



EDUCATION

- THE ARTS
- CHILD POLICY
- CIVIL JUSTICE
- EDUCATION
- ENERGY AND ENVIRONMENT
- HEALTH AND HEALTH CARE
- INTERNATIONAL AFFAIRS
- NATIONAL SECURITY
- POPULATION AND AGING
- PUBLIC SAFETY
- SCIENCE AND TECHNOLOGY
- SUBSTANCE ABUSE
- TERRORISM AND HOMELAND SECURITY
- TRANSPORTATION AND INFRASTRUCTURE
- WORKFORCE AND WORKPLACE

This PDF document was made available from www.rand.org as a public service of the RAND Corporation.

[Jump down to document](#) ▼

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world.

Support RAND

[Browse Books & Publications](#)

[Make a charitable contribution](#)

For More Information

Visit RAND at www.rand.org

Explore [RAND Education](#)

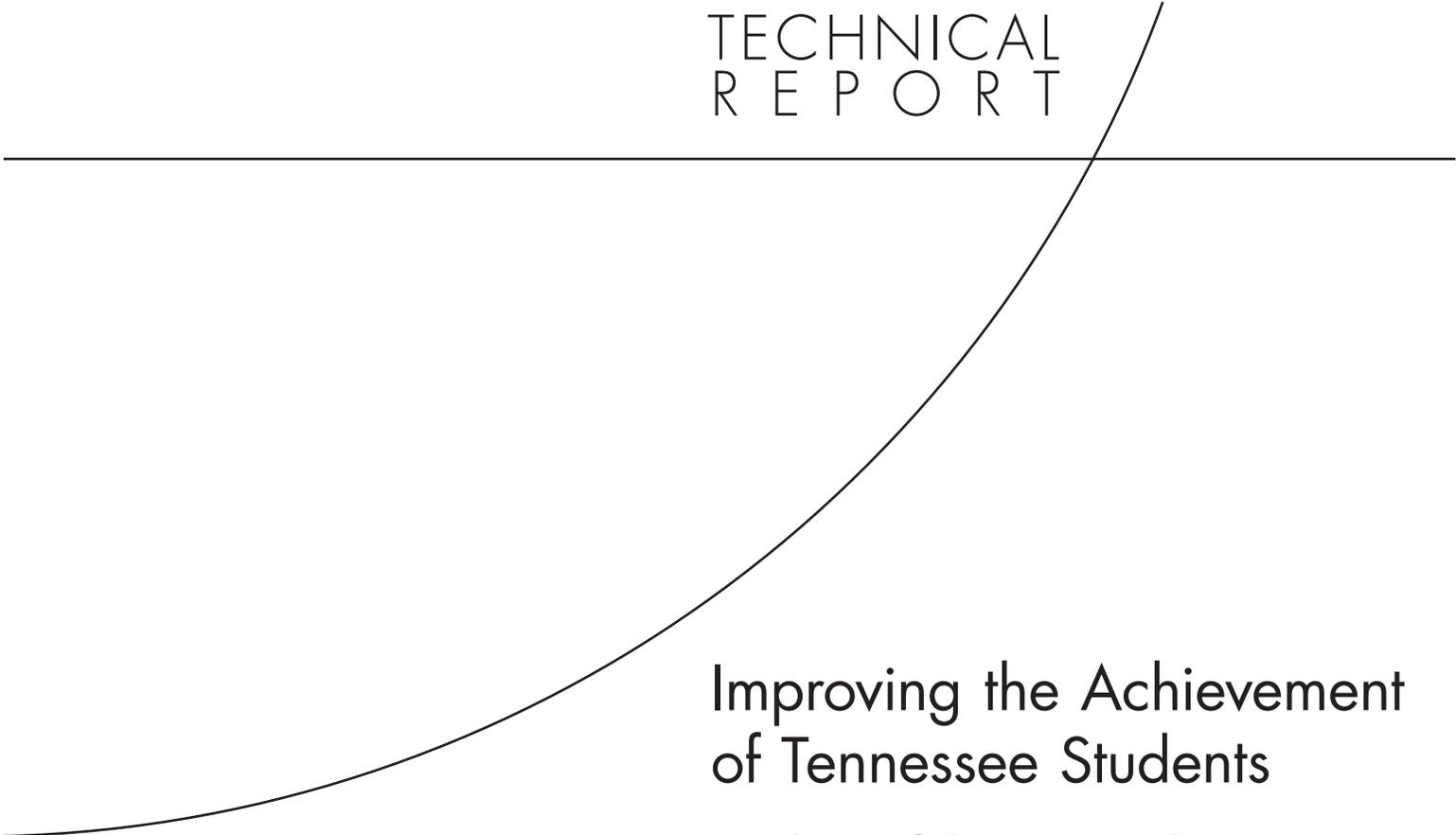
View [document details](#)

Limited Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law as indicated in a notice appearing later in this work. This electronic representation of RAND intellectual property is provided for non-commercial use only. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use.

This product is part of the RAND Corporation technical report series. Reports may include research findings on a specific topic that is limited in scope; present discussions of the methodology employed in research; provide literature reviews, survey instruments, modeling exercises, guidelines for practitioners and research professionals, and supporting documentation; or deliver preliminary findings. All RAND reports undergo rigorous peer review to ensure that they meet high standards for research quality and objectivity.

TECHNICAL
R E P O R T



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



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.

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

RAND® is a registered trademark.

© Copyright 2006 RAND Corporation

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from RAND.

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

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.