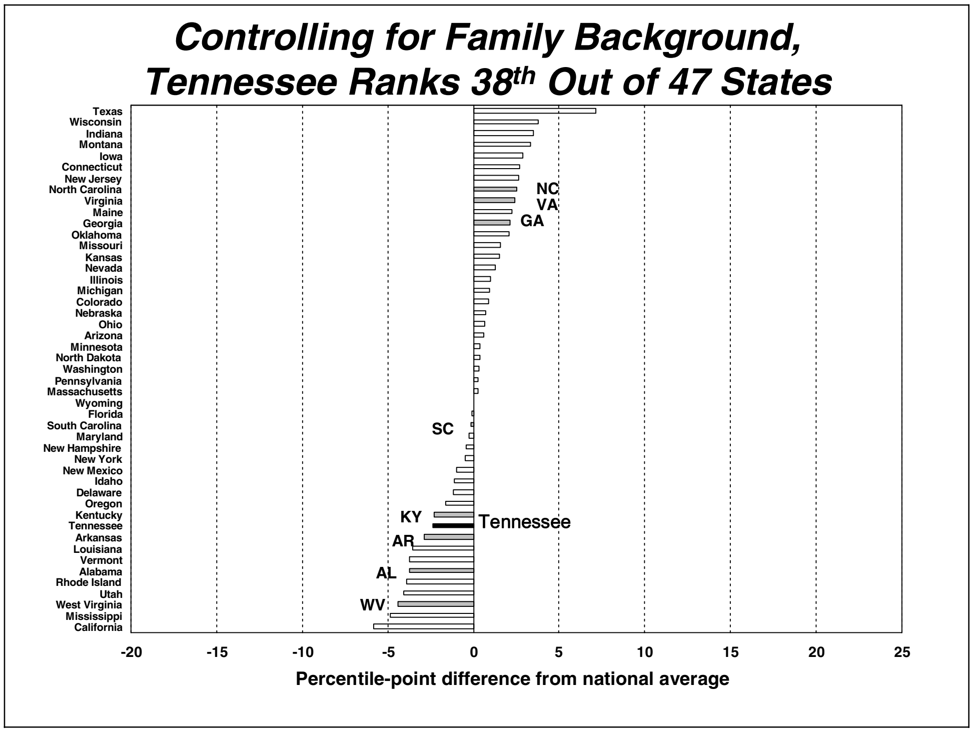
This chart shows the composite measure of educational risk from family characteristics measured as standard deviation units from the national average for all available NAEP tests used in our analysis. A high ranking on this measure indicates very low risk from family characteristics, while a low ranking indicates a high risk from family characteristics.

The weighted measure of family characteristics ranks Tennessee 36th highest in educational risk factors out of 47 states. (South Dakota is eliminated in this slide and the next because students in that state took no NAEP tests before 2003.) This ranking places Tennessee approximately in the middle of the comparison states. Virginia, West Virginia, and Kentucky have less educational risk from family factors, while North Carolina, Arkansas, Alabama, Georgia, and South Carolina have somewhat greater risk from family characteristics.

Very similar to the rankings for raw NAEP achievement scores, northern rural states have the family characteristics with lower risk for educational achievement, while the southern states have the highest risk for educational achievement. The northern urban states tend to be nearer the middle of the rankings.

These rankings are driven primarily by the percentage of minority and low-income students because such students also usually have higher risk for teen births, larger numbers of siblings, low parental education, and single-parent families. The students with less favorable family characteristics are concentrated in southern states and northern states with large cities, but tend not to be in northern rural states. This distribution of family characteristics is the primary reason northern rural states have among the highest achievement scores and southern states among the lowest.

The value of the composite measure of family characteristics is correlated with actual NAEP scores in the range of 0.6–0.8. It is the strongest variable, among those we considered, in accounting for the differences in average scores across states.



Controlling for Family Background, Tennessee Ranks 38th Out of 47 States

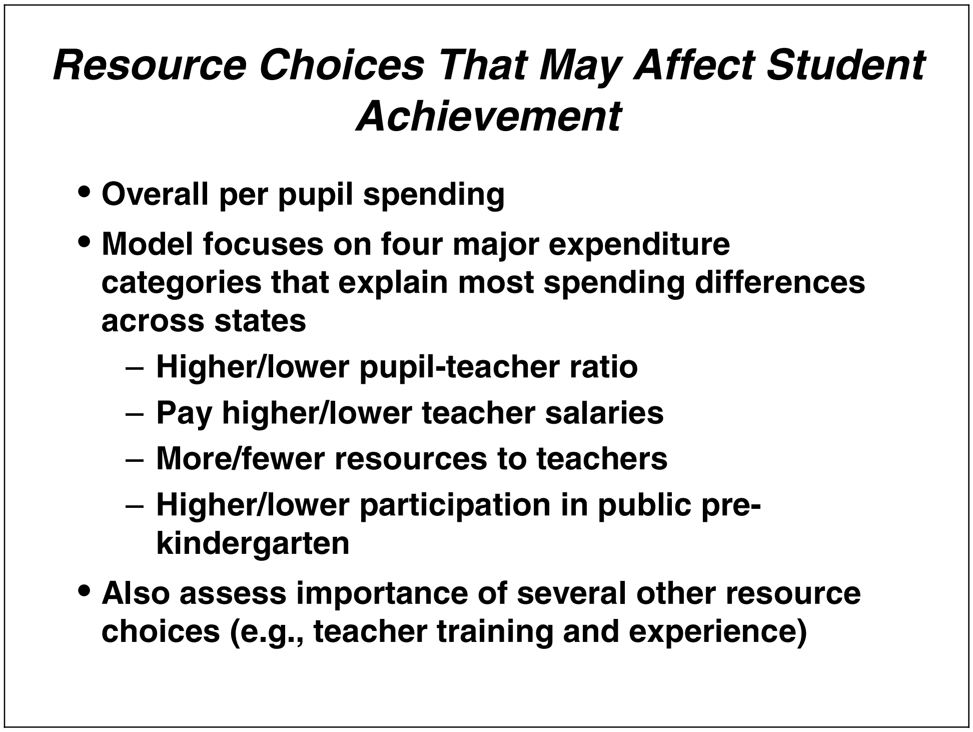
Using NAEP data across states from 1990 to 2003 for 17 tests, we made estimates of the scores for students adjusting for state-level differences in the family characteristics of students taking the NAEP tests, based on our composite measure of SES. (The models also include a control for family mobility based on the NAEP; see Grissmer, Flanagan, et al., 2000.) The resulting adjusted scores shown in this figure indicate that Texas students have the highest estimated average scores across states after adjusting for family characteristics. At the other extreme, California has the lowest adjusted scores. The difference between students in Texas and California is about 13 percentile points.

These results suggest that, even after removing differences in the family characteristics among the states, the remaining difference in achievement between students in Texas and California is 13 percentile points on average across the 17 4th and 8th grade math and reading tests used in this analysis. While our methodology will not control for all differences in family background (for example, changes over time in the characteristics of students within race and ethnic groups), the remaining difference can be more plausibly linked to the effects of differences in educational spending and policies because the analysis already controls for key differences in family characteristics.

On the measure of scores adjusted for family characteristics, Tennessee ranks 38th out of 47 states. This measure would suggest that after accounting for state-level differences in family characteristics, Tennessee's students would score about 2–3 percentile points lower than the national average. This ranking would suggest that Tennessee's educational output from the K–8 education system, measured in terms of reading and math scores at the 4th and 8th grade levels, is about the same as or somewhat below the average in other states.

Three comparison states—Virginia, Georgia, and North Carolina—have much higher scores after adjusting for differences in family characteristics. Tennessee ranks in the middle of the remaining neighboring states. This ranking would suggest that differences in the level of educational spending, the way these funds are utilized, or other educational policies may be raising the educational output of schools in these three states above Tennessee's when evaluated in terms of the NAEP reading and math scores.

Looking at all states, the pattern of scores adjusted for family characteristics no longer has the distinct north-south pattern of the raw scores. There are some southern states, such as Texas, North Carolina, Virginia, and Georgia, that have family characteristics with high educational risk but very high rankings after adjusting for differences in family characteristics. Some states with very low risk from family characteristics, such as Utah and Vermont, have very low adjusted scores, suggesting lower educational output. Some states, like Texas and California, with similar family characteristics have very different scores for students after adjusting for family background. The question is what characteristics of state educational systems might explain these differences in test scores that are not accounted for by differences in family background.



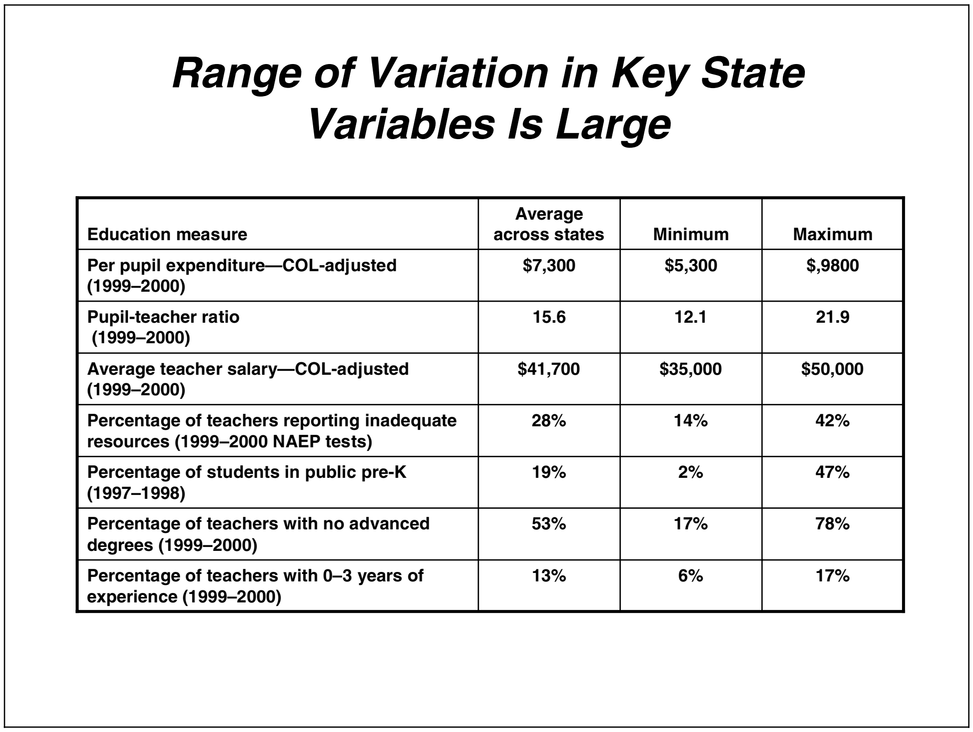
Resource Choices That May Affect Student Achievement

- **Overall per pupil spending**
- **Model focuses on four major expenditure categories that explain most spending differences across states**
  - **Higher/lower pupil-teacher ratio**
  - **Pay higher/lower teacher salaries**
  - **More/fewer resources to teachers**
  - **Higher/lower participation in public pre-kindergarten**
- **Also assess importance of several other resource choices (e.g., teacher training and experience)**

States have limited options when considering how to spend additions to educational expenditures. The educational systems across states have much similarity. All states have at least 13 years of schooling, beginning with kindergarten through high school. Teachers in classrooms deliver instruction to pupils for a fairly similar number of days per year and hours per day, and schools and school districts have administrative personnel for management and planning. Expenditures are incurred for transportation of students. Extra expenditures are mandated across all states for certain categories of special education students.

Despite these similarities, states have markedly different expenditures per pupil, ranging from $5,000 to $10,000, and our review of the literature suggests these spending differences can account for differences in achievement. Beyond overall spending, how resources are allocated may also matter. When states have higher per pupil expenditures, how do they spend additional funds? Grissmer, Flanagan, et al. (2000) showed that four major categories of expenditures, plus transportation and special education costs, accounted for almost 80 percent of the variance in per pupil expenditures across states. The four major expenditure categories were hiring more/fewer teachers to reduce the pupil-teacher ratio, paying higher/lower teacher salaries, providing teachers more/fewer resources for teaching, and adding/not adding another year of schooling at the pre-kindergarten level. States' expenditures also differ because of different transportation costs per pupil and different proportions of students in special education categories.

Our empirical analysis focuses on these four resource variables, which account for much of the variance in how states spend education funds (see Grissmer, Flanagan, et al., 2000, for a discussion of the data sources for these measures). We discuss these results now. After reviewing the results of our analyses for these specific measures, we consider other policy or reform variables that may explain any remaining variation.



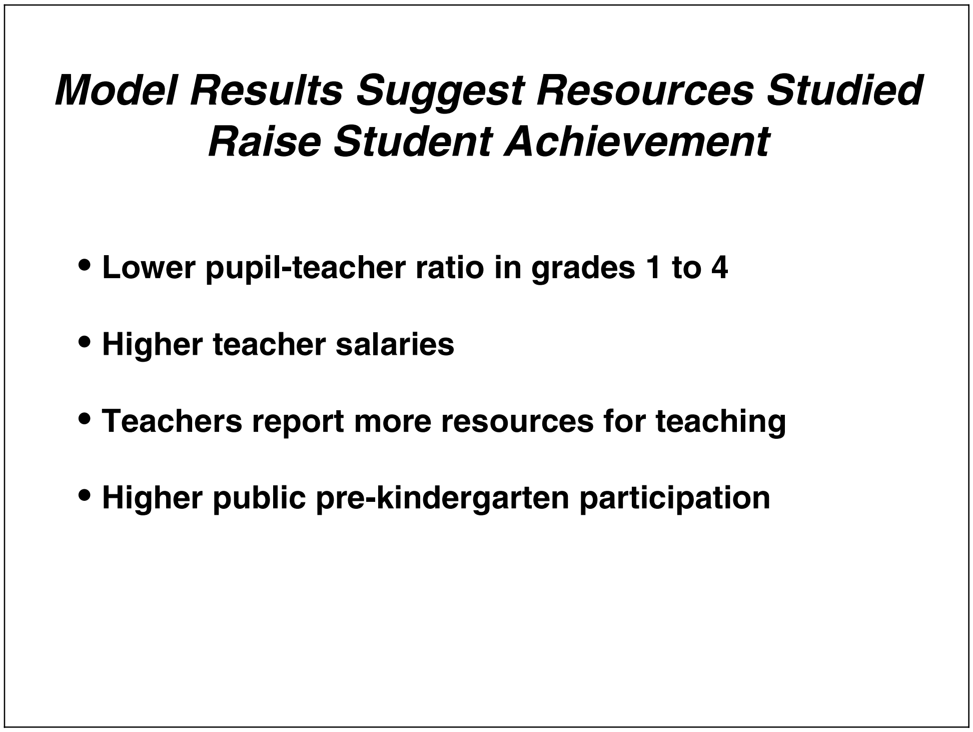
Range of Variation in Key State Variables Is Large

Education measure	Average across states	Minimum	Maximum
Per pupil expenditure—COL-adjusted (1999–2000)	$7,300	$5,300	$,9800
Pupil-teacher ratio (1999–2000)	15.6	12.1	21.9
Average teacher salary—COL-adjusted (1999–2000)	$41,700	$35,000	$50,000
Percentage of teachers reporting inadequate resources (1999–2000 NAEP tests)	28%	14%	42%
Percentage of students in public pre-K (1997–1998)	19%	2%	47%
Percentage of teachers with no advanced degrees (1999–2000)	53%	17%	78%
Percentage of teachers with 0–3 years of experience (1999–2000)	13%	6%	17%

Across the states, there is substantial variation in the amount spent per pupil on education and the way in which the funds are allocated among the four resource categories listed above. Adjusted for differences across states in the cost of living, some states in 1999–2000 spent as little as $5,300 per pupil on average (Arizona and Utah), while other states spent as much as $9,800 per pupil on average (New York and New Jersey). In terms of the pupil-teacher ratio, Vermont, Maine, and Virginia had 12–13 pupils per teacher, while Utah and California had 21–22 pupils per teacher. The cost of living-adjusted range of teacher salaries is $35,000 in Mississippi and Wyoming to $50,000 in Massachusetts and Illinois.

On a survey administered with the NAEP tests, teachers reported on the adequacy of their resources for teaching. Three categories were given. Only about 15 percent of teachers in Texas and Kentucky report the most inadequate category of resources, while about 42 percent of teachers in Rhode Island and New Mexico report inadequate resources. Finally, the share of students in public pre-kindergarten programs ranges from 2 percent in Pennsylvania to 47 percent in Oklahoma.

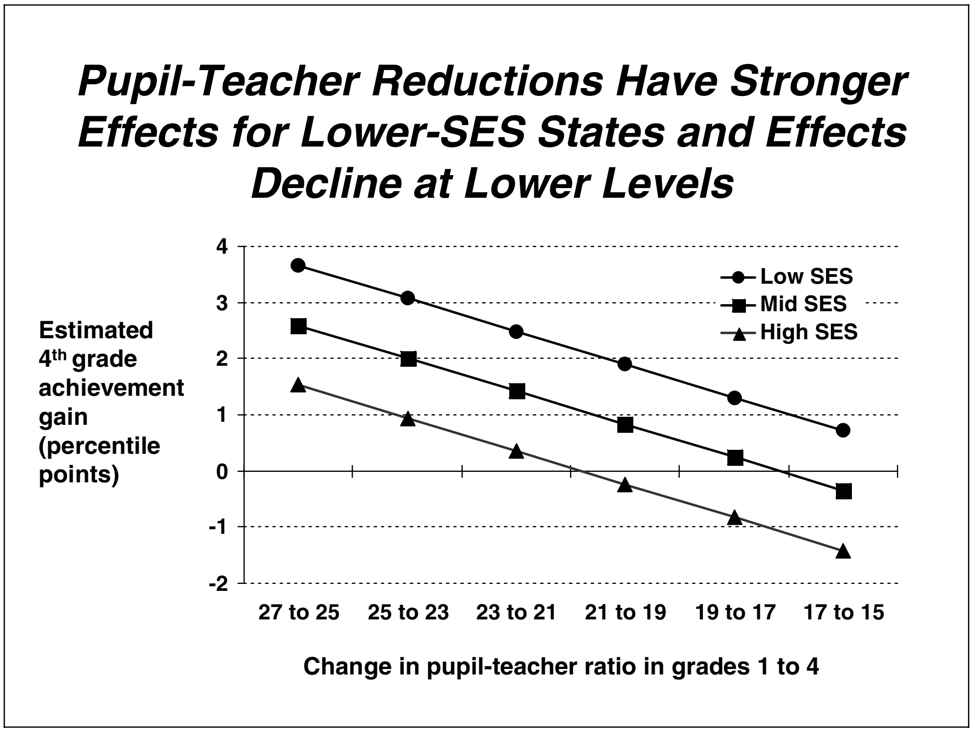
Beyond these key resource indicators, the table also shows two other potentially relevant educational measures, which we return to later. First, the average level of teacher education varies markedly across states. Only 18–21 percent of teachers in New York and Connecticut do not have advanced degrees beyond the minimum bachelor's degree, as compared with about 75 percent of teachers in North and South Dakota. The percentage of teachers with less than 4 years of experience varies from about 6–9 percent in West Virginia and Rhode Island to 17 percent in North Carolina and Vermont.



Model Results Suggest Resources Studied Raise Student Achievement

- Lower pupil-teacher ratio in grades 1 to 4
- Higher teacher salaries
- Teachers report more resources for teaching
- Higher public pre-kindergarten participation

Our statistical analysis using NAEP data from 1990 to 2003 confirms our earlier findings using a shorter time series that, holding family characteristics constant, specific resource policies affect achievement. (See the Appendix for detailed regression results.) In particular, states that focus more resources on (1) lowering the ratio of pupils to teachers in the early grades (grades 1 to 4), (2) raising teacher salaries, (3) providing teachers with adequate teaching resources, and (4) providing larger public pre-kindergarten programs—other things being equal—have higher achievement scores. All such coefficients are significant at the 95-percent confidence level. The magnitude of the effect of the pupil-teacher ratio is consistent with the size of effects from the Tennessee Standardized Testing and Reporting (STAR) class size reduction experiment.

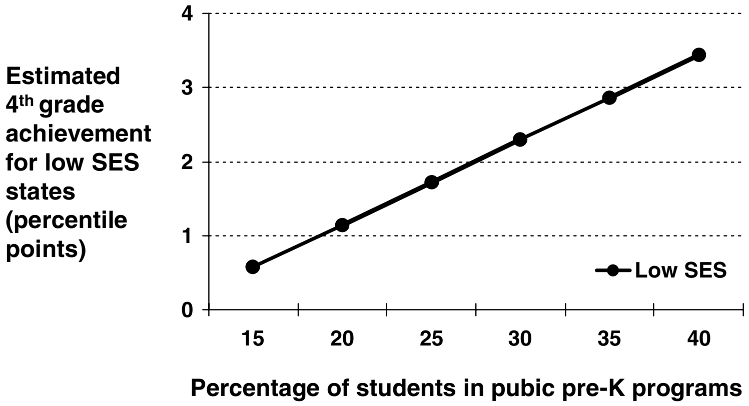


We also estimate models, for 4th grade tests only, to determine if the effect of the pupil-teacher ratio in the early grades and the public pre-kindergarten participation rate vary depending on the risk status of the state's student population (see the Appendix for model estimation details). This chart shows the estimated achievement gain statewide at the 4th grade level for different pupil-teacher reductions during grades 1 to 4 in states with different socioeconomic statuses. The extreme right data point (27 to 25) shows the estimated achievement gain for a reduction of the pupil-teacher ratio in grades 1 to 4 from 27 to 25. The upper line indicates states with among the highest proportion of disadvantaged students as measured by our statewide composite SES measure. An estimated gain of almost 4 percentile points would be predicted for a reduction of the pupil-teacher ratio from 27 to 25 in the early grades.

The middle line shows the same predictions for states with average proportions of disadvantaged students. A 2.5-percentile-point gain in 4th grade achievement would be predicted in these states for the same reduction. Finally, states with the lowest proportion of disadvantaged students would have an estimated gain of only 1.5 points.

As the starting point for the pupil-teacher ratio declines, the gains in 4th grade achievement scores from further reductions become smaller for all states. This indicates a diminishing return for reductions in the pupil-teacher ratio because of the statistically significant quadratic term included in our model. However, in the states with the poorest populations, reductions down to levels of 15 still have about an estimated 1-percentile-point gain.

Achievement Gains From Public Pre-K Participation Mainly for Low SES States



Our results also provide estimates for differences in the expected achievement of 4th grade students from an increased percentage of children enrolled in public pre-kindergarten. The results suggest that the public preschool effects at the 4th grade level occur mainly in states with low SES. In such states, most of the children attending public preschool are from lower-SES households, either because of subsidies to such families or because higher-SES families still send children to private preschools. Our results suggest that increases in participation in public pre-kindergarten from 15 to 40 percent of children correspond to an increase in average achievement statewide at the 4th grade level by an estimated 3 percentile points.

It is important to note that these findings suggesting that the effect on student achievement of at least some resources is higher for disadvantaged students, at least through the 4th grade, are based on an analysis of aggregate state-level data. The analysis does not assess whether the effects of resources are higher for specific groups of disadvantaged students within states or for disadvantaged students in districts or schools within states. Ideally, researchers would confirm that these findings hold by using further disaggregated data. Notably, however, these findings are consistent with the class size and early childhood experiments discussed earlier.

Estimated Statewide Additional Dollars Per Pupil Required to Raise 4th Grade Achievement by About 3 Percentile Points

	Low-SES states	Mid-SES states	High-SES states
Pupil-teacher (grades K–3) from high levels (25)	**250**	**550**	**>1000**
Pupil-teacher (grades K–3) from medium levels (21)	**350**	**>1000**	**>1000**
Pupil-teacher (grades K–3) from low level (17)	**600**	**>1000**	**>1000**
Pre-kindergarten	**300**	**>1000**	**>1000**
Teacher salary	**>1000**	**>1000**	**>1000**

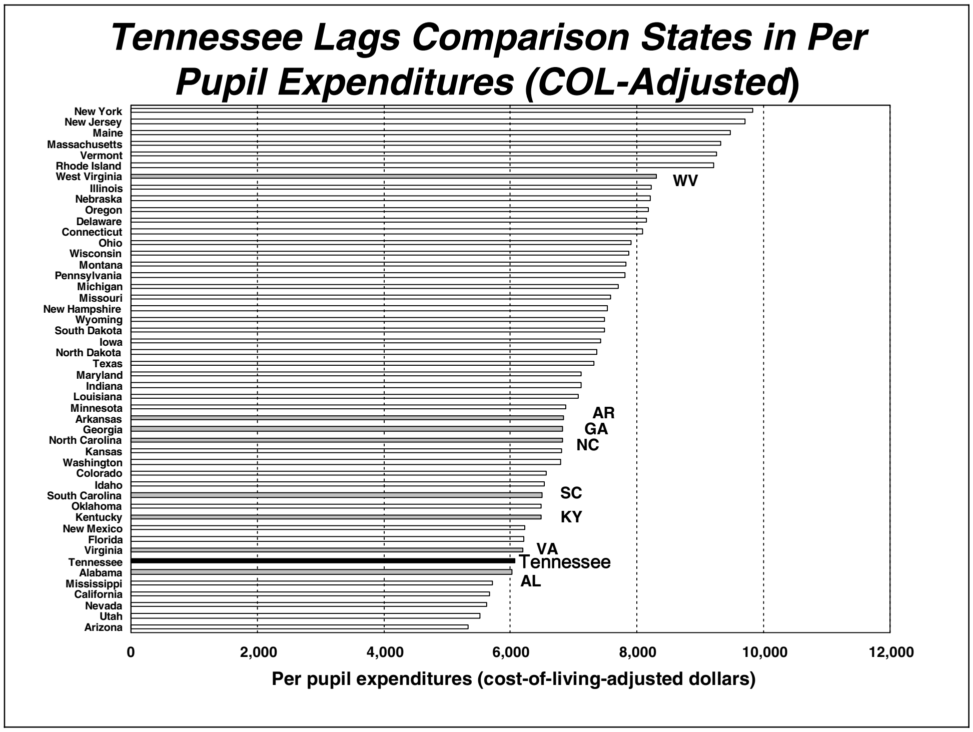
To compare the importance of the resource measures controlled for in our model, the estimates can be transformed to quantify the estimated dollars per pupil associated with an increase in student achievement of 3 percentile points at the 4th grade level. A 3-percentile-point gain for Tennessee would improve Tennessee's ranking among states by about 4–6 positions depending on the test, assuming that other states do not change resources. This chart shows estimates for low-, medium-, and high-SES states (and accounts for variations in the effect of the pupil-teacher ratio and pre-kindergarten programs estimated for states with different SES levels). The methodology is described in Grissmer, Flanagan, et al. (2000). Tennessee's family characteristics place Tennessee between the low- and medium-SES states.

The dollar figures presented in this table should be interpreted with care as they are based on a model that related historical data on 4th grade test scores with resource measures, all at the state level. It is possible that the costs of obtaining achievement gains in the future would vary as a result of economy-wide changes or changes within states that alter the historical relationship between resources and student achievement. Nevertheless, the figures can provide an approximate guide to the resources that appear to be associated with lower dollar costs to attain a given change in achievement at the 4th grade level.

In general, the results suggest that the costs would be much higher to obtain the same achievement gain as of the 4th grade in high- or medium-SES states than in low-SES states. For instance, for states with very high pupil-teacher ratios of around 25, it would cost about $250 per pupil in a low-SES state, compared with $550 per pupil in a medium-SES state and over $1,000 per pupil in a high-SES state, to obtain a statewide gain of 3 percentile points. In high-SES states where per pupil expenditures are already high, obtaining additional gains would be quite expensive regardless of where resources are directed.

In low-SES states, our estimates show that achievement gains might be obtained at even lower costs. Currently, pupil-teacher ratios overall are in the range of about 14 to 22. For low-SES states with currently high pupil-teacher ratios of 21, about $350 per pupil is required for a 3-percentile-point gain. In contrast, if the pupil-teacher ratio is already low (around 17), it would require about $600 per pupil in additional expenditures to raise achievement by the same amount. These estimates suggest a marginally decreasing return as pupil-teacher ratios are reduced; again, this is a result of the functional form of the model we estimated. For low-SES states, achieving a 3-percentile-point gain appears to be most expensive through raising teacher salaries.

If the current pupil-teacher levels are around 21, the costs of equivalent achievement gains in low-SES states are estimated to be approximately the same for public pre-kindergarten and pupil-teacher reductions. If current pupil-teacher ratios are lower, investment in public pre-kindergarten looks to be a better investment than further class size reductions. This may be the case for Tennessee, given its already fairly low pupil-teacher ratio (as shown later).



Tennessee Lags Comparison States in Per Pupil Expenditures (COL-Adjusted)

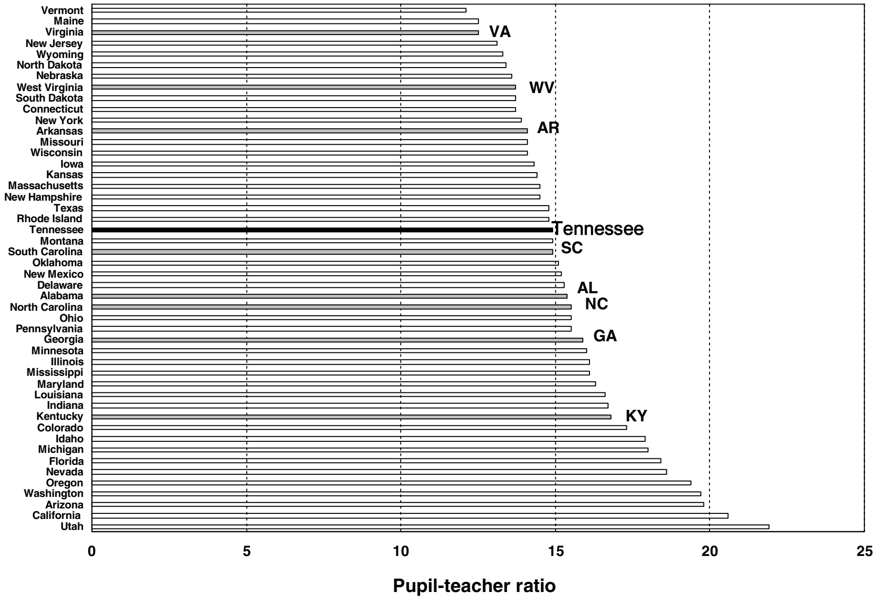
The next several slides present the characteristics of Tennessee schools as compared with other states for the education measures included in our analysis. These measures are for a specific year, in most cases 1999–2000 (see the time periods associated with these variables presented earlier in the discussion of state means). Our results suggest that the average educational inputs in the K–8 education system in Tennessee are about average or somewhat below average compared with those of other states.

Tennessee per pupil expenditures for 1999–2000 rank among the lowest compared with those of other states. Tennessee's economic status does not provide the resources present in other states to support very high levels of spending for education. States that can spend about $9,000–10,000 per pupil can more easily provide a higher-quality K–8 system than Tennessee, which spends about $6,000 per pupil. Our model shows that, other things being equal, higher spending is linked to higher scores among students after adjusting for family characteristics. But how the money is spent is an equally strong factor in higher scores after adjusting for family characteristics. Not all high-spending states have above-average scores after such adjustments, whereas some lower-spending states do. But overall, a state is more likely to have high scores adjusted for family characteristics when its spending level is high.

For instance, Texas has average per pupil spending, but very high scores after adjusting for average family characteristics. New York and Rhode Island have very high per pupil expenditures, but average or below-average scores once we make such adjustments. However, Utah and California have very low expenditures per pupil and among the lowest scores adjusted for family characteristics.

The three comparison states that have much higher adjusted scores than Tennessee (Virginia, Georgia, and North Carolina) all have higher per pupil expenditures. North Carolina and Georgia spend about $750 more per pupil and Virginia about $120 more per pupil.

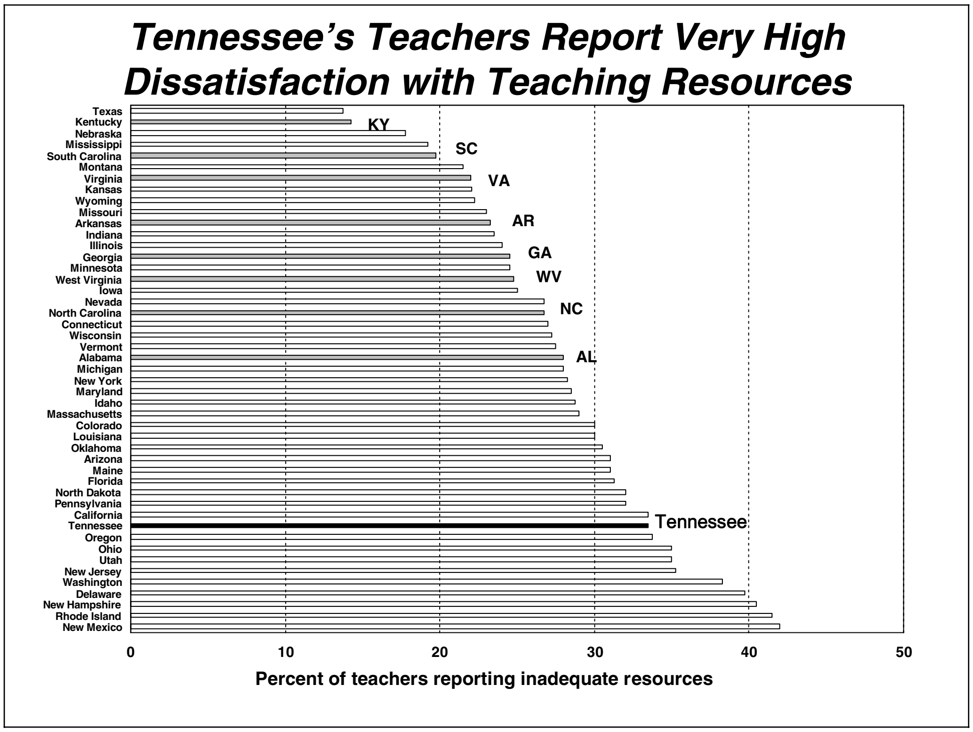
Tennessee Has the 21st Lowest Pupil-Teacher Ratio out of 48 States



This chart shows pupil-teacher ratios, which average 15.6 across the 48 states. The model results indicate that achievement scores are higher in states with lower pupil-teacher ratios, so we have ranked states from the lowest to the highest ratios in the chart.

Tennessee’s average ratio is 14.9 or a ranking of 21 out of 48 states, which places it below the national average. Utah and California have the largest pupil-teacher ratios, just over 20, while Virginia, Maine, and Vermont have the lowest ratios, between 12 and 13.

Among comparison states, Virginia has the lowest pupil-teacher ratio at 12.5, followed by West Virginia and Arkansas at 13.7 and 14.1, respectively. South Carolina’s ratio is the same as Tennessee’s, while the other comparison states have ratios ranging from about 15 to 17.



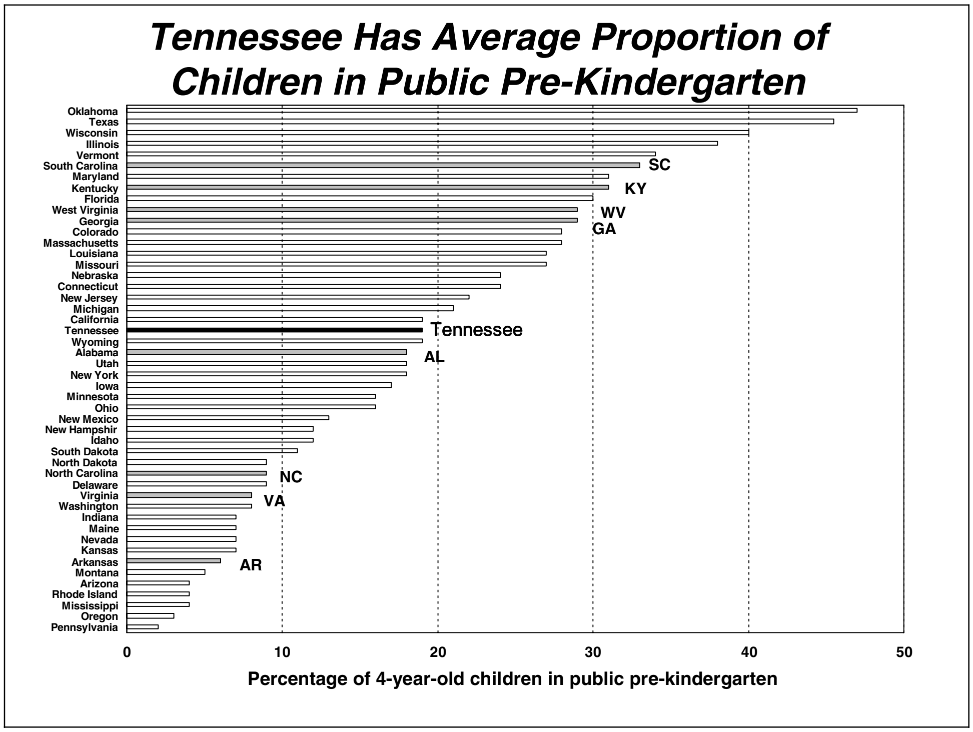
Tennessee's Teachers Report Very High Dissatisfaction with Teaching Resources

This chart shows teachers' reports of the adequacy of resources for teaching. Three answers were possible, ranging from adequate to inadequate. This chart shows the percentage of teachers in each state who reported the lowest adequacy level. According to the model, fewer teacher resources are associated with lower student achievement.

Teachers in Texas reported the lowest dissatisfaction with teaching resources, while teachers in New Mexico reported among the highest level of inadequate resources.

Tennessee has an unusually high percentage of teachers—about one-third—reporting inadequate resources, a variable that was associated with lower achievement. This is among the highest percentage across states and is much higher than in all of the neighboring comparison states. For instance, Kentucky's teachers reported among the lowest level of dissatisfaction with resources.

It is possible that Tennessee's teachers may have a lower level of resources because these resources are squeezed out of the budget by the relatively higher proportion of spending, compared to other states, on lower class sizes, teacher salaries, and pre-kindergarten (the latter two measures are shown next).



Tennessee Has Average Proportion of Children in Public Pre-Kindergarten

There has been substantial variation across the states in the investments made in publicly available pre-kindergarten programs. In 1980, just 10 states funded pre-kindergarten programs, primarily for low-income children (Gilliam and Zigler, 2004). By the 2003–2004 school year, that number had increased to 38 states, with Georgia and Oklahoma effectively providing pre-kindergarten access for all 4-year-olds (Barnett et al., 2004).

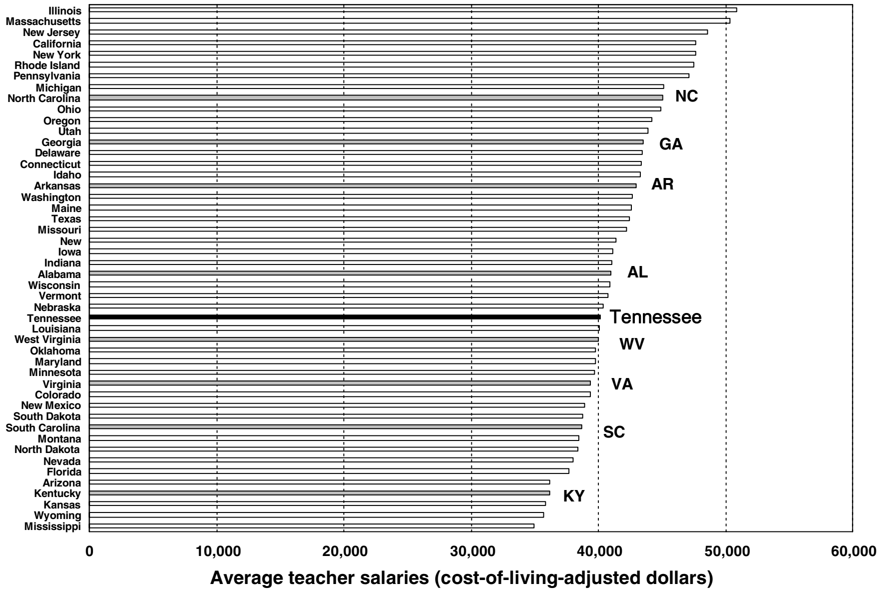
Our pre-kindergarten participation variable measures the percentage of children in public preschool for the year in which the children were 4 years old. Data for public pre-kindergarten participation is generally not uniform and sometimes not available for all states prior to about 1985. When available, our variable is the ratio of children in public pre-kindergarten to those in the first grade of public school.

Few states had high participation levels in public pre-kindergarten for the early years in which NAEP test takers were 4 years of age, though Texas and Kentucky had significant numbers in pre-kindergarten in the 1980s. Preschool enrollment has increased nationwide but, as shown in the figure, most states by 1997–1998 still had less than 20 percent of children in public pre-kindergarten. About 10 states had between 30 and 45 percent of children in public pre-kindergarten.

Among comparison states, South Carolina, Kentucky, West Virginia, and Georgia had about 30 percent of children in public pre-kindergarten, while Tennessee had an estimated 20 percent. North Carolina, Virginia, and Arkansas had less than 10 percent in public preschool.

The model results indicate that higher enrollment in public pre-kindergarten programs is associated with higher achievement scores. However, the model does not explicitly control for quality differences in the public programs funded across states, which vary considerably (Barnett et al., 2004). Tennessee's program does meet eight out of 10 research-based pre-kindergarten quality standards identified by the National Institute for Early Education Research (NIEER) and it has relatively high spending per pupil (Barnett et al., 2004). The payoff from this investment in relatively higher-quality public pre-kindergarten cannot be determined from the models estimated in this study, however.

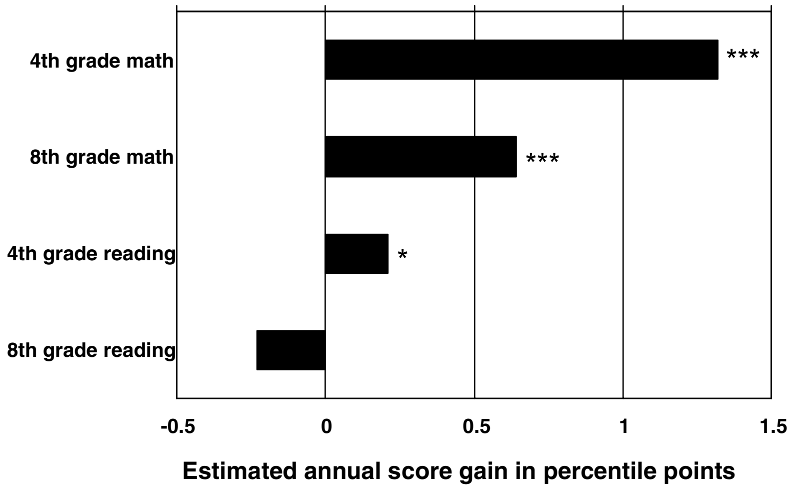
Tennessee Teacher Salaries Are Near the Average of Comparison States But Below the National Average



K–8 education is a very labor-intensive endeavor. Teacher salaries are the largest single item in education budgets, and teacher salaries and benefits account for almost two-thirds of educational expenditures per pupil. Tennessee's average teacher salaries (cost-of-living-adjusted) are below the national average and but near the average for comparison states. Tennessee’s ranking on teacher salaries is much higher than its ranking on per pupil expenditures. This suggests that Tennessee spends a higher proportion of its expenditures on teacher salaries than do other states.

As noted above, our models suggest that, on a per dollar basis, differences among states on teacher salaries correspond to smaller differences in achievement, adjusted for family characteristics, than equal per dollar differences in class size or pre-kindergarten enrollment. Thus, raising teacher salaries, where Tennessee is overspending relative to the higher-performing comparison states, might be a relatively lower-yield investment.

Math Scores Are Increasing Faster Than Reading Scores



NOTE : Statistical significance: *** 1 percent; ** 5 percent; * 10 percent.

We now turn to estimating and comparing the gains made by states across tests from 1990 to 2003. We also estimate what proportion of the gains can be attributed to resources and what proportion might be due to other educational policies that we have not examined in our analysis. (Again, see the Appendix for regression results.)

Our model estimates the average NAEP score gain across states by fitting a trend to each set of specific tests by grade and subject. The 4th grade reading gains are measured across five NAEP tests administered between 1992 and 2003. The 4th grade math gains are estimated for four tests from 1992 to 2003. The 8th grade math gains are measured by five tests from 1990 to 2003. The 8th grade reading tests are measured by three tests from 1998 to 2003.

As seen in the figure, average state math scores are increasing much faster than reading scores. There is no compelling research that explains the difference in trends between math and reading. Several hypotheses exist that need to be explored. The first is that the foundation for reading, more than for math, is thought to be in the home, whereas math is more the product of schools. Thus, if schools improve, this might be expected to have a greater effect on math scores.

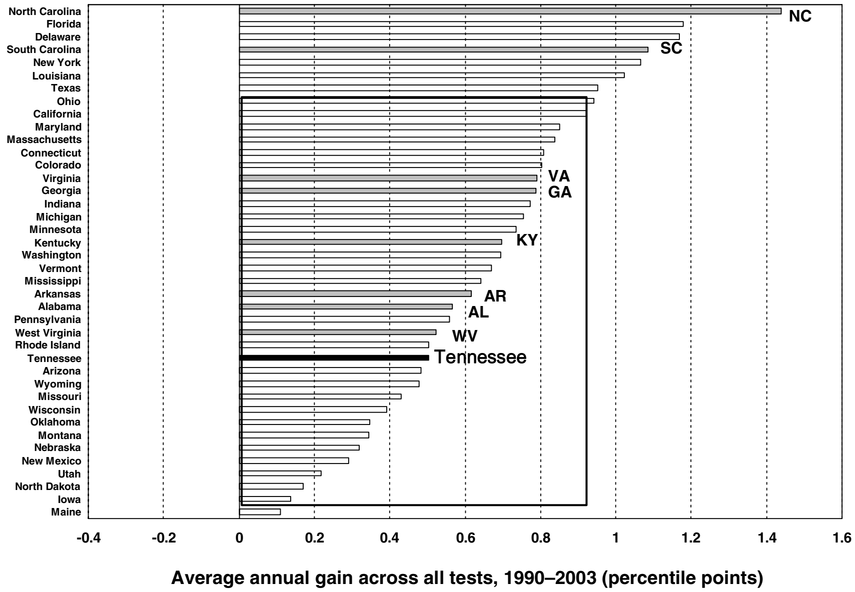
There has also been significantly more research on math than reading because of the National Science Foundation's (NSF's) significant research efforts. This research focuses only on math and science. As a result, there are more widely accepted standards for teaching math than reading, as well as the production of associated curriculum. The NSF has also focused specifically on raising the math and science scores of women and minority students.

In addition, reading appears to be a more complex process than math—at least in the early grades—and well-defined disabilities such as dyslexia significantly affect more students than do identified math disabilities. Finally, the increase in the Hispanic population and the number of other non-English-speaking immigrants tends to affect reading scores more than math scores.

In general, though, more research is needed in order to better identify to what degree these hypotheses contribute to the trends.

Both 4th grade reading and math have higher gains than 8th grade reading and math. This finding may reflect the greater resources that have been directed to lower elementary grades and preschool. Pupil-teacher ratios have declined significantly in early grades, while participation in public preschool has increased.

Tennessee Had Annual Gains of About One-Half Percentile Point a Year, 1990–2003



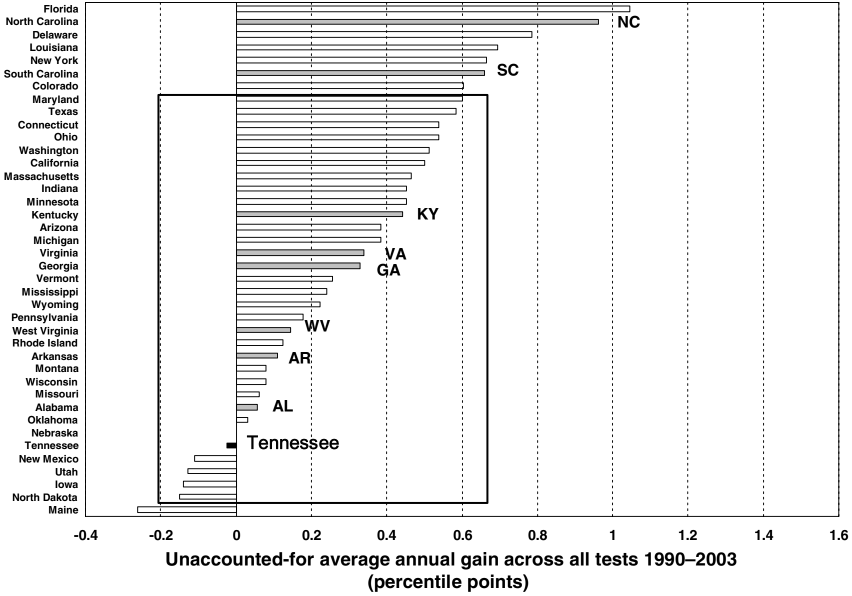
NOTE: Score gains within the box are not statistically different from Tennessee's at the 90% level of confidence.

This chart shows these same estimated annual gains between 1990 and 2003, aggregated here across all tests but disaggregated by state (see the Appendix for regression results). States that took fewer than one-half of the tests are not plotted here, namely Idaho, Illinois, Kansas, Nevada, New Hampshire, New Jersey, Oregon, and South Dakota. These gains are adjusted for demographic changes; that is, states are not penalized for high immigration or other demographic shifts because we make estimates holding family background characteristics constant using the same composite SES measure and measure of mobility.

Tennessee's gain was about one-half a percentile point per year—a below-average gain. North Carolina made the largest gain: Over the 13-year period, students in North Carolina had average gains across tests of about 18 percentile points. This is approximately equivalent to more than 1.3 grade levels. South Carolina also had among the highest gains in the nation. The gains of both North and South Carolina were statistically significantly higher than those of Tennessee at the 10-percent significance level. The remaining comparison states did not have gains that were significantly different from those of Tennessee.

An important question is whether changes in the level and use of resources across states can account for these gains.

Tennessee Has No Gains Unaccounted For By Increased Resources



NOTE: Score gains within the box are not statistically different from Tennessee's at the 90% level of confidence.

This chart shows the estimated annual gains after accounting for changes in the four key resource variables discussed earlier: pupil-teacher ratio, teacher salaries, teaching resources, and participation in public pre-kindergarten. The figure shows that a significant percentage of the estimated gains across states cannot be accounted for by resource and program changes. For instance, a lower-SES state that reduced the pupil-teacher ratio in lower grades by four (a larger-than-average reduction across states) and increased public pre-kindergarten participation by 25 percentage points (a much larger increase than the typical state) would have an estimated gain of about 7 percentile points. Even these changes would account only for less than half of the gains of those states with the largest annual gains.

As seen here, all gains shown in the prior chart for Tennessee have now been accounted for by changes in the resources we measure (i.e., the remaining test score gains in Tennessee, after accounting for resources, are essentially zero). All comparison states except Alabama have larger gains that are not accounted for by the resources we measure. Two states (North and South Carolina) have unaccounted-for gains that are statistically significantly higher than those of Tennessee. Between 1990 and 2003, North and South Carolina both gained between 9 and 13 percentile points (approximately three-quarters to a full grade equivalent) on Tennessee from factors we do not account for in our models, while Kentucky, Georgia, and Virginia gained between 4 and 5 percentile points from factors that are not accounted for (approximately one-third of a grade equivalent).

These residual gains can explain a sizeable part of the score differences between Tennessee and North and South Carolina in 2003, and for part of the higher rankings of North Carolina, Virginia, Georgia, and South Carolina on scores after controlling for family characteristics. In the early 1990s, the scores of Tennessee and North and South Carolina were much closer. It is the larger gains made by North and South Carolina, due to factors other than those we have measured, that have created the significant difference in scores that exist as of 2003.

Other Factors That Might Explain Achievement Scores

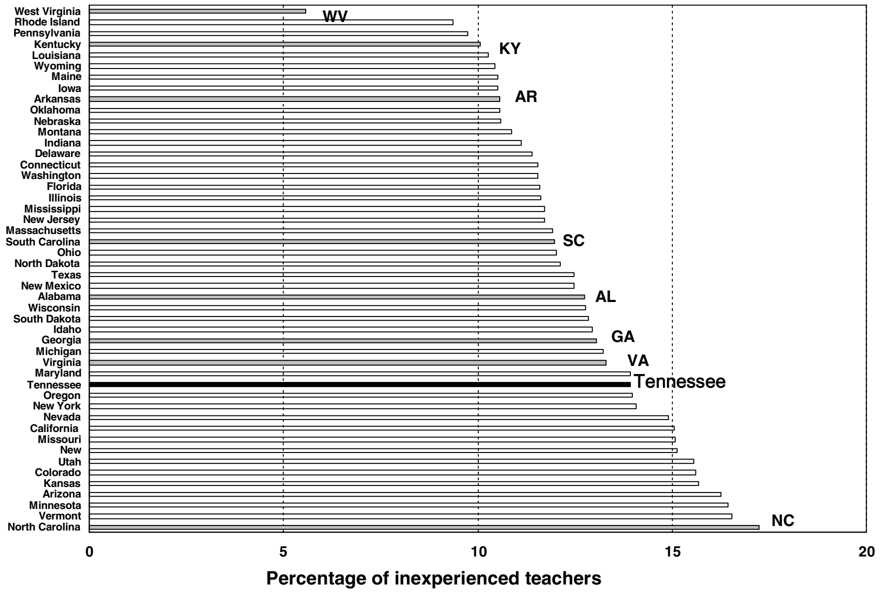
- **Other factors examined in NAEP models**
  - **More experienced teachers raises achievement**
  - **No effect from teacher advanced degrees**
- **Other factors not examined in models**
  - **Standards-based accountability reforms**
  - **Teacher pedagogical practices**
  - **Teacher preparation and training**

The size of the gains in NAEP scores across states could not be completely accounted for by changes in the measures of resource allocations included in our models. While the allocation of resources helps explain differences between states in scores at a point in time, controlling for family characteristics, the change in our resource measures over time is not enough to account for the sizable achievement gains across states. Overall, changes over time in the resources measured in our models account for about one-third of the 4th grade gains and for a smaller percentage of the 8th grade gains.

Since our models do not include all possible factors that could explain score changes over time, we conclude by assessing the potential for several other factors that we have not focused on as possible explanations for achievement score gains. Where measures of these factors across states and over time exist, they can potentially be included in our empirical models so their contribution can be assessed. In other cases, we must rely on other sources of data or other studies in the literature to make inferences about the potential importance of such factors. For this study, to glean further insights into the variation in teaching characteristics and the teaching environment across states, we analyzed teacher responses to surveys administered with the NAEP tests in various years between 1990 and 2003. We focused our analysis on the questions that could provide insight into either the standards-based accountability systems or the type of testing done in the state. We also examine evidence for differences in teaching, teacher preparation, teaching and training resources, and pedagogical techniques across comparison states and the nation.

In the remainder of this section, we focus on two factors that were included in variants of our regression models: teacher experience and training. Other factors, most notably standards-based accountability reforms, are not assessed using our empirical models, but can be considered in light of responses to the NAEP teacher surveys and other research in the literature. We also examine the NAEP survey data to look at differences across states in teacher pedagogical practices and teacher preparation and training.

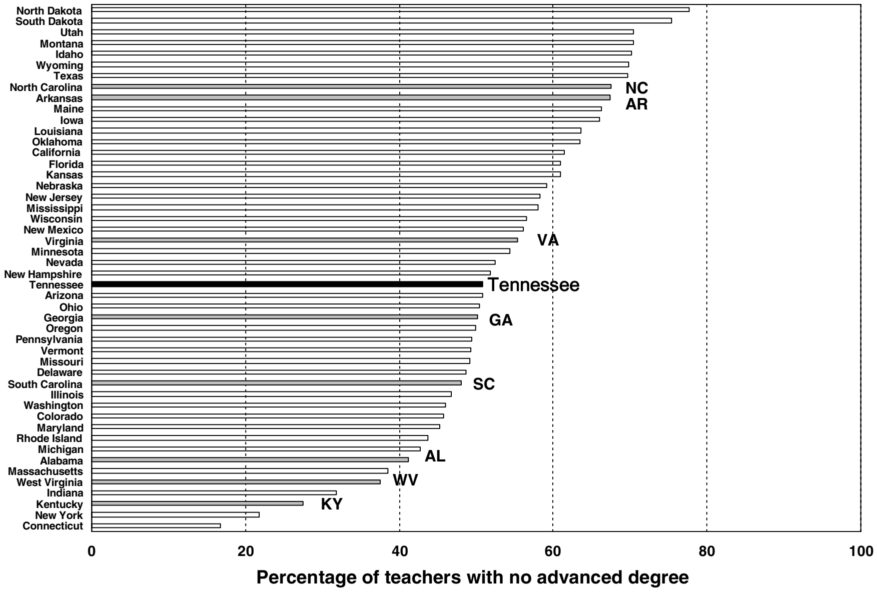
Tennessee Has a Higher Percentage of Inexperienced Teachers Than Most Comparison States



To assess the contribution of teacher experience, we included a measure of the share of teachers with three or fewer years of experience in the models (not included in the Appendix). Other things being equal, a higher share of inexperienced teachers is linked to lower achievement. However, it is important to understand the cause of higher levels of inexperienced teachers. It can indicate higher levels of hiring in response to higher enrollments or class size reductions. It can also indicate higher attrition in the teaching work force.

Tennessee has higher levels of inexperienced teachers than all comparison states except North Carolina. This may be due to reductions in pupil-teacher ratios in the late 1990s. More research is needed using teacher data from Tennessee to determine whether higher levels of entering teachers reflects high teacher attrition or higher levels of hiring to support class size reductions.

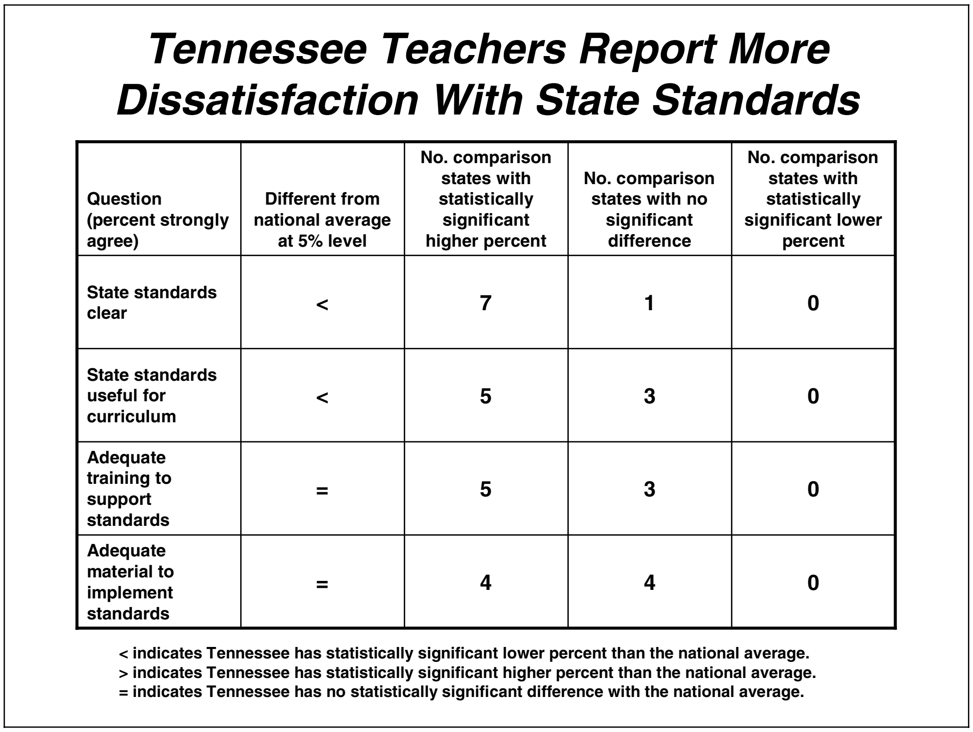
Tennessee's Teachers Have an Average Percentage of Advanced Degrees Compared to Other States



When we estimated models that included a measure of the proportion of teachers with advanced degrees, we found no effect on test scores (not included in the Appendix). This indicates that states that have higher proportions of teachers with advanced degrees, other things being equal, do not have higher achievement. Although other research has also found that, in general, advanced degrees among teachers do not correlate with higher achievement (for example, Krueger, 2000), research does suggest that subject-specific degrees (e.g., math, science, English) do lift achievement scores (Goldhaber and Brewer, 1997).

About one-half of Tennessee's teachers have advanced degrees—close to the national average. In some comparison states, such as North Carolina and Arkansas, only about 30 percent of teachers hold advanced degrees, while in Alabama, Kentucky, and West Virginia about 60 percent have advanced degrees.

The absence of effect of advanced degrees may simply reflect that the current degrees earned by teachers are not specific or specialized enough or are not directed toward the learning required to boost achievement. Further research could be done using Tennessee state data to link teacher degrees to specific schools and to state-administered tests, incorporating similar techniques to those used in this analysis. Identifying the type of degrees for different types of teachers from specific schools that lead to higher achievement should be a priority research area for Tennessee.



Tennessee Teachers Report More Dissatisfaction With State Standards

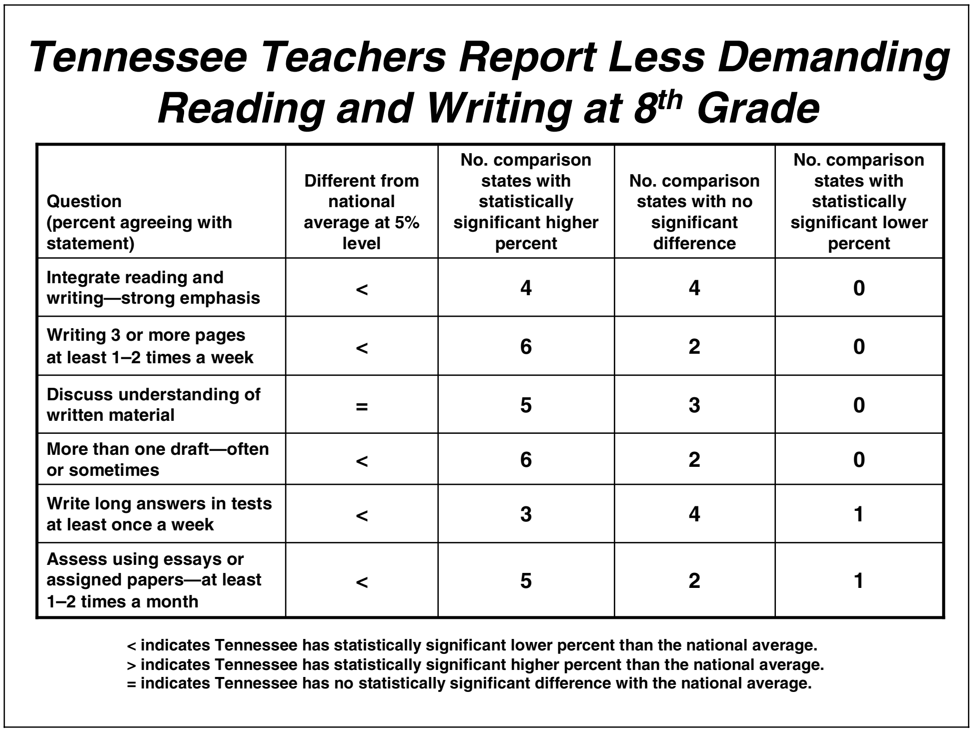
Question (percent strongly agree)	Different from national average at 5% level	No. comparison states with statistically significant higher percent	No. comparison states with no significant difference	No. comparison states with statistically significant lower percent
State standards clear	<	7	1	0
State standards useful for curriculum	<	5	3	0
Adequate training to support standards	=	5	3	0
Adequate material to implement standards	=	4	4	0

< indicates Tennessee has statistically significant lower percent than the national average.
> indicates Tennessee has statistically significant higher percent than the national average.
= indicates Tennessee has no statistically significant difference with the national average.

Over the last decade or more, states have implemented a variety of standards-based accountability systems. Grissmer and Flanagan (1998) offer examples of accountability systems for two of the states with the largest NAEP gains, North Carolina and Texas. In these states, the accountability systems were characterized by the establishment of standards, tests designed to those standards, the publication of results by school and school district, and gradually evolving types of accountability. Each state also freed districts and schools from state regulation and provided more flexibility for spending at local levels. In each state, the business community played a critical role in the design and legislative passage of the system. Each system offered rapid turnaround of assessments and used the data to diagnose problems. Tennessee's accountability system shares some of the provisions of the North Carolina and Texas systems. It has standards with tests designed to those standards and publishes results by school and school district.

The nature and timing of implementation of these accountability systems varies among states, but our model does not control for these variations in accountability systems and they could be a source the unexplained growth in NAEP scores. Although the limited and weak research on accountability systems and achievement (Carnoy and Loeb, 2002; Hanushek and Raymond, 2004) finds only small effects, teacher survey responses suggest this is a factor that might warrant investigation.

The four questions shown in this chart are the main NAEP teacher survey questions that address issues about standards-based accountability systems, and they indicate that compared to the nation and the comparison states, Tennessee teachers seem relatively less satisfied with their system. Tennessee teachers report that standards are less clear, less useful for choosing curriculum, and they report less adequate training and resources for meeting standards compared to their counterparts in most of the eight comparison states. The comparison states that have the strongest statistical differences from Tennessee include North and South Carolina, the comparison states with the largest unexplained gains in the NAEP. Several other comparison states also have clear differences from Tennessee on these measures.



Tennessee Teachers Report Less Demanding Reading and Writing at 8th Grade

Question (percent agreeing with statement)	Different from national average at 5% level	No. comparison states with statistically significant higher percent	No. comparison states with no significant difference	No. comparison states with statistically significant lower percent
Integrate reading and writing—strong emphasis	<	4	4	0
Writing 3 or more pages at least 1–2 times a week	<	6	2	0
Discuss understanding of written material	=	5	3	0
More than one draft—often or sometimes	<	6	2	0
Write long answers in tests at least once a week	<	3	4	1
Assess using essays or assigned papers—at least 1–2 times a month	<	5	2	1

< indicates Tennessee has statistically significant lower percent than the national average.
> indicates Tennessee has statistically significant higher percent than the national average.
= indicates Tennessee has no statistically significant difference with the national average.

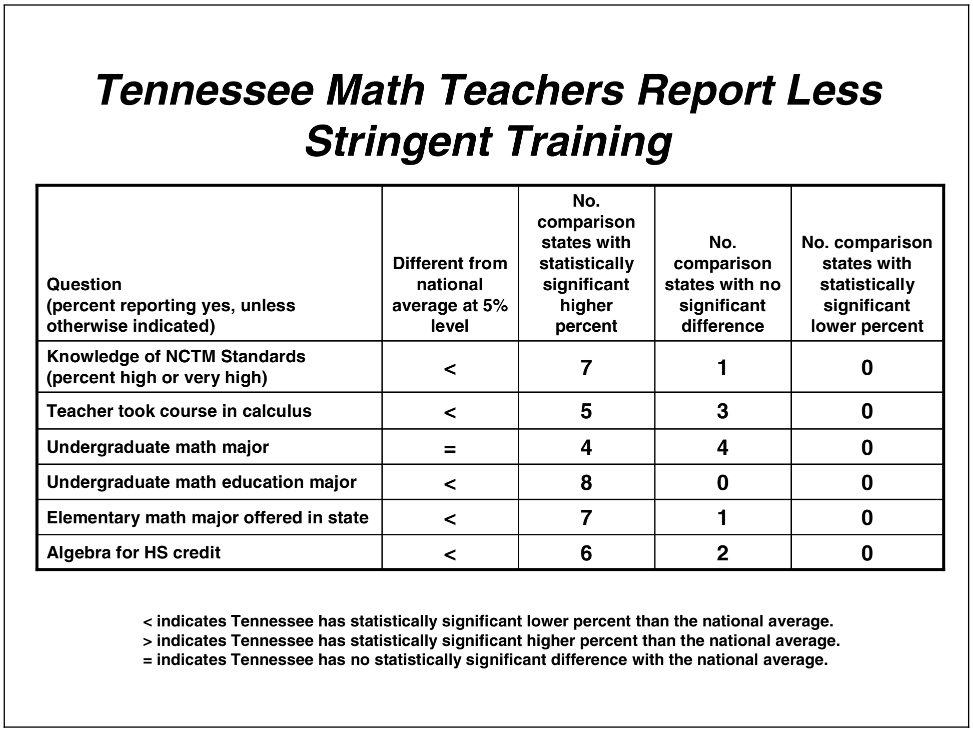
Beyond accountability reforms, other factors, such as teacher pedagogical practices or training, may also affect achievement. In these cases, the NAEP teacher survey responses also provide interesting contrasts between Tennessee and other states.

This chart shows the responses provided by 8th grade reading/English teachers in Tennessee in terms of their pedagogical practices, again with a comparison to the national average response and those for the comparison states. Tennessee teachers generally report using less stringent pedagogical approaches than teachers in several comparison states. Tennessee teachers report the following:

- Less emphasis on integrating reading and writing
- Less often requiring the writing of three or more pages
- Less often discussing interpretations of written material
- Less often requiring more than one draft
- Less often requiring long, written answers in testing
- Less often using essays or assigning papers for assessment.

The states that had significant differences from Tennessee on these measures were states that had higher raw scores and higher scores for students, controlling for family background differences. These states included North Carolina, Virginia, Georgia, and South Carolina.

These questions do not offer definitive evidence, but suggest that more advanced pedagogical techniques may not be demanded by Tennessee state standards or by the type of state testing used.



Tennessee Math Teachers Report Less Stringent Training

Question (percent reporting yes, unless otherwise indicated)	Different from national average at 5% level	No. comparison states with statistically significant higher percent	No. comparison states with no significant difference	No. comparison states with statistically significant lower percent
Knowledge of NCTM Standards (percent high or very high)	<	7	1	0
Teacher took course in calculus	<	5	3	0
Undergraduate math major	=	4	4	0
Undergraduate math education major	<	8	0	0
Elementary math major offered in state	<	7	1	0
Algebra for HS credit	<	6	2	0

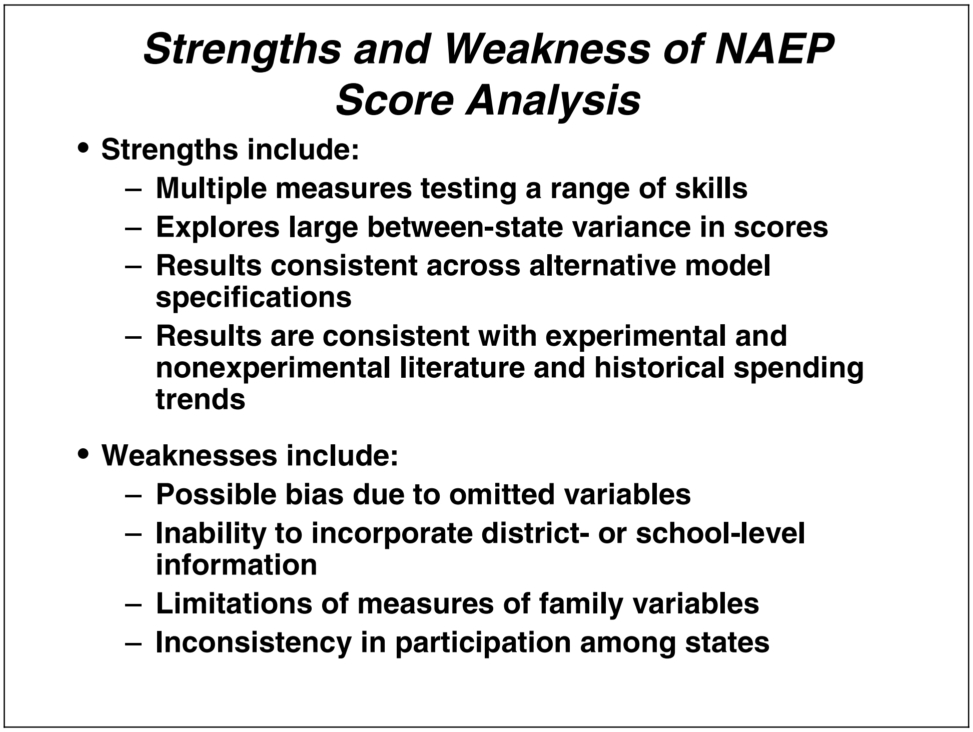
< indicates Tennessee has statistically significant lower percent than the national average.
> indicates Tennessee has statistically significant higher percent than the national average.
= indicates Tennessee has no statistically significant difference with the national average.

Eighth grade math teachers also reported on their training as part of the NAEP survey. As seen in this chart, math teachers in Tennessee report less stringent training and knowledge than teachers in many comparison states. For example, their knowledge of National Council of Teachers of Mathematics (NCTM)—the most widely utilized math standards in the nation—lags seven comparison states. Their preparation also appears to fall behind that of several comparison states if measured by whether they had taken advanced math courses (such as calculus) and whether their undergraduate major was in mathematics or mathematics education. Part of the problem the teachers indicate is the lack of an available major in math education at the state universities they attended.

Finally, a lower percentage of students in Tennessee report taking algebra in 8th grade for high school credit, while a higher percentage report taking a lower-level 8th grade math course.

The states that had better-trained teachers and a higher proportion of students taking algebra also had higher raw scores on the NAEP tests and higher scores among students, controlling for differences in family background. These states included North Carolina, Virginia, Georgia, and South Carolina.

Again, these data are only suggestive of a hypothesis that state standards and testing may not be encouraging more advanced math instruction and/or that Tennessee math teachers may be less prepared to teach such material.



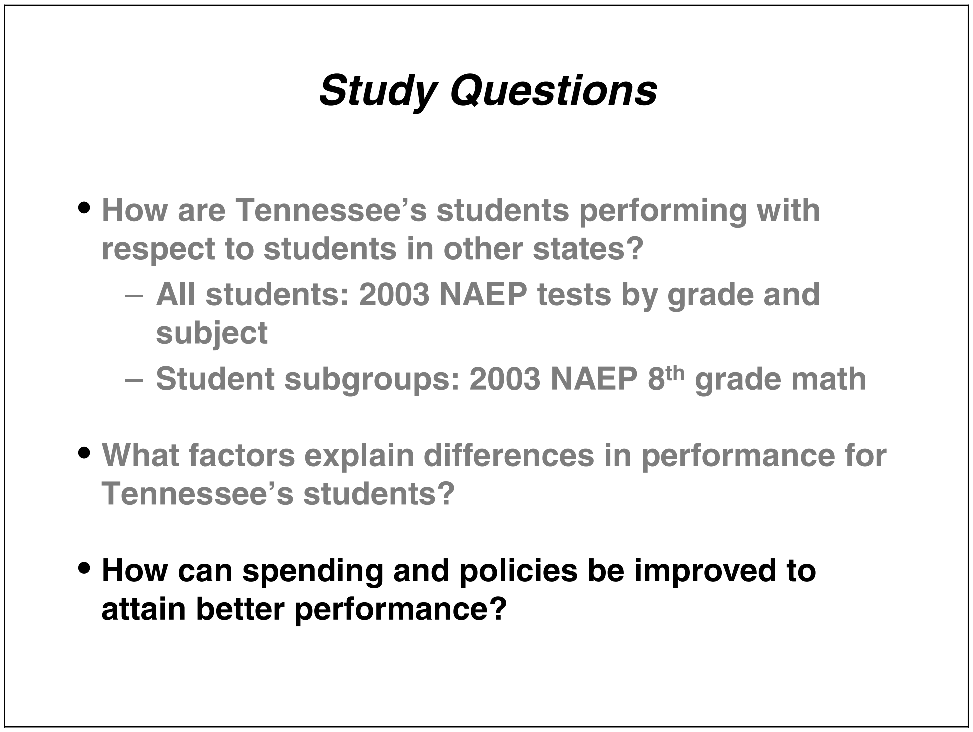
Strengths and Weakness of NAEP Score Analysis

- **Strengths include:**
  - **Multiple measures testing a range of skills**
  - **Explores large between-state variance in scores**
  - **Results consistent across alternative model specifications**
  - **Results are consistent with experimental and nonexperimental literature and historical spending trends**
- **Weaknesses include:**
  - **Possible bias due to omitted variables**
  - **Inability to incorporate district- or school-level information**
  - **Limitations of measures of family variables**
  - **Inconsistency in participation among states**

The analysis undertaken here using NAEP data across states aims to link differences in state achievement with family characteristics and educational spending and policies during an important period in American education. In this period, each state and its school districts made an effort to reform and improve its education system. The NAEP data from 1990–2003 are the only achievement data that can readily be used to validly compare state performance and attempt to explain differences at this important time in American education. Thus, NAEP data from this period must be analyzed to understand what might explain differences across states and whether reforms are working. But such results also need to be assessed with respect to the wider experimental results and nonexperimental results, and take account of the strengths and weaknesses of the analysis.

The strengths of this analysis include the following: (1) the model is based on 17 separate tests in two subjects and two grades over a 13-year period and provides over 700 observations of state achievement; (2) the NAEP evaluates not only lower-level skills through multiple-choice items, but also higher-level critical-thinking skills through open-ended items; (3) variation across states in almost all dependent variables is quite large compared to within-state district or school variation; (4) the analysis uses both random- and fixed-effects models that incorporate different statistical assumptions; (5) the model is consistent with the experimental effects of class size reductions in lower grades and pre-kindergarten programs; (6) these results also show consistency with the historical trends in achievement and spending that suggested that large achievement gains among minority and disadvantaged students occurred at the time when additional spending was directed to programs that would primarily benefit minority and disadvantaged students; and (7) none of the effects measured are inconsistent with the results of the nonexperimental literature, although because of the wide range of such measurements, this standard is not hard to meet.

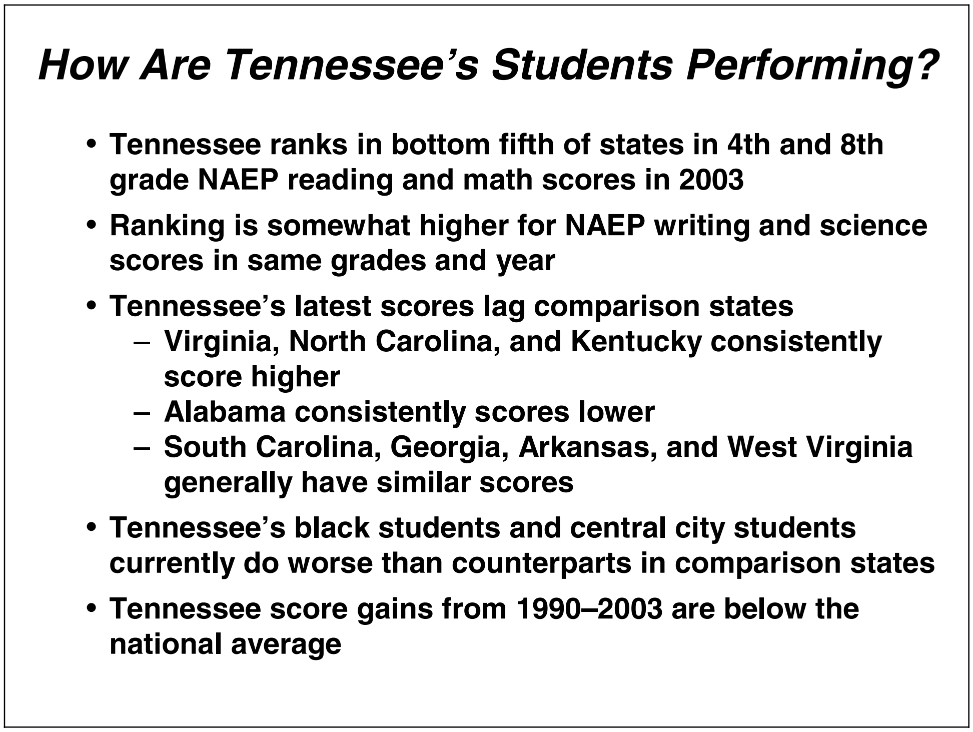
The weaknesses of the model include the following: (1) possible bias in the results from several sources, including missing variables, selectivity, and non-linearities; (2) bias resulting from the inability to incorporate district- and school-level information in the analysis (also known as the ecological fallacy); (3) the limited data on family variables available directly from the NAEP, necessitating the use of U.S. Census data and a weighting procedure for family variables using an alternative achievement test; (4) the absence of several family variables that other research has shown to be linked to achievement, but which can be collected only through parental surveys; (5) a lack of data on within-race/ethnicity changes in family characteristics across states; and (6) inconsistency in the participation of states so that data are not available for all 48 contiguous states for all 17 tests.



Study Questions

- How are Tennessee’s students performing with respect to students in other states?
  - All students: 2003 NAEP tests by grade and subject
  - Student subgroups: 2003 NAEP 8th grade math
- What factors explain differences in performance for Tennessee’s students?
- **How can spending and policies be improved to attain better performance?**

We now summarize our findings in terms of the first two study questions and suggest directions for future educational policies in Tennessee.



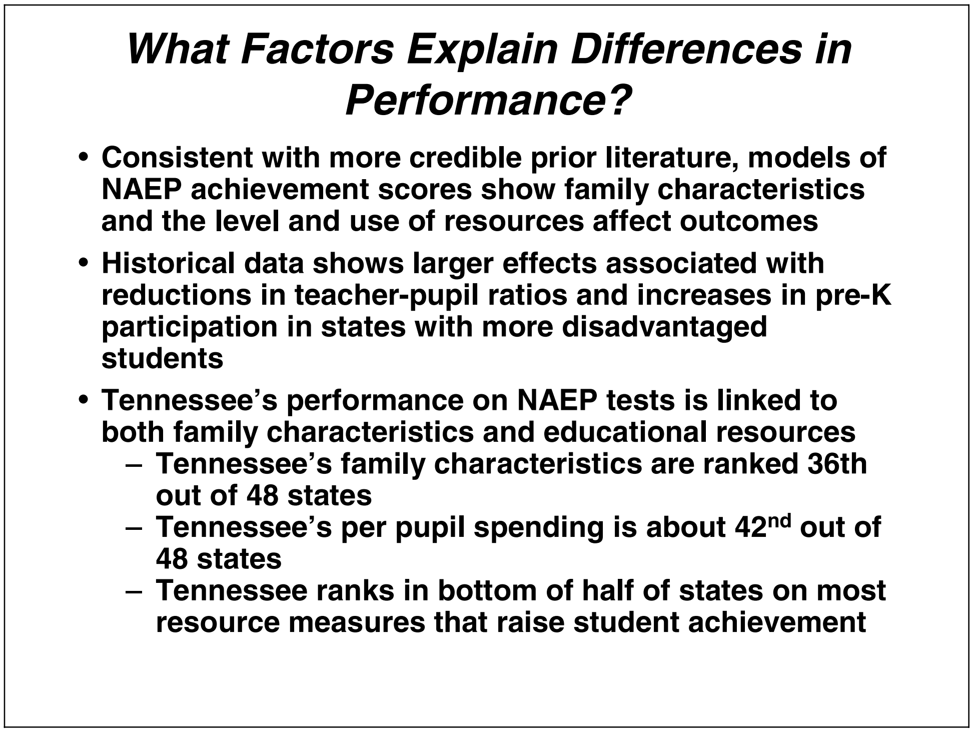
How Are Tennessee's Students Performing?

- **Tennessee ranks in bottom fifth of states in 4th and 8th grade NAEP reading and math scores in 2003**
- **Ranking is somewhat higher for NAEP writing and science scores in same grades and year**
- **Tennessee's latest scores lag comparison states**
  - **Virginia, North Carolina, and Kentucky consistently score higher**
  - **Alabama consistently scores lower**
  - **South Carolina, Georgia, Arkansas, and West Virginia generally have similar scores**
- **Tennessee's black students and central city students currently do worse than counterparts in comparison states**
- **Tennessee score gains from 1990–2003 are below the national average**

Our study aimed to assess the performance of Tennessee students on achievement tests in comparison with students in other states. For 2003, the most recent year data were available, Tennessee consistently ranked in the bottom fifth of states in 4th and 8th grade reading and math scores (as low as 42nd out of 48 states on 4th grade reading and as high as 38th out of 48 states on 8th grade reading). For the smaller set of states participating in the NAEP 4th and 8th grade writing and science tests in 2002 and 2000, respectively, Tennessee ranked somewhat higher (estimated as high as 33rd and as low as 37th, assuming all 48 states had taken these subject tests).

Of the eight comparison states we identified, three states had consistently higher statistically significant scores than Tennessee in 2003: North Carolina, Virginia, and Kentucky. Tennessee had statistically significant, consistently higher scores than Alabama. Tennessee's scores were not significantly or consistently different from scores in South Carolina, Arkansas, Georgia, and West Virginia. However, the performance of black and inner city students in Tennessee on the 2003 8th grade math test does not compare as favorably with comparison states as that of white students and students in rural areas.

Finally, between 1990 and 2003, the average annual score gain for Tennessee across all tests was 0.5 percentile points, below the national average and significantly below the gains of North Carolina and South Carolina, the highest-performing comparison states.



What Factors Explain Differences in Performance?

- Consistent with more credible prior literature, models of NAEP achievement scores show family characteristics and the level and use of resources affect outcomes
- Historical data shows larger effects associated with reductions in teacher-pupil ratios and increases in pre-K participation in states with more disadvantaged students
- Tennessee’s performance on NAEP tests is linked to both family characteristics and educational resources
  - Tennessee’s family characteristics are ranked 36th out of 48 states
  - Tennessee’s per pupil spending is about 42nd out of 48 states
  - Tennessee ranks in bottom of half of states on most resource measures that raise student achievement

The second question we addressed concerned the factors that account for Tennessee’s differential performance. To answer that question, we estimated a model using state-level data that related NAEP 4th and 8th grade math and reading scores between 1990 and 2003 to a composite measure of family background, a measure of family mobility, and several measures of educational resources that account for a large share of educational spending. The results of this model suggest that variation in achievement scores across states and over time can be partially explained by a state’s family characteristics and its pattern of educational expenditures. Our findings that a lower pupil-teacher ratio in grades 1 to 4 and a higher participation rate in public pre-kindergarten programs positively affect achievement is consistent with experimental studies that evaluate the effects of reducing class size and providing high-quality preschool. The positive effects on student achievement of raising teacher salaries estimated in our model are confirmed in other nonexperimental studies, but this has not been the subject of experimental evaluations. No previous studies are available on the effect of the adequacy of teacher resources as measured in this study.

When focusing on 4th grade achievement scores, our models indicate that the effect on achievement of lowering pupil-teacher ratios in the early grades and raising pre-kindergarten participation rates is strongest in states with a higher proportion of disadvantaged families. Again, this is consistent with the findings of experimental evaluations of class size reduction and preschool interventions, where the effects have been found to be strongest for children at risk of poor educational performance.

Given the composition of families in Tennessee and the state's allocation of educational resources, our model can be used to explain Tennessee's performance vis-à-vis the eight comparison states we identified, as well as all 48 states included in our analysis. First, in terms of family characteristics, Tennessee has a relatively higher-risk population, ranking 36th out of 48 states on our composition measure of family background. Second, the dollar resources devoted to education measured by per pupil spending are among the lowest in the country, ranking 42nd out of 48 states. Considering the specific resource measures included in our model, with the exception of the pupil-teacher ratio, Tennessee ranks in the bottom half among all the states on the resource measures that raise student achievement. On the measure of teacher resources, Tennessee ranks well below the eight comparison states. Teacher salaries and participation in public pre-kindergarten are near the average for the comparison states. In contrast, Tennessee's pupil-teacher ratio is lower than five of the other comparison states and lower than 27 of the total 48 states we examine.

Other Factors Deserve Consideration, Too

- **Estimated NAEP models do not account for all score gains between 1990 and 2003 in some states**
- **Models explored role for two other factors**
  - **Inexperienced teachers associated with lower achievement**
  - **No relationship between achievement and teachers with advanced degrees**
- **In Tennessee, NAEP teacher surveys suggest two other explanations for lack of achievement gains after accounting for resource levels**
  - **Less connection to and possible effectiveness of the standards-based accountability system**
  - **Less emphasis on and preparation to use advanced pedagogical techniques or teach advanced courses**

While family background and the specific educational resources we considered can explain some of the variation in achievement scores across states, some unexplained variation remains. When we use the model to explain the average annual score gains between 1990 and 2003, there is little unexplained variation for Tennessee. However, there are large score gains in other states, including comparison states North Carolina and South Carolina, that are not explained by the family background and educational resource measures included in the model. These and other states made large gains in achievement scores beyond what would have been expected from the resources we analyze.

Two other resource measures were included in additional regression models we explored and these might account for some of the unexplained differences among states. The results of the models indicate that a higher fraction of inexperienced teachers is associated with lower achievement. Tennessee has higher levels of inexperienced teachers than all comparison states except North Carolina. This may be the result of reductions in pupil-teacher ratios in the late 1990s. We found no effect on test scores for a measure of the proportion of teachers with advanced degrees. Although other research has also found that, in general, advanced degrees among teachers do not correlate with higher achievement, there is some evidence that subject-specific degrees can raise achievement scores.

Our analysis looked to other sources, including teacher responses on surveys administered with the NAEP, to identify other possible explanations for the residual score gains. These surveys suggest that Tennessee teachers report significantly lower positive connections to their accountability systems than their counterparts in most comparison states. They report that standards are less clear and less useful for planning curriculum, and that there are fewer resources for training and implementing the system than in most comparison states and in particular comparison states with the largest unexplained gains in NAEP scores. Tennessee English/reading teachers also report using less advanced pedagogical techniques, and Tennessee math teachers seem to have fewer credentials and less of the knowledge required to teach more advanced courses. Tennessee students also seem to less frequently take algebra for high school credit.

The data we have presented are not definitive, but only suggestive that Tennessee's accountability system and its teachers are not tightly linked in a way that might drive curriculum, pedagogy, and credentials to higher levels. Moreover, researchers have yet to demonstrate the benefits in terms of achievement scores from particular features of a standards-based accountability system.

Implications for Tennessee's Education Policy

- Tennessee already devotes significant resources to those approaches shown to have had larger favorable effects on achievement in the past
  - Lower class sizes in elementary grades
  - Expanding public pre-kindergarten programs
- Tennessee may benefit from devoting more funds to teacher resources, a factor shown to raise achievement where it lags other states
- Tennessee may be less justified in devoting further resources to raising teacher salaries, a factor that appears to be a higher-cost way to raise achievement
- Tennessee should evaluate changes in other educational policies
  - Standards-based accountability system
  - Teacher compensation, training, and pedagogy

The findings from this study have several implications for Tennessee's future educational policy. The research evidence suggests that Tennessee is justified in devoting substantial resources to lower class sizes in the elementary grades and raising the proportion of children in public pre-kindergarten programs. While our findings do not indicate the optimal level of spending in these areas, our estimates suggest that these are areas that have generated the largest returns in the past in terms of test score increases for a given dollar of investment. In the effort to expand public pre-kindergarten programs, Tennessee should continue to maintain the research-based standards associated with high-quality programs.

Given that Tennessee lags other states in teacher assessments of the adequacy of resources—another factor associated with higher achievement—the state should examine potential deficiencies in this area and consider ways to reallocate other spending toward efficient forms of teacher resources. On the other hand, while higher teacher salaries were shown to raise achievement, they do so at a relative higher cost. Given that Tennessee has salaries close to the national average, there may be less justification for using this policy lever to raise educational attainment. Since teacher salaries are the largest expenditures in education budgets, modest restraints in future salary increases may provide a source for channeling more funds into teachers' resources.

Finally, although the research base needed to guide decisionmaking is weak, Tennessee should assess the need for reforms in other areas that may be linked to improved school performance. This includes the state's standards-based accountability system, as well as its approach to teacher compensation, teacher training, and pedagogy in the classroom. For example, a useful next step would be to investigate the current standards-based accountability system in Tennessee and selected other states with the goal of discovering possible differences that might explain Tennessee's slower NAEP score improvements between 1990 and 2003, and addressing the issues that teachers have with the current system. A suggested set of objectives for such a study include the following:

- Determine the differences in improvement, particularly between Tennessee and North and South Carolina
- Assess the link between standards and curriculum to determine why teachers in Tennessee report that standards are less clear and useful for curriculum planning
- Assess whether Tennessee's standards and testing program has the appropriate balance of emphasis on basic and critical-thinking skills
- Ensure that teachers are provided appropriate training and resources to understand and support the system.

Appendix

This appendix contains brief descriptions of the methodology, data, and estimation results that support the analysis presented in the body of the report. The reader is referred to Grissmer, Flanagan, et al. (2000) for a much more complete description of the methodology and data used for the regression results.

The results of this study utilize the same methodology and data sources used in the earlier analysis, but these results reflect the addition of 10 new tests from the 1998 to 2003 to the original seven tests from the 1990–1996 period addressed in the earlier analysis. The other new addition to this analysis is the estimates of both raw scores and "full population–adjusted" scores. The full population estimates are used to control for changing exclusion rates (i.e., the fraction of sampled students who did not take the test) that became much more volatile in the 1998–2003 period. We test for the effects of changing exclusion rates by comparing the estimates using the full population score for each state and test and the actual raw (or unadjusted) NAEP scores. The full population estimates utilize the methodology in McLaughlin (2000; 2001) and a data set provided in McLaughlin that imputes scores to all excluded students. This imputation is made on the basis of information provided on each student chosen in the sample (whether included or excluded from the tests) by the student and by the teacher. This information includes an array of variables about the student, including why students are excluded. These full population estimates can, theoretically, provide estimates that are not sensitive to exclusions. In general, the coefficients for policy variables change little for estimates based on raw versus adjusted data, but the trends for a few states can have moderate changes due to their changing exclusion rates. The major weakness of the full population score methodology is that imputations sometimes are made for students with characteristics far outside the range of characteristics of students in the sample upon which the imputations are made.

In the remainder of the appendix, we first describe the basic model specifications used in the analysis. We then provide a brief overview of the data. The regression results follow in the third section.

Model Specification and Estimation

We estimate random- and fixed-effects models using a panel data set by state for the entire sample of state test observations from 1990 to 2003. We have included 48 states with up to 17 tests per state (Alaska and Hawaii are excluded due to atypical demographics). The total data set consists of 696 observations. We estimate scores using two methods to account for the gains in scores over time. The first method simply introduces dummy variables that measure the difference between a given test and the earliest test in a given grade and subject. The second method utilizes an annualized trend variable by grade and subject to account for score gains. The equation using trends (the second approach) is:

$$\begin{aligned} y_{ij} = \quad & a + g_1 T_{4math} + g_2 T_{8math} + g_3 T_{4read} + g_4 T_{8read} + g_1 T_{4math} + \\ & g_5 d_{4math} + g_6 d_{8math} + g_7 d_{8read} + \sum_k b_k F_{ijk} + \sum_k c_k G_{ijk} + u_i + e_{ij} \text{ ,} \end{aligned} \qquad (1)$$

where y_{ij} is the normalized test score (raw score or full population–adjusted score) for the *i*th state (i = 1,48) in the *j*th test (j = 1,17); $T_{4\ math}$, $T_{8\ math}$, $T_{4\ read}$, and $T_{8\ read}$ are separate trends for each test, respectively (that is, each variable equals zero for scores not from the associated test and year of testing); $d_{8\ math}$, d_{4math}, d_{8read} are dummies for each test (that is, each variable equals one for scores from the associated test and zero otherwise); F_{ijk} is the *k*th family variable; G_{ijk} is the *k*th resource variable; and b_k and c_k and g is the estimated regression coefficient, u_i is the random (fixed) effect for state i, and $e_{ij,}$ is the usual identical and independently distributed error term. We estimate three versions of each model with state trends: with no other controls; with family and demographic controls; and, finally, with family, demographic, and resource controls. The alternative method (the first approach) includes dummy gain variables rather than the four trend variables as a way of accounting for gains in scores. Some results derive from models that include interaction terms in all equations involving family variables between family and subject and grade-specific dummies that allow for different slopes for the family coefficients across subjects and grades.

The primary results presented in the body of the paper use the exclusion-adjusted scores and data from all tests. We also estimate models using two different assumptions concerning the validity of the 2003 4th grade math score. In the primary analyses, the 2003 4th grade math scores are treated as valid and are used in modeling. In alternative analyses, the 2003 4th grade math scores are treated as invalid and are not used in modeling. The sensitivity of results to the 2003 4th grade math scores is tested because this score showed an average gain of over one-half of a grade level for all students nationwide over three years between NAEP administrations. This is a gain that far surpasses the gains of earlier tests, and suggests a possible flaw in the statistical procedures used in sampling, norming, or estimating test scores. Hence, we did not want our results to be overly sensitive to this suspect test.

Data

Beginning in 1990, NAEP tests were administered to representative samples of students in participating states to allow a valid comparison of achievement performance across states. Seventeen state tests have been given in 4th and 8th grade math and reading between 1990 and 2003. Since participation was voluntary until 2003, the sample of states changes from test to test, as shown in Table A.1.

Table A.1. Number of States Participating in NAEP Tests by Grade, Subject, and Year

Year	Grade	Subject	Number of States
1990	8th	Math	38
1992	8th	Math	42
1996	8th	Math	41
2000	8th	Math	39
2003	8th	Math	50
1992	4th	Math	42
1996	4th	Math	44
2000	4th	Math	40
2003	4th	Math	50
1992	4th	Reading	42
1994	4th	Reading	39
1998	4th	Reading	30
2002	4th	Reading	41
2003	4th	Reading	50
1998	8th	Reading	35
2002	8th	Reading	43
2003	8th	Reading	50

The NAEP state tests are based on a redesign of the NAEP Main Assessment in the late 1980s; they include multiple-choice items as well as constructed-response items requiring responses of a few sentences to a few paragraphs in length. Students require approximately 90 minutes to complete NAEP testing for a given subject. Matrix sampling of questions is used to permit testing of a broad range of knowledge while limiting the time each student is tested. Balanced incomplete block spiraling of questions ensures that effects from the placement of questions within booklets and groupings of questions are minimized.

Two types of exclusions from testing are allowed: Limited English Proficiency (LEP) and Individualized Education Plan/Disabled (IEP/DS). Approximately 1–2 percent of students nationally are excluded for LEP and about 4–6 percent for IEP/DS. Exclusion rates have been increasing in recent years, and patterns of exclusion between states also shifted in the 1998–2003 tests. One reason for increasing exclusions is that NAEP does not test students who either do not take their state-administered tests or who are allowed accommodations on their state tests. Since states are adjusting at different rates in allowing accommodations on these tests, the pattern of NAEP exclusions has changed, and this might be a significant factor in explaining some of the achievement gains in certain states.

Nonparticipation also has the potential to bias results if there are significant differences across states. Both schools and individual students can choose not to participate in NAEP assessments, even if a state has agreed to participate. A higher proportion of nonparticipation comes from decisions by school officials than from individual nonparticipation decisions by the parent of a student. However, substitution of schools is attempted in the case of nonparticipation. Nationally, participation rates in the NAEP for grades 4 and 8 after substitution are approximately 85 percent. The range of variation across individual states is 74 to 100 percent. There is increasing evidence that participation rates tend to be lower in schools with lower SES (socioeconomic status) indicators, resulting in a correlation across tests in participation rates. States with lower participation in one test also tend to have lower participation in other tests. An earlier study included participation rates in regression equations and concluded it was not a significant factor in explaining trends or value-added measures of change (Grissmer, Flanagan, et al., 2000). We include a participation variable in the regression to control for changing participation rates.

The NAEP tests during this period were low-stakes tests and, therefore, not subject to many of the concerns that are involved in high-stakes testing (Hamilton, Stecher and Klein, 2002; Heubert and Hauser, 1999). NAEP individual scores are never reported back to students, teachers, or principals. It is also difficult, if not impossible, to "teach to the NAEP test," since only a few students from schools throughout the state take the tests. However, to the extent that state-administered tests are more aligned with NAEP tests in some states, it is possible that certain "preparation" practices affecting state testing would indirectly impact NAEP results to an unknown degree.

Regression Results

Tables A.2 and A.3 provide the random- and fixed-effects results, respectively, for the full sample for both raw scores and full population–adjusted scores as the dependent variable. For each dependent variable, we present three sets of results. The first model shows estimates of score gains for each test from the earliest test. For instance, for the first model using raw scores in Table A.2, the gain in the 8th grade tests from 1990 to 1992 was 0.095 standard deviation units, the gain from 1990 to 1996 was 0.198 standard deviation units, the gain from 1990 to 2000 was 0.293 standard deviation units, and the gain from 1990 to 2003 was 0.427 standard deviation units. These results do not control for changing family characteristics or educational resources or policies. The second model adds controls for family characteristics, which show strong significance at each grade and test, and, generally, the estimated gains increase. This increase is due to the changing demographics, particularly the increasing Hispanic population that—other things equal—depresses scores. The gains in this model reflect the gains that would have occurred if the demographics of the population had not changed. The third model adds the policy variables and reflects the impact of changing resources/policies on score gains. Although nearly all of the policy variables show significant effects with appropriate signs, a significant part of the gains is still present, suggesting that resources cannot account for a large part of the gains. The results are similar when the full population score is used as the outcome measure compared with the use of raw scores.

The fixed-effects models in Table A.3 generally show effects that are less significant for policy variables (except for pre-kindergarten participation, which increases in statistical significance). The fixed-effects models have fewer degrees of freedom (48 state dummies introduced), but more importantly, they reduce the size and significance of the family variables. Our view is that the assumptions in the random-effects models are more realistic than those in the fixed-effects models.

Table A.2. Random-Effects Regressions for Raw Score and Full Population–Adjusted Score Using Gain Dummies

	Raw Score			Full Population Score		
Variable	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Dummy 8th Grade Math	0.110 ***	0.164	0.284	0.128 ***	0.215	0.308 *
Dummy 4th Grade Math	0.120 ***	0.220	0.320 **	0.146 ***	0.159	0.236 *
Dummy 4th Grade Reading	0.119 ***	-0.088	0.005	0.092 ***	0.197 *	0.123
Gain 8th Grade Math 1990–1992	0.095 ***	0.103 ***	0.088 ***	0.094 ***	0.103 ***	0.089 ***
Gain 8th Grade Math 1990–1996	0.198 ***	0.221 ***	0.140 ***	0.173 ***	0.193 ***	0.122 ***
Gain 8th Grade Math 1990–2000	0.293 ***	0.318 ***	0.178 ***	0.235 ***	0.262 ***	0.141 ***
Gain 8th Grade Math 1990–2003	0.427 ***	0.459 ***	0.292 ***	0.388 ***	0.421 ***	0.274 ***
Gain 4th Grade Math 1992–1996	0.101 ***	0.103 ***	0.033 ***	0.069 ***	0.070 ***	0.012
Gain 4th Grade Math 1992–2000	0.207 ***	0.225 ***	0.108 ***	0.164 ***	0.178 ***	0.079 ***
Gain 4th Grade Math 1992–2003	0.502 ***	0.521 ***	0.380 ***	0.487 ***	0.503 ***	0.382 ***
Gain 4th Grade Reading 1992–1994	-0.068 ***	-0.093 ***	-0.124 ***	0.067 ***	0.094 ***	0.119 ***
Gain 4th Grade Reading 1992–1998	0.004	-0.028 *	-0.118 ***	0.023	0.058 ***	0.133 ***
Gain 4th Grade Reading 1992–2002	0.102 ***	0.091 ***	-0.031	0.093 ***	0.079 ***	0.026
Gain 4th Grade Reading 1992–2003	0.086 ***	0.077 ***	-0.055	0.076 ***	0.065 ***	0.049
Gain 8th Grade Reading 1998–2002	0.036 **	0.050 ***	0.018	0.040 ***	0.026 *	0.054
Gain 8th Grade Reading 1998–2003	0.009	0.039 **	-0.019	0.014	0.043 ***	0.008
Participation Rate	-0.001 *	-0.001	-0.001	0.001	0.001	0.001
Mobility x 8th Grade Math[a]		0.192	0.038		0.108	0.036
Mobility x 4th Grade Math[a]		0.155	-0.061		0.238	0.036
Mobility x 8th Grade Reading[a]		0.230	0.107		0.185	0.054
Mobility x 4th Grade Reading[a]		0.615 ***	0.409 **		0.690 ***	0.490 ***
Family (SES) x 8th Grade Math[b]		1.698 ***	1.734 ***		1.773 ***	1.813 ***
Family (SES) x 4th Grade Math[b]		1.165 ***	1.292 ***		1.151 ***	1.275 ***
Family (SES) x 8th Grade Reading[b]		0.932 ***	1.058 ***		1.020 ***	1.144 ***
Family (SES) x 4th Grade Reading[b]		1.100 ***	1.263 ***		1.153 ***	1.309 ***
Teacher Salary[c]			0.008 ***			0.007 ***
Pupil-Teacher Ratio, Grades 1 – 4[d]			-0.012 **			-0.014 **
% Teachers Reporting Low Resources[e]			-0.209 *			-0.200 *
% Teachers Reporting Med. Resources[f]			0.037			0.042
% Students in Public Pre-K[g]			0.002 **			0.002
Constant	0.016	-0.199	-0.199	0.065	0.244 *	0.108

NOTE: Statistical significance: ***1 percent; **5 percent; *10 percent.

[a] Mobility describes the stability of students' home environment and is the percentage of students reporting no change in schools in the past two years required by a change in residences. Missing 1990, 2002, and 2003 data were imputed.

[b] The Family (SES) measure is obtained from a fixed-effects regression with the following estimation equation: yij = a + bxij + uj + eij. The data are from the National Education Longitudinal Study 1988 (NELS:88) and the yij are the math and reading scores for the ith student in the jth school and the xij are a set of parent-reported family characteristics for the ith student in the jth school. In order to isolate the influence of family characteristics on test scores, fixed factors were incorporated into the model by the uj. This amounts to estimating a different intercept for each school in the NELS:88 data. The estimated regression coefficients, b, were then used to weight the same measures of family characteristics using a sample drawn from 1990 U.S. Census data for 8- to 10-year-old children (for 4th grade scores) or 12–14-year-old children (for 8th grade scores) by state. The statewide average census values and b were used to predict a state-level test score by race/ethnicity. This test score was then defined as an estimated average family characteristic score or estimated composite SES score for each racial/ethnic group in the state. The composite SES score was then adjusted by weighting each state's value by the racial and ethnic percentages of its NAEP student population for each NAEP test from 1990 to 2003. For more information on the technique used to create the composite SES score, see Grissmer, Kirby, et al. (1994) and Grissmer, Flanagan, et al. (2000).

[c] Average teacher salary is calculated to reflect the average annual teacher salary experienced by NAEP test takers by grade. Nominal average salaries were deflated to constant 2000 dollars and adjusted for cost-of-living differences between states. Cost-of-living adjustments were taken from Chambers (1996).

[d] Pupil-teacher ratio, grades 1–4, is calculated to reflect the average pupil-teacher ratio experienced by NAEP test takers (4th and 8th graders) in their first four years of schooling.

[e] Percentage of students enrolled in public pre-kindergarten was calculated as the ratio of pre-kindergarten students to students in first grade. The percentage reflects the average enrollment when NAEP test takers were of pre-kindergarten age.

[f] Percentage of teachers reporting low resources is the percentage of teachers responding, "I get some or none of the resources I need" to the question, "How well are you provided with the instructional materials and the resources you need to teach?" Missing 2002, and 2003 data were imputed.

[g] Percentage of teachers reporting medium resources is the percentage of teachers responding, "I get some of most of the resources I need" to the question, "How well are you provided with the instructional materials and the resources you need to teach?" Missing 2002 and 2003 data were imputed.

[h] The data set contains 696 state achievement scores. The earliest state scores in each test category (1990: 8th grade math; 1992: 4th grade math; 1992: 4th grade reading; and 1998: 8th grade reading) are converted to variables with a mean of zero and divided by the standard deviation of national scores at the time of the earliest test. The later tests in each category are subtracted from the mean of the earlier test and divided by the same national standard deviation. This technique maintains the test gains within each test category and allows for comparing gains across years.

[g] We test for the effects of changing exclusion rates by comparing the estimates using the "full population–estimated score" for each state and test and the actual NAEP raw scores. We obtain the full population estimates from a methodology and data set described in McLaughlin (2000; 2001). This methodology imputes scores to all excluded students. This imputation is made on the basis of information provided on each student chosen in the sample (whether included or excluded from the tests) by the student and by the teacher. This information includes an array of variables about the student, including why students are excluded. These full population estimates can, theoretically, provide estimates that are not sensitive to exclusions. The main weakness in this methodology is that the imputations are sometimes made far outside the parameter ranges of the variables.

Table A.3. Fixed Effects Regression Results for Raw Score and Full Population–Adjusted Scores Using Gain Dummies

	Raw Score			Full Population Score		
Variable	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Dummy 8th Grade Math	0.109 ***	0.138	0.264 *	0.127 ***	0.185	0.278 **
Dummy 4th Grade Math	0.119 ***	0.126	0.226 *	0.145 ***	0.055	0.125
Dummy 4th Grade Reading	0.118 ***	-0.108	-0.015	0.092 ***	-0.221 *	-0.148
Gain 8th Grade Math 1990–1992	0.095 ***	0.102 ***	0.090 ***	0.094 ***	0.101 ***	0.089 ***
Gain 8th Grade Math 1990–1996	0.199 ***	0.210 ***	0.139 ***	0.173 ***	0.181 ***	0.116 ***
Gain 8th Grade Math 1990–2000	0.295 ***	0.308 ***	0.179 ***	0.236 ***	0.251 ***	0.135 ***
Gain 8th Grade Math 1990–2003	0.426 ***	0.439 ***	0.288 ***	0.387 ***	0.398 ***	0.259 ***
Gain 4th Grade Math 1992–1996	0.101 ***	0.099 ***	0.031	0.070 ***	0.066 ***	0.007
Gain 4th Grade Math 1992–2000	0.208 ***	0.208 ***	0.102 ***	0.165 ***	0.160 ***	0.065 **
Gain 4th Grade Math 1992–2003	0.501 ***	0.500 ***	0.372 ***	0.486 ***	0.480 ***	0.364 ***
Gain 4th Grade Reading 1992–1994	-0.068 ***	-0.081 ***	-0.112 ***	-0.067 ***	-0.081 ***	-0.107 ***
Gain 4th Grade Reading 1992–1998	0.005	-0.020	-0.106 ***	-0.022	-0.050 **	-0.123 ***
Gain 4th Grade Reading 1992–2002	0.103 ***	0.080 ***	-0.030	0.094 ***	0.067 ***	-0.033
Gain 4th Grade Reading 1992–2003	0.085 ***	0.063 ***	-0.057 *	0.075 ***	0.049 **	-0.060 *
Gain 8th Grade Reading 1998–2002	0.035 *	0.041 **	0.014	-0.041 **	-0.036 *	-0.062 ***
Gain 8th Grade Reading 1998–2003	0.008	0.018	-0.030	0.012	0.021	-0.025
Participation Rate	-0.001	-0.001	0.000	-0.001	-0.001	0.000
Mobility x 8th Grade Math		0.069	-0.075		-0.027	-0.152
Mobility x 4th Grade Math		0.118	-0.057		0.198	0.041
Mobility x 8th Grade Reading		0.096	0.005		0.037	-0.069
Mobility x 4th Grade Reading		0.482 ***	0.315 **		0.544 ***	0.381 **
Family (SES) x 8th Grade Math		0.776 ***	0.790 ***		0.747 ***	0.730 ***
Family (SES) x 4th Grade Math		0.206	0.282		0.085	0.129
Family (SES) x 8th Grade Reading		-0.064	0.015		-0.089	-0.042
Family (SES) x 4th Grade Reading		0.066	0.177		0.003	0.077
Teacher Salary			0.008 ***			0.007 ***
Pupil-Teacher Ratio, Grades 1–4			-0.001			-0.006
% Teachers Reporting Low Resources			-0.258			-0.244
% Teachers Reporting Med. Resources			-0.037			-0.124
% Students in Public Pre-K			0.002 ***			0.002 **
Constant	-0.006	-0.098	-0.244	-0.088 *	0.132	-0.098

NOTE: See notes to Table A.2 for variable definitions. Statistical significance: ***1 percent; **5 percent; *10 percent.

Tables A.4 and A.5 show the results using trends rather than dummies to account for gains for random- and fixed-effects models, respectively. For instance, in the first model using raw scores, the average annual gain in 8th grade math was 0.033 standard deviation units—approximately 1 percentile point a year (see Table A.4). The estimates in the first model have family variables included with the trends, while the estimates in the second model have the added dummy for the 2003 4th grade math test. The dummy indicates that the 2003 4th grade math score was 0.215 standard deviation units above the trend line for 4th grade tests. This gain represents over a one-half grade increase over a normal gain in three years. The third and fourth models include the policy variables without and with the 2003 4th grade math test dummy. The policy coefficients show some sensitivity to accounting for gains through trends versus dummies (i.e., comparing Table A.4 versus Table A.2 and Table A.5 versus Table A.3). For instance, the pupil-teacher ratio shows much stronger effects with trends than with dummies. Our view is that the 13 dummy variables in Tables A.2 and A.3, as opposed to four trends in Table A.4 and A.5, represent a more rigorous test of the policy coefficients. The fixed-effects results (Table A.5) show the same pattern of differences with the random-effects models with respect to policy coefficients as do Tables A.2 and A.3 (again, the pre-kindergarten measure is the only one to increase in significance in the fixed-effects models).

Table A.6 shows the random-effects results using the same pattern of results as Tables A.4 and A.5, except individual trends by state are included rather than general trends. For instance, for the first model using raw scores, the state trends for Alabama show an annual gain in scores of 0.020 standard deviation units across all tests over the 13 years (1990–2003).

Table A.4. Random-Effects Regression Results for Raw Score and Full Population–Adjusted Score Using Average Annual NAEP Gains by Grade and Subject

	Raw Score				Full Population Score			
Variable	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Dummy 4th Grade Math 2003		0.215 ***		0.227***		0.255 ***		0.263 ***
Dummy 8th Grade Math	0.195	0.206	0.224	0.271	0.253 *	0.266 *	0.245	0.298 **
Dummy 4th Grade Math	0.224 *	0.188	0.268 **	0.256**	0.169	0.123	0.187	0.171
Dummy 4th Grade Reading	-0.109	-0.110	-0.065	-0.045	-0.209 **	-0.211 **	-0.186 *	-0.164
Avg. Ann. Gain 8th Grade Math	0.033 ***	0.033 ***	0.025 ***	0.023***	0.029 ***	0.029 ***	0.022 ***	0.019 ***
Avg. Ann. Gain 4th Grade Math	0.046 ***	0.028 ***	0.038 ***	0.017***	0.044 ***	0.023 ***	0.037 ***	0.014 ***
Avg. Ann. Gain 8th Grade Reading	0.006 **	0.007 **	-0.002	-0.002	0.001	0.002	-0.007 *	-0.007 *
Avg. Ann. Gain 4th Grade Reading	0.013 ***	0.013 ***	0.005	0.003	0.011 ***	0.011 ***	0.004	0.003
Participation Rate	0.000	-0.001	0.000	-0.001	0.001	0.000	0.001 *	0.000
Mobility x 8th Grade Math	0.092	0.055	0.047	-0.028	0.001	-0.052	-0.018	-0.106
Mobility x 4th Grade Math	-0.024	0.072	-0.167	-0.100	0.035	0.141	-0.093	-0.016
Mobility x 8th Grade Reading	0.160	0.136	0.070	0.038	0.108	0.070	0.004	-0.037
Mobility x 4th Grade Reading	0.458	0.437	0.316 **	0.266*	0.517 ***	0.480 **	0.384 **	0.323 *
Family (SES) x 8th Grade Math	1.697 ***	1.575 ***	1.731 ***	1.663***	1.784 ***	1.594 ***	1.804 ***	1.701 ***
Family (SES) x 4th Grade Math	1.217 ***	1.039 ***	1.348 ***	1.227***	1.222 ***	0.965 ***	1.336 ***	1.172 ***
Family (SES) x 8th Grade Reading	0.922 ***	0.787 ***	1.027 ***	0.962***	1.024 ***	0.814 ***	1.107 ***	1.007 ***
Family (SES) x 4th Grade Reading	1.108 ***	0.965 ***	1.246 ***	1.178***	1.177 ***	0.958 ***	1.291 ***	1.186 ***
Teacher Salary			0.005 *	0.006**			0.004	0.005 **
Pupil-Teacher Ratio, Grades 1–4			-0.019 ***	-0.018***			-0.023 ***	-0.022 ***
% Teachers Reporting Low Resources			0.028	-0.059			0.066	-0.034
% Teachers Reporting Med. Resources			0.105	0.086			0.024	0.001
% Students in Public Pre-K			0.002 *	0.002*			0.002	0.002
Constant	-0.247	-0.133	-0.099	-0.006	-0.379 ***	-0.233	-0.090	0.023

NOTE: See notes to Table A.2 for variable definitions. Statistical significance: ***1 percent; **5 percent; *10 percent.

Table A.5. Fixed-Effects Regression Results for Raw Score and Full Population–Adjusted Score Using Average Annual NAEP Gains by Grade and Subject

	Raw Score				Full Population Score			
Variable	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Dummy 4th Grade Math 2003		0.217 ***		0.229 ***		0.258 ***		0.265 ***
Dummy 8th Grade Math	0.152	0.164	0.178	0.229	0.199	0.215	0.181	0.241 *
Dummy 4th Grade Math	0.115	0.081	0.139	0.125	0.037	-0.004	0.032	0.015
Dummy 4th Grade Reading	-0.133	-0.135	-0.100	-0.079	-0.239 *	-0.241 **	-0.227 *	-0.203 *
Avg. Ann. Gain 8th Grade Math	0.031 ***	0.031 ***	0.026 ***	0.023 ***	0.027 ***	0.027 ***	0.022 ***	0.019 ***
Avg. Ann. Gain 4th Grade Math	0.044 ***	0.026 ***	0.038 ***	0.017 ***	0.041 ***	0.020 ***	0.036 ***	0.011 **
Avg. Ann. Gain 8th Grade Reading	0.002	0.003	-0.004	-0.004	-0.004	-0.002	-0.010 **	-0.010
Avg. Ann. Gain 4th Grade Reading	0.010 ***	0.010 ***	0.005	0.003	0.009 ***	0.009 ***	0.003	0.001
Participation Rate	0.000 ***	-0.001	0.001	-0.001	0.001 ***	0.000	0.001 ***	0.000
Mobility x 8th Grade Math	-0.038	-0.057	-0.076	-0.158	-0.155	-0.178	-0.141	-0.235
Mobility x 4th Grade Math	-0.067	0.050	-0.143	-0.078	-0.019	0.121	-0.061	0.015
Mobility x 8th Grade Reading	0.002	-0.003	-0.048	-0.088	-0.084	-0.091	-0.141	-0.187
Mobility x 4th Grade Reading	0.317 **	0.319	0.228	0.171	0.345 **	0.348 **	0.280 *	0.215
Family (SES) x 8th Grade Math	0.570	0.489 **	0.519 **	0.416 *	0.417	0.319	0.362	0.242
Family (SES) x 4th Grade Math	0.051	-0.089	0.062	-0.095	-0.192	-0.358	-0.179	-0.360
Family (SES) x 8th Grade Reading	-0.291	-0.384	-0.293	-0.393	-0.448	-0.558 **	-0.450	-0.566 **
Family (SES) x 4th Grade Reading	-0.153	-0.253	-0.139	-0.243	-0.352	-0.471 *	-0.341	-0.462 *
Teacher Salary			0.002	0.004			0.002	0.004
Pupil-Teacher Ratio, Grades 1–4			-0.014 **	-0.013 **			-0.021 ***	-0.020 ***
% Teachers Reporting Low Resources			-0.022	-0.121			0.020	-0.095
% Teachers Reporting Med. Resources			0.025	0.001			-0.068	-0.096
% Students in Public Pre-K			0.002 **	0.002 ***			0.002 *	0.002 **
Constant	-0.119	-0.017	0.054	0.140	-0.223	-0.102	0.141	0.241

NOTE: See notes to Table A.2 for variable definitions. Statistical significance: ***1 percent; **5 percent; *10 percent.

Table A.6. Random-Effects Regression Results Using Average Annual NAEP Gains by State for Raw Score and Full Population–Adjusted Score

	Raw Score				Full Population Score			
Variable	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Dummy 4th Grade Math 2003			0.255 ***	0.262 ***			0.279	0.287 ***
Dummy 8th Grade Math	0.532 ***	0.468 ***	0.605 ***	0.565 ***	0.565 ***	0.446 **	0.646	0.576 ***
Dummy 4th Grade Math	0.279 *	0.291 **	0.379 ***	0.363 ***	0.212	0.217	0.317	0.278 **
Dummy 4th Grade Reading	0.271 **	0.257 *	0.184	0.162	0.155	0.151	0.058	0.022
Participation Rate	0.000	0.000	-0.002 ***	-0.001 ***	0.001 *	0.001	-0.001	-0.001
Mobility x 8th Grade Math	-0.106	0.038	-0.389 ***	-0.391 ***	-0.127	0.094	-0.456	-0.430 ***
Mobility x 4th Grade Math	0.279 *	0.313 *	-0.209	-0.319 **	0.418 **	0.484 ***	-0.128	-0.226
Mobility x 8th Grade Reading	0.256 *	0.294 *	0.053	-0.091	0.242	0.307 *	0.004	-0.165
Mobility x 4th Grade Reading	-0.044	0.032	-0.132	-0.221 *	0.108	0.192	-0.004	-0.096
Family (SES) x 8th Grade Math	2.370 ***	2.247 ***	2.300 ***	2.101 ***	2.384 ***	2.350 ***	2.251 ***	2.048 ***
Family (SES) x 4th Grade Math	1.693 ***	1.654 ***	1.752 ***	1.633 ***	1.613 ***	1.660 ***	1.620 ***	1.502 ***
Family (SES) x 8th Grade Reading	1.389 ***	1.439 ***	1.312 ***	1.250 ***	1.440 ***	1.574 ***	1.296 ***	1.232 ***
Family (SES) x 4th Grade Reading	2.026 ***	1.987 ***	1.900 ***	1.774 ***	1.996 ***	2.000 ***	1.796 ***	1.670 ***
Teacher Salary		0.000		0.008 ***		-0.002		0.009 ***
Pupil-Teacher Ratio, Grades 1–4		-0.020 ***		-0.024 ***		-0.023 ***		-0.029 ***
% Teachers Reporting Low Resources		-0.086		-0.093		-0.067		-0.072
% Teachers Reporting Med. Resources		0.225 *		0.048		0.200		-0.030
% Students in Public Pre-K		0.000		-0.001		0.000		-0.001

Table A.6—Continued

	Raw Score				Full Population Score			
Variable	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Avg. Ann. Gain Alabama	0.020 ***	0.009 ***	0.018 ***	0.002	0.022 ***	0.014 ***	0.019 ***	0.002
Avg. Ann. Gain Arizona	0.023 ***	0.024 ***	0.018 ***	0.014 ***	0.021 ***	0.026 ***	0.016 ***	0.013 ***
Avg. Ann. Gain Arkansas	0.027 ***	0.016 ***	0.023 ***	0.007	0.025 ***	0.015 ***	0.021 ***	0.004
Avg. Ann. Gain California	0.030 ***	0.012 ***	0.029 ***	0.016 ***	0.032 ***	0.013 ***	0.031 ***	0.017 ***
Avg. Ann. Gain Colorado	0.034 ***	0.023 ***	0.028 ***	0.022 ***	0.033 ***	0.025 ***	0.027 ***	0.021 ***
Avg. Ann. Gain Connecticut	0.034 ***	0.029 ***	0.030 ***	0.021 ***	0.032 ***	0.025 ***	0.028 ***	0.018 ***
Avg. Ann. Gain Delaware	0.054 ***	0.041 ***	0.050 ***	0.038 ***	0.044 ***	0.031 ***	0.040 ***	0.027 ***
Avg. Ann. Gain Florida	0.045 ***	0.034 ***	0.041 ***	0.036 ***	0.045 ***	0.035 ***	0.040 ***	0.036 ***
Avg. Ann. Gain Georgia	0.034 ***	0.040 ***	0.029 ***	0.014 ***	0.033 ***	0.046 ***	0.027 ***	0.011 **
Avg. Ann. Gain Idaho	0.019 ***	0.015 ***	0.014 ***	0.002	0.018 ***	0.017 ***	0.013 ***	0.000
Avg. Ann. Gain Illinois	0.041 ***	0.033 ***	0.040 ***	0.024 ***	0.035 ***	0.033 ***	0.032 ***	0.016 **
Avg. Ann. Gain Indiana	0.035 ***	0.042 ***	0.029 ***	0.019 ***	0.033 ***	0.045 ***	0.026 ***	0.015 ***
Avg. Ann. Gain Iowa	0.020 ***	0.026 ***	0.012 ***	0.004	0.014 **	0.022 ***	0.005	-0.005
Avg. Ann. Gain Kansas	0.028 ***	0.022 ***	0.024 ***	0.015 ***	0.028 ***	0.024 ***	0.024 ***	0.015 ***
Avg. Ann. Gain Kentucky	0.029 ***	0.018 ***	0.026 ***	0.017 ***	0.024 ***	0.018 ***	0.020 ***	0.011 **
Avg. Ann. Gain Louisiana	0.043 ***	0.029 ***	0.040 ***	0.029 ***	0.039 ***	0.031 ***	0.035 ***	0.024
Avg. Ann. Gain Maine	0.011 **	0.020 ***	0.004	-0.007	0.011 **	0.018 ***	0.004	-0.009 *
Avg. Ann. Gain Maryland	0.039 ***	0.026 ***	0.036 ***	0.027 ***	0.033 ***	0.025 ***	0.029 ***	0.020 ***
Avg. Ann. Gain Massachusetts	0.038 ***	0.032 ***	0.034 ***	0.022 ***	0.032 ***	0.029 ***	0.029 ***	0.016 ***
Avg. Ann. Gain Michigan	0.033 ***	0.038 ***	0.029 ***	0.018 ***	0.030 ***	0.041 ***	0.026 ***	0.013 ***
Avg. Ann. Gain Minnesota	0.033 ***	0.029 ***	0.029 ***	0.020 ***	0.030 ***	0.030 ***	0.025 ***	0.015 ***
Avg. Ann. Gain Mississippi	0.029 ***	0.023 ***	0.024 ***	0.011 **	0.027 ***	0.023 ***	0.022 ***	0.008
Avg. Ann. Gain Missouri	0.025 ***	0.025 ***	0.019 ***	0.008	0.021 ***	0.026 ***	0.015 ***	0.002
Avg. Ann. Gain Montana	0.020 ***	0.021 ***	0.016 ***	0.007	0.017 ***	0.023 ***	0.012 **	0.003

Table A.6—Continued

	Raw Score				Full Population Score			
Variable	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Avg. Ann. Gain Nebraska	0.018 ***	0.009 ***	0.014 ***	0.004	0.016 ***	0.008 ***	0.011 **	0.000
Avg. Ann. Gain Nevada	0.024 ***	0.015 ***	0.021 ***	0.013 **	0.019 ***	0.012 ***	0.015 ***	0.008
Avg. Ann. Gain New Hampshire	0.024 ***	0.022 ***	0.018 ***	0.009 *	0.023 ***	0.024 ***	0.016 ***	0.006
Avg. Ann. Gain New Jersey	0.034 ***	0.035 ***	0.026 ***	0.013 **	0.032 ***	0.038 ***	0.023 ***	0.008
Avg. Ann. Gain New Mexico	0.016 ***	0.016 ***	0.013 ***	0.001	0.014 ***	0.015 ***	0.010 **	-0.004
Avg. Ann. Gain New York	0.043 ***	0.034 ***	0.040 ***	0.027 ***	0.039 ***	0.030 ***	0.036 ***	0.023 ***
Avg. Ann. Gain North Carolina	0.063 ***	0.055 ***	0.059 ***	0.044 ***	0.053 ***	0.052 ***	0.049 ***	0.033 ***
Avg. Ann. Gain North Dakota	0.014 ***	0.013 ***	0.008 *	-0.002	0.012 **	0.012 ***	0.006	-0.005
Avg. Ann. Gain Ohio	0.042 ***	0.040 ***	0.038 ***	0.024 ***	0.037 ***	0.039 ***	0.032 ***	0.018 ***
Avg. Ann. Gain Oklahoma	0.019 ***	0.022 ***	0.014 ***	0.005	0.017 ***	0.021 ***	0.012 ***	0.001
Avg. Ann. Gain Oregon	0.023 ***	0.020 ***	0.021 ***	0.014 ***	0.019 ***	0.020 ***	0.016 ***	0.010 *
Avg. Ann. Gain Pennsylvania	0.027 ***	0.030 ***	0.022 ***	0.010 *	0.025 ***	0.033 ***	0.019 ***	0.006
Avg. Ann. Gain Rhode Island	0.025 ***	0.001	0.023 ***	0.011 **	0.019 ***	-0.003	0.017 ***	0.004
Avg. Ann. Gain South Carolina	0.047 ***	0.039 ***	0.043 ***	0.029 ***	0.042 ***	0.036 ***	0.037 ***	0.022 ***
Avg. Ann. Gain Tennessee	0.020 ***	0.019 ***	0.016 ***	0.000	0.021 ***	0.023 ***	0.017 ***	-0.001
Avg. Ann. Gain Texas	0.046 ***	0.056 ***	0.040 ***	0.028 ***	0.039 ***	0.054 ***	0.032 ***	0.020 ***
Avg. Ann. Gain Utah	0.011 **	0.018 ***	0.007 *	-0.003	0.011 **	0.022 ***	0.007	-0.004
Avg. Ann. Gain Vermont	0.027 ***	0.009 ***	0.025 ***	0.012 *	0.025 ***	0.006 *	0.023 ***	0.009
Avg. Ann. Gain Virginia	0.038 ***	0.033 ***	0.034 ***	0.020 ***	0.032 ***	0.030 ***	0.027 ***	0.012 **
Avg. Ann. Gain Washington	0.025 ***	0.028 ***	0.021 ***	0.015 ***	0.029 ***	0.034 ***	0.024 ***	0.017 ***
Avg. Ann. Gain West Virginia	0.028 ***	0.005 *	0.025 ***	0.012 **	0.021 ***	-0.007 **	0.018 ***	0.005
Avg. Ann. Gain Wisconsin	0.025 ***	0.037 ***	0.018 ***	0.009 ***	0.021 ***	0.037 ***	0.013 ***	0.003
Avg. Ann. Gain Wyoming	0.021 ***	0.002	0.017 ***	0.009 **	0.020 ***	0.006 *	0.016 ***	0.008 *
Constant	-0.369 ***	-0.132	-0.058	0.176	-0.540 ***	-0.176	-0.186	0.159

NOTE: See notes to Table A.2 for variable definitions. Statistical significance: ***1 percent; **5 percent; *10 percent.

Table A.7 reports random- and fixed-effects results (4th grade tests only) for models that test the non-linearity of the pupil-teacher and pre-kindergarten effects. The first model in Table A.7 introduces both a squared term for pupil-teacher ratio and an interaction term with Family (SES). The results indicate that lowering the pupil-teacher ratio has a stronger effect in lower-SES states and that the effects have marginally diminishing returns. The second model in Table A.7 adds an interaction term between pre-kindergarten and the Family (SES) variable. This tests whether the effects of pre-kindergarten are stronger in states with a lower average SES, i.e., a state with more disadvantaged and minority students. The results indicate that effects are stronger in lower-SES states.

Table A.7. Random- and Fixed-Effects Regression Results (4th grade only) for Full Population Scores With Alternative Specifications

	Random Effects		Fixed Effects	
Variable	Model 1	Model 2	Model 1	Model 2
Dummy 4th Grade Math	0.396 ***	0.388 ***	0.325 ***	0.306 ***
Gain 4th Grade Math 1992–1996	0.020	0.023	0.045	0.039
Gain 4th Grade Math 1992–2000	0.126 ***	0.123 ***	0.173 ***	0.154 ***
Gain 4th Grade Math 1992–2003	0.425 ***	0.421 ***	0.483 ***	0.459 ***
Gain 4th Grade Reading 1992–1994	-0.122 ***	-0.118 ***	-0.090 ***	-0.090 ***
Gain 4th Grade Reading 1992–1998	-0.123 ***	-0.117 ***	-0.057 *	-0.065 *
Gain 4th Grade Reading 1992–2002	0.009	0.007	0.077 *	0.056
Gain 4th Grade Reading 1992–2003	-0.006	-0.007	0.062	0.042
Participation Rate	-0.001	-0.001	-0.001	-0.001
Mobility x 4th Grade Math	0.067	0.075	-0.204	-0.180
Mobility x 4th Grade Reading	0.570 ***	0.568 ***	0.217	0.211
Family (SES) x 4th Grade Math	-0.076	1.674 ***	-2.013 **	0.220
Family (SES) x 4th Grade Reading	-0.016	1.711 ***	-2.089 **	0.140
Teacher Salary	0.003	0.004	0.000	0.002
Pupil-Teacher Ratio, Grades 1–4	0.063	-0.014 *	0.089 **	0.008
Pupil-Teacher Ratio Squared	-0.002		-0.002 *	
Pupil-Teacher Ratio x Family (SES)	0.094 **		0.118 **	
% Teachers Reporting Low Resources	-0.334 **	-0.296 **	-0.356 **	-0.349 **
% Teachers Reporting Med. Resources	-0.207	-0.125	-0.248	-0.214
% Students in Public Pre-K	0.001	0.000	0.000	0.000
Public Pre-K x Family (SES)		-0.025 ***		-0.020 **
Constant	-0.646	-0.056	-0.733	-0.057

NOTE: See notes to Table A.2 for variable definitions. Statistical significance: ***1 percent; **5 percent; *10 percent.

Bibliography

Alexander, Karl L., Doris R. Entwisle, and Linda S. Olson, "Schools, Achievement, and Inequality: A Seasonal Perspective," *Educational Evaluation and Policy Analysis*, Vol. 23, No. 2, Summer 2001, pp. 171–191.

Barnett, W. Steven, "Long-Term Effects of Early Childhood Programs," *The Future of Children*, Vol. 5, No. 3, Winter 1995, pp. 25–50.

Barnett, W. Steven, Jason T. Hustedt, Kenneth B. Robin, and Karen L. Schulman, *The State of Preschool: 2004 State Preschool Yearbook*, New Brunswick, N.J.: National Institute for Early Education Research (NIEER), 2004. Online at http://nieer.org/yearbook2004/ (as of April 2006).

Campbell, J.R., C. M. Hombo and J. Mazzeo, *NAEP 1999 Trends in Academic Progress: Three Decades of Student Performance*, Washington, D.C.: National Center for Educational Statistics, 2000. Online at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2000469 (as of April 2006).

Carneiro, Pedro, Flavio Cunha, and James Heckman, "Interpreting the Evidence of Family Influence on Child Development," paper presented at the Economics of Early Childhood Development: Lessons for Economic Policy, conference co-hosted by the Federal Reserve Bank of Minneapolis and the McKnight Foundation in cooperation with the University of Minnesota, October 17, 2003.

Carnoy, Martin, and Susanna Loeb, "Does External Accountability Affect Student Outcomes? A Cross-State Analysis," *Education Evaluation and Policy Analysis*, Vol. 24, No. 4, Winter 2002, pp. 305–331.

Chambers, Jay G., *The Patterns of Teacher Compensation*, Washington, D.C.: National Center for Education Statistics, NCES 95-829, January 1996. Online at http://nces.ed.gov/pubs95/95829.pdf (as of April 2006).

Cook, Michael D., and William N. Evans, "Families or School? Accounting for the Convergence in White and Black Academic Performance." *Journal of Labor Economics*, Vol. 18, No. 4, October 2000, pp. 729–754.

Cook, Thomas D., "Randomized Experiments in Educational Policy Research: A Critical Examination of the Reasons the Educational Evaluation Community Has Offered for Not Doing Them," *Educational Evaluation and Policy Analysis*, Vol. 24, No. 3, Fall 2002, pp. 175–199.

Finn, Jeremy D., and Charles M. Achilles, "Tennessee's Class Size Study: Findings, Implications and Misconceptions," *Educational Evaluation and Policy Analysis*, Vol. 20, No. 2, Special Issue, Summer 1999, pp. 97–109.

Flanagan, Ann, and David Grissmer, "The Role of Federal Resources in Closing the Achievement Gaps of Minority and Disadvantaged Students," in John E. Chubb and Tom Loveless, eds., *Bridging the Achievement Gap*, Washington, D.C.: The Brookings Institution, 2002.

Flanagan, Ann, and David Grissmer, "Tracking the Improvement in State Achievement Using NAEP Data," in J. M. Ross, G. W. Bohrnstedt, and F. C. Hemphill, eds., *Instructional and Performance Consequences of High Poverty Schooling*, Washington, D.C: National Council for Educational Statistics, 2005.

Gill, Brian, Laura S. Hamilton, J. R. Lockwood, Julie A. Marsh, Ron Zimmer, Deanna Hill, and Shana Pribesh, *Inspiration, Perspiration, and Time: Operations and Achievement in Edison Schools*, Santa Monica, Calif.: RAND Corporation, MG 351-EDU, 2005. Online at http://www.rand.org/pubs/monographs/MG351/ (as of April 2006).

Gill, Brian, P. Mike Timpane, Karen E. Ross, and Dominic J. Brewer, *Rhetoric versus Reality: What We Know and What We Need To Know About Vouchers and Charter Schools*, Santa Monica, Calif.: RAND Corporation, MR-1118-EDU, 2001. Online at http://www.rand.org/pubs/monograph_reports/MR1118/ (as of April 2006).

Gilliam, Walter S., and Edward F. Zigler, *State Efforts to Evaluate the Effects of Prekindergarten: 1977 to 2003*, New Haven, Conn.: Yale University Child Study Center, 2004.

Goldhaber, Dan D., and Dominic J. Brewer, "Why Don't Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity," *Journal of Human Resources*, Vol. 32, No. 3, Summer 1997, pp. 505–523.

Greenwald, Rob, Larry V. Hedges, and Richard Laine, "The Effect of School Resources on Student Achievement," *Review of Educational Research*, Vol. 66, No. 3, Autumn 1996, pp. 361–396.

Grissmer, David, "The Continuing Use and Misuse of SAT Scores," *Journal of Psychology, Public Policy, and Law*, Vol. 6, No. 1, March 2000, pp. 223–232.

Grissmer, David, "Class Size Effects: Assessing the Evidence, Its Policy Implications, and Future Research Agenda," *Educational Evaluation and Policy Analysis*, Vol. 21, No. 2, Special Issue, Summer 1999, pp. 231–248.

Grissmer, David, and Ann Flanagan, "Exploring Rapid Score Gains in Texas and North Carolina," commissioned paper, Washington, D.C.: National Education Goals Panel, November 1998. Online at http://govinfo.library.unt.edu/negp/reports/grissmer.pdf (as of April 2006).

Grissmer, David W., Ann Flanagan, Jennifer H. Kawata, and Stephanie Williamson, *Improving Student Achievement: What Do State NAEP Scores Tell Us*, Santa Monica, Calif.: RAND Corporation, MR-924-EDU, 2000. Online at http://www.rand.org/pubs/monograph_reports/MR924/ (as of April 2006).

Grissmer, David W., Sheila Nataraj Kirby, Mark Berends, and Stephanie Williamson, *Student Achievement and the Changing American Family*, Santa Monica, Calif.: RAND Corporation, MR-488-LE, 1994.

Hamilton, Laura S., “Assessment as a Policy Tool,” *Review of Research in Education*, Vol. 27, 2003, pp. 25–68.

Hamilton, Laura S., Brian M. Stecher, and Stephen P. Klein, eds., *Making Sense of Test-Based Accountability in Education*, Santa Monica, Calif.: RAND Corporation, MR-1554-EDU, 2002. Online at http://www.rand.org/pubs/monograph_reports/MR1554/ (as of April 2006)

Hanushek, Eric A., “The Impact of Differential Expenditures on School Performance,” *Educational Researcher*, Vol. 18, No. 4, May 1989, pp. 45–51.

Hanushek, Eric A., *Making Schools Work: Improving Performance and Controlling Costs*, Washington, D.C.: The Brookings Institution, 1994.

Hanushek, Eric A., “School Resources and Student Performance,” in Gary Burtless, ed., *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, Washington, D.C.: The Brookings Institution, 1996.

Hanushek, Eric A., “Some Findings From an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Reductions,” *Educational Evaluation and Policy Analysis*, Vol. 21, No. 2, Special Issue, Summer 1999, pp. 143–164.

Hanushek, Eric A., and Dale W. Jorgenson, *Improving America’s Schools: The Role of Incentives*, Washington D.C.: National Academy Press, 1996.

Hanushek, Eric A., and Margaret E. Raymond, *Does School Accountability Lead to Improved Student Performance?* NBER Working Papers 10591, Cambridge, Mass.: National Bureau of Economic Research, June 2004. Online at http://www.nber.org/papers/w10591 (as of April 2006).

Hanushek, Eric A., Steven G. Rivkin, and Lori L. Taylor, “Aggregation and the Estimated Effects of School Resources,” *The Review of Economics and Statistics*, Vol. 78, No. 4, November 1996, pp. 611–627.

Hanushek, Eric A., “Black-White Achievement Differences and Governmental Interventions,” *American Economic Review*, Vol. 91, No. 2, May 2001, pp. 24–28.

Haveman, Robert, and Barbara Wolfe, “The Determinants of Children’s Attainments,” *Journal of Economic Literature*, Vol. 33, No.4, pp. 1829–1878.

Heckman, James J., and Dimitriy V. Masterov, “The Productivity Argument for Investing in Young Children,” Working Paper 5, Investing in Kids Working Group, Washington, D.C.: Committee for Economic Development, October 4, 2004. Online at http://www.ced.org/docs/report/report_ivk_heckman_2004.pdf (as of April 2006).

Hedges, Larry V., and Rob Greenwald, “Have Times Changed? The Relation Between School Resources and Student Performance,” in Gary Burtless, ed., *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, Washington, D.C.: The Brookings Institution, 1996.

Hedges, Larry V., Richard D. Laine, and Rob Greenwald, “Does Money Matter? Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes,” *Educational Researcher*, Vol. 23, No. 3, April 1994, pp. 5–14.

Hedges, Larry V., and Amy Nowell, “Black-White Test Score Convergence Since 1965,” in Christopher Jencks and Meredith Phillips, eds., *The Black-White Score Gap*, Washington D.C.: The Brookings Institution, 1998.

Heubert, Jay P., and Robert M. Hauser, eds., *High Stakes: Testing for Tracking, Promotion, and Graduation.* Washington, D.C.: National Academy Press, 1999.

Karoly, Lynn, Peter Greenwood, Susan Everingham, Jill Hoube, M. Rebecca Kilburn, C. Peter Rydell, Matthew Sanders, and James Chiesa, *Investing in Our Children: What We Know and Don't Know About the Costs and Benefits of Early Childhood Interventions*, Santa Monica, Calif.: RAND Corporation, MR-898-TCWF, 1998. Online at http://www.rand.org/pubs/monograph_reports/MR898/ (as of April 2006).

Karoly, Lynn A., M. Rebecca Kilburn, and Jill S. Cannon, *Early Childhood Interventions: Proven Results, Future Promise*, Santa Monica, Calif.: RAND Corporation, MG-341-PNC, 2005. Online at http://www.rand.org/pubs/monographs/MG341/ (as of April 2006).

Krueger, Alan B., “Reassessing the View that American Schools Are Broken,” *Economic Policy Review, Federal Reserve Bank of New York*, Vol. 4, No. 1, March 1998, pp. 29–43. Online at http://www.ny.frb.org/research/epr/98v04n1/9803krue.pdf (as of April 2006).

Krueger, Alan B., “Experimental Estimates of Education Production Functions,” *Quarterly Journal of Economics*, Vol. 114, No. 2, May 1999, pp. 497–532.

Krueger, Alan B., “An Economist’s View of Class Size Research,” The CEIC Review, Vol. 9, No.2, March 2000. Online at http://www.temple.edu/lss/pdf/ceicreviews/CEICVol9No2.pdf (as of April 2006).

Krueger, Alan B., “Economic Considerations and Class Size,” *Economic Journal*, Vol. 113, No. 485, February 1, 2003, pp. 34–63.

Krueger, Alan B., and Diane Whitmore, “The Effects of Attending a Small Class in the Early Grades on College Test Taking and Middle School Test Results: Evidence from Project STAR,” *Economic Journal*, Vol. 111, No. 468, January 2001, pp. 1–28

Krueger, Alan B., and Diane Whitmore, “Would Smaller Classes Help Close the Black-White Score Gap,” in John E. Chubb and Tom Loveless, eds., *Bridging the Achievement Gap*, Washington, D.C.: The Brookings Institution, 2002.

Lee, Valerie E., and David T. Burkam, *Inequality at the Starting Gate: Social Background Differences in Achievement as Children Begin School*, Washington, D.C.: Economic Policy Institute, September 2002.

Lynch, Robert G., *Exceptional Returns: Economic, Fiscal, and Social Benefits of Investment in Early Childhood Development*, Washington, D.C.: Economic Policy Institute, 2004. Online at http://www.epinet.org/books/exceptional/exceptional_returns_(full).pdf (as of April 2006).

Masse, Leonard, and W. Steven Barnett, *A Benefit Cost Analysis of the Abecedarian Early Childhood Intervention*, New Brunswick, N.J.: National Institute for Early Education Research, 2002. Online at http://nieer.org/docs/index.php?DocID=57 (as of April 2006).

McLaughlin, D. H., Protecting State NAEP Trends from Changes in SD/LEP Inclusion Rates, paper presented at the National Institute of Statistical Sciences workshop on NAEP inclusion strategies. Research Triangle Park, N.C., July 2000.

McLaughlin, D. H., *Exclusions and Accommodations Affect State NAEP Gain Statistics: Mathematics, 1996 to 2000*, report to the NAEP Validity Studies Panel, Palo Alto, Calif.: American Institutes for Research, 2001.

Molnar, Alex, Philip Smith, John Zahorik, Amanda Palmer, Anke Halbach, and Karen Ehrle, “Evaluating the SAGE Program: A Pilot Program in Targeted Pupil-Teacher Ratio in Wisconsin,” *Educational Evaluation and Policy Analysis*, Vol. 21, No. 2, Special Issue, Summer 1999, pp. 165–178.

National Commission on Excellence in Education, *A Nation at Risk: The Imperative for Educational Reform*, Washington, D.C.: U.S. Government Printing Office, April 1983. Online at http://www.ed.gov/pubs/NatAtRisk/index.html (as of April 2006).

Nye, Barbara, Larry V. Hedges, and Spyros Konstantopoulos, “The Long-Term Effects of Small Classes: A Five-Year Follow-Up of the Tennessee Class Size Experiment,” *Educational Evaluation and Policy Analysis*, Vol. 21, No. 2, Special Issue, Summer 1999, pp. 127–142.

O’Day, Jennifer A., and Marshall S. Smith, “Systemic Reform and Educational Opportunity,” in Susan H. Fuhrman, ed., *Designing Coherent Educational Policy*, San Francisco, Calif.: Jossey-Bass, 1993, pp. 250–312.

Phelps, Richard P., Thomas M. Smith, and Nabeel Alsalam, *Education in States and Nations*, 2nd ed., NCES 96-160, Washington, D.C.: Government Printing Office, 1996.

Reynolds, Arthur J., Judy A. Temple, Dylan L. Robertson, and Emily A. Mann, “Age 21 Cost-Benefit Analysis of the Title I Chicago Child-Parent Centers,” *Educational Evaluation and Policy Analysis*, Vol. 24, No. 4, Winter 2002, pp. 267–303.

Rothstein, Richard, *Class and Schools: Using Social, Economic, and Educational Reform to Close the Black-White Achievement Gap*. Washington, D.C.: Economic Policy Institute, 2004.

Rothstein, Richard, and Karen H. Miles, *Where's the Money Gone? Changes in the Level and Composition of Education Spending*, Washington, D.C.: Economic Policy Institute, 1995.

Schweinhart, L. J., J. Montie, Z. Xiang, W. S. Barnett, C. R. Belfield, and M. Nores, *Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40.* Monograph 14, Ypsilanti, Mich.: High/Scope Press, 2005.

Smith, Marshall S., and Jennifer O'Day, "Systemic School Reform," in Susan Fuhrman and Betty Malem, eds., *The Politics of Curriculum and Testing: The 1990 Yearbook of the Politics of Education Association*, London: Falmer, 1991.

Stecher, Brian M., "Consequences of Large-Scale, High-Stakes Testing on School and Classroom Practices," in Laura S. Hamilton, Brian M. Stecher, and Stephen P. Klein, eds., *Making Sense of Test-Based Accountability in Education*, Santa Monica, Calif.: RAND Corporation, MR-1554-EDU, 2002. Online at http://www.rand.org/pubs/monograph_reports/MR1554/ (as of April 2006).

Stecher, Brian, and George W. Bohrnstedt, *Class Size Reduction in California: Findings from 1999–00 and 2000–01*, Palo Alto, Calif.: CSR Research Consortium, American Institute for Research, 2002.

U.S. Supreme Court, *San Antonio School District v. Rodriguez*, No. 71-1332, 411 US 1 93 S. Ct. 1278, 1973.

Vinovskis, Maris, "An Analysis of the Concept and Uses of Systemic Educational Reform," *American Educational Research Journal*, Vol. 33, No. 1, Spring 1996, pp. 53–85.

West, Jerry, Kristin Denton, and Elvira Germino-Hausken, *America's Kindergartners*, NCES 2000-070, Washington, D.C.: U.S. Department of Education, Office of Educational Research and Improvement, February 2000. Online at http://nces.ed.gov/pubs2000/2000070.pdf (as of April 2006).

Zimmer, Ron, Richard Buddin, Derrick Chau, Glenn A. Daley, Brian Gill, Cassandra M. Guarino, Laura S. Hamilton, Cathy Krop, Daniel F. McCaffrey, Melinda Sandler, and Dominic J. Brewer, *Charter School Operations and Performance: Evidence from California*, Santa Monica, Calif.: RAND Corporation, MR-1700-EDU, 2003. Online at http://www.rand.org/pubs/monograph_reports/MR1700/ (as of April 2006).