Estimating Effects of Non-Participation on State NAEP Scores Using Empirical Methods

David Grissmer
RAND

Commissioned by the NAEP Validity Studies (NVS) Panel
June 2007

George W. Bohrnstedt, Panel Chair
Frances B. Stancavage, Project Director

The NAEP Validity Studies (NVS) Panel was formed in 1995 to provide a technical review of NAEP plans and products and to identify technical concerns and promising techniques worthy of further study and research. The members of the panel have been charged with writing focused studies and issue papers on the most salient of the identified issues.

Panel Members:

Albert E. Beaton
Boston College

Peter Behuniak
University of Connecticut

George W. Bohrnstedt
American Institutes for Research

James R. Chromy
Research Triangle Institute

Phil Daro
University of California, Berkeley

Lizanne DeStefano
University of Illinois

Richard P. Durán
University of California, Santa Barbara

David Grissmer
RAND

Larry Hedges
University of Chicago

Gerunda Hughes
Howard University

Robert Linn
University of Colorado

Donald M. McLaughlin
American Institutes for Research

Ina V.S. Mullis
Boston College

Jeffrey Nellhaus
Massachusetts State Department of Education

P. David Pearson
University of California, Berkeley

Lorrie Shepard
University of Colorado

David Thissen
University of North Carolina-Chapel Hill

Project Director:

Frances B. Stancavage
American Institutes for Research

Project Officer:

Janis Brown
National Center for Education Statistics

For Information:

NAEP Validity Studies (NVS)
American Institutes for Research
1070 Arastradero Road, Suite 200
Palo Alto, CA 94304-1334
Phone: 650/ 843-8192
Fax: 650/ 858-0958

CONTENTS

LIST OF EXHIBITS

Summary

Description and Objectives

The NAEP Validity Studies (NVS) Panel provides advice and recommendations to help insure the "validity" of the National Assessment of Educational Progress (NAEP) test scores. The primary objectives of NAEP tests are to accurately monitor the progress of defined groups of students over time and to measure valid differences in scores between student groups at a single point in time. In this context, valid scores reflect differences in scores that are linked to "real" differences in student knowledge as measured on achievement tests.

The NVS Panel previously sponsored analysis directed toward estimating the potential bias from changing exclusion rates in NAEP tests. Students are excluded either because teachers and test administrators judge that language fluency is insufficient (limited English proficiency) or that a disability exists that would prevent a valid score. Bias can occur in state and national scores if teachers and test administrators do not apply exactly the same exclusion criteria across schools and states. Different criteria can arise because states have different criteria for classifying students and provide different test accommodations for students. Many states began providing accommodations for their state administered tests before NAEP provided such accommodations. Students accommodated on state tests often were excluded from NAEP tests if NAEP did not offer a similar accommodation. These differences in state and NAEP accommodations were one factor in driving NAEP exclusion rates higher in the late 1990s and early 2000. NAEP exclusion rates for many states changed significantly, and the NVS Panel undertook analysis to determine the extent to which changing exclusion rates caused bias in state NAEP scores.

The NVS Panel analysis utilized similar data collected on included and excluded students, including data for each excluded student specifying the reason for the exclusion. Such data allowed a statistical analysis that imputed scores for excluded students and generated associated estimates of the amount of bias in state scores from non-uniformly applied exclusion criteria. Adjustments to state scores were made and published for all state NAEP tests from 1990-2003.

This study takes up a second threat to the validity of scores arising from differential and changing participation rates of schools and students in NAEP testing. Non-participation can arise either from the absence or refusal to participate of a student chosen in the sample (student non-participation) or from a decision of a principal to refuse to allow the school to participate (school non-participation). School participation was voluntary until the 2003 test when participation of sampled schools became mandatory by federal statute. However, student participation continues to be voluntary.

NAEP administrators have established criteria for student and school participation for states that attempt to limit the potential for bias due to non-participation. Such criteria have changed somewhat from 1990-2003, but generally a state's student participation needed to be above either 80 or 85 percent and a school participation needed to be above 70 percent for their score to be reported. Almost all states maintained participation rates well above these minimum requirements. Student participation rates generally were in the range of 90-97 percent for state NAEP, while school participation was generally in the range of 85-100 percent. However, even at these rates, the NVS Panel thought a more thorough analysis of the potential for bias was needed.

In the case of non-participation, the causal mechanisms are not as well studied and known as those for exclusions. Decisions not to participate are made by individuals outside the test administration process, and there is no data collected that would allow exploration of why such decisions are made. Student non-participation is assumed to be caused by a legitimate absence on the day of testing (illness, etc.) unrelated to NAEP testing or by a decision by a parent or student not to participate, which is related to NAEP testing. School non-participation is normally due to a school principal's refusal to participate.[1] However, school principals can also be influenced by state and district policymakers who may try to change the principal's decision in order to achieve high participation rates state-wide. Some states have near perfect records of school participation across all NAEP testing suggesting the influence of state-wide policies. Other states have consistently low levels of participation, perhaps indicating principals' complete discretion in the decision whether to participate. Student non-participation due to legitimate absence is more similar to exclusion since an underlying mechanism (illness) is present throughout states that drive non-participation. However, non-participation by student or parent choice, or by some combination of principal or state policymaker choice, may not have an underlying common mechanism allowing predictions across states. This difference makes non-participation bias potentially more difficult to estimate.

NAEP does not routinely collect information about the reasons that students or schools do not participate. Some sampling characteristics of non-participating schools and students are collected that allow adjustments to be made for non-participation assuming that non-participating schools and students would score similarly to schools and students within the same sampling frame. However, the question is whether this adjustment accounts adequately for potential bias. The students and schools who do not participate may have special circumstances or characteristics that make them far different than participating students and schools in the same sampling frame. Without information about the reasons for non-participation, imputation would not likely improve on the normal adjustment made for non-participation.

An initial study sponsored by the panel made estimates of bias based on a set of worst case assumptions. These assumptions were that nonparticipating students and schools were drawn from the extremes of the test score distribution. Under these conditions, the empirical levels of student and school non-participation represented a significant threat to the validity of NAEP scores. However, unlike bias due to exclusion rates, the scores of non-participating students and schools across all states could not be accurately imputed by equations estimated from detailed information such as that collected during the exclusion process. The decision to exclude leaves a rich paper trail that can be used to estimate potential bias at the student level. No such paper trail exists for decisions related to student or school non-participation.

Although it is likely that these worst case assumptions are not accurate, a method was needed to help determine whether the available empirical data supported the worst case assumption estimates. The current study was sponsored in order to explore the extent to which common mechanisms might exist across states that explain patterns of non-participation, and if such mechanisms exist, to estimate their potential for bias. If common

[1] In some cases, substitution occurs at the school level based on matching some characteristics of the non-participating school with the substitute school. This process likely would introduce less bias than not substituting for such schools if fairly similar schools can be found. However, substitution occurs in a small proportion of such cases, leaving significant school non-participation.

mechanisms are operating within each state that partially determine the level of non-participation, then non-participation becomes a factor in explaining the pattern of NAEP scores across states. Empirically estimated models that explain the pattern of state scores across 17 tests from 1990-2003 might then provide an estimate of the effect on state scores of changing patterns of non-participation.

We use two methods in this report to make such estimates. The first method is to treat the 4th and 8th grade state reading test scores given in 2002 and 2003 as a "natural experiment" to test for the presence of bias due to school non-participation. Reading scores and factors that can change reading scores would be expected to change little within a single year, but school non-participation changed dramatically for many states due to the mandate in 2003 for all school to participate. If changing school participation rates cause substantial bias, it may be detected in differences in 2002-2003 reading test scores.

We also developed empirical models that attempt to explain the pattern of scores across 17 state NAEP tests from 1990-2003 (696 observations). We test whether the pattern of school non-participation introduces bias by including this variable in the analysis. The coefficient and statistical significance of the non-participation variable in such a model can be used to estimate the potential bias. Such a method relies on a properly specified model of what causes the pattern of state NAEP scores. Since there are always some differences among researchers about what constitutes a properly specified model, this analysis uses a wide variety of specifications to determine whether the non-participation rate is sensitive to different specifications. Such an analysis can often establish some reasonable empirical limits to the effects of non-participation.

Specifically, this study has the following objectives:

- To compile and examine student and school non-participation rates across state NAEP tests from 1990-2003 and assess whether common factors are present that might explain non-participation patterns across states and their potential for bias;
- To treat the 2002-2003 4th and 8th grade state scores as a natural experiment to estimate the extent of possible bias;
- To develop statistical models that account for the pattern of state NAEP scores for 696 state scores from 1990-2003, and to assess whether the pattern of non-participation is a significant explanatory factor in this pattern of state NAEP scores; and
- To compare estimates of bias from these methods to the bias from worst case scenarios estimated by McLaughlin, 2004.

Results

Student non-participation

These data suggest that normal absences are a strong contributor to NAEP student non-participation at both 4th and 8th grade and that over one-third of the variance in state student non-participation can be accounted for by normal student absences. As the 4th grade the data show, student non-participation at the national level (5.1%) is only about 1 percentage point above the estimated level of normal absences (4.0%). At 8th grade, there is a larger gap between estimated absence (4.5%) and non-participation (8.0%). The analysis would also suggest that 4th grade non-participation is not correlated with NAEP scores, but that 8th grade non-participation is weakly correlated with NAEP scores, with higher non-participation in lower scoring states. These data suggest that a significant part of the student non-participation at 8th grade may be due to specific decisions by parents or students not to participate in NAEP rather than routine absences. Exploring the causes and possible bias of the higher student non-participation at 8th grade seems warranted, especially given that 12th grade state tests, which may occur in the future, may show similar causes and could possibly show even higher student non-participation.

The threat of bias due to 4th grade student non-participation appears minimal. Since normal absences are not draw from the extremes of the test distribution, the worst case assumptions are not viable. The evidence from simple analysis of the patterns of scores and non-participation across states would suggest no linkage between 4th grade NAEP scores and non-participation. At 8th grade, more than one-half of the non-participation seems attributable to normal absence, and therefore poses little threat to validity. However, somewhat less than one-half of the non-participation seems attributable to specific decisions not to participate either by students or parents. Thus non-participation at 8th grade poses a somewhat greater threat to validity. However, the weak correlation between non-participation and NAEP scores across states would suggest that the threat to validity is not consistent with worst-case assumptions, especially given the relatively narrow range of variation in student non-participation across states.

School non-participation

The precise factors involved in school non-participation decisions remain largely unknown. However, the pattern of participation across states shows modest consistency across the 17 tests. Some states have near perfect participation across all tests, and such participation seems unlikely unless state policymakers have adopted explicit or implicit policies that mandate school participation or the principals have characteristics that would predict high compliance. States with a pattern of very high participation are more often lower scoring states. Other states have fairly consistent lower levels of participation. This pattern might be consistent with the absence of state influence, thereby allowing principals more discretion in NAEP decisions, and/or having principals with characteristics that would more often lead to noncompliance. States with patterns of lower participation are more often higher scoring states. This analysis would suggest that one avenue for further research would be to explore a possible relationship between characteristics of principals and schools and participation decisions, using school level data. Such data could be constructed by merging state personnel data sets with NAEP data.

The question is the extent to which the differences in school participation among states results in a systematic bias in NAEP scores related to these differences. The two methods utilized to test this hypothesis lead to similar conclusions. The evidence evaluated here would suggest that patterns of school non-participation may cause relatively small bias in scores—much smaller than worst case estimates—representing only a marginal threat to validity. However, there appears to be no reliable method to make adjustments to scores in the 1990-2002 period. After 2002, school participation is mandatory and thereby no threat to future validity is present.

1. Introduction

The NAEP Validity Studies (NVS) Panel provides advice and recommendations to help insure the "validity" of the National Assessment of Educational Progress (NAEP) test scores. The primary objectives of NAEP tests are to accurately monitor the progress of defined groups of students over time and to measure valid differences in scores between student groups at a single point in time. In this context, valid scores reflect differences in scores that are linked to "real" differences in student knowledge as measured on achievement tests.

Several threats exist that can change scores so that differences are not linked to actual student knowledge. Scores can vary due to a variety of more or less random factors linked to sampling, administration, student motivation, question selection and distribution among test booklets, success at guessing, and other factors. These factors are not particularly problematic if they are caused by truly random events, particularly if such random variation is captured and estimated by standard errors. The more problematic changes in scores are those caused by factors that can systematically bias the scores among student groups taking the test at a single point in time, or to bias the scores of student groups taking tests at two points in time. Two of the main threats that might cause such bias are *differential student exclusion* from NAEP testing and *differential participation of selected schools and students* in NAEP testing.

Student exclusion criteria have been established by NAEP for two categories of students: limited English proficient (LEP) and individualized education plan and/or learning disabled students (IEP/SD). These latter students include those who have individualized education programs or who are receiving special services as a result of section 504 of the Rehabilitation Act. The criteria used in NAEP's assessments state that IEP/SD students can be excluded only if they are mainstreamed in academic subjects less than 50 percent of the time and/or are judged to be incapable of participating meaningfully in the assessment. Furthermore, LEP students can be excluded if they are native speakers of a language other than English, enrolled in an English speaking school for less than 2 years, and judged to be incapable of taking part in the assessment. These criteria for exclusion applied across states have resulted in approximately 3-4 percent of public school students excluded for LEP and 5-7 percent for IEP/SD in NAEP tests from 1990-2003.

The potential bias from excluded students depends primarily on three factors: the uniformity of the application of exclusion criteria across states and over time; the amount of variance in the exclusion rates across states and over time, and the difference in scores between included and excluded students. Exclusion rates differ considerably by state and have changed markedly over time. LEP rates are 1 percent or less in many states, but can reach 15 percent in a few states. IEP rates tend to vary between 3-8 percent across states in any single test. Overall exclusion rates have risen from approximately 5 percent in the early 1990s to as high as 8 percent in the 2000-2002 period. The validity threat to scores from excluded students is heightened because these students would have usually been among the lower scoring students had they been tested.

The estimated bias arising in state NAEP scores from differential and changing exclusion rates across states was the subject of earlier work by the NVS panel (McLaughlin, 2000; McLaughlin, 2001). The conclusion of the analysis was that differential and changing exclusion rates could bias scores sufficiently to represent a significant threat to the validity of the scores. Accordingly, estimated adjustments to the scores have been made and published in later NAEP documents. McLaughlin, 2000 and McLaughlin, 2001 utilized an imputation methodology to make these adjustments from information provided on each student flagged for possible exclusion in the sample. Such information is provided by the teacher and includes an array of variables about the student, including why the student was excluded. Such data are used to impute scores to students excluded from tests, thus allowing a "full population" estimate of scores. These full population estimates can, theoretically, provide estimates that are not sensitive to exclusions.

The magnitude of estimated adjustments to each state NAEP score was less than a single NAEP point (about .03 standard deviation) for most tests, but could be as large as 3-4 NAEP points. Such adjustments can shift scores sufficiently to change a state's rankings in NAEP considerably and/or to make state gains from consecutive tests shift from being statistically significant to non-significant or vice-versa. Thus making adjustments becomes an important part of making valid interpretations from NAEP scores.

This study takes up a second threat to the validity of scores arising from differential and changing participation rates of schools and students in NAEP testing. Non-participation can arise either from the absence of a student chosen in the sample (student non-participation) or from a decision of a principal to refuse to allow the school to participate (school non-participation). School participation was voluntary until the 2003 test when participation of sampled schools became mandatory by federal statute. However, student participation continues to be voluntary.

Non-participation can introduce bias into test results if those schools and students not participating would potentially have different average scores compared to those participating. The threat of bias from non-participation is somewhat different from exclusion rates. Participation rates generally have a greater variance across states than do exclusion rates. School non-participation rates can vary from 0 to over 20 percent across states, and like exclusion rates, have increased from 1990-2002. However, there is likely less potential bias threat from each *non-participating* student than each *excluded* student since non-participating students likely have a wider distribution of scores whose average is closer to scores of participating students than is the case for excluded students. A final factor posing a validity threat from non-participation is the abrupt transition in 2003 to virtually 100 percent participation due to legal mandates in the No Child Left Behind (NCLB) legislation. This transition suddenly changed participation rates that had more smaller and more continuous variation from 1990-2002 to a one-time abrupt and large variation for many states. A few states went from about 70 percent participation to 100 percent in a single year. Such changes could make bias in the 2003 tests sufficiently large to require adjustments.

There are key differences between exclusion and non-participation that needs to be considered in estimating bias. One difference is that the decision to exclude is theoretically guided by a well defined process that is more or less uniformly applied across the states to determine whether a student should be tested, and the decision is made by test administrators and teachers applying these guidelines. The important question in determining exclusion bias is whether criteria developed to assess the extent of disability or language facility is uniformly applied across states and student groups. Non uniformity can occur if state specific factors are used in evaluation and exclusion.[2] A second difference is that the decision whether to include or exclude identified students with disabilities or lack of facility with language is well documented leaving a rich paper trail that can be used to estimate potential bias at the student level. No such paper trail exists for decisions related to non-participation.

In the case of non-participation, the causal mechanisms are not as well studied. Decisions not to participate are made by individuals outside the test administration process, and there are no data collected that would allow exploration of why such decisions are made. Student non-participation is assumed to be caused by a legitimate absence on the day of testing (illness, etc.) unrelated to NAEP testing or by a decision by a parent or student not to participate, which is related to NAEP testing. School non-participation is normally caused by a school principal's refusal to participate.[3] However, school principals can also be influenced by state and district policymakers who may try to change the principal's decision to achieve high participation rates state-wide. Some states have near perfect records of school participation across all NAEP testing suggesting the influence of state-wide policies. Student non-participation due to legitimate absence is more similar to exclusion since an underlying mechanism (i.e., illness) is present throughout states that drive non-participation. However, non-participation by student or parent choice or by some combination of principal or state policymaker choice may not have an underlying common mechanism or uniform process nation-wide. This difference makes non-participation bias potentially more difficult to estimate.

NAEP data does not contain information about students who are chosen in the sample, but who do not participate either because their school principal chooses not to participate (school non-participation) or the student or parent in a participating school is absent from school due to illness or opts out of NAEP testing (student non-participation). There is also no information from principals, students, or parents regarding their motivation for not participating. Thus straightforward imputation of scores to non-participants as a way of determining bias cannot be done as is the case for student exclusions.

Two other methodologies can be used to assess possible bias from non-participation. The first method assigns scores to non-participating students based on assumptions about their score distribution. McLaughlin, 2004 has made such estimates of bias under a series of "worst case" scenarios. It was assumed that non-participants would have been among the lowest scoring students taking each test or were enrolled in a school with traditionally low performance on NAEP. With these assumptions, estimates of bias were made under a variety of scenarios. Since state non-participation ranges from 0 to almost 20 percent, these estimates show that non-participation could possibly be a significant problem with equal or greater threat to the validity of scores as exclusions, provided that the worst-case scenario assumptions are reasonably accurate. Although it is likely that these worst case assumptions are not accurate, a method was needed to help determine a more reasonable set of assumptions.

---

2 In practice, state criteria for defining IEP and LEP could vary, creating non-uniformity in the application of exclusion across states. Also, in later years, exclusion was predicated on whether a student received accommodations on state administered tests, and wide differences in accommodations across states made exclusions dependent on idiosyncratic state specific factors.

3 In some cases, substitution occurs at the school level based on matching some characteristics of the non-participating school with the substitute school. This process likely would introduce less bias than not substituting for such schools if fairly similar schools can be found. However, substitution occurs in a small proportion of such cases, leaving significant school non-participation.

In this report, we first use the available non-participation data by state for 17 tests from 2000-2003 to determine possible common mechanisms across states that can explain the pattern of non-participation and the extent to which these factors might cause bias. If non-participating students or schools are systematically driven by common processes throughout the nation resulting in bias, then a pattern of bias will be present in state scores that will depend on the level by state of student and state non-participation. This pattern can likely be detected empirically using two methods. The first method is to treat the 4th and 8th grade state reading test scores given in 2002 and 2003 as a "natural experiment" to test for the presence of bias due to school non-participation. Reading scores and factors that can change reading scores would be expected to change little within a single year, but school non-participation changed dramatically for many states due to the mandate in 2003 for all schools to participate. If changing school participation rates cause substantial bias, it may be detected in differences between 2002 and 2003 reading test scores.

We also developed empirical models that attempt to explain the pattern of scores across 17 state NAEP tests from 1990-2003 (696 observations). We test whether the pattern of school non-participation introduces bias by including this variable in the analysis. The coefficient and statistical significance of the non-participation variable in such a model can be used to estimate the potential bias. Such a method relies on a properly specified model of what causes the pattern of state NAEP scores. Since there are always some differences among researchers about what constitutes a properly specified model, this analysis uses a wide variety of specifications to determine whether the non-participation rate is sensitive to different specifications. Such an analysis can often establish some reasonable empirical limits to the effects of non-participation.

Specifically, this study has the following objectives:

- To compile and examine student and school non-participation rates across state NAEP tests from 1990-2003, and to assess whether common factors are present that might explain non-participation patterns across states and their potential for bias;
- To treat the 2002-2003 4th and 8th grade state scores as a natural experiment to estimate the extent of possible bias;
- To develop statistical models that account for the pattern of state NAEP scores for 696 state scores from 1990-2003, and to assess whether the pattern of non-participation is a significant explanatory factor in this pattern of state NAEP scores; and
- To compare estimates of bias from these methods to the bias from worst case scenarios estimated by McLaughlin, 2004.

Chapter 2 of this report provides the analysis of the basic pattern of student and school non-participation for states across 17 state NAEP tests from 1990-2003, as well as an analysis to explore possible common mechanisms that would cause non-participation. Chapter 3 summarizes the results of the earlier study of potential bias from non-participation based on worst-case scenarios. Chapter 4 presents the methodology for the estimates of potential bias from analysis of the 2002-2003 "natural experiment" and the analysis of the 696 observations from 1990-2003. Chapter 5 presents the results comparing empirically estimated bias estimates to the worst case simulation estimates. Chapter 6 presents conclusions.

2. Historical NAEP Non-participation Rates and Possible Common Causes

This chapter describes the rates of student and school non-participation by states on 17 NAEP tests given from 1990-2003 (see Exhibit 1) and undertakes a preliminary analysis directed at identifying the common mechanisms leading to non-participation.

Exhibit 1. NAEP tests by grade, subject, and year from 1990-2003

Year	Grade	Subject	Number of States
1990	8th	Mathematics	38
1992	8th	Mathematics	42
1996	8th	Mathematics	41
2000	8th	Mathematics	39
2003	8th	Mathematics	50
1992	4th	Mathematics	42
1996	4th	Mathematics	44
2000	4th	Mathematics	40
2003	4th	Mathematics	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

Non-participation Rates

School non-participation occurs when principals of schools selected in the sample refuse to participate in NAEP testing. Student non-participation refers to students who fail to take the test due to absence on the day of the test and have no later make-up test, or students who are present on the day of the test, but who refuse to participate. NAEP administrators carry out several activities that attempt to achieve high levels of school and student participation and/or reduce potential bias from non-participation. These activities include:

- Taking non-participation into account when weighting the data;
- Setting criteria for levels of required school and student participation in order for a state's scores to be included in reports;
- Holding make-up sessions if there are 5 or more student absences;

- Doing periodic studies to assess potential problems[4]; and
- Making school participation mandatory in 2003.

The major analytical activity used to reduce any bias from non-participation is to use non-participation data when estimating individual level weights. The assumption is made that non-participants would have scored at the average level for participants in their sampling cell. Also, NAEP administrators establish guidelines for participation by states in an attempt to encourage high participation and minimize potential bias. Exhibit 2 shows NAEP guidelines established in 1990 to attempt to limit the threats to validity from non-participation. States receive a notation in published NAEP reports if any of the guidelines presented in Exhibit 2 are not met. If a jurisdiction fails to meet all of the requirements, their results are not reported. The results for several states have been excluded over time due to low participation, and the results for several states in each test contain a notation for low participation. The NAEP guidelines presented in Exhibit 2 governed the 1992 state mathematics assessment. There have been a few modification and consolidations to the guidelines. In 1996, the weighted student response rate within participating schools could not be below 80 percent compared to below 85 percent in pervious assessments. In this study we exclude states that have non-participation rates that prevented the publication of their scores, but include states that receive notations to their scores.

Exhibit 2. NAEP guidelines governing school and student participation Rrates: 1992 state mathematics assessment

1. Both the state's weighted participation rate for the initial sample of schools below 85 percent <u>and</u> a weighted school participation rate after substitution below 90 percent; <u>or</u> a weighted school participation rate of the initial sample of schools below 70 percent.
2. The non-participating schools includes a class of schools with similar characteristics, which together account for more than five percent of the state's total fourth- or eighth-grade weighted sample of public schools.
3. A weighted student response rate within participating schools below 85 percent.
4. The non-responding students within participating schools include a class of students with similar characteristics, who together comprised more than five percent of the state's weighted assessable student sample (Mullis, 1993).

Exhibit 3 shows average school and student participation rates for the 17 state NAEP tests given from 1990 to 2003. Appendix A provides the data for school and student participation for all states and tests. For the five 8th grade mathematics tests, school participation declined from about 97 percent in 1990, to 89 percent in 2000. School participation rates jumped to over 99 percent in the 2003 tests with the legislative mandate for all state schools (but not students) to participate. Only three 8th grade reading tests have been given, starting in 1998. School participation was 92 percent for the 1998 and 2002 tests, and then jumped to over 99 percent for the 2003 test. The four 4th grade mathematics tests show a similar pattern to the 8th grade mathematics tests with declining rates from 95 percent in 1992 to 89 percent in 2000, and then jumping to over 99 percent for the 2003 test. The five 4th grade reading tests show a slightly declining trend from 1992 to 2002 and

[4] See for instance Spencer, Bruce, 1994 and McLaughlin et al, 2004

the expected jump to over 99 percent in the 2003 test. The standard deviation across states of school participation suggests fairly large variation in rates across states from 1990-2002, and such variation has increased over time for all tests until 2003. Such data suggests much greater school compliance in the early years than the later years, with the 2000 4th and 8th grade mathematics tests having the largest non-participation and variance.

Exhibit 3. Average school and student participation rates by state NAEP test

Year of Test	Subject	Grade	Average School Participation Across States	Standard Deviation	Average Student Participation Across States	Standard Deviation
1990	Mathematics	8th	96.6	5.0	94.3	1.1
1992	Mathematics	8th	95.4	5.4	93.7	1.3
1996	Mathematics	8th	92.2	8.0	91.1	0.9
2000	Mathematics	8th	89.0	10.6	91.7	1.4
2003	Mathematics	8th	99.3	1.5	91.9	1.9
1998	Reading	8th	92.5	8.7	91.1	1.3
2002	Reading	8th	92.4	9.6	91.6	1.7
2003	Reading	8th	99.3	1.5	91.8	1.8
1992	Mathematics	4th	94.9	6.5	95.6	.7
1996	Mathematics	4th	93.4	6.9	95.0	1.1
2000	Mathematics	4th	89.2	11.4	95.0	0.8
2003	Mathematics	4th	99.4	0.9	94.5	1.1
1992	Reading	4th	94.9	6.2	95.3	1.8
1994	Reading	4th	94.2	6.7	95.4	0.9
1998	Reading	4th	93.4	8.2	94.5	0.9
2002	Reading	4th	92.6	9.3	94.5	1.2
2003	Reading	4th	99.4	0.9	94.5	1.2

Student participation rates have either been stable or slightly declining for all tests. Average rates have varied only from about 96 to 92 percent. The standard deviation of student rates across states is also markedly reduced and much more stable than for the school participation rates from 1990-2002. Student participation is generally lower for 8th grade students than 4th grade students in both reading and mathematics by 2-4 percentage points. In contrast, school participation does not have a distinctive pattern that differs by grade level.

Exhibit 4 shows student participation rates by state, averaged across 17 NAEP tests from 1990-2003. Average student rates vary across states from 96 to 91 percent. With the exception of a few states, the standard deviations are approximately 1-2 percentage points, indicating a fair amount of stability in student participation rates across tests.

Exhibit 4. Average student participation rates and standard deviations by state across 17 NAEP tests from 1990-2003

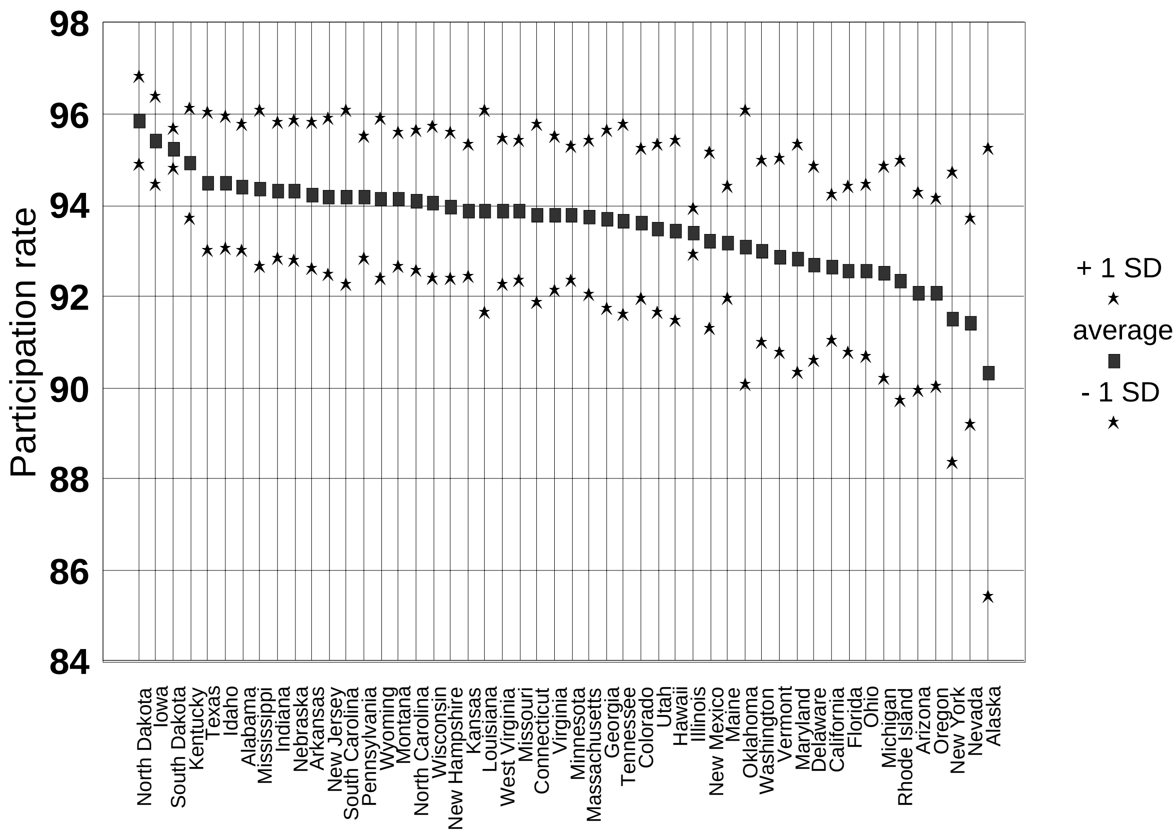
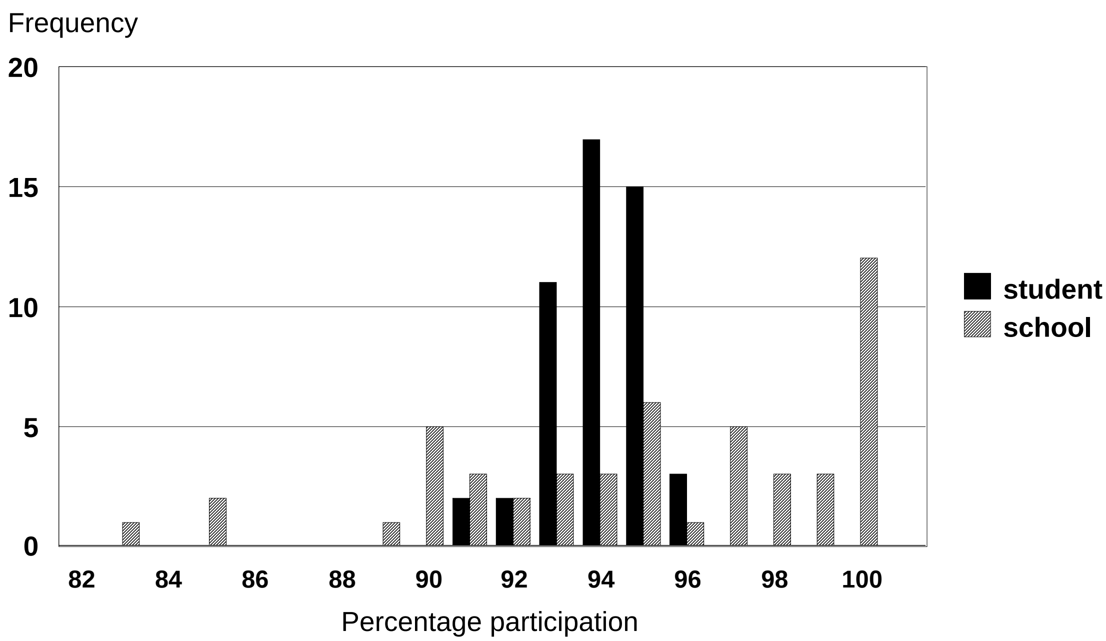


Exhibit 4 shows similar data for school participation rates. Average school participation rates after substitution for states range from 100 to 83 percent. The standard deviation across states varies markedly from less than 1 percentage point to over 10 percentage points. In Exhibit 4, the means + one standard deviation sometimes indicate values over 100 percent. Since the maximum participation rate is 100 percent, such values indicate only that the distribution is not normal. School participation rates by state have considerably more variance than do student rates, and the average rates across states vary more than do student rates. Thus, school non-participation may pose a greater threat to validity than does student non-participation.

The differences in the pattern of student and school participation are also illustrated in Exhibit 5, which shows the frequency distribution of states for the average student and school state participation across all 17 NAEP tests. The student participation shows a peaked distribution with narrow standard deviation. There are no extreme state outliers for student participation. School participation shows a much wider distribution with several extreme outlier states. Twelve states have had average school participation above 99 percent, but a few states have had average school participation below 85 percent. These data would suggest that a common underlying random process such as normal student absence across states may account for a significant part of the variance in student participation, but the school participation data suggest the possibility of much more idiosyncratic factors. For instance, the states with consistently high levels of school participation may be driven by implicit state-wide policies for participation in NAEP, while other states may leave the decision to local principals. We explore these hypotheses further in the next chapter.

Exhibit 5. Frequency distribution of average school and student participation across 17 NAEP tests from 1990-2003.



Explaining Participation Rates

The NAEP tests during the 1990-2002 period were low-stakes tests and, therefore, not subject to many of the concerns of potential bias 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 also is difficult, if not impossible, to "teach to the NAEP test" since only a few students from schools throughout the state take the tests. Since 2003, there have been very modest state incentives linked to NAEP scores. However, it seems unlikely that such small incentives would trigger any systematic actions that could bias scores, and such actions would be very hard to coordinate across schools in a state. Therefore, it seems unlikely that either student or school non-participation would be manipulated to cause bias. Rather bias might be an unintended consequence of other actions that could indirectly cause bias.

In this chapter, we explore whether the variance in student participation rates across states is linked to the normal absence rates of students as reported by school administrators, and whether the level of normal absences are correlated with NAEP scores or SES.[5] We also explore whether the school participation rates are correlated with reported NAEP scores and the SES of states.

Exhibits 6 and 7 present correlation matrices among the variables of interest for 4th and 8th grade students respectively. The coefficients suggest that higher levels of 4th grade student non-participation are moderately to strongly positively correlated with the estimate of daily absences and not significantly correlated with the 2003 mathematics scores or SES. Eighth-grade student non-participation is also moderately to strongly correlated with school estimated absences and weakly negatively correlated with SES and NAEP scores. Student and school non-participation are essentially uncorrelated.

Exhibit 6. Correlation matrix for average 4th-grade student and school nonparticipation across states with average NAEP scores, state SES and teacher reported absences

Fourth Grade					
	Average state student non-participation	Average state school non-participation	State SES[1]	Teacher reported absence	2003 NAEP math score
Student non-participation		0.14	0.04	0.57	-0.09
School non-participation			0.42	-0.30	0.32
SES				-0.22	0.66
Teacher reported absence					-0.39

[1] SES measure from Grissmer et al, 2000 (see Appendix E)

[2] Teacher estimated absences from the 2003 teacher surveys on the 2003 NAEP tests

School non-participation shows a somewhat different pattern. Both 4th and 8th grade school non-participation is positively correlated with the 2003 mathematics score and SES and somewhat surprisingly negatively correlated with teacher reported absences. The data also shows the expected strong positive correlation between SES and NAEP scores, and moderate to weak negative correlations between school estimated normal absence and NAEP scores and SES.

[5] School administrators in participating NAEP schools are surveyed and asked to estimate the level of normal absences in four categories (0-2%, 3-5%, 6-10% and >10%). These estimates are then weighted to represent state-wide percentages of absent students. We use the midpoint of each category (and 10 for the final category) to estimate percentage absences by state. Estimates are made separately for fourth and eighth grade schools.

Exhibit 7. Correlation matrix for 8th-grade student and school nonparticipation across states with average NAEP scores, state SES and teacher reported absences

Eighth Grade					
	Average state student non-participation	Average state school non-participation	State SES[1]	Teacher reported absence[2]	2003 NAEP math score
Student non-participation		0.05	-0.18	0.61	-0.27
School non-participation			0.20	-0.32	0.27
SES				-0.18	0.80
Teacher reported absence					-0.36

[1] SES measure from Grissmer et al, 2000 (see Appendix E)
[2] Teacher estimated absences from the 2003 teacher surveys on the 2003 NAEP tests

This data suggest that normal absences are a strong contributor to NAEP student non-response at both 4th and 8th grade. The data suggests that over one-third of the variance in student non-participation can be accounted for by normal student absences. Exhibit 8 compares the estimated level of national normal absences with NAEP student non-participation. At the 4th grade the data show that nationally student non-participation (5.1%) is only about one percentage point above the estimated level of normal absences (4.0%). At 8th grade there is a larger gap between estimated absence (4.5%) and non-participation (8.0%). This data suggests that a significant part of the student non-participation at 8th grade may be due to specific decisions by parents or students not to participate in NAEP.

Exhibit 8. Comparison of estimated normal rates of absence and student nonparticipation in NAEP

	Estimated normal absences	NAEP student nonparticipation
4th grade	4.0%	5.1%
8th grade	4.5%	8.0%

We further search for possible mechanisms that might explain student and school non-participation through regressions as shown in Exhibit 9. The results for both 4th and 8th grade school non-participation show significant relationships (10 percent level or better) between non-participation and NAEP scores. States with higher NAEP scores have higher non-participation. Since school non-participation involves decisions by principals/state policymakers, the data suggests less cooperation with NAEP in higher scoring states. The regression results with student non-participation show a significant relationship (10 percent or better) between student non-participation and NAEP scores at 8th grade, but not at 4th grade.

Exhibit 9. Regression results for 4th and 8th grade school and student nonparticipation

	Constant	2003 Mathematics Score	R-squared	N
8th-grade school non-participation	4.9	.2(1.8)	.07	43
4th-grade school non-participation	5.3	.2 (2.2)	.11	43
8th-grade student non-participation	7.9	-.05(-1.8)	.07	43
4th-grade student non-participation	5.1	-.01(-.6)	.01	43

Analysis at the state level can only suggest mechanisms that might explain non-participation. The mechanism suggested for school non-participation is that principals in higher scoring schools may more often refuse participation in NAEP, or that state policymakers in lower scoring states mandate participation more often. The mechanism suggested for student non-participation is student absences. Student absences at 4th grade look to be uncorrelated with NAEP scores, but 8th grade non-participation is higher in lower scoring states. If this pattern also occurs within states, then lower scoring students would be more likely to be absent or refuse to participate at 8th grade, with the potential to bias NAEP scores. Perhaps more importantly, the data suggest that if these mechanisms occur within states, the potential bias from student and school non-participation may be partially offsetting particularly at 8th grade. Higher scoring states appear to have higher school non-participation, but lower student non-participation.

These regression results suggest a potential for bias since a relationship exists between non-participation and NAEP scores. However, a possible confounding influence is state SES. SES is highly correlated with state NAEP scores, but is often correlated with non-participation. Exhibit 10 shows regression results linking 4th and 8th grade student non-participation with SES and NAEP scores for school non-participation and teacher absences for student non-participation.[6] Both 4th and 8th grade results for student non-participation show an expected highly significant predictive effect from estimated absences, but no additional statistically significant effect from SES or NAEP scores. If the additional students above normal absences at 8th grade were drawn highly disproportionately from high or low scoring students, we would expect stronger coefficients for the SES and/or NAEP score variables. The lack of such significance probably suggests that such students are not drawn from the extremes of the distribution.

[6] The OLS regression has student non-participation by state as the dependent variable with SES, estimated absences, and the 2003 NAEP score as independent variables.

Exhibit 10. Regression Results for 4th and 8th grade school and student non-participation

	Constant	SES	Teacher estimated absence rate	2003 mathematics score	R-squared	N
8th grade school non-participation	4.9	-1.6 (-.1)		.2 (1.1)	.07	43
4th-grade school non-participation	5.3	19.3 (1.9)		.06 (.5)	.18	43
8th-grade student non-participation	7.9	1.6(.4)		-.06(-1.4)	.08	43
4th-grade student non-participation	5.1	1.4(.9)		-.02(-1.0)	.03	43
8th-grade student non-participation	1.4	-1.1 (.3)	1.4 (4.4)	.00 (.00)	.38	43
4th-grade student non-participation	.93	1.1 (.8)	1.0 (4.5)	.01 (.3)	.36	43

For school non-participation, both SES and NAEP schools do not reach statistical significance at 8th grade; and SES, but not NAEP scores, is significant at 4th grade. Overall these regressions show little evidence that non-participation is strongly systematically linked to NAEP scores in a way that would be present if worst case assumptions—presented in the next chapter—were accurate.

3. Results from Earlier Estimates of Bias from Non-participation

McLaughlin et al, 2004 made estimates of the potential bias from non-participation by making assumptions about the scores of nonparticipating students and schools, and simulating the assumptions made by the Educational Testing Service (ETS) to correct for non-participation. Nonparticipating students were assigned either the lowest reported or predicted scores for students in the same school. Nonparticipating schools were assigned the lowest reported or predicted school scores in the same state. These assumptions were meant to determine how much bias might be present under "worst case" scenarios. The correction made for non-participation was simulated by imputing the scores of non-participating schools and students using the demographic characteristics of participating students and schools.[7] Estimates were made for between 1 to 5 students non-participating in each school, and from 5 to 25 percent school non-participation.

We have used these estimates of bias to estimate bias given the levels of school and student non-participation in each test.[8] For school non-participation, we estimated the bias between the states with the highest and lowest participation rates for each test. Using the highest and lowest rates provides a second set of "worst case" bias estimates for differences in scores between states. Exhibit 11 shows these estimates. For tests before 2003, the estimates show that the score differences between two states with the maximum difference in non-participation would be approximately 3-4 NAEP points at 8th grade and 2-3 NAEP points at 4th grade. For 2003 tests in which school participation was "mandatory", bias is reduced to less than 1 NAEP point. (Bias in present even in 2003 since not all states achieved 100 percent school participation.)

---

7 McLaughlin et al, 2004 also provides bias estimates using predictions from state administered tests, however, ETS does not have these scores available, and these estimates would not be accurate simulations of the corrections made by ETS.

8 We have used the bias estimates from McLaughlin et al, 2004 that correspond to using reported scores as the basis for non-participation and the linear equating method for corrections.

Exhibit 11. "Worst case" school nonparticipation bias estimates (using the highest and lowest state participation rates) in each NAEP test

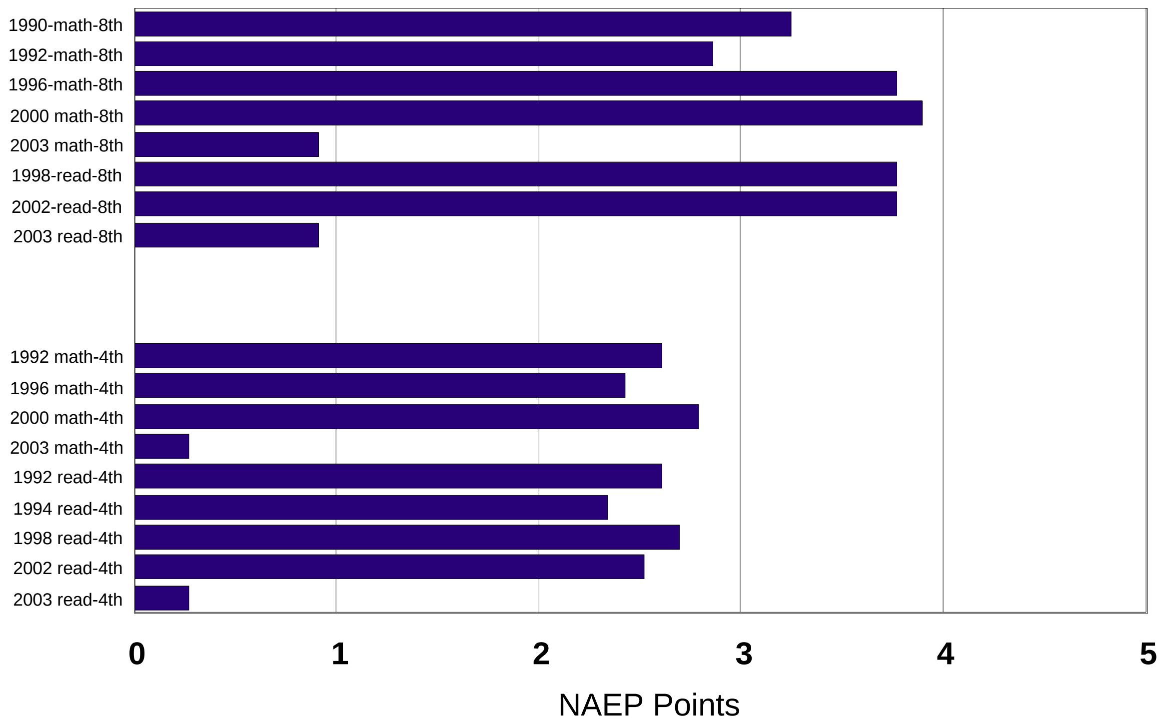


Exhibit 12 shows similar estimates for student non-participation. Most bias estimates are between 1-2 NAEP points. The larger estimates are due to very low participation rates of one or a few states. The student non-participation bias estimates are generally lower at 4th than 8th grade. Exhibit 13 shows the highest estimated bias between any two national scores given in each category between 1990-2003. The maximum bias for school non-participation is usually between the 2003 and the 2002 or 2000 scores due to the large increase in participation in 2003. Again the 4th grade bias estimates are always much lower than the 8th grade estimates.

Exhibit 12. Worst case student nonparticipation bias estimates (using the highest and lowest state student participation rates for each test)

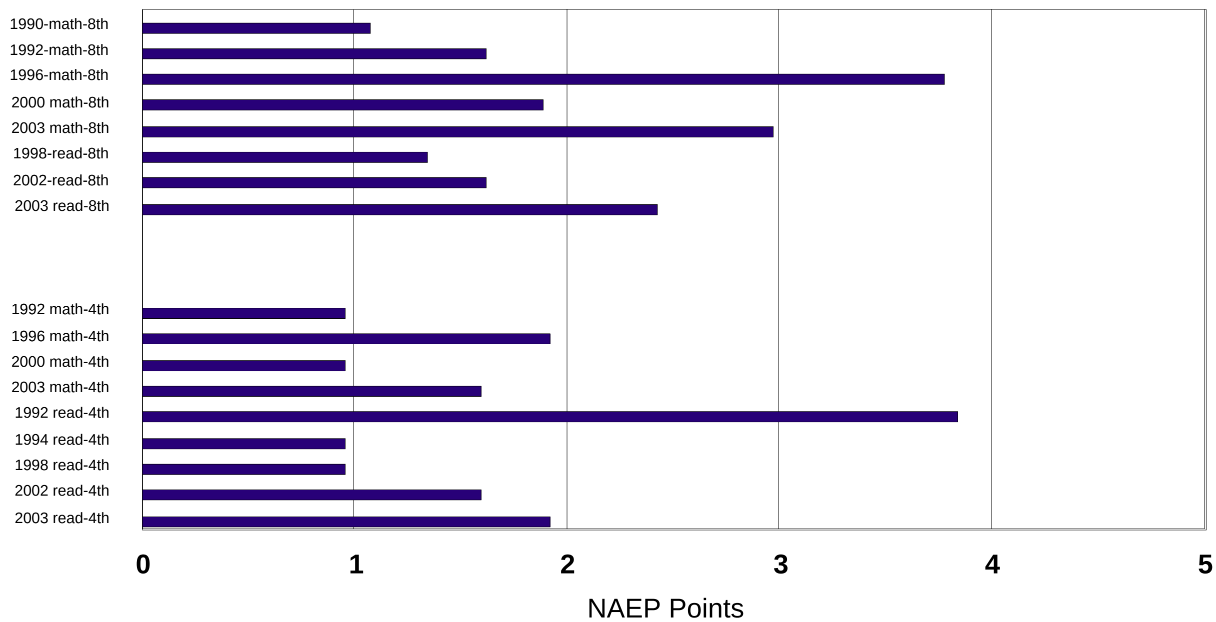
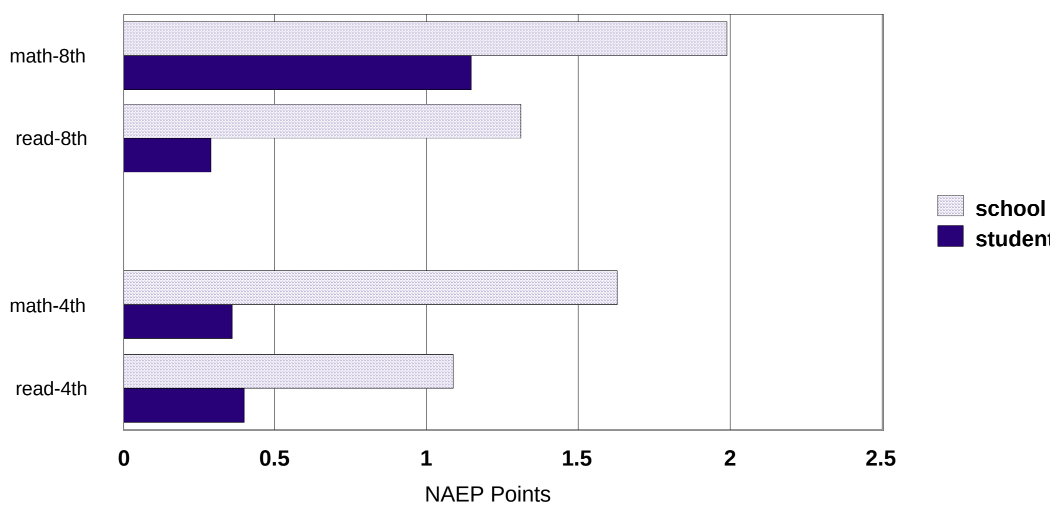


Exhibit 13. Maximum bias estimates of the average national scores between any two tests given between 1990-2003



These worst case bias estimates after corrections, if present in reported data, are certainly large enough to require adjustments. The question is whether there is evidence to suggest that the pattern of non-participation is due to mechanisms that would be consistent with these worst case assumptions. In the next chapters, we make estimates of bias using the empirical pattern of scores and non-participation to determine whether the evidence supports a worst case scenario. We focus on school non-participation because the previous analysis suggests that student non-participation is largely linked to legitimate student absences, and the score distribution of absences is not consistent with worst case scenarios.

4. Methodology for Empirical Estimates

Two methods are used in this report to link school participation and state NAEP scores and thereby explore the potential for bias. The first simple method is to treat the 2002-2003 4th and 8th grade state reading scores and the associated large changes in school participation for many states as a "natural experiment". The second method is to estimate empirical models that "explain" the pattern of state scores across 17 tests from 1990-2003 (696 observations) and to include participation rate as a possible explanatory variable.

The rationale for the first method is that the trend in reading scores have been fairly flat and predictable from 1990-2003, suggesting that little changes would be expected in scores in a single year- between 2002 and 2003 scores (Grissmer and Flanagan, 2006). However, between 2002-2003, school participation rates changed dramatically for many states—as much as 30 percentage points—due to the federal statute mandating all schools to participate in 2003. If changing school participation creates large systematic bias in scores, it should be evident in the changes in the 2002-2003 reading scores.

The rationale for the second method is that large systematic bias in scores from non-participation should also be evident in empirical models "explaining" the pattern of scores over all 17 tests and states. Since non-participation is only one of many factors that might "explain" differences in state scores, such a model has to incorporate non-participation along with these other factors.

Exploiting the 2002-2003 "Natural Experiment"

National public school average 4th grade reading scores declined by 1.4 NAEP points between 2002 and 2003, while 8th grade scores were virtually unchanged. Between 2002-2003, 10 states had changes in both 4th and 8th grade school non-participation of over 20 percentage points; 5-6 states had changes of between 10-20 percentage points; and 17-20 states had changes of less than 5 percentage points. Based on the worst case scenarios discussed the last chapter, the score bias between states with little or no change in non-participation compared with states with 20 or more point changes would be 2-4 NAEP points. Such score changes should leave a measurable imprint on the pattern of score changes between 2002-2003. We estimate two simple models to determine whether such imprints are present in the data. The first model simply regresses the change in scores against the change in non-participation. Since the long-term pattern in NAEP scores is for states with lower scores and SES to have larger score gains, the second model includes SES in the regression as a control variable.

Estimating Regression Models with State Data from 1990-2009[9]

We estimate random and fixed effect models using a panel data set by state for the entire sample of state test observations from 1990-2003. We have 48 states included with up to 17 tests per state (Alaska and Hawaii are excluded due to atypical demographics). The total data set is 696 observations. Previous analysis of the pattern of state scores shows a strong dependence on the characteristics of students and families in the state as captured by an SES

[9] The description of the methodology is substantially taken from Grissmer and Flanagan, 2006.

variable; the characteristics of the educational system, including per pupil spending levels; pupil-teacher ratio; levels of pre-school participation; teacher salaries and teacher reported adequacy of teaching resources; and a significant positive trend that varies considerably by subject, grade, and state (Grissmer et al, 2000; Grissmer and Flanagan, 2006). The trend variables remain strong and significant for 4th and 8th grade mathematics even after accounting for changes in SES and educational variables, suggesting a possible effect from effect due to structural and persistent state factors outside of changing resources. 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, and also by state to account for these score gains.
For instance, the estimated equation using subject and grade trends (the second approach) is:

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

where y_{ij} is the normalized test score (reported score or full population adjusted score) for the ith state (i = 1,48) in the jth 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 kth family variable; G_{ijk} is the kth resource variable; and b_k and c_k and the g's are estimated regression coefficients, 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 subject and grade 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 by subject and grade, rather than the four trend variables, as a way of accounting for gains in scores. We also estimate scores using trends by state, grade and subject to account for the wide differences in state gains in scores. In all equations involving family variables, we include interaction terms between family and subject and grade specific dummies that allow for different slopes for the family coefficients across subjects and grades. We incorporate a variable for the school participation rate by state and test in the above equations in order to test for its significance—other things equal—in accounting for the pattern of scores.

Besides using three different methods that account for trends, we make estimates under a variety of estimation techniques and assumptions to determine whether the coefficient of the participation variable is sensitive to such changes. We test for the effects of changing exclusion rates by comparing—for each state and test—the estimates using the "full population estimated score" (which incorporate adjustments for exclusions) and the reported NAEP score. We obtain the full population estimates utilizing a methodology and data set developed by Don McLaughlin that imputes scores to all excluded students (McLaughlin, 2000; McLaughlin, 2001). We make estimates using both random and fixed effect assumptions. We also estimate models using two different assumptions concerning the validity of the 2003 4th grade mathematics score. In the primary analyses, the 2003 4th grade mathematics scores are treated as valid and used in modeling. In alternative analyses, a dummy variable is introduced for the 2003 4th grade mathematics scores that allows for an error in this test. The sensitivity of results to the 2003 4th grade mathematics scores is tested because this score showed an average gain of over one-half of a grade level for all students nation-wide over 3 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. Therefore, we did not want our results to be overly sensitive to this test.

Data

We briefly describe the data sources and variable construction below. More complete descriptions are given in Grissmer, et al, 2000.

Achievement scores

The published data set contains 696 state achievement scores. The earliest state scores in each test category (1990: 8th-grade mathematics; 1992: 4th-grade mathematics; 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.

Family variables

We described three sets of family variables extensively in an earlier report (Grissmer et al., 2000). In that report, the results were almost always insensitive to which of the three sets of family variables we used. In this analysis we utilize only the SES-FE variable consistently across equations.

This variable is constructed using the National Education Longitudinal Study (NELS) data and 1990 Census data. The former is the largest nationally representative data collection containing both achievement and family data. The NELS tested over 25,000 eighth-grade students and collected data from their *parents* on family characteristics. We develop equations from NELS relating reading and mathematics achievement to eight family characteristics; highest education level of each parent, family income, family size, family type, age of mother at child's birth, mother labor force status and race/ethnicity (see Grissmer et al. [2000] for equations).

These equations essentially develop weights for the influence of each family characteristic and estimate how much of the difference in scores can be attributable to family influence. We want the family control variables to reflect the influence of family only. However, the NELS equations containing family variables only may still reflect some influence of school variables due to the correlation between family and school variables. To address this, we add school fixed-effects. This equation introduces a dummy variable for the approximately 1,000 schools in NELS that further reduces any influence of school variables in the family coefficients.

We also extract from the 1990 Census a sample of families from each state with children aged 12-14 or 8-10 and obtain the same eight family characteristics for each family.[10] We then use the NELS equation to project a score for each child given his or her eight family characteristics. We implicitly assume here that the family equations are similar for fourth- and eighth-grade students. We estimate the mean SES measure for White, Black, and Hispanic families within each state, and develop a weighted state average SES by using the racial/ethnic percentage taking each of the 11 NAEP tests in each state. For instance, the average SES value by racial/ethnic group in Indiana might be -.7 for Blacks, -.2 for Hispanics and .10 for White students. If 80% of NAEP students in Indiana for the 1990 test were White, 12% Black, and 8% Hispanic, the SES composite is (.80)x(.10)+(.12)x(-.7)+(.08) x (-.20).

This method provides family variables that partially reflect the changing characteristics of NAEP test-takers due to changing exclusion rates, participation rates, and population shifts over time, as well as normal sampling variation. To the extent that these factors shift the race/ethnicity of students taking NAEP, our variables will reflect such changes. However, it will not reflect that part of changing family characteristics that affects within-race/ethnicity changes since the SES measures for each racial group do not change over time. The 2000 Census will allow us to develop improved SES variables that better track family changes over time by racial/ethnic group.

Mobility

The mobility variable is the percentage of students reporting no change in schools in the past 2 years required by a change in residence. This variable is taken from the NAEP student survey. Missing 1990 data were imputed by utilizing data on the percentage of students reporting living in the same house for 2 consecutive years (1990-91).

Participation rate

This variable is the reported overall school participation rate for the state in each test.

Educational resource measures

We include 4 variables that account for over 80 percent of state education budgets. These variables are the “big ticket” items in education budgets. They are: average state teacher salaries (cost of living adjusted); pupil-teacher ratios; proportion of children at age 4 in public pre-kindergarten; and teacher reported adequacy of resources (dummy variable for the lowest and next lowest reported adequacy of resources). These variables almost always enter the regression with the appropriate sign. That is, higher teacher salaries, lower pupil-teacher ratios, higher pre-kindergarten attendance, and more teacher reported resources would be predicted to have higher achievement, other things equal.

Dummy control variables

We introduce dummy variables for each subject and grade to allow for differences in test structure and norming. In some specifications we introduce dummy variables to account for gains in each test.

---

10 Census data derived for families with children of similar ages as NAEP test-takers is a different sample than the families of NAEP test-takers. State NAEP excludes private school students, some disabled students and Limited English Proficiency students, and non-participants— all of whom are sampled on Census files. 1990 Census data also will not reflect the demographic changes in the NAEP test-taking population from 1990-1996. In addition the NAEP sample will reflect normal sampling variation from test to test.

Results

The 2002-2003 "Natural Experiment"

Exhibit 14 and 15 show the regression results for 4th and 8th grade reading respectively. The regression has the change in the 2002-2003 score as the dependent variable with the change in participation and/or SES as independent variables. The results show that the coefficient of the variable measuring the change in participation is never statistically significant, either when entered alone or with the SES variable.

Exhibit 14. Regression results for changes in 2002-2003 4th grade reading scores with change in non-participation

	Constant	Change in non-participation	SES	R-squared	N
Model 1	-.52	-.037 (-1.2)		.03	42
Model 2	-.78		-7.15(-2.2)**	.11	42
Model 3	-.60	-.025(-.8)	-6.64(-2.0)**	.12	42

** statistically significant at 5 percent level
* statistically significant at the 10 percent level

Exhibit 15. Regression results for changes in 2002-2003 8th grade reading score with change in non-participation

	Constant	Change in non-participation	SES	R-squared	N
Model 1	-.98	-.015 (-.5)		.01	40
Model 2	-1.08		3.11(.9)	.9	40
Model 3	-.95	-.018(-.6)	3.31(.99)	.03	40

In all regressions, the coefficients of participation are negative—other things equal—suggesting that increasing participation lowers scores. The lack of statistical significance may be due to the lack of power because of the small sample sizes. In this case, the coefficients could still possibly reflect significant bias. To check this out, we have taken the coefficients from model 3 in each case, and compared the estimated bias by state inferred by these coefficients with the bias that would be estimated from the simulated worst case scenarios described in chapter 3. Exhibits 16 and 17 compare the bias implied by the two estimates for 4th and 8th grade reading. More precisely, the bias estimates are the adjustments that would need to be made to the change in 4th and 8th grade scores between 2002 and 2003 to account for differing school participation rates across states. Both figures show that the empirical estimates of bias are substantially less than the simulated worst case bias estimates. The 4th grade empirical estimates are about one-third of the worst case estimates, and the 8th grade estimates are about one-sixth of the worst case estimates. The bias adjustments for 8th grade are at most .5 NAEP points and .75 NAEP points for 4th grade, and these adjustments are for about 5-10 states that had the lowest participation in 2002. Most adjustments are one-quarter of a point or less.

Exhibit 16. Comparison of simulated "worst case" bias estimates with empirical estimates for 8th grade reading

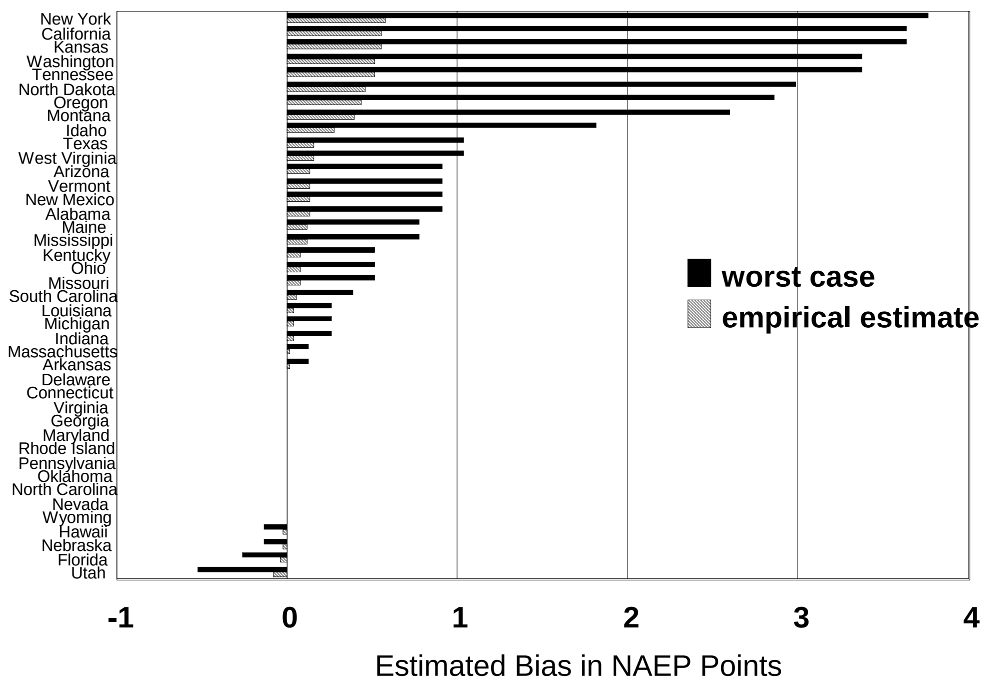
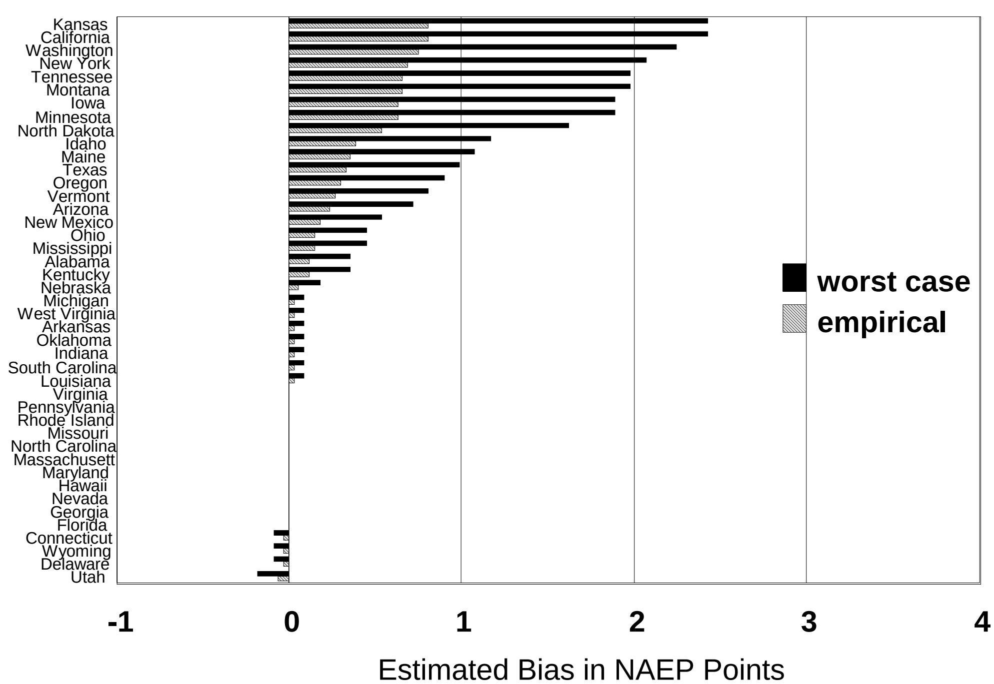


Exhibit 17. Comparison of simulated "worst case" bias estimates with empirical estimates for 4th grade reading



Exhibits 18 and 19 show the reported and adjusted score differences (for changing participation) based on the empirical estimates for 4th and 8th grade reading score differences between 2002 and 2003. For instance, Massachusetts had a 2.4 NAEP point gain in 4th grade reading from 2002-2003, and the adjusted score is identical because school participation was near 100 percent in both years. But New York had a score gain of 1.4 NAEP points that was adjusted to .8 NAEP points as a result of its school participation rate changing from 71 percent in 2002 to 100 percent in 2003. New York's rank would change from 4th highest score change to 7th highest as a result of the adjustment. Generally there are a handful of states whose rank declines by 2-3 positions. At the 4th grade, Kansas changed from 32nd to 38th as a result of the adjustment.

Exhibit 18. Comparing reported and adjusted differences in 2002-2003 8th grade reading NAEP scores

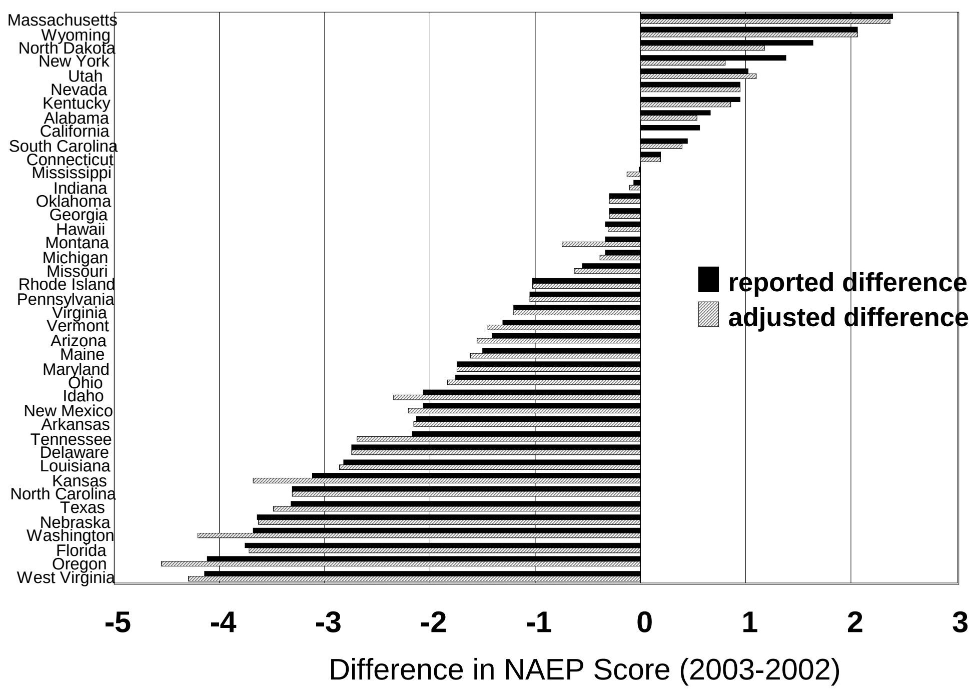
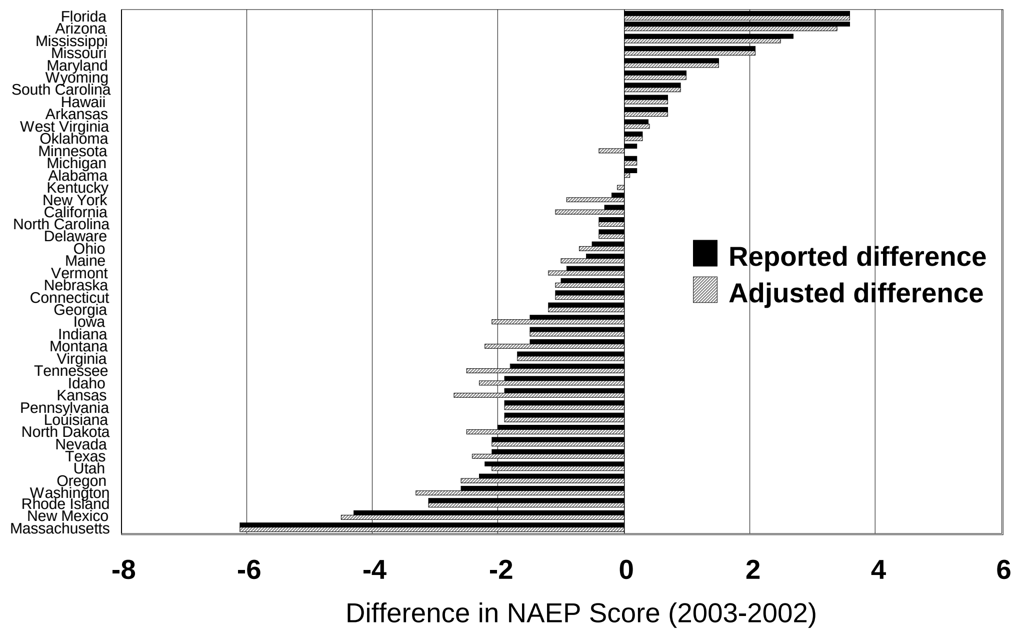


Exhibit 19. Comparing reported and adjusted differences in 2002-2003 4th grade reading NAEP scores



The adjustment of scores for the 4th and 8th grade 2002-2003 test score differences would result in only one change in the estimate of a statistically significant gain or loss from 2002-2003. For the 8th grade test, six states had statistically significant declines in scores, and no states had statistically significant gains. For the adjusted scores, Kansas would be added to the six states having statistically significant declines. For the 4th grade scores, only one state, Massachusetts had a statistically significant change—a loss in score, and the adjusted scores would show the same pattern. Thus the empirically estimated gains show only minor changes in the interpretation of scores for reading tests given in 2002 and 2003.

Regression Results for Complete Set of State Observations

Appendix B shows the complete set of regressions using all state scores (696 observations) from 1990-2003. We show the first set of results from the appendix (TableB.1) in Exhibit 20 below in order to illustrate the results. Exhibit 20 provides the random effect results respectively for the full sample for both reported scores and full population adjusted scores as the dependent variable. The score gains are accounted for in these estimates by dummy variables. For each dependent variable, we present three sets of results. The first model includes these dummy gain variables and shows estimates of score gains for each test from the earliest test. For instance, for the first model using reported scores in Exhibit 20, the gains in the 8th grade tests from 1990-1992 were 0.095 standard deviation units; the gain from 1990-1996 were 0.198 standard deviation units; the gains from 1990-2000 were 0.293 standard deviation units; and the gains from 1990-2003 were 0.427 standard deviation units. For 4th grade mathematics, the gains from 1992-1996 were .101 standard deviation units; the gains from 1992 to 2000 were .207 standard deviation units, and the gains from 1992-2003 were .502 standard deviation units. Gains in reading are substantially smaller.

The second model adds controls for family characteristics interacted with each grade and subject test. All show strong significance at each grade and test and generally the estimated gains increase from model 1. 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. All of the policy variables show significant effects with appropriate signs. Other things equal, lower pupil-teacher ratio, higher teacher salaries, a higher proportion of children in pre-kindergarten and more adequate teacher reported resources are associated with higher scores. However, a significant part of the gains are still present suggesting that resources and demographics alone cannot account for a large part of the gains. Increasing resources seem to account for about one-third of the gains. A plausible candidate for the remaining gains is the impact of standard's based systems- but this analysis does not link empirically to this variable.[11] The analysis using the full population scores differs from those using the reported scores mainly in lowering the size of score gains somewhat. This effect would indicate that rising exclusion rates over time might account for about 5-20 percent of the gains in scores.

The coefficient of the school participation rate variable (-.001) is statistically significant in model 1 using reported scores. Its sign would indicate that higher participation may be linked to lower state scores- other things equal. However, the coefficient for the remaining models in Exhibit 20 are not statistically significant, and the coefficient ranges from +.001 to -.001. The most relevant coefficients are in model 3 which has controls for SES and educational resources. Since the scores are in standard deviation units, the range in NAEP points is about +.03 to -.03. These estimates are very similar to those obtained from the previous analysis of the 2002-2003 reading differences.

[11] Links between standard's based accountability variables and NAEP gains have been the subject of other research (See Hanushek and Raymond, 2004; Hanushek and Raymond, 2003a; Hanushek and Raymond, 2003b; Carnoy and Loeb, 2002)

Exhibit 20. 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 mathematics	0.110	***	0.164		0.284		0.128	***	0.215		0.308	*
Dummy 4th-grade mathematics	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 mathematics 1990-1992	0.095	***	0.103	***	0.088	***	0.094	***	0.103	***	0.089	***
Gain 8th-grade mathematics 1990-1996	0.198	***	0.221	***	0.140	***	0.173	***	0.193	***	0.122	***
Gain 8th-grade mathematics 1990-2000	0.293	***	0.318	***	0.178	***	0.235	***	0.262	***	0.141	***
Gain 8th-grade mathematics 1990-2003	0.427	***	0.459	***	0.292	***	0.388	***	0.421	***	0.274	***
Gain 4th-grade mathematics 1992-1996	0.101	***	0.103	***	0.033	***	0.069	***	0.070	***	0.012	
Gain 4th-grade mathematics 1992-2000	0.207	***	0.225	***	0.108	***	0.164	***	0.178	***	0.079	***
Gain 4th-grade mathematics 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	

(continued on next page)

Exhibit 20. Random effects regressions for raw score and full population adjusted score using gain dummies (continued)

	Raw Score			Full Population Score		
Variable	**Model 1**	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
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 mathematics[a]		0.192	0.038		0.108	0.036
Mobility x 4th-grade mathematics[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 mathematics[b]		1.698 ***	1.734 ***		1.773 ***	1.813 ***
Family (SES) x 4th-grade mathematics[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 report low resources[e]			-0.209 *			-0.200 *
% Teachers report medium resources[f]			0.037			0.042
% Students in public PreK[g]			0.002 **			0.002
Constant	0.016	-0.199	-0.199	0.065	0.244 *	0.108

NOTE: See Table B.1 for variable definitions. Statistical significance denoted by; ***1 percent; **5 percent; *10 percent.

Exhibit 21 shows the coefficients of school participation for the full range of estimated models in Appendix B. Table B.2 is similar to B.1 except fixed effect rather than random effect estimates are made. Tables B.3 and B.4 show the results using trends rather than dummies to account for gains for random and fixed effects models respectively. In these tables estimates are made both assuming the 2003 4th mathematics score is accurate, and that the score is atypical due to its showing about a one-half grade gain over 3 years. Table B.5 shows the random effect results using the same pattern of results as Tables B.3 and B.4 except individual trends by state are included rather than general trends by subject and grade.

Exhibit 21. Coefficients of the school participation variable across alternative models

	Reported Scores				Full Population Scores			
Table A.1	-.001*	-.001	-.001		.001	.001	.001	
Table A.2	-.001	-.001	.000		-.001	-.001	.000	
Table A.3	.000	-.001	.000	-.001	.001	.000	.001*	*.000*
Table A.4	.000***	-.001	.001	-.001	.001***	.000	.001***	*.000*
Table A.5	.000	.000	-.002***	-.001***	.001*	.001	-.001	-.001

Except for one estimated coefficient, the range of coefficients consistently lies between .001 and -.001. The coefficient is statistically significant for about one-quarter of the coefficients. The range when translated into NAEP points is about -.03 to +.03. That is, the suggested bias in a state score would be +/- .03 NAEP points for each percentage point change in school participation. For the few states whose participation changed by 25 points in the 2003 tests from the last test with voluntary participation, the estimated bias would be about .8 NAEP points. This bias would be present between states at the extreme values of participation within each of the 13 tests before 2002. A state like New York, which had one of the lowest average participation rates from 1990-2002 of 79 percent, would have a consistent score bias with respect to those states with near perfect participation of about .6 NAEP points.

The magnitude of these estimates, if accurate, would make non-participation substantially less than the worst case estimates in chapter 3, and overall a marginal threat to validity. The magnitude of the estimated adjustments are generally smaller than those made for changing exclusion rates. However, since there is no methodology to estimate accurate adjustments for non-participation, the question of making adjustments becomes moot.

Strengths and Weaknesses of the Analyses

The strengths of this analysis include: (1) the model is based on 17 separate tests in 2 subjects and 2 grades over a 13-year period that provides over 700 observations of state achievement; (2) the NAEP tests address not only lower level skills through multiple-choice items, but more critical thinking skills with open-ended items; (3) variation across states in almost all dependent variables is quite large compared with within state district or school variation; (4) the analysis uses both random and fixed effect models that incorporates 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 for 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 non-experimental literature, although because of the wide range of such measurements, this standard is not hard to meet.

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

Conclusions

The following conclusions seem consistent with the analysis:

- The level and pattern of *student* non-participation across states at the 4th grade level is predicted by the level and pattern of normal absences, and analysis shows no evidence that the level and pattern of student non-participation is related to NAEP scores or SES.
- The level of student non-participation at 8th grade is much higher than predicted by normal absences, but an analysis of the pattern across states shows a weak relationship with NAEP scores. Higher levels of non-participation are more likely in lower scoring states. However, exploring the causes and possible bias of the higher student non-participation at 8th grade seems warranted, especially given that the future 12th grade state tests may show similar causes and could possibly show even higher student non-participation and potential bias.
- There is no evidence that the student non-participation follows a worst case set of assumptions given that normal absences are not drawn from the extremes of the distribution. In the case of 8th grade non-participation, the weak relationship with NAEP scores implies a potential bias that is substantially smaller than worst case estimates, and represents only a minor threat to validity.
- The precise factors involved in a school non-participation decision remain largely unknown. It is clear that some states with near perfect participation records have either adopted implicit or explicit policies that mandate all schools participate or principals in these states have characteristics that predict compliance. These states tend to be lower scoring states. Higher scoring states appear more likely to allow principals to make voluntary decisions, and states with lower participation tend to be higher scoring states
- The evidence evaluated here would suggest that school non-participation may cause relatively small bias in scores- much smaller than worst case estimates- representing only a marginal threat to validity. However, there appears to be no reliable method to make adjustments to scores in the 1990-2002 period. After 2002, school participation is mandatory and thereby no threat to future validity is present.

Bibliography

Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, *24*(4), 305-331.

Chambers, J. (January 1996). *The patterns of teacher compensation*. NCES 95-829. Washington, D.C.: National Center for Education Statistics.

Grissmer, D.W., Flanagan, A., Kawata, J., & Williamson, S. (2000). *Improving student achievement: What do state NAEP scores tell us?* Santa Monica, CA: RAND

Grissmer, D.W. & Flanagan, A. (May 2006). *Improving the achievement of Tennessee students: An analysis of the National Assessment of Educational Progress*. TR-381-EDU. Santa Monica, CA: RAND.

Grissmer, D. W., Kirby, S. N., Berends, M., & Williamson, S. (1994). *Student achievement and the changing American family*. Santa Monica, CA: RAND.

Hamilton, L.S., Stecher, B.M., & Klein, S.P. (2002). *Making sense of test-based accountability in education* (1554-EDU). Santa Monica, CA: RAND.

Hanushek, E. & Raymond, M. (September 2004). *Does school accountability lead to improved student performance?* NBER xxxx.

Hanushek, E. & Raymond, M. (forthcoming, 2003a). *The effect of school accountability systems on the level and distribution of student achievement.* Journal of the European Economic Association.

Hanushek, E. & Raymond, M. (2003b). "Lessons About the Design of State Accountability Systems, in No Child Left Behind?" *The Politics and Practice of Accountability*, Paul E. Peterson and Martin West, Eds. Washington DC: Brookings, 127-151.

Heubert, J.P., & Hauser, R.M. (1999). *High stakes: Testing for tracking, promotion and graduation*. Washington D.C.: National Academy Press.

McLaughlin, D.H. (2000, July). *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, NC, July 2000.

McLaughlin, D.H. (2001, November). *Exclusions and accommodations affect state NAEP gain statistics: Mathematics, 1996 to 2000.* Report to the NAEP Validity Studies Panel. Palo Alto, CA: American Institutes for Research.

McLaughlin, D.H. , L. Gallagher, F. Stancavage (2004). Evaluation of Bias Correction Methods for "Worst Case" Selective Non-participation in NAEP , American Institute for Research, Palo Alto, CA

Mullis, I.V.S., Dossey, J.A., Owen, E.H., and Phillips, G.W. NAEP 1992 Mathematics Report Card for the Nation and the States: Data from the National and Trail State Assessments, Washington, DC: National Center for Education Statistics, 1993.

Spencer, B.D. (1996). School and Student Sampling in the 1994 Trial State Assessment: An Evaluation, in Glaser, R. and Linn, R. (Eds.). *Quality and Utility: The 1994 Trial State Assessment in Reading, Background Studies,* Stanford, CA.

Appendix A:

Student and School Participation Rates for 17 State NAEP Tests from 1990-2003

Appendix A. Student and School Participation Rates for 17 State NAEP tests from 1990-2003

Table A.1 and A.2 show school and student participation rates for the 17 state NAEP tests given from 1990-2003. Blank entries indicate that states did not take the particular tests or that their participation rates did not meet NAEP guidelines and their scores were not reported.

Table A.1 School participation rates by state NAEP test and state

	1990	1992	1992	1992	1994	1996	1996	1998	1998	2000	2000	2002	2002	2003	2003	2003	2003		
	Mathematics			Reading		Reading		Reading		Mathematics		Reading				Mathematics			
State	8th	4th	8th	4th	4th	4th	8th	8th	4th	4th	8th	4th	8th	4th	8th	4th	8th	Avg.	Std. Dev.
Alabama	97	97	92	97	93	93	90	91	91	94	91	96	93	100	100	100	100	94.4	4.4
Alaska						91	92							97	94	97	94	94.2	2.3
Arizona	97	100	99	99	99	87	87	97	98	88	76	91	93	99	100	99	100	93.9	8.1
Arkansas	100	99	97	96	94	78	71	97	97	87	87	99	99	100	100	100	100	93.6	8.9
California	94	97	98	97	91	94	94	84	80	76	72	72	71	99	99	99	99	88.5	11.6
Colorado	100	100	100	100	100	99	100	97	95					100	100	100	100	99.3	1.5
Connecticut	100	99	99	99	96	100	100	99	98	100	99	100	100	99	100	99	100	98.5	2.3
Delaware	100	92	100	92	100	100	100	100	100			100	100	99	100	99	100	98.8	2.7
Florida	98	100	100	100	100	100	100	100	99			100	100	100	98	100	98	99.5	0.8
Georgia	100	100	99	100	99	98	99	100	99	99	99	100	100	100	100	100	100	98.7	2.8
Hawaii	100	100	100	100	99	100	100	100	100	99	91	100	100	100	99	100	99	98.4	4.4
Idaho	97	97	91	96						75	78	87	86	100	100	100	100	91.6	9.8
Illinois	75									74	75			100	100	100	100	87.7	14.3
Indiana	94	91	94	92	92	91	91			71	73	99	98	100	100	100	100	91.9	10.0
Iowa	91	100	99	100	99	87	84		84	70		77		98	97	98	97	91.3	9.8
Kansas								71	70	71	71	73	72	100	100	100	100	81.9	14.9
Kentucky	100	96	98	97	96	96	92	87	92	94	95	96	96	100	100	100	100	95.7	4.0
Louisiana	100	100	100	100	100	100	100	100	100	100	100	99	98	100	100	100	100	99.0	2.5
Maine		71	84	71	97	87	90	97	96	86	84	88	94	100	100	100	100	89.6	9.9
Maryland	100	99	91	99	96	93	86	85	88	100	98	100	93	100	93	100	93	94.1	5.0
Massachusetts		97	95	97	97	97	92	89	88	99	99	100	98	100	99	100	99	95.9	3.7
Michigan	97	90	94	90		88	86		90	85	81	99	98	100	100	100	100	92.2	8.2
Minnesota	93	94	92	94	95	93	88	74	86	83	74	77		98	100	98	100	89.3	9.5
Mississippi		100	100	100	99	97	95	92	94	98	98	95	94	100	100	100	100	96.8	3.4
Missouri		97	99	97	98	99	96	97	99	96	94	100	96	100	100	100	100	97.3	3.4
Montana	90				89	81	75	78	78	77	75	75	76	97	96	97	96	83.6	9.8

Table A.1 School participation rates by state NAEP test and state (continued)

	1990	1992	1992	1992	1994	1996	1996	1998	1998	2000	2000	2002	2002	2003	2003	2003	2003		
	Mathematics			Reading		Reading		Reading		Mathematics		Reading				Mathematics			
State	8th	4th	8th	4th	4th	4th	8th	8th	4th	4th	8th	4th	8th	4th	8th	4th	8th	Avg.	Std. Dev.
Nebraska	94	87	85	87	77	100	100			97	99	95	99	97	98	97	98	93.1	6.4
Nevada						86		99	100	100	100	100	100	100	100	100	100	97.4	4.5
New Hampshire	97	80	92	81	79				70					98	100	98	100	89.5	10.4
New Jersey	94	82	78	82	91	73								100	99	100	99	89.8	9.7
New Mexico	100	90	94	91	100	100	100	96	99	93	91	93	93	99	100	99	100	95.5	5.4
New York	86	83	83	84	91	86	80	77	84	71	70	77	71	100	100	100	100	84.2	11.1
North Carolina	100	99	98	99	99	97	100	100	99	100	99	100	100	100	100	100	100	98.7	2.3
North Dakota	100	90	97	91	91	96	95			88	90	82	77	100	100	100	100	92.7	7.2
Ohio	98	91	90	91						82	91	95	96	100	100	100	100	93.5	6.9
Oklahoma	99	98	98	98				100	100	100	99	99	100	100	100	100	100	98.5	2.3
Oregon	100					90	92	88	94	74	75	88	78	98	100	98	100	89.4	11.0
Pennsylvania	93	95	94	95	84	86						100	100	100	100	100	100	95.6	5.4
Rhode Island	97	96	100	96	86	99	90	100	100	100	100	100	100	100	100	100	100	97.1	4.2
South Carolina		99	97	99	97	88	87	95	97	97	92	99	97	100	100	100	100	95.9	4.7
South Dakota														98	100	98	100	99.0	1.0
Tennessee		93	91	94	74	94	92	89	97	97	91	78	74	100	100	100	100	90.7	8.7
Texas	97	98	99	97	93	97	95	96	97	99	96	89	92	100	100	100	100	96.1	3.5
Utah		99	100	99	100	100	100	100	100	100	100	100	100	98	96	98	96	98.3	2.4
Vermont						81	74			70	82	90	91	99	98	99	98	87.2	11.5
Virginia	99	99	97	99	99	100	100	100	100	100	100	100	100	100	100	100	100	98.9	2.1
Washington					100	99	95	86	89			75	74	100	100	100	100	92.6	9.7
West Virginia	100	100	100	100	100	100	100	100	100	100	100	99	92	100	100	100	100	98.7	2.7
Wisconsin	99	100	100	99	86	94	78	73		69	73			100	100	100	100	92.1	11.3
Wyoming	100	97	99	97	98	100	100	95	100	100	100	100	100	99	100	99	100	98.4	2.1
Average	*96.6*	*94.9*	*95.4*	*94.9*	*94.2*	*93.4*	*92.2*	*92.5*	*93.4*	*85.3*	*89.2*	*92.6*	*92.4*	*99.4*	*99.3*	*99.4*	*99.3*		
Std. Deviation	*4.9*	*6.4*	*5.3*	*6.2*	*6.7*	*6.9*	*8.0*	*8.7*	*8.2*	*10.7*	*11.4*	*9.3*	*9.6*	*0.9*	*1.6*	*0.9*	*1.6*		

Table A.2 Student participation rates by state NAEP test and state

	1990	1992	1992	1992	1994	1996	1996	1998	1998	2000	2000	2002	2002	2003	2003	2003	2003		
	Mathematics			Reading		Mathematics		Reading		Mathematics		Reading				Mathematics			
State	**8th**	**4th**	**8th**	**4th**	**4th**	**4th**	**8th**	**8th**	**4th**	**4th**	**8th**	**4th**	**8th**	**4th**	**8th**	**4th**	**8th**	**Avg.**	**Std. Dev**
Alabama	95	95	95	96	96	96	93	93	96	95	92	95	93	95	92	95	93	94.4	1.4
Alaska						91	80							94	90	95	92	90.3	4.9
Arizona	93	95	93	95	94	95	91	91	94	94	91	91	88	91	89	92	89	92.1	2.2
Arkansas	95	96	94	96	96	96	92	92	95	95	93	94	91	96	93	95	93	94.2	1.6
California	93	94	92	94	94	94	90	91	93	94	91	95	90	94	91	94	91	92.7	1.6
Colorado	94	95	93	95	94	95	91	91	94					95	91	96	93	93.6	1.6
Connecticut	95	96	94	95	96	96	91	91	94	96	92	95	92	95	91	95	91	93.8	1.9
Delaware	93	95	92	95	96	94	90	91	94			94	90	94	90	94	89	92.7	2.1
Florida	92	95	91	95	94	94	91	89	94			95	91	93	91	93	91	92.6	1.8
Georgia	94	95	93	96	95	95	90	90	96	95	90	95	93	95	93	95	93	93.7	2.0
Hawaii	93	95	90	95	95	95	91	91	95	94	90	96	93	96	92	95	93	93.5	1.9
Idaho	95	97	95	96						96	93	95	93	95	93	94	92	94.4	1.4
Illinois	93									94	93			94	93	94	93	93.3	0.7
Indiana	95	96	94	96	96	96	93			95	93	94	91	94	93	96	93	94.3	1.6
Iowa	96	96	95	96	96	97	93		96	95		95		96	94	96	95	95.4	1.0
Kansas								92	93	96	92	96	93	95	93	95	94	93.8	1.3
Kentucky	95	96	96	96	97	95	94	93	96	95	94	96	94	96	93	95	93	94.9	1.2
Louisiana	94	95	93	96	96	95	89	91	95	96	90	96	93	96	92	96	93	93.9	2.2
Maine		95	92	95	94	94	92	92	93	95	91	94	92	93	92	94	93	93.1	1.2
Maryland	94	96	93	95	95	96	91	89	95	95	90	93	90	94	89	94	89	92.8	2.4
Massachusetts		95	94	96	95	95	92	91	95	96	93	95	93	94	91	94	91	93.8	1.7
Michigan	95	94	94	94		94	90		93	94	88	92	88	95	91	95	91	92.5	2.3
Minnesota	95	95	94	96	95	94	92	93	94	94	93	95		94	90	95	92	93.8	1.5
Mississippi		97	95	97	97	96	93	92	95	95	92	95	93	94	93	94	92	94.4	1.7
Missouri		96	95	95	95	95	91	92	95	95	92	94	91	95	94	94	93	93.9	1.5
Montana	96				96	96	92	92	95	95	92	95	94	94	93	95	93	94.1	1.5

Table A.2 Student participation rates by state NAEP test and state (continued)

	1990	1992	1992	1992	1994	1996	1996	1998	1998	2000	2000	2002	2002	2003	2003	2003	2003		
	Mathematics			Reading		Mathematics		Reading		Mathematics		Reading				Mathematics			
State	**8th**	**4th**	**8th**	**4th**	**4th**	**4th**	**8th**	**8th**	**4th**	**4th**	**8th**	**4th**	**8th**	**4th**	**8th**	**4th**	**8th**	**Avg.**	**Std. Dev**
Nebraska	95	96	96	96	95	95	91			94	92	96	92	95	94	94	94	94.3	1.7
Nevada						92		91	94	94	92	93	88	93	88	93	88	91.5	2.3
New Hampshire	94	96	94	96	96				93					94	92	94	91	94.0	1.6
New Jersey	94	96	94	96	95	95								95	91	95	91	94.2	1.7
New Mexico	94	95	93	95	95	94	90	90	94	95	89	94	92	95	93	95	92	93.2	1.9
New York	93	96	92	95	95	94	91	88	95	94	90	91	88	91	86	92	85	91.5	3.2
North Carolina	95	95	94	96	96	96	91	92	94	95	92	94	93	96	93	95	93	94.2	1.6
North Dakota	96	96	96	97	97	96	94			96	95	96	94	97	95	97	96	95.8	1.0
Ohio	93	95	93	96						95	91	93	90	92	91	92	90	92.6	1.9
Oklahoma	80	84	80	85				91	95	95	93	95	92	96	93	96	93	90.4	5.5
Oregon						95	90	89	95	93	90	94	91	94	90	93	91	92.3	2.0
Pennsylvania		96	94	95	94	95						94	92	96	92	95	93	94.2	1.3
Rhode Island	93	95	93	95	95	95	89	88	94	95	91	94	89	94	88	93	89	92.3	2.7
South Carolina		97	94	96	96	95	89	93	95	96	93	95	93	95	92	95	93	94.2	1.9
South Dakota														95	95	96	95	95.3	0.4
Tennessee		96	94	95	96	96	91	90	94	96	90	96	92	94	93	94	92	93.8	2.0
Texas	96	96	94	96	96	96	92	93	95	96	93	95	93	95	93	96	92	94.5	1.5
Utah		96	94	96	95	95	91	90	95	94	92	94	92	95	92	94	91	93.4	1.8
Vermont						96	93			95	92	95	92	94	90	93	89	92.9	2.1
Virginia	94	95	94	96	95	95	91	91	95	96	92	95	92	95	92	95	92	93.7	1.7
Washington					94	94	90	91	94			95	90	95	92	96	92	93.0	2.0
West Virginia	94	96	94	96	96	95	92	91	94	95	92	96	92	94	92	94	93	93.8	1.7
Wisconsin	94	96	94	96	96	95	92	92			92			95	92	95	92	93.8	1.7
Wyoming	96	96	95	96	96	96	93	91	95	95	93	95	92	94	92	95	91	94.2	1.8
Average	*93.9*	*95.3*	*93.4*	*95.3*	*95.4*	*95.0*	*91.1*	*91.1*	*94.5*	*95.0*	*91.7*	*94.5*	*91.6*	*94.5*	*91.8*	*94.5*	*91.9*		
Std. Deviation	*2.6*	*1.9*	*2.5*	*1.8*	*0.9*	*1.1*	*2.2*	*1.3*	*0.9*	*0.8*	*1.4*	*1.2*	*1.7*	*1.2*	*1.8*	*1.1*	*1.9*		

Appendix B

Regression Results

Appendix B. Regression Results[12]

Tables B.1 and B.2 provide the random and fixed effect results respectively for the full sample for both reported 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 reported scores in Table B.1 the gains in the 8th grade tests from 1990-1992 was 0.095 standard deviation units, the gain from 1990-1996 was 0.198 standard deviation units, the gain from 1990-2000 was 0.293 standard deviation units and the gain from 1990-2003 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 are still present suggesting that resources cannot account for a large part of the gains. The results are similar with the full population score is used as the outcome measure compared with using reported scores.

The fixed effect models in Table B.2 generally show effects that are less significant for policy variables (except for pre-kindergarten which increases in statistical significance). The fixed effect models have fewer degrees of freedom (48 state dummies introduced), but more importantly, reduce the size and significance of the family variables. Our view is that the assumptions in the random effects models are more realistic than the fixed effect models.

[12] These results are reproduced from Grissmer and Flanagan, 2006

Table B.1 Random effects regressions for reported 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 mathematics	0.110 ***	0.164	0.284	0.128 ***	0.215	0.308 *
Dummy 4th-grade mathematics	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 mathematics 1990-1992	0.095 ***	0.103 ***	0.088 ***	0.094 ***	0.103 ***	0.089 ***
Gain 8th-grade mathematics 1990-1996	0.198 ***	0.221 ***	0.140 ***	0.173 ***	0.193 ***	0.122 ***
Gain 8th-grade mathematics 1990-2000	0.293 ***	0.318 ***	0.178 ***	0.235 ***	0.262 ***	0.141 ***
Gain 8th-grade mathematics 1990-2003	0.427 ***	0.459 ***	0.292 ***	0.388 ***	0.421 ***	0.274 ***
Gain 4th-grade mathematics 1992-1996	0.101 ***	0.103 ***	0.033 ***	0.069 ***	0.070 ***	0.012
Gain 4th-grade mathematics 1992-2000	0.207 ***	0.225 ***	0.108 ***	0.164 ***	0.178 ***	0.079 ***
Gain 4th-grade mathematics 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

Table B.1 Random effects regressions for reported score and full population adjusted score using gain dummies (continued)

	Raw Score			Full Population Score		
Variable	**Model 1**	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
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 mathematics[a]		0.192	0.038		0.108	0.036
Mobility x 4th-grade mathematics[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 mathematics[b]		1.698 ***	1.734 ***		1.773 ***	1.813 ***
Family (SES) x 4th-grade mathematics[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 report low resources[e]			-0.209 *			-0.200 *
% Teachers report medium resources[f]			0.037			0.042
% Students in public PreK[g]			0.002 **			0.002
Constant	0.016	-0.199	-0.199	0.065	0.244 *	0.108

NOTE: Statistical significance denoted by; ***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 effect regression with the following estimation equation: $y_{ij} = a + bx_{ij} + u_j + e_{ij}$. The data are from the NELS:88 and the y_{ij} are the mathematics and reading scores for the ith student in the jth school and the x_{ij} 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, the bs were then used to weight the same measures of family characteristics using a sample drawn from 1990 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 the bs 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 on each NAEP test from 1990 to 2003. For more information on the technique used to create the composite SES score, see Grissmer et al. (1994) and Grissmer 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 of 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: eighth grade math; 1992: fourth grade math; 1992: fourth 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 reported NAEP scores. We obtain the full population estimates from a methodology and data set described in McLaughlin (2000) and McLaughlin (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 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 parameters ranges of the variables.

Table B.2 Fixed Effects Regression Results for Reported 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 mathematics	0.109 ***	0.138	0.264 *	0.127 ***	0.185	0.278 **
Dummy 4th-grade mathematics	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 mathematics 1990-1992	0.095 ***	0.102 ***	0.090 ***	0.094 ***	0.101 ***	0.089 ***
Gain 8th-grade mathematics 1990-1996	0.199 ***	0.210 ***	0.139 ***	0.173 ***	0.181 ***	0.116 ***
Gain 8th-grade mathematics 1990-2000	0.295 ***	0.308 ***	0.179 ***	0.236 ***	0.251 ***	0.135 ***
Gain 8th-grade mathematics 1990-2003	0.426 ***	0.439 ***	0.288 ***	0.387 ***	0.398 ***	0.259 ***
Gain 4th-grade mathematics 1992-1996	0.101 ***	0.099 ***	0.031	0.070 ***	0.066 ***	0.007
Gain 4th-grade mathematics 1992-2000	0.208 ***	0.208 ***	0.102 ***	0.165 ***	0.160 ***	0.065 **
Gain 4th-grade mathematics 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 *

Table B.2 Fixed Effects Regression Results for Reported Score and Full Population Adjusted Scores Using Gain Dummies (continued)

	Raw Score			Full Population Score		
Variable	**Model 1**	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
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 mathematics[a]		0.069	-0.075		-0.027	-0.152
Mobility x 4th-grade mathematics[a]		0.118	-0.057		0.198	0.041
Mobility x 8th-grade reading[a]		0.096	0.005		0.037	-0.069
Mobility x 4th-grade reading[a]		0.482 ***	0.315 **		0.544 ***	0.381 **
Family (SES) x 8th-grade mathematics[b]		0.776 ***	0.790 ***		0.747 ***	0.730 ***
Family (SES) x 4th-grade mathematics[b]		0.206	0.282		0.085	0.129
Family (SES) x 8th-grade reading[b]		-0.064	0.015		-0.089	-0.042
Family (SES) x 4th-grade reading[b]		0.066	0.177		0.003	0.077
Teacher salary[c]			0.008 ***			0.007 ***
Pupil-teacher ratio, grades 1–4[d]			-0.001			-0.006
% Teachers report low resources[e]			-0.258			-0.244
% Teachers report medium resources[f]			-0.037			-0.124
% Students in public PreK[g]			0.002 ***			0.002 **
Constant	-0.006	-0.098	-0.244	-0.088 *	0.132	-0.098

NOTE: See Table B.1 notes for variable definitions. Statistical significance denoted by; ***1 percent; **5 percent; *10 percent.

Tables B.3 and B.4 show the results using trends rather than dummies to account for gains for random and fixed effects models respectively. For instance, for the first model using reported scores, the average annual gain in 8th grade mathematics was 0.033 standard deviation units- approximately one percentile point a year (see Table B.3). 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 mathematics test. The dummy indicates that the 4th grade 2003 score was 0.215 standard deviation 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 mathematics test dummy. The policy coefficients show some sensitivity to accounting for gains through trends vs. dummies (i.e., comparing Table B.3 versus Table B.1 and Table B.4 versus Table B.2). For instance, pupil-teacher ratio shows much stronger effects with trends than with dummies. Our view is that the 13 dummy variables in Tables B.1 and B.2, as opposed to four trends in Table B.3 and B.4, represent a more rigorous test of the policy coefficients. The fixed effect results (Table B.4) show the same pattern of differences with random effects models with respect to policy coefficients as do Tables B.1 and B.2 (again the pre-kindergarten measure is the only one to increase in significance in the fixed effects models).

Table B.5 shows the random effect results using the same pattern of results as Tables B.3 and B.4 except individual trends by state are included rather than general trends. For instance, for the first model using reported scores, the state trends for Alabama show an annual gain in scores of 0.020 standard deviation units across all tests over the thirteen years (1990-2003).

Table B.3 Random effects results for reported score and full population adjusted score using average annual NAEP gains by grade and subject

	Reported Score				Full Population Score			
Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Dummy 4th-Grade Mathematics 2003		0.215 ***		0.227 ***		0.255 ***		0.263 ***
Dummy 8th-Grade Mathematics	0.195	0.206	0.224	0.271	0.253 *	0.266 *	0.245	0.298 **
Dummy 4th-Grade Mathematics	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. Annual Gain 8th-Grade Mathematics	0.033 ***	0.033 ***	0.025 ***	0.023 ***	0.029 ***	0.029 ***	0.022 ***	0.019 ***
Avg. Annual Gain 4th-Grade Mathematics	0.046 ***	0.028 ***	0.038 ***	0.017 ***	0.044 ***	0.023 ***	0.037 ***	0.014 ***
Avg. Annual Gain 8th-Grade Reading	0.006 **	0.007 **	-0.002	-0.002	0.001	0.002	-0.007 *	-0.007 *
Avg. Annual 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 Mathematics	0.092	0.055	0.047	-0.028	0.001	-0.052	-0.018	-0.106
Mobility x 4th-Grade Mathematics	-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 *

Table B.3 Random effects results for reported score and full population adjusted score using average annual NAEP gains by grade and subject (continued

	Reported Score								Full Population Score							
Variable	**Model 1**		**Model 2**		**Model 3**		**Model 4**		**Model 1**		**Model 2**		**Model 3**		**Model 4**	
Family (SES) x 8th-Grade Mathematics	1.697	***	1.575	***	1.731	***	1.663	***	1.784	***	1.594	***	1.804	***	1.701	***
Family (SES) x 4th-Grade Mathematics	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 Report Low Resources					0.028		-0.059						0.066		-0.034	
% Teachers Report Medium Resources					0.105		0.086						0.024		0.001	
% Students in Public PreK					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 Table B.1 notes for variable definitions. Statistical significance denoted by; ***1 percent; **5 percent; *10 percent.

Table B.4 Fixed Effects Regression Results for Reported Score and Full Population Adjusted Score Using Average Annual NAEP Gains by Grade and Subject

	Reported Score								Full Population Score							
Variable	**Model 1**		**Model 2**		**Model 3**		**Model 4**		**Model 1**		**Model 2**		**Model 3**		**Model 4**	
Dummy 4th-Grade Mathematics 2003			0.217	***			0.229	***			0.258	***			0.265	***
Dummy 8th-Grade Mathematics	0.152		0.164		0.178		0.229		0.199		0.215		0.181		0.241	*
Dummy 4th-Grade Mathematics	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. Annual Gain 8th-Grade Mathematics	0.031	***	0.031	***	0.026	***	0.023	***	0.027	***	0.027	***	0.022	***	0.019	***
Avg. Annual Gain 4th-Grade Mathematics	0.044	***	0.026	***	0.038	***	0.017	***	0.041	***	0.020	***	0.036	***	0.011	**
Avg. Annual Gain 8th-Grade Reading	0.002		0.003		-0.004		-0.004		-0.004		-0.002		-0.010	**	-0.010	
Avg. Annual 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 Mathematics	-0.038		-0.057		-0.076		-0.158		-0.155		-0.178		-0.141		-0.235	
Mobility x 4th-Grade Mathematics	-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	

Table B.4 Fixed Effects Regression Results for Reported Score and Full Population Adjusted Score Using Average Annual NAEP Gains by Grade and Subject (continued)

	Reported Score				Full Population Score			
Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Family (SES) x 8th-Grade Mathematics	0.570	0.489 **	0.519 **	0.416 *	0.417	0.319	0.362	0.242
Family (SES) x 4th-Grade Mathematics	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 Report Low Resources			-0.022	-0.121			0.020	-0.095
% Teachers Report Medium Resources			0.025	0.001			-0.068	-0.096
% Students in Public PreK			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 Table B.1 notes for variable definitions. Statistical significance denoted by; ***1 percent; **5 percent; *10 percent.

Table B.5. Random Effects Regression Results Using Average Annual NAEP Gains by State for Reported Score and Full Population, Adjusted Score

	Reported Score				Full Population Score			
Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Dummy 4th-Grade Mathematics 2003			0.255 ***	0.262 ***			0.279	0.287 ***
Dummy 8th-Grade Mathematics	0.532 ***	0.468 ***	0.605 ***	0.565 ***	0.565 ***	0.446 **	0.646	0.576 ***
Dummy 4th-Grade Mathematics	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 Mathematics	-0.106	0.038	-0.389 ***	-0.391 ***	-0.127	0.094	-0.456	-0.430 ***
Mobility x 4th-Grade Mathematics	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 Mathematics	2.370 ***	2.247 ***	2.300 ***	2.101 ***	2.384 ***	2.350 ***	2.251 ***	2.048 ***
Family (SES) x 4th-Grade Mathematics	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 ***

Table B.5. Random Effects Regression Results Using Average Annual NAEP Gains by State for Reported Score and Full Population, Adjusted Score (continued)

	Reported Score				Full Population Score			
Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 1**	**Model 2**	**Model 3**	**Model 4**
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 Report Low Resources		-0.086		-0.093		-0.067		-0.072
% Teachers Report Medium Resources		0.225 *		0.048		0.200		-0.030
% Students in Public PreK		0.000		-0.001		0.000		-0.001
Average Annual Gain by State								
Alabama	0.020 ***	0.009 ***	0.018 ***	0.002	0.022 ***	0.014 ***	0.019 ***	0.002
Arizona	0.023 ***	0.024 ***	0.018 ***	0.014 ***	0.021 ***	0.026 ***	0.016 ***	0.013 ***
Arkansas	0.027 ***	0.016 ***	0.023 ***	0.007	0.025 ***	0.015 ***	0.021 ***	0.004
California	0.030 ***	0.012 ***	0.029 ***	0.016 ***	0.032 ***	0.013 ***	0.031 ***	0.017 ***
Colorado	0.034 ***	0.023 ***	0.028 ***	0.022 ***	0.033 ***	0.025 ***	0.027 ***	0.021 ***
Connecticut	0.034 ***	0.029 ***	0.030 ***	0.021 ***	0.032 ***	0.025 ***	0.028 ***	0.018 ***
Delaware	0.054 ***	0.041 ***	0.050 ***	0.038 ***	0.044 ***	0.031 ***	0.040 ***	0.027 ***
Florida	0.045 ***	0.034 ***	0.041 ***	0.036 ***	0.045 ***	0.035 ***	0.040 ***	0.036 ***
Georgia	0.034 ***	0.040 ***	0.029 ***	0.014 ***	0.033 ***	0.046 ***	0.027 ***	0.011 **
Idaho	0.019 ***	0.015 ***	0.014 ***	0.002	0.018 ***	0.017 ***	0.013 ***	0.000
Illinois	0.041 ***	0.033 ***	0.040 ***	0.024 ***	0.035 ***	0.033 ***	0.032 ***	0.016 **
Indiana	0.035 ***	0.042 ***	0.029 ***	0.019 ***	0.033 ***	0.045 ***	0.026 ***	0.015 ***
Iowa	0.020 ***	0.026 ***	0.012 ***	0.004	0.014 **	0.022 ***	0.005	-0.005

Table B.5. Random Effects Regression Results Using Average Annual NAEP Gains by State for Reported Score and Full Population, Adjusted Score (continued)

	Reported Score				Full Population Score			
Variable	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Average Annual Gain by State								
Kentucky	0.029 ***	0.018 ***	0.026 ***	0.017 ***	0.024 ***	0.018 ***	0.020 ***	0.011 **
Kansas	0.028 ***	0.022 ***	0.024 ***	0.015 ***	0.028 ***	0.024 ***	0.024 ***	0.015 ***
Louisiana	0.043 ***	0.029 ***	0.040 ***	0.029 ***	0.039 ***	0.031 ***	0.035 ***	0.024
Maine	0.011 **	0.020 ***	0.004	-0.007	0.011 **	0.018 ***	0.004	-0.009 *
Maryland	0.039 ***	0.026 ***	0.036 ***	0.027 ***	0.033 ***	0.025 ***	0.029 ***	0.020 ***
Massachusetts	0.038 ***	0.032 ***	0.034 ***	0.022 ***	0.032 ***	0.029 ***	0.029 ***	0.016 ***
Michigan	0.033 ***	0.038 ***	0.029 ***	0.018 ***	0.030 ***	0.041 ***	0.026 ***	0.013 ***
Minnesota	0.033 ***	0.029 ***	0.029 ***	0.020 ***	0.030 ***	0.030 ***	0.025 ***	0.015 ***
Mississippi	0.029 ***	0.023 ***	0.024 ***	0.011 **	0.027 ***	0.023 ***	0.022 ***	0.008
Missouri	0.025 ***	0.025 ***	0.019 ***	0.008	0.021 ***	0.026 ***	0.015 ***	0.002
Montana	0.020 ***	0.021 ***	0.016 ***	0.007	0.017 ***	0.023 ***	0.012 **	0.003
Nebraska	0.018 ***	0.009 ***	0.014 ***	0.004	0.016 ***	0.008 ***	0.011 **	0.000
Nevada	0.024 ***	0.015 ***	0.021 ***	0.013 **	0.019 ***	0.012 ***	0.015 ***	0.008
New Hampshire	0.024 ***	0.022 ***	0.018 ***	0.009 *	0.023 ***	0.024 ***	0.016 ***	0.006
New Jersey	0.034 ***	0.035 ***	0.026 ***	0.013 **	0.032 ***	0.038 ***	0.023 ***	0.008
New Mexico	0.016 ***	0.016 ***	0.013 ***	0.001	0.014 ***	0.015 ***	0.010 **	-0.004
New York	0.043 ***	0.034 ***	0.040 ***	0.027 ***	0.039 ***	0.030 ***	0.036 ***	0.023 ***
North Carolina	0.063 ***	0.055 ***	0.059 ***	0.044 ***	0.053 ***	0.052 ***	0.049 ***	0.033 ***
North Dakota	0.014 ***	0.013 ***	0.008 *	-0.002	0.012 **	0.012 ***	0.006	-0.005

Table B.5. Random Effects Regression Results Using Average Annual NAEP Gains by State for Reported Score and Full Population, Adjusted Score (continued)

	Reported Score								Full Population Score							
Variable	**Model 1**		**Model 2**		**Model 3**		**Model 4**		**Model 1**		**Model 2**		**Model 3**		**Model 4**	
Average Annual Gain by State																
Ohio	0.042	***	0.040	***	0.038	***	0.024	***	0.037	***	0.039	***	0.032	***	0.018	***
Oklahoma	0.019	***	0.022	***	0.014	***	0.005		0.017	***	0.021	***	0.012	***	0.001	
Oregon	0.023	***	0.020	***	0.021	***	0.014	***	0.019	***	0.020	***	0.016	***	0.010	*
Rhode Island	0.025	***	0.001		0.023	***	0.011	**	0.019	***	-0.003		0.017	***	0.004	
South Carolina	0.047	***	0.039	***	0.043	***	0.029	***	0.042	***	0.036	***	0.037	***	0.022	***
Tennessee	0.020	***	0.019	***	0.016	***	0.000		0.021	***	0.023	***	0.017	***	-0.001	
Texas	0.046	***	0.056	***	0.040	***	0.028	***	0.039	***	0.054	***	0.032	***	0.020	***
Utah	0.011	**	0.018	***	0.007	*	-0.003		0.011	**	0.022	***	0.007		-0.004	
Vermont	0.027	***	0.009	***	0.025	***	0.012	*	0.025	***	0.006	*	0.023	***	0.009	
Virginia	0.038	***	0.033	***	0.034	***	0.020	***	0.032	***	0.030	***	0.027	***	0.012	**
Washington	0.025	***	0.028	***	0.021	***	0.015	***	0.029	***	0.034	***	0.024	***	0.017	***
West Virginia	0.028	***	0.005	*	0.025	***	0.012	**	0.021	***	-0.007	**	0.018	***	0.005	
Wisconsin	0.025	***	0.037	***	0.018	***	0.009	***	0.021	***	0.037	***	0.013	***	0.003	
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 Table B.1 notes for variable definitions. Statistical significance denoted by; ***1 percent; **5 percent; *10 percent.