Research Report

# Evaluating the Effectiveness of a Full-Population Estimation Method

Henry Braun

Jinming Zhang

Sailesh Vezzu

April 2008

ETS RR-08-18

Listening. Learning. Leading.®

# Evaluating the Effectiveness of a Full-Population Estimation Method

Henry Braun[1]
Boston College, Chestnut Hill, MA

Jinming Zhang and Sailesh Vezzu
ETS, Princeton, NJ

April 2008

## Abstract

At present, although the percentages of students with disabilities (SDs) and/or students who are English language learners (ELL) excluded from a NAEP administration are reported, no statistical adjustment is made for these excluded students in the calculation of NAEP results. However, the exclusion rates for both SD and ELL students vary substantially across jurisdictions at a given administration, and, in some cases, have changed substantially over time within a jurisdiction. Consequently, comparisons of performance based on reported NAEP scores may indeed be biased by differential exclusion and identification practices.

Using only NAEP data, this report investigates plausible explanations for the observed heterogeneity among jurisdictions in exclusion rates. It also examines the operating characteristics of a particular class of methods that carry out statistical adjustments to NAEP's reported scores to address the possible bias due to differential exclusion rates. The final results of such adjustments are termed full-population estimates (FPEs). The conclusions are that there is both a strong likelihood of bias and that neither the current NAEP procedure nor the FPE methodologies constitutes an ideal solution. The former because it assumes that all excluded students could not meaningfully participate in NAEP, and the latter because they implicitly assume that all students could obtain a proper NAEP score.

Key words: Excluded students, full-population estimates (FPEs), indirect standardization, NAEP

## Table of Contents

## List of Tables

## 1. Introduction

The purpose of the National Assessment of Education Progress (NAEP; also known as the Nation's Report Card) is to document the achievement of American students in a number of academic disciplines at the national and state levels, overall and by subgroups defined in terms of various student characteristics. One of the most important uses of NAEP is to track the achievement trajectories over time of the different groups. Both comparisons at a given point of time and comparisons of changes over time are affected by how well the assessed students represent the populations of interest. Although the schools and students within schools originally selected for NAEP are indeed representative of the total population (in a strict, statistical sense), the students actually assessed may not be. One of the reasons is that schools or students within schools may refuse to participate. In addition, some students may be willing to participate but happen to be absent on the day of the assessment. Currently, NAEP employs a number of strategies to minimize the effects of these occurrences.

A different concern arises because students with disabilities (SD) and/or English language learners (ELL) can be excluded from the assessment if, in the considered judgment of school officials, they cannot meaningfully participate in the assessment, even with the accommodations provided. At present, no adjustment is made for these excluded students for NAEP reporting. Now, if all excluded students indeed could not meaningfully participate in NAEP, it would be appropriate to make no adjustment. That is, the current NAEP procedure is intended to provide estimates for the population of students who could meaningfully participate in NAEP, which is a subset of all students enrolled in a particular grade.

However, exclusion rates do vary substantially across jurisdictions at a given administration and, in some cases, have changed substantially over time within a jurisdiction. It is likely, therefore, that exclusion decisions are not being made according to uniform procedures and that there are systematic differences among jurisdictions. Accordingly, there is a reasonable concern that, for some jurisdictions, NAEP's estimates of achievement can be biased sufficiently to lead to incorrect inferences. Obviously, estimates of differences among some pairs of jurisdictions would be biased as well. These concerns are heightened with the greater prominence of NAEP results following passage of the No Child Left Behind (NCLB) legislation.

The purpose of this paper is twofold: The first is to determine if there are plausible explanations for the observed heterogeneity among jurisdictions in exclusion rates. The second is to examine the operating characteristics of a particular class of methods that carry out statistical adjustments to NAEP's reported scores to address the possible bias due to these differences in exclusion rates. The final results of such adjustments are termed *full-population estimates* (FPEs), since they are intended to mimic the estimates that would have been obtained had every student in the NAEP sample actually sat for the assessment. That is, the FPE estimates what would have been observed had all students in the grade (irrespective of disability status or English language proficiency) taken the NAEP assessment. It is important to recognize, therefore, that the target populations for the current NAEP procedure and for the FPE are qualitatively different.

For present purposes the term *FPE* refers to a family of methods in which plausible values (PV) are imputed for excluded students in order to construct a complete data file. These PV imputed for excluded students are called *pseudo-plausible values* (PPV) to distinguish them from the PV of assessed students produced during the NAEP operational analysis.

One FPE method developed by McLaughlin (2000, 2001, 2003) employs a regression adjustment based on the estimated relationship between achievement and student characteristics for the SD or ELL students who were assessed. The McLaughlin approach has the advantage of generating results that are easily accommodated within NAEP's current reporting protocols, and can produce adjusted estimates of virtually any statistic that NAEP has reported in the past. The Department of Education decided to include in an appendix to the report of the NAEP 2002 Reading Assessment results based on McLaughlin's method. This decision highlights the importance of examining the technical merits of McLaughlin's or similar methods and the validity of the results obtained thereby.

McLaughlin (2000) divided classified students into two mutually exclusive groups: SD and ELL but not with disabilities (*ELL-only*). That is, students with disabilities and who are also ELL are included in the SD group, which is called *SD- all* in this report.[2] The analyses were carried out separately for the two groups, SD-all and ELL-only. McLaughlin's method involves building a linear regression model that links the mean PV to some set of student characteristics and estimating the parameters of that model employing data from classified students who were assessed. Available student characteristics include demographics (e.g., gender, race/ethnicity), variables such as degree of disability and/or years of study in English, as well as grade level of instruction. Then he builds and estimates a variance model that captures, albeit approximately, the different components of uncertainty. Finally, for each excluded student, PPV are drawn from normal distributions whose means are derived from the fitted regression model and whose variances are derived from the variance models.

Over the years, McLaughlin has also carried out auxiliary analyses to support his contention that excluded students (SD or ELL) differ systematically from included students (SD or ELL). These analyses employed data from certain states where state test data were available for substantial numbers of excluded (from NAEP) and included SD/ELL students. In every case, he observed that the mean state score for excluded SD/ELL students was lower than the mean state test score for included SD/ELL students. For further details, see McLaughlin (2005, pp. 14–21). This finding is consistent with the hypothesis that excluded students are not missing completely at random, with the implication that the current NAEP method would yield biased estimates for the population of all enrolled students.

The McLaughlin approach was evaluated by means of simulations developed and implemented by HumRRO (Wise, Hoffman, & Becker, 2006). Wise et al. (2006) began by applying a particular hot-deck procedure to complete the original NAEP data set by generating PPV for excluded students.[3] Using the completed data set as a starting point, they then constructed three sets of simulations representing three different levels of selection bias, which were denoted as Conditions 1, 2, and 3. The results indicated that, if the target population is taken to be all enrolled students, then the McLaughlin method is superior (in the sense of lower mean squared error) to the current NAEP procedure. Again, this is to be expected, since the current NAEP procedure essentially ignores the excluded students. Not surprisingly, the reduction in mean squared error is greater when the selection bias is greater.

ETS also proposed an FPE method based on slight modifications of the McLaughlin method. One modification is that a missing category is used if a background variable is missing for an excluded student. Initially, McLaughlin's method involved imputing missing values. More recently, McLaughlin (2005) proposed a different method, which involves coding the levels of each variable based on the mean PV of the included students at each level. Another modification is that a different variance component formula is proposed for generating PPV. The ETS approach is described in details in section 3. Not surprisingly, the HumRRO simulation results are quite similar for the two methods (see Wise et al., 2006). At the same time, a number of concerns have been expressed concerning the implications of employing FPEs as the official results in the Nation's

Report Card. These concerns, juxtaposed against HumRRO's empirical findings, indicate the desirability of a more thorough look at FPEs.

In section 2, we explore the differences among states in exclusion rates, with the goal of determining the extent to which they can be explained by differences in student populations. (This is critical inasmuch as the principal motivation for resorting to FPEs is the interpretation of the observed heterogeneity in exclusion rates as evidence of "gaming the system" on the part of some states.) In section 3, we describe the results of an independent simulation carried out at ETS. The rationale for this simulation is that it adopts a different design strategy so that the findings complement (rather than simply replicate) HumRRO's findings. Moreover, its design affords the possibility of enhancing our understanding of how FPEs work. In the final section, we summarize the findings, present key issues around the implementation of FPEs, and sketch out some further research that should be carried out before a policy decision can be made.

It must be emphasized that the investigations reported here do not bear on the advantages and disadvantages of the operational methodology employed by NAEP to obtain PV. That methodology is taken as given and the PV so obtained constitute the input for the FPE methodologies described below.[4]

## 2. Investigating Differences in Exclusion Rates Among States

As indicated in the introduction, the principal justification for employing FPEs in place of the current NAEP estimates is that the heterogeneity in exclusion rates among states in a particular administration, as well as substantial differences in some states in exclusion rates over administrations, may signal the presence of bias in the comparisons that are at the heart of NAEP reports. The term *bias* implies that (a) the observed differences in exclusion rates do not reflect true differences in the proportions of the student populations that can meaningfully participate in NAEP but, rather, are the result of systematic differences in the application of the protocols by which school officials are instructed to determine whether SD and/or ELL students can meaningfully participate in NAEP[5] and/or other differences among states, such as the range of accommodations they offer, and (b) these systematic differences change the expected values of the estimates of population quantities because the comparability of the assessed samples across states is compromised. The idea behind the introduction of the FPE is to create a more level playing field for state-to-state comparisons.

It is impossible to determine retrospectively what were the true differences in exclusion rates and to what extent they were tracked by the observed differences. On the other hand, it is possible to accumulate relevant circumstantial evidence that should be considered in formulating policy with respect to the issue. This is the subject of the current section.

The statewide exclusion rates reported by NAEP employ the full originally selected student sample in that state as the denominator for the calculation. That is, they estimate the proportion of students in the state sampling frame excluded from NAEP. For the 42 states included in this analysis based on the NAEP Reading Assessment for grade 4, for example, state exclusion rates vary from 0.02 to 0.17 (see the second column of Table 1).

It is instructive to examine the exclusion rates separately for each group; that is, employ the number of students in the group as the denominator for the calculation. Observed exclusion rates are the proportions of excluded students in the SD or ELL groups and are presented in Table 1 (columns 4 and 7). Note that students who are classified as both SD and ELL are included in what we refer to as the *SD-all* group.

The 42 states listed in the table are those that are employed in the simulation described in the next section. Columns 1 and 2 display the state name and the overall exclusion rate; column 3 contains the total number of SD students, column 4 contains the exclusion rate for the SD-all category; columns 6 and 7 present analogous results for ELL students. (Columns 5 and 8 will be described shortly.)

Comparing the category-specific exclusion rates (columns 4 and 7) with the reported state-level exclusion rate (column 2), it is evident that the former are much more variable than the latter. Note that for the category-specific rates, the denominators are the total in that category, while for the state-level exclusion rate the denominator is the number of students sampled. Thus, the numbers in column 2 are likely to be more stable. From column 4 we see that exclusion rates for SD-all students vary from 0.12 to 0.64 with a median of 0.30. For ELL students (column 7), they vary from 0.03 to 0.60 with a median of 0.29. Evidently, there is substantial heterogeneity among states that is partially masked when only statewide rates are reported. These data are worrisome in the absence of a plausible explanation for such heterogeneity. What form might such an explanation take?

The simulation in the next section exploits the fact that exclusion rates vary with student characteristics. In fact, we employ a pair of student characteristics derived from the NAEP questionnaire to classify SD-all students into one of 10 categories.[6] Another pair of student characteristics, also derived from the NAEP questionnaire, is used to classify ELL-only students into 1 of 10 categories.[7] Tables 2 and 3 indicate how the different combinations of levels (i.e., cross-categories) of the two characteristics have been assigned the labels 1 to 10, for SD-all students and ELL-only students, respectively. For completeness, Tables 4 and 5 present the aggregate (i.e., pooled over states) counts of total, assessed and excluded by category, for SD-all and ELL-only students, respectively. The last panel of each table presents the corresponding category-specific exclusion rates. Evidently, and not unexpectedly, there are substantial differences in the category-specific exclusion rates. For future reference, we note that the data in the top panels of Tables 4 and 5 serve as the basis for the simulation described in the next section.

In this section, however, interest centers on the category-specific exclusion rates for each state. These are presented in Tables 6 and 7 for SD-all and ELL-only students, respectively. Summary statistics are presented at the bottom of each table. Clearly, there can be substantial uncertainty attached to the rates for some state-category combinations that are based on small samples. Principal interest focuses, however, on summary statistics across states that are relatively insensitive to the variability of individual rates.

Examination of the tables reveals that for each state the exclusion rates are rather different from category to category, with the pattern generally conforming to what one would expect given the definitions of the characteristics. For example, for SD-all students with a moderate level of disability and receiving grade-level instruction (Category 2), the median exclusion rate (across the 42 states) is 0.11. On the other hand, for SD-all students with a profound level of disabilities and receiving instruction two or more levels below grade (Category 9), the median exclusion rate is 0.75.

This observation leads to one possible explanation for the between-state heterogeneity in the aggregate exclusion rates by category: Suppose that states are indeed appropriately and uniformly implementing the exclusion protocols; however, the distribution of SD (ELL) students' characteristics varies substantially across states and, accordingly, the proportions of students falling in each of the 10 categories also varies across states. Consequently, the observed differences in the aggregate SD (ELL) exclusion rates across states are (mostly) due to the differences in student characteristics and not to systematic differences in states' implementation of the policies governing exclusions. Were that the case, it would cast the data presented in Table 1 in a rather different light.

**Table 1**

***Counts of Assessed and Excluded SD-All and ELL-Only Students and the Corresponding Exclusion Rates, Reading, Grade 4, by State: 2003***

| State[a] | Overall state-level exclusion rate | SD-all | | | ELL-only | | |
|---|---|---|---|---|---|---|---|
| | | Total [b] | Observed exclusion rate | Standardized exclusion rate | Total [b] | Observed exclusion rate | Standardized exclusion rate |
| Alabama | 0.02 | 400 | 0.17 | 0.37 | 0 | 0.42 | 0.31 |
| Alaska | 0.03 | 400 | 0.14 | 0.34 | 300 | 0.03 | 0.15 |
| Arizona | 0.08 | 400 | 0.47 | 0.36 | 700 | 0.16 | 0.23 |
| Arkansas | 0.06 | 400 | 0.39 | 0.36 | 100 | 0.34 | 0.19 |
| California | 0.06 | 1,000 | 0.26 | 0.35 | 3,000 | 0.09 | 0.19 |
| Colorado | 0.03 | 400 | 0.2 | 0.36 | 200 | 0.18 | 0.27 |
| Connecticut | 0.05 | 400 | 0.3 | 0.32 | 100 | 0.44 | 0.31 |
| Delaware | 0.12 | 600 | 0.64 | 0.32 | 100 | 0.45 | 0.28 |
| Florida | 0.05 | 600 | 0.19 | 0.33 | 300 | 0.21 | 0.27 |
| Georgia | 0.03 | 600 | 0.24 | 0.32 | 100 | 0.27 | 0.31 |
| Hawaii | 0.04 | 400 | 0.24 | 0.37 | 200 | 0.26 | 0.27 |
| Idaho | 0.04 | 400 | 0.24 | 0.36 | 200 | 0.16 | 0.17 |
| Illinois | 0.09 | 800 | 0.35 | 0.34 | 500 | 0.33 | 0.32 |
| Indiana | 0.04 | 500 | 0.28 | 0.28 | 100 | 0.2 | 0.21 |
| Kansas | 0.03 | 400 | 0.19 | 0.34 | 100 | 0.29 | 0.26 |
| Louisiana | 0.06 | 600 | 0.3 | 0.26 | 0 | 0.42 | 0.24 |
| Maine | 0.07 | 500 | 0.39 | 0.36 | 0 | 0.14 | 0.29 |
| Maryland | 0.08 | 500 | 0.48 | 0.34 | 100 | 0.45 | 0.3 |
| Massachusetts | 0.06 | 800 | 0.19 | 0.33 | 300 | 0.36 | 0.3 |
| Michigan | 0.07 | 400 | 0.57 | 0.38 | 200 | 0.23 | 0.18 |
| Minnesota | 0.04 | 500 | 0.22 | 0.34 | 200 | 0.13 | 0.21 |
| Mississippi | 0.06 | 400 | 0.61 | 0.35 | 0 | 0.56 | 0.25 |
| Missouri | 0.08 | 600 | 0.46 | 0.31 | 100 | 0.6 | 0.19 |
| Nevada | 0.1 | 500 | 0.42 | 0.39 | 500 | 0.32 | 0.27 |
| New Hampshire | 0.04 | 600 | 0.21 | 0.32 | 100 | 0.33 | 0.3 |
| New Jersey | 0.05 | 500 | 0.27 | 0.33 | 100 | 0.45 | 0.29 |

*(Table continues)*

Table 1 (continued)

| State[a] | Overall state-level exclusion rate | SD-all | | | ELL-only | | |
|---|---|---|---|---|---|---|---|
| | | Total [b] | Observed exclusion rate | Standardized exclusion rate | Total [b] | Observed exclusion rate | Standardized exclusion rate |
| New Mexico | 0.08 | 600 | 0.25 | 0.36 | 700 | 0.15 | 0.21 |
| New York | 0.08 | 600 | 0.32 | 0.33 | 300 | 0.57 | 0.31 |
| North Carolina | 0.07 | 900 | 0.35 | 0.32 | 200 | 0.32 | 0.26 |
| North Dakota | 0.04 | 500 | 0.26 | 0.27 | 100 | 0.03 | 0.16 |
| Ohio | 0.09 | 700 | 0.58 | 0.39 | 100 | 0.51 | 0.33 |
| Oregon | 0.09 | 600 | 0.4 | 0.37 | 400 | 0.26 | 0.24 |
| Rhode Island | 0.05 | 600 | 0.17 | 0.3 | 200 | 0.24 | 0.26 |
| South Carolina | 0.08 | 600 | 0.45 | 0.3 | 100 | 0.44 | 0.26 |
| Tennessee | 0.05 | 500 | 0.31 | 0.35 | 100 | 0.27 | 0.25 |
| Texas | 0.17 | 1,000 | 0.54 | 0.36 | 1,100 | 0.48 | 0.3 |
| Utah | 0.05 | 500 | 0.24 | 0.34 | 300 | 0.18 | 0.2 |
| Vermont | 0.07 | 500 | 0.38 | 0.36 | 0 | 0.28 | 0.22 |
| Virginia | 0.11 | 500 | 0.59 | 0.31 | 200 | 0.48 | 0.25 |
| Washington | 0.06 | 500 | 0.33 | 0.38 | 200 | 0.19 | 0.27 |
| Wisconsin | 0.06 | 500 | 0.34 | 0.35 | 200 | 0.3 | 0.21 |
| Wyoming | 0.02 | 400 | 0.12 | 0.34 | 100 | 0.1 | 0.17 |

*Note.* The SD-all category includes students classified as students with disabilities (SD) and students classified as both SD and English language learners (ELL). The ELL-only category includes students classified as English language learners only. ELL-only totals for California and Texas are exceptionally large due to state-specific immigration patterns. SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment. Authors' calculations.

[a] Forty-two states with the state achievement test score as a school-level sampling variable were included in the study. [b] The counts presented in the table are rounded to the nearest 100.

**Table 2**

***Structure of Stratification for SD-All Analyses: 2003***

| Grade level of instruction in reading/language arts | Degree of disability | | | |
|---|---|---|---|---|
| | Mild | Moderate | Profound | Other |
| At or above grade level | (1) | (2) | (3) | † |
| 1 year below grade level | (4) | (5) | (6) | † |
| 2 or more years below grade level | (7) | (8) | (9) | † |
| Other | † | † | † | (10) |

*Note.* † = not applicable. Numbers in parentheses represent cross-categories of years of receiving instruction in English and grade level of instruction. The SD-all category includes students classified as students with disabilities (SD) and students classified as both SD and English language learners (ELL).

**Table 3**

***Structure of Stratification for ELL-Only Analyses: 2003***

| Grade level of instruction in reading/language arts | Years of receiving instruction in English | | | |
|---|---|---|---|---|
| | 4 or more years | 2 or 3 years | 1 year | Other |
| At or above grade level | (1) | (2) | (3) | † |
| 1 year below grade level | (4) | (5) | (6) | † |
| 2 or more years below grade level | (7) | (8) | (9) | † |
| Other | † | † | † | (10) |

*Note.* † = not applicable. Numbers in parentheses represent cross-categories of years of receiving instruction in English and grade level of instruction. The ELL-only category includes students classified as English language learners (ELL) only.

**Table 4**

***Counts of Students Identified as SD-All, Assessed, and Excluded, and the Proportion of Excluded Students, Reading, Grade 4, by Degree of Disability and Grade Level of Instruction: 2003***

| Grade level of instruction in reading/language arts | Total | Degree of disability | | | |
|---|---|---|---|---|---|
| | | Mild | Moderate | Profound | Other |
| Identified | | | | | |
| Total | 23,095 | 9,133 | 7,996 | 1,713 | 4,253 |
| At or above grade level | 7,017 | 4,633 | 2,098 | 286 | † |
| 1 year below grade level | 4,943 | 2,598 | 2,137 | 208 | † |
| 2 or more years below grade level | 6,882 | 1,902 | 3,761 | 1,219 | † |
| Other | 4,253 | † | † | † | 4,253 |
| Assessed | | | | | |
| Total (regular & accommodated) | 15,267 | 7,047 | 4,990 | 654 | 2,576 |
| At or above grade level | 6,137 | 4,148 | 1,769 | 220 | † |
| 1 year below grade level | 3,566 | 1,910 | 1,526 | 130 | † |
| 2 or more years below grade level | 2,988 | 989 | 1,695 | 304 | † |
| Other | 2,576 | † | † | † | 2,576 |
| Excluded | | | | | |
| Total | 7,828 | 2,086 | 3,006 | 1,059 | 1,677 |
| At or above grade level | 880 | 485 | 329 | 66 | † |
| 1 year below grade level | 1,377 | 688 | 611 | 78 | † |
| 2 or more years below grade level | 3,894 | 913 | 2,066 | 915 | † |
| Other | 1,677 | † | † | † | 1,677 |
| Proportion of excluded students | | | | | |
| Total | 0.34 | 0.23 | 0.38 | 0.62 | 0.39 |
| At or above grade level | 0.13 | 0.10 | 0.16 | 0.23 | † |
| 1 year below grade level | 0.28 | 0.26 | 0.29 | 0.38 | † |
| 2 or more years below grade level | 0.57 | 0.48 | 0.55 | 0.75 | † |
| Other | 0.39 | † | † | † | 0.39 |

*Note.* † = not applicable. The SD-all category includes students classified as SD and students classified as both SD and ELL. SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment. Authors' calculations.

**Table 5**

***Counts of Students Identified as ELL-Only, Assessed, and Excluded, and the Proportion of Excluded Students, Reading, Grade 4, by Years of Receiving Instruction in English and Grade Level of Instruction: 2003***

| Grade level of instruction in reading/language arts | Total | Years of receiving instruction in English: 4 or more years | 2 or 3 years | 1 year | Other |
|---|---|---|---|---|---|
| Identified | | | | | |
| Total | 11,888 | 4,246 | 2,415 | 1,504 | 3,723 |
| At or above grade level | 5,014 | 2,990 | 1,332 | 692 | † |
| 1 year below grade level | 1,644 | 806 | 584 | 254 | † |
| 2 or more years below grade level | 1,507 | 450 | 499 | 558 | † |
| Other | 3,723 | † | † | † | 3,723 |
| Assessed | | | | | |
| Total (regular and accommodated) | 9,040 | 3,993 | 1,860 | 769 | 2,418 |
| At or above grade level | 4,509 | 2,883 | 1,151 | 475 | † |
| 1 year below grade level | 1,287 | 726 | 442 | 119 | † |
| 2 or more years below grade level | 826 | 384 | 267 | 175 | † |
| Other | 2,418 | † | † | † | 2,418 |
| Excluded | | | | | |
| Total | 2,848 | 253 | 555 | 735 | 1,305 |
| At or above grade level | 505 | 107 | 181 | 217 | † |
| 1 year below grade level | 357 | 80 | 142 | 135 | † |
| 2 or more years below grade level | 681 | 66 | 232 | 383 | † |
| Other | 1,305 | † | † | † | 1,305 |
| Proportion of excluded students | | | | | |
| Total | 0.24 | 0.06 | 0.23 | 0.49 | 0.35 |
| At or above grade level | 0.10 | 0.04 | 0.14 | 0.31 | † |
| 1 year below grade level | 0.22 | 0.10 | 0.24 | 0.53 | † |
| 2 or more years below grade level | 0.45 | 0.15 | 0.46 | 0.69 | † |
| Other | 0.35 | † | † | † | 0.35 |

*Note.* † = not applicable. The ELL-only category includes students classified as ELL-only.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment. Authors' calculations.

**Table 6**

***Category-Specific Exclusion Rates for SD-All Students, Reading, Grade 4, by State: 2003***

| | Cross-categories of degree of disability and grade level of instruction | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| State[a] | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Alabama | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.17 | 0.32 | 0.65 | 0.18 |
| Alaska | 0.01 | 0.00 | 0.00 | 0.11 | 0.00 | 0.25 | 0.14 | 0.23 | 0.74 | 0.17 |
| Arizona | 0.08 | 0.41 | 0.50 | 0.38 | 0.39 | 1.00 | 0.86 | 0.69 | 0.70 | 0.36 |
| Arkansas | 0.18 | 0.27 | 0.29 | 0.28 | 0.25 | 0.00 | 0.45 | 0.57 | 0.81 | 0.51 |
| California | 0.04 | 0.07 | 0.07 | 0.20 | 0.11 | 0.23 | 0.37 | 0.51 | 0.77 | 0.26 |
| Colorado | 0.02 | 0.07 | 0.00 | 0.00 | 0.06 | 0.67 | 0.22 | 0.27 | 0.83 | 0.33 |
| Connecticut | 0.03 | 0.11 | 0.25 | 0.14 | 0.24 | 0.33 | 0.86 | 0.48 | 0.73 | 0.51 |
| Delaware | 0.38 | 0.54 | 0.60 | 0.64 | 0.76 | 0.50 | 0.79 | 0.89 | 0.95 | 0.59 |
| Florida | 0.10 | 0.02 | 0.14 | 0.05 | 0.06 | 0.00 | 0.44 | 0.27 | 0.47 | 0.36 |
| Georgia | 0.06 | 0.10 | 0.14 | 0.16 | 0.24 | 0.00 | 0.46 | 0.38 | 0.47 | 0.30 |
| Hawaii | 0.02 | 0.00 | 0.29 | 0.10 | 0.19 | 0.00 | 0.22 | 0.37 | 0.82 | 0.33 |
| Idaho | 0.02 | 0.03 | 0.00 | 0.16 | 0.18 | 0.00 | 0.33 | 0.40 | 0.71 | 0.30 |
| Illinois | 0.10 | 0.18 | 0.40 | 0.22 | 0.29 | 0.10 | 0.43 | 0.58 | 0.70 | 0.47 |
| Indiana | 0.13 | 0.16 | 0.00 | 0.31 | 0.19 | 0.00 | 0.49 | 0.61 | 0.70 | 0.33 |
| Kansas | 0.01 | 0.03 | 0.00 | 0.11 | 0.03 | 0.29 | 0.22 | 0.44 | 0.72 | 0.18 |
| Louisiana | 0.24 | 0.24 | 0.38 | 0.26 | 0.63 | 0.50 | 0.34 | 0.49 | 0.63 | 0.33 |
| Maine | 0.05 | 0.19 | 0.40 | 0.12 | 0.20 | 0.38 | 0.60 | 0.63 | 0.85 | 0.60 |
| Maryland | 0.05 | 0.23 | 0.50 | 0.34 | 0.53 | 1.00 | 0.65 | 0.81 | 0.74 | 0.59 |
| Massachusetts | 0.03 | 0.03 | 0.27 | 0.11 | 0.12 | 0.25 | 0.27 | 0.35 | 0.65 | 0.29 |
| Michigan | 0.23 | 0.29 | 0.38 | 0.48 | 0.44 | 0.10 | 0.82 | 0.81 | 0.87 | 0.60 |
| Minnesota | 0.04 | 0.04 | 0.17 | 0.04 | 0.14 | 0.33 | 0.14 | 0.43 | 0.67 | 0.34 |
| Mississippi | 0.29 | 0.13 | 0.00 | 0.76 | 0.88 | 0.00 | 0.90 | 0.96 | 1.00 | 0.27 |
| Missouri | 0.16 | 0.41 | 0.00 | 0.49 | 0.51 | 1.00 | 0.63 | 0.69 | 0.86 | 0.56 |
| Nevada | 0.05 | 0.10 | 0.00 | 0.31 | 0.33 | 0.60 | 0.63 | 0.62 | 0.82 | 0.35 |
| New Hampshire | 0.02 | 0.11 | 0.22 | 0.00 | 0.05 | 0.22 | 0.17 | 0.28 | 0.81 | 0.50 |
| New Jersey | 0.10 | 0.14 | 0.00 | 0.24 | 0.30 | 0.17 | 0.58 | 0.39 | 0.50 | 0.30 |
| New Mexico | 0.04 | 0.09 | 0.00 | 0.18 | 0.25 | 0.50 | 0.19 | 0.41 | 0.77 | 0.24 |
| New York | 0.13 | 0.31 | 0.00 | 0.30 | 0.24 | 0.25 | 0.24 | 0.57 | 0.68 | 0.34 |
| North Carolina | 0.17 | 0.16 | 0.25 | 0.22 | 0.32 | 0.33 | 0.60 | 0.54 | 0.70 | 0.46 |
| North Dakota | 0.05 | 0.21 | 0.36 | 0.37 | 0.43 | 0.33 | 0.36 | 0.53 | 0.61 | 0.25 |

*(Table continues)*

Table 6 (continues)

| State[a] | Cross-categories of degree of disability and grade level of instruction | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Ohio | 0.30 | 0.24 | 0.50 | 0.49 | 0.58 | 1.00 | 0.67 | 0.68 | 0.94 | 0.59 |
| Oregon | 0.09 | 0.06 | 0.38 | 0.29 | 0.24 | 0.00 | 0.58 | 0.62 | 0.79 | 0.48 |
| Rhode Island | 0.04 | 0.00 | 0.33 | 0.12 | 0.15 | 0.00 | 0.19 | 0.38 | 0.89 | 0.26 |
| South Carolina | 0.07 | 0.12 | 0.43 | 0.58 | 0.64 | 0.56 | 0.79 | 0.92 | 0.90 | 0.58 |
| Tennessee | 0.05 | 0.15 | 0.14 | 0.26 | 0.10 | 0.50 | 0.29 | 0.56 | 0.65 | 0.38 |
| Texas | 0.29 | 0.22 | 0.20 | 0.67 | 0.54 | 0.90 | 0.80 | 0.80 | 0.82 | 0.50 |
| Utah | 0.03 | 0.09 | 0.50 | 0.11 | 0.07 | 0.00 | 0.32 | 0.49 | 0.76 | 0.20 |
| Vermont | 0.00 | 0.04 | 0.00 | 0.26 | 0.16 | 0.23 | 0.55 | 0.72 | 0.91 | 0.53 |
| Virginia | 0.22 | 0.38 | 0.60 | 0.72 | 0.79 | 0.44 | 0.90 | 0.88 | 0.95 | 0.52 |
| Washington | 0.03 | 0.14 | 0.00 | 0.27 | 0.21 | 0.50 | 0.37 | 0.51 | 0.78 | 0.33 |
| Wisconsin | 0.06 | 0.08 | 0.00 | 0.09 | 0.27 | 0.33 | 0.59 | 0.52 | 0.62 | 0.59 |
| Wyoming | 0.02 | 0.00 | 0.00 | 0.02 | 0.19 | 0.00 | 0.10 | 0.22 | 0.33 | 0.21 |
| Summary statistics | | | | | | | | | | |
| Mean | 0.10 | 0.15 | 0.21 | 0.26 | 0.29 | 0.33 | 0.47 | 0.54 | 0.74 | 0.39 |
| Median | 0.05 | 0.11 | 0.18 | 0.23 | 0.24 | 0.27 | 0.44 | 0.52 | 0.75 | 0.35 |
| Standard deviation | 0.10 | 0.13 | 0.20 | 0.20 | 0.22 | 0.31 | 0.24 | 0.20 | 0.14 | 0.14 |
| Minimum | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.10 | 0.22 | 0.33 | 0.17 |
| Maximum | 0.38 | 0.54 | 0.60 | 0.76 | 0.88 | 1.00 | 0.90 | 0.96 | 1.00 | 0.60 |
| Correlation with $p_k$ | 0.77 | 0.74 | 0.47 | 0.89 | 0.84 | 0.48 | 0.86 | 0.93 | 0.63 | 0.70 |

*Note.* The SD-all category includes students classified as SD and students classified as both SD and ELL. $p_k$ is the actual exclusion rate of SD-all in state *k*. SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment. Authors' calculations.
[a]Forty-two states with the state achievement test score as a school-level sampling variable were included in the study.

**Table 7**

***Category-Specific Exclusion Rates for ELL-Only Students, Reading, Grade 4, by State: 2003***

| | Cross-categories of years of receiving instruction in English and grade level of instruction | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| State[a] | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Alabama | 0.40 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.67 | 0.50 |
| Alaska | 0.00 | 0.00 | 0.13 | 0.04 | 0.00 | 0.50 | 0.00 | 0.09 | 0.80 | 0.04 |
| Arizona | 0.04 | 0.09 | 0.33 | 0.08 | 0.33 | 0.67 | 0.00 | 0.21 | 0.69 | 0.17 |
| Arkansas | 0.05 | 0.22 | 0.00 | 0.25 | 0.50 | 1.00 | 1.00 | 0.67 | 0.67 | 0.67 |
| California | 0.01 | 0.05 | 0.26 | 0.01 | 0.16 | 0.50 | 0.01 | 0.17 | 0.37 | 0.15 |
| Colorado | 0.00 | 0.00 | 0.00 | 0.05 | 0.11 | 0.08 | 0.00 | 0.42 | 0.62 | 0.36 |
| Connecticut | 0.20 | 0.14 | 0.33 | 0.00 | 0.17 | 0.50 | 0.00 | 0.55 | 1.00 | 0.67 |
| Delaware | 0.00 | 0.40 | 0.67 | 0.00 | 0.38 | 1.00 | 0.00 | 0.40 | 1.00 | 0.60 |
| Florida | 0.00 | 0.11 | 0.35 | 0.00 | 0.08 | 0.83 | 0.00 | 0.27 | 0.82 | 0.23 |
| Georgia | 0.00 | 0.08 | 1.00 | 0.00 | 0.17 | 1.00 | 0.17 | 0.15 | 0.85 | 0.31 |
| Hawaii | 0.00 | 0.22 | 0.17 | 0.13 | 0.44 | 0.40 | 0.30 | 0.43 | 0.36 | 0.38 |
| Idaho | 0.01 | 0.00 | 0.25 | 0.03 | 0.38 | 0.50 | 0.09 | 0.50 | 1.00 | 0.36 |
| Illinois | 0.15 | 0.28 | 0.58 | 0.33 | 0.44 | 0.96 | 1.00 | 0.67 | 1.00 | 0.23 |
| Indiana | 0.00 | 0.00 | 0.67 | 0.00 | 0.00 | 1.00 | 0.00 | 0.50 | 1.00 | 0.31 |
| Kansas | 0.28 | 0.20 | 0.67 | 0.00 | 0.00 | 0.57 | 0.00 | 0.60 | 1.00 | 0.21 |
| Louisiana | 0.14 | 0.33 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.50 |
| Maine | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| Maryland | 0.17 | 0.33 | 0.73 | 0.56 | 0.38 | 0.50 | 0.29 | 0.57 | 1.00 | 0.40 |
| Massachusetts | 0.00 | 0.25 | 0.17 | 0.06 | 0.16 | 0.30 | 0.00 | 0.56 | 0.69 | 0.56 |
| Michigan | 0.01 | 0.00 | 0.80 | 0.20 | 0.50 | 0.75 | 0.21 | 0.63 | 1.00 | 0.31 |
| Minnesota | 0.00 | 0.17 | 0.00 | 0.00 | 0.22 | 1.00 | 0.00 | 0.64 | 0.77 | 0.13 |
| Mississippi | 1.00 | 0.50 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.25 |
| Missouri | 0.40 | 0.63 | 1.00 | 0.50 | 0.83 | 0.00 | 0.00 | 0.00 | 1.00 | 0.45 |
| Nevada | 0.05 | 0.32 | 0.43 | 0.14 | 0.32 | 0.80 | 0.36 | 0.74 | 0.51 | 0.30 |
| New Hampshire | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 1.00 | 0.75 | 0.67 | 0.92 | 0.19 |
| New Jersey | 0.27 | 0.24 | 0.45 | 0.00 | 0.00 | 0.87 | 0.67 | 0.40 | 0.83 | 0.62 |
| New Mexico | 0.04 | 0.08 | 0.12 | 0.08 | 0.36 | 0.08 | 0.12 | 0.75 | 0.31 | 0.24 |
| New York | 0.26 | 0.38 | 0.73 | 0.53 | 0.65 | 0.60 | 0.50 | 1.00 | 0.96 | 0.56 |
| North Carolina | 0.15 | 0.14 | 0.45 | 0.00 | 0.38 | 1.00 | 0.00 | 0.50 | 0.88 | 0.34 |
| North Dakota | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.05 |

*(Table continues)*

Table 7 (continued)

| State[a] | Cross-categories of years of receiving instruction in English and grade level of instruction | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Ohio | 0.33 | 0.22 | 0.17 | 0.00 | 0.40 | 0.50 | 0.00 | 0.67 | 1.00 | 0.57 |
| Oregon | 0.04 | 0.04 | 0.33 | 0.10 | 0.00 | 0.83 | 0.37 | 0.52 | 0.78 | 0.35 |
| Rhode Island | 0.00 | 0.25 | 0.60 | 0.17 | 0.14 | 0.00 | 0.00 | 0.33 | 0.61 | 0.41 |
| South Carolina | 0.00 | 0.75 | 0.60 | 0.00 | 0.50 | 0.00 | 0.00 | 0.00 | 0.75 | 0.38 |
| Tennessee | 0.00 | 0.00 | 0.88 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.25 |
| Texas | 0.07 | 0.16 | 0.24 | 0.10 | 0.08 | 0.08 | 0.00 | 0.71 | 0.27 | 0.80 |
| Utah | 0.02 | 0.04 | 0.43 | 0.09 | 0.09 | 1.00 | 0.26 | 0.43 | 1.00 | 0.29 |
| Vermont | 0.00 | 0.40 | 0.00 | 0.09 | 0.20 | 0.00 | 0.00 | 1.00 | 0.67 | 0.25 |
| Virginia | 0.14 | 0.27 | 0.50 | 0.53 | 0.77 | 1.00 | 0.38 | 0.95 | 1.00 | 0.35 |
| Washington | 0.00 | 0.07 | 0.20 | 0.03 | 0.13 | 0.33 | 0.22 | 0.33 | 0.58 | 0.27 |
| Wisconsin | 0.20 | 0.05 | 0.08 | 0.31 | 0.18 | 0.50 | 0.00 | 0.57 | 0.67 | 0.57 |
| Wyoming | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.00 | 1.00 | 0.80 | 0.14 |
| Summary statistics | | | | | | | | | | |
| Mean | 0.11 | 0.19 | 0.39 | 0.10 | 0.27 | 0.53 | 0.16 | 0.49 | 0.77 | 0.35 |
| Median | 0.02 | 0.15 | 0.33 | 0.03 | 0.17 | 0.50 | 0.00 | 0.51 | 0.81 | 0.33 |
| Standard deviation | 0.18 | 0.19 | 0.33 | 0.16 | 0.28 | 0.39 | 0.27 | 0.31 | 0.25 | 0.18 |
| Minimum | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Maximum | 1.00 | 0.75 | 1.00 | 0.56 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.80 |
| Correlation with $p_k$ | 0.62 | 0.74 | 0.37 | 0.43 | 0.32 | 0.05 | 0.19 | 0.08 | 0.12 | 0.70 |

*Note.* The ELL-only category includes students classified as ELL-only. $p_k$ is the actual exclusion rate of ELL-only in state *k*. SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment. Authors' calculations.

[a] Forty-two states with the state achievement test score as a school-level sampling variable were included in the study.

For example, Minnesota has an SD-all exclusion rate of 0.22 and Michigan an SD-all exclusion rate of 0.57. Suppose the SD-all students in Michigan are more likely to possess characteristics that place them in categories with typically high exclusion rates while SD-all students in Minnesota are more likely to possess characteristics that place them in categories with typically low exclusion rates. Then these differences could account for the discrepancy in overall SD-all exclusion rates of $0.33 = 0.57 - 0.22$. Of course, an analogous scenario could pertain to ELL-only students.

This reasoning leads naturally to consideration of indirect standardization as a diagnostic tool (Mosteller & Tukey, 1977). Typically, with indirect standardization, there are a number of units of interest—states in our case. The population in each unit is stratified with respect to one or more characteristics. In our situation, there are 2 characteristics leading to 10 strata (categories). A set of standard category-specific rates are somehow obtained. (Here, these will be category-specific exclusion rates pooled over states.) Then, for each unit, these standard rates are applied to the population in each category, eventually yielding what is termed an indirectly standardized exclusion rate for the unit. It is important to note that if the category-specific rates in each unit are generally close to the chosen standard rates, then the indirectly standardized rates will be close to the observed rates.

The observed aggregate state exclusion rate is a weighted average of the category-specific exclusion rates (in that state), with the weights being the proportions of the sample (in that state) falling in the different categories. As indicated above, indirect standardization requires that we replace the category-specific exclusion rates for the state by a set of standard or pooled exclusion rates, derived from the experience of all the states. These pooled exclusion rates are then combined with the same weights as before to yield an indirectly standardized overall exclusion rate for the state. Differences among states in these indirectly standardized rates are entirely due to differences in the distributions of student characteristics in the state samples. To the extent that the indirectly standardized rates track the observed rates, the scenario described above offers an explanation for the data in Table 1.

There are a number of ways to obtain standardized category-specific exclusion rates. One would be to simply compute for each category the average exclusion rate across states. Another would be to compute a ratio estimator of the aggregate rate. We tried both and obtained very similar results. In the interest of brevity, we only present the latter approach. (Note that since no state contributes a disproportionate amount of data, any reasonable standardized rate will do.) The calculation is carried out separately for SD-all and ELL-only students.

Let $k$ index states and $i$ index categories in the classification of students (either SD all or ELL-only). Further, let

$n_{ik}$ = the number of SD-all or ELL-only students in category $i$ of state $k$,

$m_{ik}$ = the number of excluded SD-all or ELL-only students in category $i$ of state $k$,

$N_k = \sum_i n_{ik}$, and

$M_k = \sum_i m_{ik}$.

Then the aggregate observed state exclusion rate for that group of students is $p_k = M_k / N_k$. But this rate can also be expressed as

$$p_k = \frac{\sum_i p_{ik} n_{ik}}{N_k},$$

where $p_{ik} = m_{ik} / n_{ik}$.

Now let

$$\lambda_i = \frac{\sum_k m_{ik}}{\sum_k n_{ik}}.$$

Then, $\lambda_i$ is a standardized (or pooled) rate for category $i$. We define the aggregate indirect standardization exclusion rate for state $k$ as

$$\tilde{p}_k = \frac{\sum_i \lambda_i n_{ik}}{N_k}.$$

Interest centers on comparing $\{p_k\}$ and $\{\tilde{p}_k\}$. The results for SD-all and ELL-only students are presented in Table 1, columns 4 and 5 and columns 7 and 8, respectively. The corresponding scatter-plots are found in Figures 1 and 2.

From the summary statistics in the table, as well as the scatter-plots, it is evident that the indirectly standardized exclusion rates are much less variable than are the original rates. Recall that for SD-all students, the observed exclusion rates range from 0.12 to 0.64; however, the indirectly standardized rates range from 0.26 to 0.39. The ratio of the interquartile ranges is

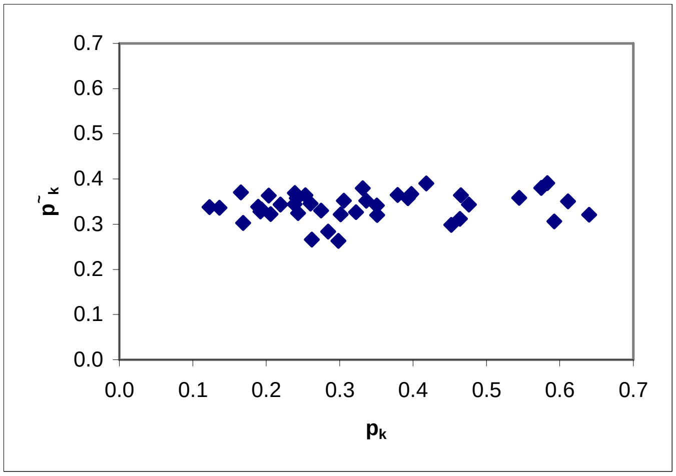


***Figure 1.* Plot of indirectly standardized exclusion rates $\tilde{p}_k$ vs. observed exclusion rates $p_k$ for 42 states, SD-all students, Reading, Grade 4: 2003.**

*Note.* The SD-all category includes students classified as SD and students classified as both SD and ELL. SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment. Authors' calculations.

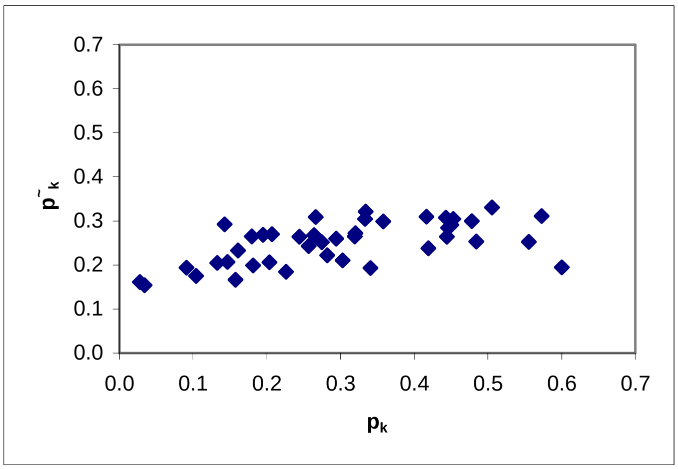


***Figure 2.* Plot of indirectly standardized exclusion rates $\tilde{p}_k$ vs. observed exclusion rates $p_k$ for 42 states, ELL-only students, Reading, Grade 4: 2003.**

*Note.* The ELL-only category includes students classified as ELL-only. SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment. Authors' calculations.

0.21/0.04 = 5.2.[8] For ELL-only students, the observed exclusion rates range from 0.03 to 0.60; however, the indirectly standardized rates range from 0.15 to 0.33. The ratio of the interquartile ranges is 0.26/0.08 = 3.2. These ratios indicate that states are more homogeneous with respect to the characteristics of their SD students than of their ELL students.[9] This accounts, at least in part, for the fact that the correlation between the observed and indirectly standardized rates is 0.17 for the SD-all group and 0.58 for the ELL-only group. With regard to the latter, it is evident that states with low observed exclusion rates also tend to have relatively low indirectly standardized rates.

In the present context, the variability in the standardized rates is more critical than the correlation with the original rates. The substantially reduced heterogeneity among indirectly standardized rates means that differences among states in the characteristics of their SD-all or ELL-only students can only account for a small part of the differences in the aggregate rates. That is, for both SD-all and ELL-only groups, differences in category-specific exclusion rates appear to be the major contributor to the heterogeneity among states in aggregate exclusion rates. This conclusion is further supported by inspection of each column in Tables 6 and 7. There is substantial variability across states in the category-specific exclusion rates. This impression is supported by examination of the minimum, maximum, and standard deviation for each column. (Admittedly, some of the variation is a consequence of the small sample sizes in some of the categories for some states.) In sum, we cannot account for the differences in state exclusion rates by appealing to the differences in their SD/ELL student populations.[10]

A follow-up analysis can shed light on the practical import of the heterogeneity across states in the category-specific exclusion rates. For example, suppose that states' category-specific exclusion rates tend to differ most from the pooled exclusion rates for categories with small numbers of students. Then the difference between (a) the actual number of excluded students in the state and (b) the number that would have been excluded had the indirectly standardized rates been in force would be relatively small. Although this is an unlikely scenario, it cannot be entirely ruled out.

Accordingly, let $E_{ik} = m_{ik} - \lambda_i n_{ik}$. Then $E_{ik}$ is the difference between the actual and the expected number excluded in category *i* of state *k*, where the expectation is based on the indirectly standardized rate. Then define

$$Q_k = \frac{\sum_i |E_{ik}|}{\sum_i m_{ik}},$$

which provides a relative measure of the difference in exclusions for state $k$ under the two scenarios. (We assume that at least one of the $m_{ik}$ is different from 0.) Clearly, the statistic $Q$ has a lower limit of 0, which is obtained when the actual and expected numbers are identical in every category. The upper limit is finite and depends in a complicated way on the patterns of counts. Typically, the values of $Q$ are less than 1. Table 8 presents the results for both the SD-all and ELL-only groups. We note that for the former, the median value of $\{Q_k\}$ is 0.34. That is, for half of the 42 states the $Q_k$ exceeds 0.34 (i.e., the relative impact is at least one-third of the observed number of exclusions). For the latter, the median value of $\{Q_k\}$ is 0.47. That is, for half of the 42 states the relative impact is almost one-half or greater of the observed number of exclusions. We conclude that the heterogeneity across states in the category-specific exclusion rates is both large and substantively important: That is, the departures (in numbers of students excluded) from what one would expect if states had homogeneous category-specific exclusion rates are serious and merit consideration.

## 3. Simulation

### *3.1 Data Source*

Data used in this simulation are drawn from the 2003 NAEP Grade 4 Reading Assessment. The analyses that follow were carried out separately for the two groups, SD-all and ELL-only, as McLaughlin did before 2005. Forty-two states with the state achievement test score as a school-level sampling variable were included in this study. The basic data elements have already been presented in Table 1.

There are different ways to get a complete sample for a simulation study. Wise et al. (2006) created a complete sample from the 2003 NAEP Grade 4 Reading Assessment by filling in missing values, PV, and background information, for all excluded students using a hot-deck procedure. Then, a systematic random sample (with unequal probabilities of selection) was drawn to identify those students to be excluded in order to obtain a set of simulated data. The selection process was designed to yield simulated data that had a similar missing-value pattern to the original data set. Finally, an FPE method was used to fill in the missing values for the simulated data. The results (e.g., mean scores and their SEs or percent proficient) from the imputed-complete data were compared with the corresponding values from the complete data set constructed by Wise et al. (2006).

**Table 8**

***Relative Differences Between Actual Numbers of SD-All and ELL-Only Students Excluded and the Numbers Expected Under Uniform Category-Specific Exclusion Rates $Q_k$, Reading, Grade 4, by State: 2003***

| State [a] | SD-all | ELL-only | State [a] | SD-all | ELL-only |
|---|---|---|---|---|---|
| Alabama | 1.24 | 0.48 | Mississippi | 0.51 | 0.71 |
| Alaska | 1.47 | 3.61 | Missouri | 0.33 | 0.68 |
| Arizona | 0.27 | 0.53 | Nevada | 0.17 | 0.32 |
| Arkansas | 0.14 | 0.46 | New Hampshire | 0.75 | 0.54 |
| California | 0.34 | 1.13 | New Jersey | 0.25 | 0.47 |
| Colorado | 0.91 | 0.50 | New Mexico | 0.45 | 0.52 |
| Connecticut | 0.26 | 0.37 | New York | 0.20 | 0.46 |
| Delaware | 0.50 | 0.44 | North Carolina | 0.14 | 0.22 |
| Florida | 0.72 | 0.47 | North Dakota | 0.29 | 4.81 |
| Georgia | 0.33 | 0.34 | Ohio | 0.33 | 0.39 |
| Hawaii | 0.59 | 0.26 | Oregon | 0.15 | 0.27 |
| Idaho | 0.48 | 0.34 | Rhode Island | 0.83 | 0.39 |
| Illinois | 0.12 | 0.49 | South Carolina | 0.41 | 0.44 |
| Indiana | 0.15 | 0.40 | Tennessee | 0.16 | 0.65 |
| Kansas | 0.79 | 0.52 | Texas | 0.34 | 0.56 |
| Louisiana | 0.35 | 0.50 | Utah | 0.49 | 0.31 |
| Maine | 0.28 | 2.12 | Vermont | 0.34 | 0.42 |
| Maryland | 0.32 | 0.33 | Virginia | 0.48 | 0.48 |
| Massachusetts | 0.71 | 0.35 | Washington | 0.17 | 0.41 |
| Michigan | 0.36 | 0.45 | Wisconsin | 0.25 | 0.53 |
| Minnesota | 0.57 | 0.82 | Wyoming | 1.75 | 1.01 |

*Note.* The SD-all category includes students classified as SD and students classified as both SD and ELL. The ELL-only category includes students classified as ELL-only. SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment. Authors' calculations.

[a] Forty-two states with the state achievement test score as a school-level sampling variable were included in the study.

Another approach is to treat the students who actually took the NAEP assessment as the complete data set and then, according to some design, exclude some of those students by deleting their cognitive data. The remaining students then constitute the set of assessed students for purposes of the simulation. This strategy avoids having to fill in missing data as a first step and, consequently, obviates the question of to what degree do the simulation results depend on how the missing data are constructed. It also has the advantage that the PPV generated can be compared with the corresponding actual PV. As we shall see, this capability can yield new insights into the operating characteristics of FPEs. The obvious disadvantage, of course, is that the size of the assessed sample is smaller than the one created by Wise et al. (2006).

For this study, we use a deletion mechanism corresponding to Condition 1 of Wise et al. (2006).[11] Since the purpose of our simulation study is to carry out a preliminary evaluation of the effectiveness of the FPE method, this approach should work well enough. To recap, the data of assessed students from the 2003 NAEP Grade 4 Reading Assessment are treated as a complete data set in our simulation study.

The exclusion rates of SD-all and ELL-only samples for each state are presented (again) in the columns 2 and 4 of Table 9. The maximum rate among these 42 states is 0.64 (Delaware) and the minimum is 0.12 (Wyoming) for the SD-all students, while for the ELL-only students, the maximum rate is 0.60 (Missouri) and the minimum is 0.03 (Alaska, North Dakota). In order to prevent the simulated sample of included students from becoming too small, we kept the minimum rate at 0.12 and 0.03 and reduced the maximum rate from 0.64 and 0.60 to 0.50 for SD-all and ELL-only students, respectively. Fixing the two points for the SD-all or the ELL-only group, a linear equation passing through these two points was established and the simulation exclusion rates for the 42 states were obtained. These simulation exclusion rates of SD-all and ELL-only students are listed in the third and the last columns of Table 9, respectively. The simulation exclusion rate of a state multiplies the number of SD-all (or ELL-only) students in that state in the simulation to obtain the target number of excluded SD-all (or ELL-only) students for that state. The simulation target number of excluded students is denoted as $m_k$.

As discussed in the previous section, for each group (ELL-only or SD-all), two student characteristics that were strongly correlated with exclusion rates within states were selected. The category-specific exclusion rates by state for the 10 categories formed from these two pairs of student characteristics for the SD-all and ELL-only students have been presented in Tables 6 and 7.

**Table 9**

***Observed and Simulation Exclusion Rates for SD-All and ELL-Only Students, Reading, Grade 4, by State: 2003***

| | SD-all | | ELL-only | |
|---|---|---|---|---|
| State[a] | Observed exclusion rate | Simulation exclusion rate | Observed exclusion rate | Simulation exclusion rate |
| Alabama | 0.17 | 0.15 | 0.42 | 0.35 |
| Alaska | 0.14 | 0.13 | 0.03 | 0.03 |
| Arizona | 0.47 | 0.37 | 0.16 | 0.14 |
| Arkansas | 0.39 | 0.32 | 0.34 | 0.29 |
| California | 0.26 | 0.22 | 0.09 | 0.08 |
| Colorado | 0.20 | 0.18 | 0.18 | 0.15 |
| Connecticut | 0.30 | 0.25 | 0.44 | 0.37 |
| Delaware | 0.64 | 0.50 | 0.45 | 0.37 |
| Florida | 0.19 | 0.17 | 0.21 | 0.18 |
| Georgia | 0.24 | 0.21 | 0.27 | 0.22 |
| Hawaii | 0.24 | 0.21 | 0.26 | 0.22 |
| Idaho | 0.24 | 0.21 | 0.16 | 0.13 |
| Illinois | 0.35 | 0.29 | 0.33 | 0.28 |
| Indiana | 0.28 | 0.24 | 0.20 | 0.17 |
| Kansas | 0.19 | 0.17 | 0.29 | 0.25 |
| Louisiana | 0.30 | 0.25 | 0.42 | 0.35 |
| Maine | 0.39 | 0.32 | 0.14 | 0.12 |
| Maryland | 0.48 | 0.38 | 0.45 | 0.38 |
| Massachusetts | 0.19 | 0.17 | 0.36 | 0.30 |
| Michigan | 0.57 | 0.45 | 0.23 | 0.19 |
| Minnesota | 0.22 | 0.19 | 0.13 | 0.11 |
| Mississippi | 0.61 | 0.48 | 0.56 | 0.46 |
| Missouri | 0.46 | 0.37 | 0.60 | 0.50 |
| Nevada | 0.42 | 0.34 | 0.32 | 0.27 |
| New Hampshire | 0.21 | 0.18 | 0.33 | 0.28 |
| New Jersey | 0.27 | 0.23 | 0.45 | 0.38 |
| New Mexico | 0.25 | 0.22 | 0.15 | 0.13 |
| New York | 0.32 | 0.27 | 0.57 | 0.48 |

*(Table continues)*

Table 9 (continued)

| State[a] | SD-all Observed exclusion rate | SD-all Simulation exclusion rate | ELL-only Observed exclusion rate | ELL-only Simulation exclusion rate |
|---|---|---|---|---|
| North Carolina | 0.35 | 0.29 | 0.32 | 0.27 |
| North Dakota | 0.26 | 0.22 | 0.03 | 0.03 |
| Ohio | 0.58 | 0.46 | 0.51 | 0.42 |
| Oregon | 0.40 | 0.32 | 0.26 | 0.22 |
| Rhode Island | 0.17 | 0.16 | 0.24 | 0.21 |
| South Carolina | 0.45 | 0.36 | 0.44 | 0.37 |
| Tennessee | 0.31 | 0.26 | 0.27 | 0.23 |
| Texas | 0.54 | 0.43 | 0.48 | 0.40 |
| Utah | 0.24 | 0.21 | 0.18 | 0.15 |
| Vermont | 0.38 | 0.31 | 0.28 | 0.24 |
| Virginia | 0.59 | 0.47 | 0.48 | 0.40 |
| Washington | 0.33 | 0.27 | 0.19 | 0.17 |
| Wisconsin | 0.34 | 0.28 | 0.30 | 0.25 |
| Wyoming | 0.12 | 0.12 | 0.10 | 0.09 |

*Note.* The SD-all category includes students classified as SD and students classified as both SD and ELL. The ELL-only category includes students classified as ELL-only. SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment. Authors' calculations.

[a]Forty-two states with the state achievement test score as a school-level sampling variable were included in the study.

Tables 4 and 5 present the counts of assessed and excluded students in these 10 categories and the proportions of excluded students for the state-aggregate SD-all sample and the state-aggregate ELL-only sample, respectively.

Recall that $p_{ik}$ denotes the category-specific exclusion rate given category $i$ in state $k$, $i = 1, 2, \ldots, 10$ and $k = 1, 2, \ldots, 42$. Typically, the entries in the table of { $p_{ik}$, $i = 1, 2, \ldots, 10$ } are quite heterogeneous. Some of the variation is due to small sample fluctuations. In order to generate more stable estimates, we employed an empirical Bayes-type approach. Specifically, we smoothed these category-specific exclusion rates by using $p_k$, the observed SD-all or ELL-only exclusion rate in state $k$ (i.e., column 5 or column 10 of Table 1), as follows: Let

$$\alpha_{ik} = \frac{(p_{ik} - p_k)^2}{\sum_{i=1}^{10} (p_{ik} - p_k)^2}.$$

Then, truncate $\alpha_{ik}$ at $0.5$ from below. That is, if $\alpha_{ik} < 0.5$, let $\alpha_{ik} = 0.5$. (Thus, the final $\alpha_{ik}$ is at least $0.5$, which will constrain the amount that the category-specific exclusion rate can be shifted toward overall exclusion rate.) The smoothed category-specific exclusion rate is defined as

$$\hat{p}_{ik} = \alpha_{ik} p_{ik} + (1 - \alpha_{ik}) p_k.$$

Simulated data are then generated such that the number of simulated excluded students in state $k$ is $m_k$ (the target number), and the probability of exclusion of an SD or ELL student in category $i$ is proportional to $\hat{p}_{ik}$, defined above. Following the methodology of Wise et al. (2006), we generate four replicate data files for the simulation, with the replicates consisting of systematic random samples designed to have minimal overlap in the sets of excluded students.

The key steps are as follows:

- Each student in state $k$ is placed in a particular category based on her/his characteristics. Let $j$ index students and $\hat{p}_{ik}(j)$ denote the exclusion probability for the category associated with student $j$. For each state, calculate the sum of the probabilities of exclusion across SD-all or ELL-only students and then divide by the target number of excluded SD-all or ELL-only students for that state, which is the sampling threshold, denoted as $\theta_k$. That is,

$$\theta_k = \frac{\sum_j \hat{p}_{ik}(j)}{m_k}.$$

The sampling threshold determines the step size employed in systematic sampling.

- Four starting values are selected, say 1/8, 3/8, 5/8, and 7/8, one for each of the four replications. The starting value is denoted as $\theta_0$.

- For each state, the SD-all students or the ELL-only students are sorted by race, accommodation, and $\hat{p}_{ik}(j)$, respectively. The following $m_k$ (possibly $\pm 1$) students are then excluded: Student $j$ is excluded if, and only if,

$$\theta_0 + \sum_{l=1}^{j-1} \hat{p}_{ik}(l) < m\theta_k < \theta_0 + \sum_{l=1}^{j} \hat{p}_{ik}(l)$$

for $m = 1, 2, \ldots, m_k$, where $m_k$ is the target number of simulated excluded students in state $k$. Thus, we obtained four replicate data files, with simulated included and excluded students for each state. This process parallels the procedure followed by Wise et al. (2006), which contains further details on the selection of a systematic random sample with unequal probabilities of selection.

### *3.2 Method*

In order to obtain an FPE, we establish a linear regression model that relates the PV of included SD or ELL students to a number of student characteristics to generate (predict) PPV for the excluded students. The process has two phases.

*Phase 1: Variable selection for regression models.* In this simulation, complete data (i.e., all assessed SD-all or ELL-only students in the original NAEP sample) were used to select a common collection of predictors for the linear regression models for all four replicates. The first columns of Tables 10 and 11 present the 19 characteristics available for the models for SD-all and ELL-only students, respectively. These characteristics are a combination of those available in advance of sample selection and those obtained for each classified student. Each discrete characteristic generates a set of dummy variables as predictors in the linear regression model. These predictors are centered separately within each state, so that they can be pooled across states when fitting a regression model. The school-level achievement score is also standardized separately within each state.

The dependent variable in the model is the deviation of the student's PV from the corresponding unweighted mean of the PV of the students in the corresponding group (SD- all or ELL-only) in the state. Since there are five PV generated for each student, a regression analysis can be replicated over five sets of PV, where a set denotes a full sample of students (SD-all or ELL-only, pooled over states) together with one plausible value for each. A backward selection method is used to determine which of the 19 sets of variables should be included in the model. Note that we either keep or delete all dummy variables associated with one discrete characteristic in the regression model. When deleting stepwise a set of dummy variables, we use a maxmin rule. That is, we delete a set of dummy variables associated with a particular characteristic if all these dummy variables are not significant and their minimum $p$-value is the maximum among all sets of dummy variables that are not significant.

The last five columns of Tables 10 and 11 indicate whether or not the variables are retained in the corresponding regression model, based on each of the five sets of PV. For the final regression model used to generate the PPV, predictors are discarded if the corresponding coefficients are statistically significant in two or fewer of the five regression models. For the SD-all case, we deleted three variables: participation in the same curriculum content as nondisabled students receiving the same grade level of instruction in language arts, percent Hispanic, and percent American Indian. For the ELL-only case, we deleted two variables: school enrollment and percent Asian.

*Phase 2: Fitting a model.* After selecting the common set of independent variables (i.e., the predictors), we independently estimate the regression coefficients for each of the five sets of PV for each of four replicates. Thus, there are 20 sets of estimated regression coefficients in all.

Let $x_{jk}$ be the vector of selected characteristics of excluded student $j$ in state $k$ for a fixed replicate. Let $b$ be the vector of estimated regression coefficients from a particular set of PV within a replicate. Then, the prediction from the model is

$$\tilde{y}_{jk} = \bar{y}_k + b' x_{jk},$$

where $\bar{y}_k$ is the unweighted mean of (simulated) included SD-all or ELL-only students in state $k$. The PPV of excluded students are generated according to

$$\hat{y}_{jk} = \tilde{y}_{jk} + \varepsilon_{jk},$$

where $\varepsilon_{jk} \sim N(0, V_{jk})$ and $V_{jk} = V_k^{(1)} + V_k^{(2)} + V_{jk}^{(3)}$ is intended to capture the full uncertainty associated with generating PPV.[12] The $\{\varepsilon_{jk}\}$ are generated independently for each $j$ and $k$, for each PPV, and across four replications.

**Table 10**

***Variables Retained in Regression Models for Each Set of Plausible Values for SD-All, Reading, Grade 4: 2003***

| | | Model selection | | | | |
|---|---|---|---|---|---|---|
| Variable | Description | PV 1 | PV 2 | PV 3 | PV 4 | PV 5 |
| x012101 | Student's primary disability | √ | √ | √ | √ | √ |
| x012201 | Degree of student's disability | √ | √ | √ | √ | √ |
| x015101 | Grade level student receiving in reading/language arts | √ | √ | √ | √ | √ |
| x015201 | Participation in the same curriculum as nondisabled reading/language arts | √ | √ | × | × | × |
| x013001 | Adaptation used for achievement testing | √ | √ | √ | √ | √ |
| slunch1 | National school lunch eligibility | √ | √ | √ | √ | √ |
| tol3 | Type of location (3 categories) | √ | √ | √ | √ | √ |
| senrol4 | School enrollment | √ | × | √ | √ | √ |
| lep | Limited English proficiency | √ | √ | √ | √ | √ |
| title1 | Receiving Title I funding | √ | √ | √ | √ | √ |
| srace | Race/ethnicity (from school records) | √ | √ | √ | √ | √ |
| pctasn | Percent Asian | √ | × | √ | √ | × |
| pctblk | Percent Black | √ | √ | √ | √ | √ |
| pcthsp | Percent Hispanic | × | × | × | × | × |
| pctind | Percent American Indian/Alaska Native | × | × | × | × | × |
| accom2 | Accommodated | √ | √ | √ | √ | √ |
| new_x[a] | Missing category for achvmed variable | √ | √ | √ | √ | √ |
| achvmed | Achievement or median income | √ | √ | √ | √ | √ |
| Dsex | Gender | √ | √ | √ | √ | √ |

*Note.* The SD-all category includes students classified as SD and students classified as both SD and ELL. PV = plausible value. √ = variable retained in regression model for the corresponding set of PV. × = variable not retained in regression model for the corresponding set of PV.

**Table 11**

***Variables Retained in Regression Models for Each of the Plausible Values for ELL-Only, Reading, Grade 4: 2003***

| | | Model selection | | | | |
|---|---|---|---|---|---|---|
| Variable | Description | PV 1 | PV 2 | PV 3 | PV 4 | PV 5 |
| x014201 | This year percent academic instruction native language | √ | √ | √ | √ | √ |
| x015601 | Years receiving academic instruction in English | √ | √ | √ | √ | √ |
| x015701 | Grade level receiving reading/language arts | √ | √ | √ | √ | √ |
| x015901 | How participate in NAEP reading/language arts | √ | √ | √ | √ | √ |
| x013801 | Student's first or native language | √ | √ | √ | √ | × |
| b018201 | Language other than English spoken in home | √ | √ | √ | √ | √ |
| slunch1 | National school lunch eligibility | √ | √ | √ | √ | √ |
| tol3 | Type of location (3 categories) | √ | × | √ | √ | √ |
| senrol4 | School enrollment | × | × | √ | √ | × |
| srace | Race/ethnicity (from school records) | √ | √ | √ | √ | √ |
| pctasn | Percent Asian | × | × | × | × | × |
| pctblk | Percent Black | √ | √ | √ | √ | √ |
| pcthsp | Percent Hispanic | √ | √ | √ | √ | √ |
| pctind | Percent American Indian/Alaska Native | √ | √ | √ | √ | √ |
| accom2 | Accommodated | √ | √ | √ | √ | √ |
| new_x[a] | Missing category for achvmed variable | √ | √ | √ | √ | √ |
| achvmed | Achievement or median income | √ | √ | √ | √ | √ |
| title1 | Receiving Title I funding | √ | √ | √ | √ | √ |
| Dsex | Gender | √ | √ | √ | √ | √ |

*Note.* The ELL-only category includes students classified as ELL-only. PV = plausible value. √ = variable retained in regression model for the corresponding set of PV. × = variable not retained in regression model for the corresponding set of PV.

[a]Indicator of missing values in the achvmed variable.

Here $V_k^{(1)}$ is the ordinary NAEP estimate of the variance of the mean of included SD-all or ELL-only students in state $k$, $V_k^{(2)}$ is the MSE of included SD-all or ELL-only students in state $k$ based on the estimated regression model, and $V_{jk}^{(3)}$ is the estimated variance of $\tilde{y}_{jk}$. Specifically,

$$V_k^{(2)} = \frac{1}{n_k} \sum_{i=1}^{n_k} (y_{ik} - \tilde{y}_{ik})^2,$$

where $n_k$ is the number of included SD-all or ELL-only students in state $k$, $y_{ik}$ is a PV of an included SD or ELL student in state $k$, and $\tilde{y}_{ik} = \bar{y}_k + b' x_{ik}$. (Note: Here $i$ is used to index included SD-all or ELL-only students while $j$ is used to index excluded SD-all or ELL-only students. That is, these indices are different from those in the last section.) Finally,

$$V_{jk}^{(3)} = s^2 x'_{jk} (Z'WZ)^{-1} x_{jk},$$

where $s^2$ is the residual mean square from the regression based on the state-aggregate sample (i.e., centered data for each state, pooled across states), $Z$ is the design matrix of the corresponding aggregate regression, $W = \text{diag}(w_1, \ldots, w_n)$, $w_l$ is the weight of student $l$ ($l = 1, \ldots, n$), and $n$ is the number of included SD-all or ELL-only students in the state-aggregate sample.

Once the PPV are generated, they are inserted into the data base, which is now complete; that is, all students have entries in the five columns designated for plausible values. Thus, all derived statistics can be calculated using standard NAEP methods employing PV for assessed students and PPV for excluded students.

### *3.3 Results*

There are three kinds of results (e.g., means and variances) for the simulation data. They are obtained by using three different methods, labeled as target, NAEP-like, and FPE in this report.

- Target results are based on the PV of the complete data.

- NAEP-like results are based on the current NAEP procedure, which employs the PV of assessed students only.
- FPE results are based on both the PV of assessed students and the PPV of excluded students.

The following tables present results for each of the four replicates for all 42 states in the simulation. Tables 12 and 13 display the target mean, the bias in the FPE method, and the bias in the NAEP-like method for SD-all and ELL-only students, respectively. Recall that the bias is calculated with respect to the target mean, which differs from the estimand of the NAEP-like method. The results for SD-all students in Table 12 indicate that, for most states, the bias of the FPE tends to be small (i.e., less than one scale score point) and takes on both positive and negative values across replications. By contrast, the bias of the NAEP-like estimator tends to be large (i.e., greater than one scale score point) and, for the most part, takes positive values. This is consistent with expectation, since excluded students typically have lower PV than assessed students. The results for ELL-only students in Table 13 follow a similar pattern, except that the pattern in signs is somewhat less clear-cut and the magnitude of the biases is greater. Both of these findings can be ascribed to the smaller sample sizes in the ELL analysis. For those states with especially small numbers of ELL students, the results can be quite exotic.

Tables 14 and 15 present for SD-all and ELL-only students, respectively, the target variance components for the mean estimator based on the complete data, and the differences between the FPE variance components and the corresponding target variance components. For the SD-all data, we observe that the estimated variance component for measurement error of the FPE tends to be greater (and often substantially greater) than the corresponding estimate from the complete data. On the other hand, the estimated variance component for sampling error of the FPE tends to be smaller than the corresponding estimate from the complete data. The results for ELL-only students in Table 14 follow a similar pattern, although the magnitudes of the differences are substantially greater.

Tables 16 and 17 display the variances of the estimators based on the complete data (denoted as the target variances), the differences between the FPE mean squared errors (MSE) and the target variances, and the differences between the NAEP-like MSE and the target variances for SD-all and ELL-only students, respectively. To obtain the results displayed in Tables 16 and 17, for each method the square of the bias and the total variance were combined for each replication in each state. Then the difference between the MSE of the method and the estimated variance based on the complete data was computed. For both methods, the MSE is almost always greater than the variance, and this is reflected in the columns that present the average differences across replications for each state. Notably, for nearly all states, the NAEP-like method performs more poorly than the FPE, often by a substantial amount. Again, the patterns are similar for both groups of students with the magnitudes of the excess in MSE greater for ELL-only students than for SD-all students.

Tables 18, 19, and 20 are the results based on what constitutes the complete sample in each state (i.e., regular students and classified students combined) for purposes of the simulation. Table 18 presents the bias in the FPE method and the bias in the NAEP-like method. Table 19 displays the variance components of the target state mean and the differences between the variance components of the FPE mean of each simulation replicate and the corresponding target variance components. Table 20 presents the target variance, the difference between the FPE MSE and the target variance, and the difference between the NAEP-like MSE and the target variance for each state. Since these tables present results for the full sample in each state, they are analogous to those presented for Condition 1 in Wise et al. (2006). Examination of Table 18 reveals that the bias in the FPE is small and about equally divided into positive and negative values. The bias of the NAEP-like method is typically between one and two scale score points and always positive. From Table 19, we see that the measurement variance component of the FPE tends to be greater than the measurement variance component of the complete data, but that the sampling variance component of the FPE tends to be smaller than the sampling variance component of the complete data. The comparison of the total variances then depends on the magnitude and signs of the variance component differences. In general, they are very close.

The results displayed in Table 20 indicate that the FPE performs nearly as well as—and, for a few states, slightly better than—the estimate based on the complete data. By contrast, the NAEP-like method always displays an excess in MSE, typically in the range of one to two units (squared scale score points).

**Table 12**

***Bias in FPE Mean and NAEP-Like Mean for SD-All Students in NAEP Reporting Scale, Reading, Grade 4, by State: 2003***

| State[a] | Target mean | *N* | [Bias = FPE mean – target mean] Average | Rep 1 | Rep 2 | Rep 3 | Rep 4 | [Bias = NAEP-like mean – target mean] Average | Rep 1 | Rep 2 | Rep 3 | Rep 4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Average | † | † | -0.18 | -0.05 | -0.34 | -0.28 | -0.04 | 1.48 | 1.61 | 1.40 | 1.52 | 1.40 |
| Alabama | 158.28 | 333 | 0.46 | 0.02 | 0.06 | 0.68 | 1.10 | 2.01 | 1.79 | 1.84 | 2.61 | 1.81 |
| Alaska | 177.12 | 387 | -0.52 | -0.25 | 0.06 | -1.07 | -0.82 | 0.74 | 1.16 | 0.96 | 0.69 | 0.15 |
| Arizona | 177.20 | 238 | -0.43 | -0.48 | -0.56 | 0.35 | -1.02 | 1.89 | 0.09 | 0.45 | 3.27 | 3.74 |
| Arkansas | 164.21 | 266 | -0.48 | -0.34 | -0.49 | -1.08 | -0.02 | 1.46 | 1.60 | 1.06 | 1.40 | 1.77 |
| California | 175.89 | 716 | 0.80 | 0.77 | 0.45 | 0.66 | 1.33 | 1.14 | 0.37 | 0.94 | 1.24 | 2.01 |
| Colorado | 185.27 | 313 | 0.00 | 0.44 | 0.16 | 0.69 | -1.29 | 1.04 | 1.64 | 0.55 | 1.52 | 0.43 |
| Connecticut | 191.95 | 292 | -0.52 | 1.04 | 0.13 | -1.39 | -1.85 | 1.37 | 3.33 | 1.85 | 0.81 | -0.50 |
| Delaware | 204.75 | 207 | 0.44 | 4.63 | 0.67 | -3.56 | 0.01 | 1.97 | 6.57 | 3.30 | -1.20 | -0.77 |
| Florida | 184.13 | 492 | -0.10 | -1.10 | -0.30 | 0.11 | 0.88 | 0.94 | 0.41 | 0.99 | 1.32 | 1.04 |
| Georgia | 181.40 | 481 | 0.04 | -1.27 | 0.26 | 0.81 | 0.38 | 1.31 | -0.39 | 2.19 | 1.93 | 1.52 |
| Hawaii | 162.10 | 328 | -0.18 | 0.12 | -1.16 | 0.27 | 0.08 | 1.00 | 1.14 | 0.41 | 1.65 | 0.78 |
| Idaho | 175.40 | 317 | -0.17 | 0.64 | -2.15 | 1.12 | -0.28 | 1.21 | 1.68 | 0.17 | 1.99 | 0.99 |
| Illinois | 182.90 | 530 | 0.19 | 0.84 | -0.74 | 0.92 | -0.25 | 1.92 | 2.85 | 1.37 | 3.29 | 0.16 |
| Indiana | 187.82 | 360 | 0.21 | 0.87 | 0.43 | 0.23 | -0.70 | 1.40 | 1.56 | 1.58 | 1.81 | 0.63 |
| Kansas | 184.96 | 330 | -0.11 | 0.70 | -0.88 | 0.18 | -0.43 | 1.33 | 2.20 | 0.50 | 1.50 | 1.11 |
| Louisiana | 172.28 | 428 | 0.00 | -0.49 | -0.35 | 1.03 | -0.19 | 0.48 | -0.38 | 1.19 | 1.22 | -0.11 |
| Maine | 195.42 | 321 | -0.23 | 2.46 | -1.10 | -1.98 | -0.28 | 1.39 | 4.66 | 1.10 | -1.07 | 0.89 |
| Maryland | 191.48 | 252 | -0.89 | -1.93 | -3.40 | 0.74 | 1.03 | 2.25 | 1.87 | -1.17 | 4.42 | 3.88 |
| Massachusetts | 199.91 | 668 | 0.43 | 0.73 | 0.34 | 0.32 | 0.32 | 1.01 | 1.44 | 1.07 | 0.56 | 0.95 |
| Michigan | 185.98 | 180 | -1.40 | -3.00 | -3.31 | 1.20 | -0.49 | 1.87 | -1.31 | 0.80 | 6.33 | 1.65 |
| Minnesota | 184.62 | 370 | -0.21 | -0.84 | -0.57 | -0.93 | 1.49 | 1.47 | 0.42 | 1.16 | 1.33 | 2.99 |
| Mississippi | 190.60 | 140 | 0.59 | 4.14 | 3.52 | -2.99 | -2.33 | 3.93 | 8.65 | 6.63 | -0.38 | 0.80 |
| Missouri | 195.83 | 318 | -0.67 | -1.97 | -1.26 | -0.29 | 0.83 | 1.87 | 0.27 | 1.42 | 3.45 | 2.33 |

*(Table continues)*

Table 12 (continued)

| State[a] | Target mean | *N* | [Bias = FPE mean – target mean] | | | | | [Bias = NAEP-like mean – target mean] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Average | Rep 1 | Rep 2 | Rep 3 | Rep 4 | Average | Rep 1 | Rep 2 | Rep 3 | Rep 4 |
| Nevada | 172.19 | 263 | 0.14 | -0.57 | 0.65 | -0.32 | 0.78 | 2.91 | 2.59 | 1.72 | 3.53 | 3.81 |
| New Hampshire | 193.53 | 451 | 0.11 | 0.23 | -0.23 | 0.01 | 0.43 | 0.80 | 0.99 | 0.64 | 0.54 | 1.04 |
| New Jersey | 195.96 | 359 | -0.57 | -1.29 | 0.69 | -0.29 | -1.41 | 0.60 | -0.09 | 2.03 | 0.99 | -0.53 |
| New Mexico | 180.89 | 412 | -0.24 | -1.19 | 0.54 | 0.59 | -0.90 | 1.06 | 0.39 | 1.94 | 1.72 | 0.20 |
| New York | 192.64 | 437 | -0.34 | -0.46 | 0.33 | -0.89 | -0.33 | 1.02 | 0.05 | 2.29 | 1.21 | 0.52 |
| North Carolina | 194.34 | 567 | -0.37 | 0.49 | -0.06 | -0.84 | -1.08 | 1.07 | 2.11 | 1.35 | 0.22 | 0.62 |
| North Dakota | 189.61 | 332 | -0.61 | -0.73 | -0.85 | -1.65 | 0.77 | 1.02 | 1.05 | 0.98 | 0.20 | 1.85 |
| Ohio | 173.96 | 295 | 1.31 | 1.71 | -0.34 | 0.23 | 3.62 | 1.77 | 1.32 | 2.68 | 0.73 | 2.34 |
| Oregon | 187.93 | 344 | -0.13 | -0.53 | -0.40 | 1.53 | -1.12 | 1.66 | 0.84 | 1.92 | 2.22 | 1.66 |
| Rhode Island | 190.09 | 518 | -0.45 | -0.69 | 0.17 | -1.21 | -0.09 | 0.65 | 0.67 | 1.30 | -0.10 | 0.71 |
| South Carolina | 193.41 | 337 | -1.32 | -1.91 | -0.69 | -1.98 | -0.68 | 1.19 | 1.67 | 0.89 | 0.30 | 1.89 |
| Tennessee | 180.14 | 352 | -1.01 | -1.71 | -1.03 | -1.06 | -0.24 | 0.88 | 0.25 | -0.24 | 1.85 | 1.64 |
| Texas | 190.73 | 434 | 0.61 | 1.18 | -0.14 | -0.05 | 1.45 | 2.15 | 2.60 | 1.38 | 1.57 | 3.05 |
| Utah | 178.76 | 389 | -0.54 | -0.27 | -1.32 | 0.42 | -0.98 | 1.56 | 1.63 | 1.20 | 2.32 | 1.09 |
| Vermont | 202.54 | 300 | -0.31 | -0.29 | -0.25 | -0.39 | -0.31 | 1.62 | 1.47 | 2.48 | 0.69 | 1.85 |
| Virginia | 200.56 | 207 | -0.50 | 1.84 | -2.06 | -1.89 | 0.12 | 3.01 | 7.20 | 0.48 | -0.18 | 4.52 |
| Washington | 188.24 | 353 | 0.18 | -2.29 | 1.92 | 0.48 | 0.60 | 1.85 | -0.31 | 3.31 | 3.95 | 0.45 |
| Wisconsin | 181.18 | 300 | -0.41 | -0.81 | -1.09 | 0.34 | -0.10 | 1.86 | 1.79 | 1.21 | 1.92 | 2.50 |
| Wyoming | 183.99 | 350 | -0.27 | -0.77 | 0.07 | -0.67 | 0.28 | 0.69 | -0.08 | 1.07 | 0.58 | 1.19 |

*Note.* † = not applicable. The SD-all category includes students classified as SD and students classified as both SD and ELL. FPE = full-population estimation. *N* = number of students. Rep = replicate. SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment. Authors' calculations.

[a]Forty-two states with the state achievement test score as a school-level sampling variable were included in the study.

**Table 13**

***Bias in FPE Mean and NAEP-Like Mean for ELL-Only Students in NAEP Reporting Scale, Reading, Grade 4, by State: 2003***

| State[a] | Target mean | $N$ | [Bias = FPE mean – target mean] | | | | | [Bias = NAEP-like mean – target mean] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Average | Rep 1 | Rep 2 | Rep 3 | Rep 4 | Average | Rep 1 | Rep 2 | Rep 3 | Rep 4 |
| Average | † | † | -0.31 | -0.37 | -0.16 | 0.25 | -0.94 | 0.32 | 1.09 | 0.52 | 0.30 | -0.62 |
| Alabama | 181.69 | 14 | -2.53 | -5.98 | -6.91 | -5.16 | 7.92 | -2.69 | -1.89 | -7.17 | -5.39 | 3.69 |
| Alaska | 184.04 | 310 | 0.04 | 0.65 | -0.08 | 0.05 | -0.48 | -0.02 | 0.41 | -0.25 | -0.16 | -0.09 |
| Arizona | 178.47 | 595 | -0.35 | -0.47 | -0.71 | -0.68 | 0.48 | 0.17 | -0.20 | 0.14 | 0.05 | 0.69 |
| Arkansas | 205.82 | 60 | -1.34 | -0.17 | 0.61 | -3.81 | -2.01 | 0.01 | 2.88 | 1.87 | -3.74 | -0.98 |
| California | 186.31 | 2,712 | 0.08 | -0.08 | 0.61 | -0.34 | 0.14 | 0.23 | 0.03 | 0.62 | 0.07 | 0.20 |
| Colorado | 194.91 | 201 | -0.62 | -0.65 | -0.65 | 0.39 | -1.58 | 0.14 | 0.75 | -0.44 | 1.27 | -1.02 |
| Connecticut | 183.99 | 49 | -2.79 | -3.82 | -0.23 | -3.73 | -3.39 | -1.01 | -0.55 | 4.78 | -4.25 | -4.04 |
| Delaware | 194.25 | 36 | -0.76 | 5.27 | -1.90 | -3.77 | -2.64 | -0.98 | 7.91 | -2.82 | -4.14 | -4.88 |
| Florida | 205.22 | 257 | 0.32 | 0.13 | 1.51 | 1.21 | -1.57 | 0.83 | 0.87 | 2.39 | 1.03 | -0.96 |
| Georgia | 182.18 | 99 | 0.52 | -1.01 | -0.74 | 1.77 | 2.06 | 0.42 | 0.85 | -1.72 | 1.20 | 1.36 |
| Hawaii | 171.02 | 142 | -1.75 | -1.33 | -1.52 | 0.14 | -4.28 | -0.30 | 2.09 | -0.72 | -0.04 | -2.54 |
| Idaho | 194.52 | 171 | -0.19 | 0.22 | -0.48 | -0.81 | 0.33 | 0.58 | 0.77 | -0.20 | 1.15 | 0.59 |
| Illinois | 183.06 | 341 | -0.49 | -0.09 | -1.42 | -1.13 | 0.69 | 0.32 | 1.44 | -0.53 | -0.29 | 0.65 |
| Indiana | 195.06 | 47 | 0.55 | -0.31 | -1.92 | 1.89 | 2.53 | 1.25 | 2.57 | -3.69 | 2.60 | 3.52 |
| Kansas | 193.93 | 60 | 0.49 | -1.30 | 2.76 | -2.11 | 2.60 | 0.53 | -0.25 | 2.77 | -2.13 | 1.73 |
| Louisiana | 210.46 | 18 | -0.88 | 6.57 | -3.25 | 2.05 | -8.87 | 0.01 | 9.30 | 1.74 | -0.53 | -10.46 |
| Maine | 213.06 | 18 | 0.68 | 0.60 | -5.22 | 3.74 | 3.60 | 0.93 | 0.69 | -2.06 | 1.50 | 3.59 |
| Maryland | 195.41 | 70 | -0.16 | -4.18 | -1.62 | 5.01 | 0.14 | 0.91 | -5.35 | -1.26 | 6.98 | 3.27 |
| Massachusetts | 196.63 | 217 | 1.19 | 2.48 | 0.85 | 0.11 | 1.35 | 0.22 | 2.31 | -0.79 | -0.35 | -0.28 |
| Michigan | 206.84 | 130 | -0.38 | 1.40 | -1.40 | -1.93 | 0.40 | 1.27 | 2.47 | -0.19 | 0.63 | 2.17 |
| Minnesota | 178.95 | 183 | 0.03 | 1.06 | -0.76 | 0.06 | -0.24 | 0.53 | 1.90 | 0.61 | -0.03 | -0.34 |
| Mississippi | 192.59 | 4 | 7.25 | -4.35 | 18.98 | 29.60 | -15.22 | 4.18 | 1.61 | 11.19 | 12.37 | -8.44 |
| Missouri | 220.73 | 22 | 0.29 | 0.62 | 4.47 | -0.13 | -3.79 | 1.38 | 9.26 | 2.56 | -3.33 | -2.98 |

*(Table continues)*

Table 13 (continued)

| State[a] | Target mean | *N* | [Bias = FPE mean – target mean] Average | Rep 1 | Rep 2 | Rep 3 | Rep 4 | [Bias = NAEP-like mean – target mean] Average | Rep 1 | Rep 2 | Rep 3 | Rep 4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nevada | 181.26 | 327 | 0.23 | -0.58 | 1.44 | 0.63 | -0.59 | 1.12 | 0.97 | 1.39 | 1.51 | 0.60 |
| New Hampshire | 205.22 | 54 | -0.12 | 2.49 | 0.00 | -0.57 | -2.40 | 0.72 | 4.87 | 0.56 | -1.89 | -0.69 |
| New Jersey | 187.20 | 72 | -1.04 | 3.47 | -0.89 | -2.95 | -3.79 | -0.45 | 3.54 | 1.13 | -4.07 | -2.41 |
| New Mexico | 185.65 | 575 | -0.15 | -0.35 | 0.15 | -0.33 | -0.08 | 0.25 | 0.05 | 0.63 | -0.09 | 0.41 |
| New York | 194.50 | 123 | -0.30 | -4.29 | -2.24 | 4.17 | 1.18 | -0.09 | -3.33 | 0.69 | 4.39 | -2.13 |
| North Carolina | 204.67 | 147 | -1.34 | -2.33 | -1.41 | -1.59 | -0.03 | -0.23 | -1.27 | -0.42 | -0.22 | 1.01 |
| North Dakota | 189.75 | 70 | -0.12 | -0.26 | 0.26 | 0.18 | -0.65 | 0.13 | 0.00 | 0.68 | 0.65 | -0.82 |
| Ohio | 208.11 | 44 | -2.57 | 3.25 | -10.05 | -5.79 | 2.29 | -1.30 | 0.12 | -8.62 | 0.81 | 2.50 |
| Oregon | 188.18 | 272 | -0.52 | -1.57 | -0.27 | -0.58 | 0.33 | 0.80 | 0.37 | 1.69 | 0.17 | 0.98 |
| Rhode Island | 177.85 | 167 | 0.08 | 0.37 | -0.04 | -0.04 | 0.04 | -0.02 | 0.17 | -0.05 | -0.08 | -0.12 |
| South Carolina | 185.06 | 30 | 0.37 | -3.62 | 6.62 | 3.01 | -4.52 | 1.78 | 1.26 | 9.23 | 6.10 | -9.45 |
| Tennessee | 203.89 | 37 | -1.67 | 3.52 | -2.93 | -3.16 | -4.10 | 0.21 | 4.69 | -0.23 | -1.91 | -1.70 |
| Texas | 189.69 | 562 | 0.41 | 1.37 | -0.11 | -1.87 | 2.25 | 1.02 | 1.93 | 0.31 | -0.87 | 2.69 |
| Utah | 193.49 | 275 | -0.28 | -0.69 | -0.06 | -1.11 | 0.74 | 0.60 | -0.11 | 1.03 | -0.07 | 1.54 |
| Vermont | 214.87 | 28 | -2.27 | -3.41 | -2.16 | 1.05 | -4.57 | -2.07 | -4.13 | -3.42 | 0.74 | -1.47 |
| Virginia | 204.25 | 114 | -1.16 | -2.66 | 1.23 | -0.99 | -2.21 | 1.86 | 0.19 | 7.27 | 2.03 | -2.05 |
| Washington | 187.17 | 186 | -0.61 | -0.59 | 0.11 | -1.07 | -0.88 | 0.55 | 0.67 | 1.44 | 0.45 | -0.36 |
| Wisconsin | 201.62 | 115 | -0.66 | -4.05 | 1.75 | -0.22 | -0.13 | -0.04 | -3.24 | 1.17 | 0.78 | 1.12 |
| Wyoming | 198.45 | 86 | -0.31 | -0.82 | 0.83 | -0.72 | -0.53 | -0.18 | -0.66 | 0.67 | -0.45 | -0.28 |

*Note.* † = not applicable. The ELL-only category includes students classified as ELL-only. FPE = full-population estimation. *N* = number of students. Rep = replicate. SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment. Authors' calculations.

[a]Forty-two states with the state achievement test score as a school-level sampling variable were included in the study.

**Table 14**

***Variance Components for Target Data and Differences in Variance Components Between FPE and Target, for SD-All Students, Reading, Grade 4, by State: 2003***

| | Target variance | | | [FPE variance components – target variance components] | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Replicate 1 | | | Replicate 2 | | | Replicate 3 | | | Replicate 4 | | |
| State[a] | Total | Meas | Samp | Total | Meas | Samp | Total | Meas | Samp | Total | Meas | Samp | Total | Meas | Samp |
| Alabama | 12.42 | 2.22 | 10.20 | 0.73 | 2.87 | -2.13 | 1.32 | 0.56 | 0.76 | 4.52 | 2.59 | 1.93 | 5.61 | 4.18 | 1.43 |
| Alaska | 12.08 | 4.13 | 7.95 | 0.81 | 2.76 | -1.95 | 2.67 | 2.79 | -0.11 | -0.23 | -0.42 | 0.19 | 3.09 | 3.94 | -0.85 |
| Arizona | 16.35 | 5.86 | 10.49 | 4.98 | 5.20 | -0.22 | -2.99 | -3.34 | 0.35 | 2.13 | 0.46 | 1.67 | 7.45 | 6.79 | 0.66 |
| Arkansas | 18.20 | 2.82 | 15.38 | 4.78 | 5.15 | -0.36 | -1.74 | 0.56 | -2.30 | -1.25 | 2.56 | -3.81 | 2.67 | 6.38 | -3.71 |
| California | 5.68 | 0.53 | 5.15 | 1.81 | 1.37 | 0.44 | -1.26 | 0.12 | -1.38 | 3.59 | 2.63 | 0.96 | 1.66 | 0.68 | 0.99 |
| Colorado | 8.73 | 2.20 | 6.53 | -1.35 | -0.84 | -0.51 | 3.42 | 4.51 | -1.09 | 2.43 | 2.03 | 0.40 | 1.74 | 1.64 | 0.09 |
| Connecticut | 10.57 | 0.89 | 9.68 | 3.54 | 3.78 | -0.24 | -1.40 | 0.10 | -1.50 | -0.47 | -0.14 | -0.33 | -0.27 | 1.04 | -1.31 |
| Delaware | 12.28 | 1.53 | 10.75 | -3.28 | 2.07 | -5.35 | -3.37 | 2.30 | -5.67 | 3.83 | 5.26 | -1.43 | 0.64 | -0.97 | 1.61 |
| Florida | 5.54 | 0.36 | 5.18 | -0.26 | 0.10 | -0.36 | -1.04 | -0.15 | -0.89 | -1.05 | -0.18 | -0.87 | 0.53 | 0.36 | 0.18 |
| Georgia | 7.39 | 0.82 | 6.57 | 1.95 | 2.13 | -0.18 | 3.46 | 1.88 | 1.57 | 0.98 | 1.32 | -0.35 | 0.80 | 0.83 | -0.03 |
| Hawaii | 7.47 | 2.17 | 5.30 | 4.03 | 3.25 | 0.78 | 3.60 | 2.45 | 1.15 | 0.20 | -0.88 | 1.07 | 2.34 | 2.29 | 0.05 |
| Idaho | 6.72 | 1.59 | 5.13 | 3.88 | 3.08 | 0.80 | 0.37 | -0.19 | 0.56 | 3.34 | 2.26 | 1.08 | 0.25 | 1.30 | -1.05 |
| Illinois | 14.59 | 1.52 | 13.08 | -0.60 | 1.80 | -2.40 | -2.75 | 0.91 | -3.65 | -3.02 | 0.31 | -3.33 | 0.78 | 2.24 | -1.46 |
| Indiana | 9.66 | 1.96 | 7.70 | 0.34 | 1.77 | -1.43 | 0.52 | -1.69 | 2.21 | 3.65 | 4.04 | -0.40 | 0.71 | 1.64 | -0.93 |
| Kansas | 6.83 | 0.19 | 6.64 | 2.00 | 2.28 | -0.28 | 1.72 | 1.25 | 0.48 | 0.33 | 0.91 | -0.59 | 2.14 | 1.16 | 0.98 |
| Louisiana | 10.09 | 0.64 | 9.44 | -0.54 | 2.14 | -2.68 | -0.91 | 2.35 | -3.26 | -0.42 | 0.06 | -0.48 | 1.29 | 0.99 | 0.30 |
| Maine | 4.32 | 1.12 | 3.20 | 0.84 | -0.06 | 0.90 | 2.70 | 2.53 | 0.17 | 1.92 | 2.23 | -0.31 | 0.42 | 0.53 | -0.11 |
| Maryland | 14.48 | 2.50 | 11.99 | 2.11 | 4.50 | -2.39 | 2.78 | 2.54 | 0.25 | 2.87 | 3.75 | -0.88 | 0.09 | 1.68 | -1.59 |
| Massachusetts | 4.67 | 0.89 | 3.78 | -0.26 | -0.49 | 0.23 | 1.41 | 0.48 | 0.94 | 0.98 | 1.13 | -0.15 | -0.96 | 0.02 | -0.98 |
| Michigan | 19.49 | 0.91 | 18.58 | -1.65 | 1.21 | -2.86 | 6.92 | 3.09 | 3.82 | 8.51 | 0.43 | 8.09 | -1.43 | 4.61 | -6.04 |
| Minnesota | 4.21 | 0.17 | 4.04 | -0.05 | -0.08 | 0.03 | 2.99 | 2.14 | 0.85 | 0.74 | -0.10 | 0.84 | 0.20 | 0.59 | -0.38 |
| Mississippi | 12.80 | 2.99 | 9.80 | 27.87 | 26.92 | 0.95 | 29.86 | 23.51 | 6.35 | 9.04 | 6.48 | 2.56 | 22.74 | 14.57 | 8.17 |
| Missouri | 10.87 | 0.70 | 10.16 | 0.40 | 2.93 | -2.54 | 1.42 | 0.82 | 0.61 | 1.19 | 1.65 | -0.45 | -0.51 | 0.21 | -0.73 |

*(Table continues)*

Table 14 (continued)

| State[a] | Target variance | | | [FPE variance components – target variance components] | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Replicate 1 | | | Replicate 2 | | | Replicate 3 | | | Replicate 4 | | |
| | Total | Meas | Samp | Total | Meas | Samp | Total | Meas | Samp | Total | Meas | Samp | Total | Meas | Samp |
| Nevada | 13.22 | 1.38 | 11.84 | 4.39 | 1.47 | 2.92 | 3.29 | 6.15 | -2.85 | 4.66 | 6.53 | -1.87 | 10.02 | 8.05 | 1.97 |
| New Hampshire | 6.06 | 0.78 | 5.27 | 0.87 | 0.93 | -0.06 | 0.01 | 0.85 | -0.84 | 1.59 | 1.19 | 0.40 | -0.36 | 0.79 | -1.16 |
| New Jersey | 6.70 | 0.22 | 6.48 | 1.39 | 0.62 | 0.77 | 1.35 | 0.91 | 0.44 | -0.19 | 0.72 | -0.91 | 0.01 | 0.16 | -0.15 |
| New Mexico | 11.68 | 0.88 | 10.79 | -0.32 | 0.81 | -1.13 | -1.34 | 0.26 | -1.60 | -0.65 | 2.11 | -2.76 | -1.62 | 1.17 | -2.79 |
| New York | 6.62 | 0.32 | 6.30 | 0.99 | 0.85 | 0.13 | 0.87 | 2.22 | -1.35 | -0.07 | 0.50 | -0.57 | 2.05 | 2.33 | -0.28 |
| North Carolina | 6.66 | 1.02 | 5.64 | -0.73 | 0.65 | -1.38 | 2.81 | 1.66 | 1.15 | 1.58 | 1.99 | -0.41 | -1.16 | -0.49 | -0.67 |
| North Dakota | 5.77 | 2.36 | 3.40 | 3.57 | 2.36 | 1.21 | 0.99 | 0.83 | 0.15 | 0.34 | 0.53 | -0.19 | 1.04 | 1.73 | -0.69 |
| Ohio | 15.19 | 1.72 | 13.47 | -5.13 | 1.28 | -6.41 | 2.30 | 4.89 | -2.59 | 2.28 | 5.59 | -3.31 | -0.81 | 1.44 | -2.25 |
| Oregon | 7.43 | 1.47 | 5.97 | 1.44 | 2.42 | -0.98 | 2.63 | 0.94 | 1.70 | -0.52 | 0.17 | -0.69 | 1.21 | 2.09 | -0.88 |
| Rhode Island | 6.59 | 0.28 | 6.32 | 0.40 | 1.64 | -1.24 | -0.87 | 0.32 | -1.19 | -0.60 | 0.56 | -1.16 | -0.67 | -0.03 | -0.64 |
| South Carolina | 8.64 | 0.69 | 7.95 | -0.76 | 1.67 | -2.43 | 2.70 | 3.43 | -0.74 | -0.74 | 1.68 | -2.42 | 1.21 | 4.27 | -3.06 |
| Tennessee | 22.04 | 1.10 | 20.94 | -1.17 | -0.67 | -0.50 | -5.25 | 0.22 | -5.48 | 2.35 | 2.35 | 0.00 | 1.77 | 2.00 | -0.23 |
| Texas | 11.10 | 2.53 | 8.57 | -3.52 | -0.07 | -3.45 | -1.03 | -1.76 | 0.73 | 5.61 | 4.26 | 1.35 | 6.36 | 5.21 | 1.14 |
| Utah | 5.79 | 1.28 | 4.51 | -0.88 | 0.24 | -1.12 | 0.93 | 0.29 | 0.64 | -0.09 | 0.23 | -0.32 | 0.96 | 0.05 | 0.91 |
| Vermont | 6.81 | 0.85 | 5.96 | 4.78 | 2.09 | 2.69 | 0.08 | 1.25 | -1.17 | -0.48 | 1.69 | -2.17 | 4.18 | 4.80 | -0.62 |
| Virginia | 17.72 | 3.05 | 14.67 | 3.32 | -1.14 | 4.46 | 2.56 | 1.26 | 1.29 | 2.52 | 6.31 | -3.79 | -5.75 | -0.16 | -5.59 |
| Washington | 5.85 | 0.10 | 5.76 | -0.50 | 1.47 | -1.97 | 1.84 | 2.09 | -0.25 | 1.10 | 2.27 | -1.17 | 3.05 | 2.31 | 0.75 |
| Wisconsin | 9.22 | 2.32 | 6.90 | 2.75 | 3.55 | -0.80 | 3.92 | 4.70 | -0.78 | 1.32 | 0.78 | 0.53 | 0.54 | -0.78 | 1.32 |
| Wyoming | 3.94 | 0.48 | 3.46 | -0.68 | -0.12 | -0.56 | 1.01 | 0.86 | 0.15 | 1.75 | 1.14 | 0.61 | 1.41 | 1.56 | -0.15 |

*Note.* The SD-all category includes students classified as SD and students classified as both SD and ELL. FPE = full-population estimation. Meas = measurement variance. Samp = sampling variance. SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment. Authors' calculations.

[a] Forty-two states with the state achievement test score as a school-level sampling variable were included in the study.

**Table 15**

***Variance Components for Target Data and Differences in Variance Components Between FPE and Target, for ELL-Only Students, Reading, Grade 4, by State: 2003***

| | | | | [FPE variance components – target variance components] | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Target variance | | | Replicate 1 | | | Replicate 2 | | | Replicate 3 | | | Replicate 4 | | |
| State[a] | Total | Meas | Samp | Total | Meas | Samp | Total | Meas | Samp | Total | Meas | Samp | Total | Meas | Samp |
| Alabama | 93.58 | 7.76 | 85.82 | -3.07 | 14.34 | -17.42 | -18.39 | 13.69 | -32.08 | 20.62 | 61.05 | -40.43 | 63.55 | 48.54 | 15.01 |
| Alaska | 22.41 | 4.02 | 18.40 | -0.30 | -0.50 | 0.21 | 0.19 | 0.08 | 0.11 | -0.58 | 0.22 | -0.80 | -0.20 | 0.47 | -0.67 |
| Arizona | 6.24 | 0.52 | 5.72 | -0.17 | -0.03 | -0.14 | 1.76 | 1.01 | 0.75 | -0.13 | 0.22 | -0.36 | 0.26 | 1.29 | -1.03 |
| Arkansas | 18.04 | 8.66 | 9.38 | -5.82 | -4.62 | -1.21 | 5.66 | 6.27 | -0.61 | -0.22 | 2.08 | -2.30 | 14.09 | 10.46 | 3.64 |
| California | 2.25 | 0.55 | 1.70 | 0.04 | 0.18 | -0.14 | 0.45 | 0.65 | -0.20 | 0.41 | 0.40 | 0.01 | 0.34 | 0.19 | 0.14 |
| Colorado | 7.38 | 1.42 | 5.96 | 2.18 | 1.89 | 0.29 | 1.46 | 1.40 | 0.07 | 0.40 | 0.32 | 0.08 | 0.48 | -0.76 | 1.24 |
| Connecticut | 22.22 | 5.74 | 16.49 | 32.06 | 24.79 | 7.27 | 13.37 | 13.29 | 0.08 | 34.40 | 26.59 | 7.81 | 4.68 | -1.19 | 5.87 |
| Delaware | 31.56 | 3.62 | 27.94 | -12.33 | 0.69 | -13.02 | 1.23 | 10.34 | -9.10 | 51.40 | 49.65 | 1.75 | 5.13 | 1.92 | 3.21 |
| Florida | 9.57 | 1.45 | 8.12 | 0.64 | 0.80 | -0.16 | 0.31 | 0.68 | -0.37 | 1.74 | 2.35 | -0.60 | -2.13 | 0.46 | -2.60 |
| Georgia | 44.57 | 2.14 | 42.43 | 4.12 | 4.15 | -0.02 | 15.93 | 6.05 | 9.89 | 15.33 | 2.34 | 12.99 | -14.97 | 7.63 | -22.61 |
| Hawaii | 33.73 | 8.46 | 25.28 | 23.75 | 11.27 | 12.48 | -5.46 | 2.23 | -7.69 | -6.86 | 2.43 | -9.29 | -7.69 | 1.15 | -8.83 |
| Idaho | 12.11 | 2.01 | 10.10 | -1.18 | -0.95 | -0.23 | -0.42 | 1.02 | -1.44 | 0.38 | 0.54 | -0.16 | 1.98 | 3.71 | -1.73 |
| Illinois | 11.21 | 1.40 | 9.81 | -1.13 | 1.13 | -2.26 | -2.62 | 0.63 | -3.25 | -0.43 | 2.66 | -3.09 | -1.25 | 1.22 | -2.47 |
| Indiana | 42.25 | 7.07 | 35.17 | 6.86 | -2.58 | 9.44 | -1.35 | 19.68 | -21.03 | 20.13 | 6.68 | 13.46 | -3.55 | 5.62 | -9.17 |
| Kansas | 27.08 | 2.82 | 24.27 | 7.40 | 6.34 | 1.06 | -1.26 | 2.22 | -3.48 | 15.57 | 20.17 | -4.60 | -1.54 | -1.74 | 0.19 |
| Louisiana | 237.57 | 43.99 | 193.58 | 12.41 | 11.80 | 0.61 | 100.05 | -30.91 | -69.14 | 78.59 | 55.14 | 23.45 | -24.18 | 134.44 | 158.62 |
| Maine | 60.01 | 9.00 | 51.01 | 7.00 | 0.04 | 6.95 | -7.09 | -2.13 | -4.96 | -7.10 | 1.13 | -8.23 | -17.19 | -2.07 | -15.12 |
| Maryland | 79.61 | 4.87 | 74.74 | -29.91 | 11.92 | -41.83 | -3.75 | 27.89 | -31.65 | -9.39 | -0.94 | -8.46 | -29.88 | 12.25 | -42.13 |
| Massachusetts | 14.08 | 5.86 | 8.22 | 9.10 | 6.39 | 2.71 | 0.78 | 1.58 | -0.79 | 2.19 | 0.06 | 2.13 | 0.16 | -3.48 | 3.64 |
| Michigan | 49.44 | 4.40 | 45.03 | -6.12 | 8.26 | -14.38 | 18.41 | 5.31 | 13.10 | -8.18 | 1.27 | -9.45 | 4.36 | 7.22 | -2.86 |
| Minnesota | 7.16 | 0.93 | 6.23 | 0.70 | -0.53 | 1.22 | 0.08 | -0.35 | 0.43 | 1.37 | 0.79 | 0.58 | 0.30 | 1.11 | -0.81 |
| Mississippi | 382.21 | [illegible]14.33 | 267.88 | 14.50 | 1.43 | 13.07 | 266.88 | -27.18 | 239.69 | 53.58 | 284.84 | 231.25 | 496.50 | [illegible]58.11 | 438.39 |
| Missouri | 56.49 | 17.34 | 39.15 | 0.82 | 5.84 | -5.02 | 40.13 | 44.43 | -4.30 | 44.15 | 18.46 | 25.69 | 17.15 | 37.59 | -20.43 |

*(Table continues)*

Table 15 (continued)

| State[a] | Target variance | | | [FPE variance components – target variance components] | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Replicate 1 | | | Replicate 2 | | | Replicate 3 | | | Replicate 4 | | |
| | Total | Meas | Samp | Total | Meas | Samp | Total | Meas | Samp | Total | Meas | Samp | Total | Meas | Samp |
| Nevada | 15.32 | 3.09 | 12.23 | -3.77 | -1.71 | -2.06 | -6.19 | -1.96 | -4.22 | -1.92 | -0.80 | -1.12 | 3.32 | 3.41 | -0.09 |
| New Hampshire | 31.13 | 11.64 | 19.50 | -4.12 | 1.48 | -5.59 | -7.26 | 1.02 | -8.28 | -1.07 | -4.72 | 3.65 | 0.95 | -5.74 | 6.69 |
| New Jersey | 20.96 | 0.82 | 20.15 | 3.41 | 1.80 | 1.61 | 13.31 | 12.01 | 1.30 | -1.53 | 2.41 | -3.94 | 12.54 | 14.94 | -2.40 |
| New Mexico | 8.24 | 1.72 | 6.52 | -1.50 | -0.06 | -1.43 | -2.22 | -0.92 | -1.30 | -1.54 | -1.09 | -0.46 | -0.79 | -0.34 | -0.45 |
| New York | 21.31 | 3.74 | 17.57 | 0.36 | 5.93 | -5.57 | 16.42 | 17.85 | -1.43 | 7.85 | 9.25 | -1.39 | 13.73 | 15.19 | -1.46 |
| North Carolina | 19.02 | 4.46 | 14.56 | 3.24 | 8.30 | -5.06 | -1.93 | -3.18 | 1.25 | 5.95 | 9.37 | -3.42 | 0.32 | 7.90 | -7.58 |
| North Dakota | 20.99 | 4.64 | 16.35 | 5.63 | 1.17 | 4.46 | 3.15 | 1.78 | 1.37 | 5.67 | 2.62 | 3.05 | 1.39 | 1.54 | -0.16 |
| Ohio | 126.04 | 51.79 | 74.25 | 37.30 | 25.67 | 11.63 | 77.27 | 01.78 | -24.52 | 18.51 | 8.21 | 10.30 | 76.28 | -40.38 | -35.90 |
| Oregon | 7.35 | 1.09 | 6.27 | 1.14 | 1.84 | -0.70 | -1.10 | 0.96 | -2.06 | -1.24 | 0.31 | -1.55 | 1.71 | 0.67 | 1.03 |
| Rhode Island | 23.43 | 3.32 | 20.11 | -1.74 | -1.03 | -0.72 | 2.00 | 7.15 | -5.14 | 5.41 | 2.57 | 2.84 | -5.02 | 1.67 | -6.69 |
| South Carolina | 59.66 | 11.80 | 47.87 | 15.70 | 27.62 | -11.91 | 85.07 | 32.52 | 52.55 | 19.72 | 8.93 | 10.79 | 8.82 | 16.70 | -7.87 |
| Tennessee | 76.51 | 13.18 | 63.33 | 49.35 | 28.28 | 21.08 | 14.19 | 24.82 | -10.63 | -19.27 | -2.59 | -16.68 | 13.05 | 7.82 | -20.87 |
| Texas | 8.50 | 1.13 | 7.37 | -1.48 | 0.96 | -2.45 | -0.63 | 0.35 | -0.98 | -1.02 | 0.61 | -1.63 | 1.31 | 3.36 | -2.05 |
| Utah | 9.19 | 0.69 | 8.50 | -0.94 | 1.23 | -2.17 | 2.32 | -0.02 | 2.34 | 5.84 | 0.26 | 5.59 | -0.85 | -0.45 | -0.40 |
| Vermont | 80.93 | 1.82 | 79.11 | 15.21 | 13.04 | -28.25 | -7.36 | 10.38 | -17.74 | 3.53 | 9.08 | -5.56 | 13.36 | 5.44 | 7.93 |
| Virginia | 21.08 | 6.82 | 14.26 | -6.72 | 1.57 | -8.29 | -4.75 | -2.95 | -1.80 | 5.30 | 6.01 | -0.72 | 7.65 | 7.25 | 0.40 |
| Washington | 7.98 | 2.15 | 5.83 | 1.02 | 1.83 | -0.80 | 8.21 | 7.47 | 0.74 | 1.49 | 0.20 | 1.29 | 1.24 | 1.97 | -0.73 |
| Wisconsin | 24.26 | 10.14 | 14.12 | 6.67 | 5.43 | 1.23 | 6.47 | 8.79 | -2.32 | 8.34 | 15.73 | -7.39 | -2.73 | -5.73 | 3.00 |
| Wyoming | 12.70 | 4.24 | 8.46 | 1.90 | 0.11 | 1.79 | 5.57 | 5.86 | -0.28 | 1.93 | 0.01 | 1.92 | -1.25 | -1.32 | 0.07 |

*Note.* The ELL-only category includes students classified as English language learners (ELL) only. FPE = full-population estimation. Meas = measurement variance. Samp = sampling variance. Table entries for Louisiana and Mississippi are exceptionally large due to small sample fluctuations. SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment. Authors' calculations.

[a] Forty-two states with the state achievement test score as a school-level sampling variable were included in the study.

**Table 16**

***Differences Between FPE MSE and Target Variance and Differences Between NAEP-Like MSE and Target Variance, for SD-All Students, Reading, Grade 4, by State: 2003***

| State[a] | Target variance | $N$ | [MSE (FPE) – variance (target)] | | | | | [MSE (NAEP-like) – variance (target)] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Average | Rep 1 | Rep 2 | Rep 3 | Rep 4 | Average | Rep 1 | Rep 2 | Rep3 | Rep 4 |
| Average | † | † | 3.17 | 3.77 | 3.11 | 2.91 | 2.87 | 7.20 | 9.10 | 6.32 | 7.39 | 5.99 |
| Alabama | 12.42 | 333 | 3.46 | 0.74 | 1.32 | 4.98 | 6.82 | 5.73 | 2.53 | 6.66 | 9.39 | 4.33 |
| Alaska | 12.08 | 387 | 2.06 | 0.87 | 2.68 | 0.92 | 3.75 | 0.81 | -0.81 | 1.54 | 1.11 | 1.41 |
| Arizona | 16.35 | 238 | 3.32 | 5.21 | -2.67 | 2.25 | 8.48 | 13.80 | 13.02 | 8.93 | 14.79 | 18.45 |
| Arkansas | 18.20 | 266 | 1.50 | 4.89 | -1.50 | -0.08 | 2.67 | 7.70 | 10.48 | 5.17 | 7.90 | 7.23 |
| California | 5.68 | 716 | 2.20 | 2.40 | -1.06 | 4.03 | 3.43 | 3.09 | 2.23 | 1.59 | 3.42 | 5.12 |
| Colorado | 8.73 | 313 | 2.15 | -1.16 | 3.45 | 2.91 | 3.41 | 2.45 | 2.44 | 2.07 | 3.16 | 2.12 |
| Connecticut | 10.57 | 292 | 1.96 | 4.62 | -1.38 | 1.46 | 3.14 | 5.33 | 14.71 | 4.80 | 1.37 | 0.43 |
| Delaware | 12.28 | 207 | 8.10 | 18.14 | -2.92 | 16.52 | 0.64 | 23.19 | 45.05 | 15.92 | 17.09 | 14.70 |
| Florida | 5.54 | 492 | 0.07 | 0.95 | -0.95 | -1.04 | 1.30 | 1.62 | 0.96 | 1.11 | 2.17 | 2.24 |
| Georgia | 7.39 | 481 | 2.41 | 3.56 | 3.52 | 1.63 | 0.95 | 4.39 | 0.25 | 8.32 | 5.69 | 3.28 |
| Hawaii | 7.47 | 328 | 2.90 | 4.04 | 4.96 | 0.27 | 2.34 | 3.73 | 3.21 | 1.33 | 7.30 | 3.07 |
| Idaho | 6.72 | 317 | 3.55 | 4.29 | 4.98 | 4.60 | 0.33 | 4.26 | 3.43 | 2.58 | 8.02 | 3.00 |
| Illinois | 14.59 | 530 | -0.85 | 0.11 | -2.20 | -2.17 | 0.84 | 8.13 | 11.25 | 4.94 | 12.75 | 3.60 |
| Indiana | 9.66 | 360 | 1.67 | 1.09 | 0.71 | 3.70 | 1.20 | 4.00 | 2.96 | 5.99 | 5.51 | 1.53 |
| Kansas | 6.83 | 330 | 1.92 | 2.49 | 2.49 | 0.36 | 2.33 | 2.90 | 6.23 | 1.57 | 1.16 | 2.62 |
| Louisiana | 10.09 | 428 | 0.22 | -0.30 | -0.79 | 0.64 | 1.32 | 2.00 | 2.57 | 2.43 | 1.77 | 1.23 |
| Maine | 4.32 | 321 | 4.29 | 6.91 | 3.91 | 5.85 | 0.50 | 7.59 | 21.95 | 3.54 | 3.81 | 1.05 |
| Maryland | 14.48 | 252 | 6.18 | 5.82 | 14.32 | 3.41 | 1.16 | 12.50 | 5.02 | 5.59 | 23.43 | 15.98 |
| Massachusetts | 4.67 | 668 | 0.51 | 0.27 | 1.53 | 1.09 | -0.86 | 1.41 | 2.13 | 2.58 | 0.89 | 0.03 |
| Michigan | 19.49 | 180 | 8.50 | 7.35 | 17.87 | 9.96 | -1.19 | 21.57 | 11.13 | 8.86 | 59.15 | 7.12 |
| Minnesota | 4.21 | 370 | 2.00 | 0.65 | 3.31 | 1.61 | 2.42 | 4.19 | 1.97 | 2.30 | 2.13 | 10.36 |
| Mississippi | 12.80 | 140 | 33.36 | 44.99 | 42.28 | 17.98 | 28.18 | 40.42 | 90.56 | 58.89 | 0.91 | 11.33 |
| Missouri | 10.87 | 318 | 2.19 | 4.28 | 3.01 | 1.28 | 0.18 | 8.35 | -0.16 | 7.08 | 16.93 | 9.57 |

*(Table continues)*

Table 16 (continued)

| State[a] | Target variance | *N* | [MSE (FPE) – variance (target)] | | | | | [MSE (NAEP-like) – variance (target)] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Average | Rep 1 | Rep 2 | Rep 3 | Rep 4 | Average | Rep 1 | Rep 2 | Rep3 | Rep 4 |
| Nevada | 13.22 | 263 | 5.95 | 4.71 | 3.71 | 4.76 | 10.63 | 15.36 | 15.92 | 5.85 | 17.71 | 21.98 |
| New Hampshire | 6.06 | 451 | 0.60 | 0.92 | 0.06 | 1.59 | -0.18 | 0.97 | 1.22 | -0.78 | 1.91 | 1.54 |
| New Jersey | 6.70 | 359 | 1.69 | 3.04 | 1.83 | -0.10 | 1.99 | 2.75 | 0.73 | 6.17 | 2.09 | 2.02 |
| New Mexico | 11.68 | 412 | -0.26 | 1.11 | -1.05 | -0.30 | -0.81 | 2.71 | 2.08 | 4.49 | 4.91 | -0.65 |
| New York | 6.62 | 437 | 1.27 | 1.20 | 0.98 | 0.73 | 2.15 | 3.28 | 2.23 | 5.60 | 2.39 | 2.91 |
| North Carolina | 6.66 | 567 | 1.15 | -0.49 | 2.82 | 2.28 | 0.00 | 3.40 | 5.68 | 7.21 | 1.50 | -0.80 |
| North Dakota | 5.77 | 332 | 2.62 | 4.10 | 1.70 | 3.05 | 1.63 | 2.41 | 3.47 | 1.70 | 0.33 | 4.14 |
| Ohio | 15.19 | 295 | 3.71 | -2.20 | 2.41 | 2.33 | 12.31 | 9.12 | 7.94 | 9.19 | 5.10 | 14.25 |
| Oregon | 7.43 | 344 | 2.20 | 1.72 | 2.79 | 1.82 | 2.47 | 5.21 | 1.93 | 7.66 | 7.34 | 3.91 |
| Rhode Island | 6.59 | 518 | 0.06 | 0.88 | -0.84 | 0.85 | -0.66 | 0.94 | 1.25 | 1.56 | 0.36 | 0.61 |
| South Carolina | 8.64 | 337 | 2.73 | 2.89 | 3.18 | 3.18 | 1.67 | 3.34 | 4.09 | 5.20 | 0.03 | 4.05 |
| Tennessee | 22.04 | 352 | 0.72 | 1.74 | -4.19 | 3.48 | 1.83 | 3.48 | -0.04 | -3.00 | 11.26 | 5.70 |
| Texas | 11.10 | 434 | 2.73 | -2.12 | -1.01 | 5.61 | 8.46 | 7.60 | 6.06 | 4.14 | 5.30 | 14.88 |
| Utah | 5.79 | 389 | 0.97 | -0.81 | 2.67 | 0.09 | 1.91 | 4.14 | 4.64 | 2.81 | 6.44 | 2.66 |
| Vermont | 6.81 | 300 | 2.24 | 4.86 | 0.14 | -0.33 | 4.27 | 4.92 | 7.81 | 7.47 | -1.09 | 5.46 |
| Virginia | 17.72 | 207 | 3.47 | 6.72 | 6.81 | 6.09 | -5.73 | 26.82 | 59.81 | 17.55 | 8.01 | 21.90 |
| Washington | 5.85 | 353 | 3.75 | 4.73 | 5.51 | 1.33 | 3.42 | 8.92 | 0.22 | 14.22 | 17.21 | 4.05 |
| Wisconsin | 9.22 | 300 | 2.62 | 3.40 | 5.11 | 1.43 | 0.55 | 6.38 | 4.34 | 2.71 | 7.32 | 11.16 |
| Wyoming | 3.94 | 350 | 1.15 | -0.10 | 1.02 | 2.20 | 1.49 | 1.56 | -0.09 | 1.82 | 2.39 | 2.12 |

*Note.* † = not applicable. The SD-all category includes students classified as SD and students classified as both SD and ELL. FPE = full-population estimation. MSE = mean square error. *N* = number of students. Rep = replicate. Table entries for Mississippi are exceptionally large due to small sample fluctuations. SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment. Authors' calculations.

[a] Forty-two states with the state achievement test score as a school-level sampling variable were included in the study.

**Table 17**

***Differences Between FPE MSE and Target Variance and Differences Between NAEP-Like MSE and Target Variance, for ELL-Only Students, Reading, Grade 4, by State: 2003***

| State[a] | Target variance | *N* | [MSE (FPE) – variance (target)] | | | | | [MSE (NAEP-like) – variance (target)] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Average | Rep 1 | Rep 2 | Rep 3 | Rep 4 | Average | Rep 1 | Rep 2 | Rep 3 | Rep 4 |
| Average | † | † | 32.82 | 10.59 | 13.43 | 34.96 | 72.30 | 30.37 | 25.35 | 16.77 | 27.63 | 51.72 |
| Alabama | 93.58 | 14 | 58.89 | 32.72 | 29.38 | 47.20 | 126.25 | 93.64 | 4.49 | 149.24 | 126.14 | 94.68 |
| Alaska | 22.41 | 310 | -0.05 | 0.13 | 0.19 | -0.58 | 0.04 | 1.06 | 0.50 | 0.20 | 1.35 | 2.20 |
| Arizona | 6.24 | 595 | 0.79 | 0.05 | 2.27 | 0.33 | 0.49 | 0.84 | 0.67 | 1.28 | 0.20 | 1.18 |
| Arkansas | 18.04 | 60 | 8.17 | -5.79 | 6.03 | 14.31 | 18.12 | 11.51 | 8.69 | 8.82 | 16.91 | 11.61 |
| California | 2.25 | 2,712 | 0.44 | 0.05 | 0.82 | 0.53 | 0.35 | 0.16 | -0.02 | 0.36 | 0.03 | 0.24 |
| Colorado | 7.38 | 201 | 2.00 | 2.60 | 1.89 | 0.55 | 2.97 | 1.78 | 1.33 | 0.38 | 3.41 | 1.99 |
| Connecticut | 22.22 | 49 | 31.13 | 46.64 | 13.42 | 48.29 | 16.18 | 33.71 | 23.51 | 26.90 | 25.65 | 58.78 |
| Delaware | 31.56 | 36 | 24.49 | 15.42 | 4.85 | 65.60 | 12.09 | 50.54 | 60.66 | 22.16 | 72.56 | 46.79 |
| Florida | 9.57 | 257 | 1.70 | 0.66 | 2.59 | 3.21 | 0.34 | 3.77 | 4.17 | 5.96 | 4.89 | 0.07 |
| Georgia | 44.57 | 99 | 7.35 | 5.15 | 16.49 | 18.47 | -10.72 | 5.05 | 7.90 | 20.26 | -0.18 | -7.80 |
| Hawaii | 33.73 | 142 | 6.55 | 25.53 | -3.14 | -6.84 | 10.66 | 3.31 | 14.23 | 1.38 | -4.71 | 2.35 |
| Idaho | 12.11 | 171 | 0.45 | -1.14 | -0.19 | 1.03 | 2.08 | 1.64 | 3.04 | 0.13 | 2.82 | 0.56 |
| Illinois | 11.21 | 341 | -0.42 | -1.12 | -0.62 | 0.85 | -0.78 | 0.70 | 5.27 | -0.55 | -1.85 | -0.07 |
| Indiana | 42.25 | 47 | 8.96 | 6.95 | 2.32 | 23.69 | 2.86 | 15.99 | 18.12 | 20.09 | 17.90 | 7.83 |
| Kansas | 27.08 | 60 | 10.16 | 9.08 | 6.33 | 20.02 | 5.23 | 13.51 | 7.59 | 13.89 | 19.39 | 13.19 |
| Louisiana | 237.57 | 18 | 25.88 | 55.59 | -89.47 | 82.81 | 54.59 | 153.65 | 250.16 | 168.44 | 119.48 | 76.54 |
| Maine | 60.01 | 18 | 7.54 | 7.36 | 20.16 | 6.86 | -4.23 | 7.29 | 6.62 | 12.11 | 9.70 | 0.72 |
| Maryland | 79.61 | 70 | -6.93 | -12.47 | -1.14 | 15.75 | -29.86 | 31.89 | 31.43 | 15.43 | 83.34 | -2.66 |
| Massachusetts | 14.08 | 217 | 5.23 | 15.23 | 1.51 | 2.20 | 1.97 | 4.52 | 5.11 | 0.26 | 6.51 | 6.20 |
| Michigan | 49.44 | 130 | 4.07 | -4.17 | 20.37 | -4.45 | 4.52 | 7.01 | 6.93 | 5.03 | 4.44 | 11.65 |
| Minnesota | 7.16 | 183 | 1.05 | 1.81 | 0.66 | 1.37 | 0.36 | 1.28 | 3.68 | 0.16 | -0.44 | 1.71 |
| Mississippi | 382.21 | 4 | 946.14 | 33.39 | 93.18 | 929.93 | 728.05 | 418.88 | 172.21 | 237.82 | 321.22 | 419.91 |
| Missouri | 56.49 | 22 | 34.24 | 1.21 | 60.11 | 44.17 | 31.49 | 76.40 | 118.07 | 83.40 | 96.10 | 8.02 |

*(Table continues)*

Table 17 (continued)

| State[a] | Target variance | *N* | [MSE (FPE) – variance (target)] Average | Rep 1 | Rep 2 | Rep 3 | Rep 4 | [MSE (NAEP-like) – variance (target)] Average | Rep 1 | Rep 2 | Rep 3 | Rep 4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nevada | 15.32 | 327 | -1.35 | -3.44 | -4.10 | -1.52 | 3.67 | 2.76 | 4.42 | -1.50 | 2.43 | 5.69 |
| New Hampshire | 31.13 | 54 | 0.20 | 2.08 | -7.26 | -0.75 | 6.72 | 12.77 | 25.66 | -7.29 | 15.21 | 17.50 |
| New Jersey | 20.96 | 72 | 15.91 | 15.46 | 14.10 | 7.17 | 26.89 | 17.57 | 18.22 | 19.79 | 19.24 | 13.02 |
| New Mexico | 8.24 | 575 | -1.45 | -1.38 | -2.20 | -1.44 | -0.78 | 0.36 | 1.15 | -0.53 | 1.35 | -0.53 |
| New York | 21.31 | 123 | 20.13 | 18.73 | 21.46 | 25.22 | 15.12 | 19.91 | 14.40 | 13.05 | 33.30 | 18.91 |
| North Carolina | 19.02 | 147 | 4.39 | 8.69 | 0.07 | 8.48 | 0.32 | 2.74 | 3.15 | -0.50 | 8.68 | -0.37 |
| North Dakota | 20.99 | 70 | 4.11 | 5.70 | 3.21 | 5.70 | 1.81 | 1.83 | 0.39 | 1.93 | 2.61 | 2.37 |
| Ohio | 126.04 | 44 | 51.79 | 47.88 | 178.26 | 52.06 | -71.03 | 127.55 | 178.12 | 120.89 | 36.84 | 174.35 |
| Oregon | 7.35 | 272 | 0.87 | 3.61 | -1.03 | -0.90 | 1.81 | 2.74 | 3.81 | 3.16 | 0.89 | 3.11 |
| Rhode Island | 23.43 | 167 | 0.20 | -1.60 | 2.01 | 5.41 | -5.02 | -1.62 | -0.19 | -0.54 | -0.43 | -5.32 |
| South Carolina | 59.66 | 30 | 53.91 | 28.83 | 128.86 | 28.75 | 29.22 | 77.40 | 0.11 | 135.99 | 84.99 | 88.52 |
| Tennessee | 76.51 | 37 | 19.74 | 61.72 | 22.77 | -9.29 | 3.78 | 4.29 | 44.21 | -3.34 | -13.47 | -10.22 |
| Texas | 8.50 | 562 | 2.15 | 0.40 | -0.61 | 2.47 | 6.35 | 4.89 | 6.18 | 0.55 | 1.84 | 10.99 |
| Utah | 9.19 | 275 | 2.15 | -0.47 | 2.32 | 7.07 | -0.31 | 1.93 | 0.12 | 3.91 | 1.99 | 1.71 |
| Vermont | 80.93 | 28 | 8.15 | -3.55 | -2.71 | 4.62 | 34.23 | 25.08 | -10.45 | 21.35 | 21.03 | 68.38 |
| Virginia | 21.08 | 114 | 3.99 | 0.36 | -3.25 | 6.29 | 12.54 | 21.81 | -5.41 | 57.96 | 13.16 | 21.52 |
| Washington | 7.98 | 186 | 3.56 | 1.37 | 8.22 | 2.64 | 2.02 | 2.48 | 0.71 | 4.98 | 2.41 | 1.83 |
| Wisconsin | 24.26 | 115 | 9.57 | 23.07 | 9.54 | 8.39 | -2.71 | 10.64 | 23.02 | 16.03 | -0.94 | 4.45 |
| Wyoming | 12.70 | 86 | 2.58 | 2.56 | 6.26 | 2.45 | -0.97 | 2.11 | 2.62 | 0.75 | 4.36 | 0.70 |

*Note.* † = not applicable. The ELL-only category includes students classified as ELL-only. FPE = full-population estimation. MSE = mean square error. *N* = number of students. Rep = replicate. Table entries for Alabama, Louisiana, Mississippi, Ohio, and South Carolina are exceptionally large due to small sample fluctuations. SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment. Authors' calculations.

[a]Forty-two states with the state achievement test score as a school-level sampling variable were included in the study.

**Table 18**

***Bias in FPE Mean and NAEP-Like Mean for All Students, on NAEP Reporting Scale, Reading, Grade 4, by State: 2003***

| State[a] | Target mean | *N* | [Bias = FPE mean – target mean] | | | | | [Bias = NAEP-like mean – target mean] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Average | Rep 1 | Rep 2 | Rep 3 | Rep 4 | Average | Rep 1 | Rep 2 | Rep 3 | Rep 4 |
| Average | † | † | -0.03 | -0.03 | -0.03 | -0.04 | -0.02 | 1.22 | 1.23 | 1.22 | 1.23 | 1.21 |
| Alabama | 207.08 | 3,495 | 0.04 | -0.02 | -0.02 | 0.05 | 0.14 | 0.96 | 0.89 | 0.90 | 1.07 | 0.96 |
| Alaska | 211.55 | 2,712 | -0.07 | 0.05 | 0.00 | -0.14 | -0.18 | 0.89 | 1.02 | 0.82 | 0.92 | 0.80 |
| Arizona | 208.87 | 3,776 | -0.08 | -0.10 | -0.15 | -0.08 | 0.01 | 1.60 | 1.46 | 1.52 | 1.61 | 1.79 |
| Arkansas | 213.62 | 3,162 | -0.07 | -0.03 | -0.03 | -0.17 | -0.04 | 1.56 | 1.64 | 1.55 | 1.49 | 1.56 |
| California | 205.63 | 8,297 | 0.09 | 0.04 | 0.20 | -0.03 | 0.14 | 1.11 | 0.98 | 1.18 | 1.12 | 1.18 |
| Colorado | 223.66 | 3,466 | -0.04 | 0.00 | -0.03 | 0.09 | -0.22 | 1.03 | 1.08 | 0.98 | 1.15 | 0.91 |
| Connecticut | 228.34 | 3,207 | -0.09 | 0.04 | 0.01 | -0.18 | -0.22 | 1.22 | 1.35 | 1.32 | 1.15 | 1.08 |
| Delaware | 223.93 | 2,959 | 0.02 | 0.39 | 0.02 | -0.29 | -0.03 | 0.89 | 1.15 | 0.90 | 0.74 | 0.78 |
| Florida | 218.01 | 3,502 | 0.01 | -0.14 | 0.07 | 0.10 | 0.01 | 1.15 | 1.05 | 1.26 | 1.24 | 1.06 |
| Georgia | 213.60 | 5,353 | 0.02 | -0.15 | 0.01 | 0.12 | 0.08 | 0.96 | 0.86 | 1.00 | 1.03 | 0.95 |
| Hawaii | 208.26 | 3,493 | -0.09 | -0.05 | -0.17 | 0.03 | -0.18 | 1.31 | 1.39 | 1.23 | 1.40 | 1.22 |
| Idaho | 218.26 | 3,262 | -0.03 | 0.07 | -0.24 | 0.07 | -0.01 | 1.19 | 1.24 | 1.08 | 1.27 | 1.16 |
| Illinois | 216.30 | 4,864 | 0.00 | 0.09 | -0.15 | 0.05 | 0.00 | 1.76 | 1.88 | 1.69 | 1.89 | 1.58 |
| Indiana | 220.41 | 3,624 | 0.03 | 0.08 | 0.02 | 0.05 | -0.03 | 1.01 | 1.01 | 0.94 | 1.10 | 1.00 |
| Kansas | 220.14 | 3,020 | 0.00 | 0.05 | -0.04 | -0.02 | 0.00 | 0.93 | 0.96 | 0.90 | 0.92 | 0.96 |
| Louisiana | 204.73 | 2,864 | -0.01 | -0.03 | -0.08 | 0.17 | -0.09 | 1.35 | 1.28 | 1.42 | 1.48 | 1.20 |
| Maine | 223.86 | 2,735 | -0.02 | 0.31 | -0.18 | -0.22 | -0.01 | 1.31 | 1.61 | 1.29 | 1.06 | 1.28 |
| Maryland | 218.67 | 3,431 | -0.07 | -0.23 | -0.29 | 0.16 | 0.08 | 1.13 | 1.03 | 0.92 | 1.30 | 1.26 |
| Massachusetts | 227.60 | 4,396 | 0.10 | 0.18 | 0.08 | 0.05 | 0.09 | 1.14 | 1.30 | 1.06 | 1.05 | 1.14 |
| Michigan | 218.79 | 3,675 | -0.08 | -0.11 | -0.22 | 0.00 | -0.01 | 0.96 | 0.90 | 0.90 | 1.08 | 0.97 |
| Minnesota | 222.61 | 3,407 | -0.02 | -0.04 | -0.10 | -0.10 | 0.15 | 1.29 | 1.29 | 1.26 | 1.24 | 1.35 |
| Mississippi | 205.46 | 3,269 | 0.03 | 0.17 | 0.17 | -0.10 | -0.12 | 0.42 | 0.52 | 0.49 | 0.32 | 0.34 |
| Missouri | 222.26 | 3,347 | -0.06 | -0.19 | -0.09 | -0.03 | 0.05 | 1.13 | 1.06 | 1.09 | 1.22 | 1.13 |

*(Table continues)*

Table 18 (continued)

| State[a] | Target mean | *N* | [Bias = FPE mean – target mean] | | | | | [Bias = NAEP-like mean – target mean] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Average | Rep 1 | Rep 2 | Rep 3 | Rep 4 | Average | Rep 1 | Rep 2 | Rep 3 | Rep 4 |
| Nevada | 206.96 | 3,108 | 0.03 | -0.11 | 0.20 | 0.04 | 0.01 | 2.12 | 2.14 | 2.06 | 2.13 | 2.14 |
| New Hampshire | 227.79 | 3,182 | 0.01 | 0.07 | -0.03 | -0.01 | 0.02 | 1.11 | 1.22 | 1.09 | 1.03 | 1.09 |
| New Jersey | 225.07 | 3,497 | -0.08 | -0.06 | 0.05 | -0.09 | -0.22 | 1.06 | 1.08 | 1.17 | 1.02 | 0.97 |
| New Mexico | 203.19 | 2,787 | -0.07 | -0.25 | 0.11 | 0.02 | -0.15 | 1.45 | 1.33 | 1.62 | 1.49 | 1.36 |
| New York | 222.19 | 4,325 | -0.04 | -0.16 | -0.03 | 0.03 | 0.00 | 1.21 | 1.07 | 1.34 | 1.29 | 1.15 |
| North Carolina | 221.22 | 4,810 | -0.08 | -0.01 | -0.04 | -0.14 | -0.12 | 1.13 | 1.18 | 1.18 | 1.07 | 1.07 |
| North Dakota | 221.64 | 2,922 | -0.07 | -0.09 | -0.09 | -0.18 | 0.07 | 0.95 | 0.93 | 0.97 | 0.89 | 1.02 |
| Ohio | 221.87 | 4,631 | 0.08 | 0.14 | -0.07 | -0.01 | 0.27 | 1.70 | 1.69 | 1.59 | 1.70 | 1.84 |
| Oregon | 217.61 | 3,176 | -0.06 | -0.18 | -0.07 | 0.12 | -0.10 | 1.82 | 1.73 | 1.88 | 1.83 | 1.85 |
| Rhode Island | 216.49 | 3,162 | -0.07 | -0.09 | 0.03 | -0.20 | -0.01 | 1.27 | 1.26 | 1.37 | 1.16 | 1.29 |
| South Carolina | 214.81 | 3,403 | -0.12 | -0.22 | -0.01 | -0.16 | -0.11 | 0.99 | 1.02 | 1.02 | 0.93 | 0.98 |
| Tennessee | 211.95 | 3,533 | -0.12 | -0.13 | -0.13 | -0.14 | -0.07 | 0.92 | 0.90 | 0.87 | 1.00 | 0.92 |
| Texas | 214.81 | 5,067 | 0.09 | 0.22 | -0.02 | -0.19 | 0.33 | 2.00 | 2.07 | 1.93 | 1.88 | 2.12 |
| Utah | 219.27 | 3,668 | -0.08 | -0.08 | -0.14 | -0.04 | -0.05 | 1.39 | 1.33 | 1.39 | 1.43 | 1.40 |
| Vermont | 226.12 | 2,734 | -0.06 | -0.07 | -0.05 | -0.03 | -0.09 | 1.01 | 0.99 | 1.09 | 0.95 | 1.02 |
| Virginia | 223.34 | 3,308 | -0.07 | 0.03 | -0.09 | -0.16 | -0.07 | 1.17 | 1.29 | 1.19 | 1.06 | 1.15 |
| Washington | 221.10 | 3,635 | -0.02 | -0.26 | 0.19 | -0.01 | 0.01 | 1.40 | 1.27 | 1.53 | 1.56 | 1.24 |
| Wisconsin | 220.83 | 3,048 | -0.07 | -0.23 | -0.04 | 0.03 | -0.01 | 1.48 | 1.36 | 1.47 | 1.51 | 1.60 |
| Wyoming | 222.08 | 2,716 | -0.05 | -0.13 | 0.04 | -0.11 | 0.02 | 0.79 | 0.66 | 0.86 | 0.79 | 0.84 |

*Note.* † = not applicable. FPE = full-population estimation. *N* = number of students. Rep = replicate. Detail may not sum to totals because of rounding. SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment. Authors' calculations.

[a] Forty-two states with the state achievement test score as a school-level sampling variable were included in the study.

**Table 19**

***Variance Components for Target Data and Differences in Variance Components Between FPE and Target, for All Students, Reading, Grade 4, by State: 2003***

| | | | | [FPE variance components – target variance components] | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Target variance | | | Replicate 1 | | | Replicate 2 | | | Replicate 3 | | | Replicate 4 | | |
| State[a] | Total | Meas | Samp | Total | Meas | Samp | Total | Meas | Samp | Total | Meas | Samp | Total | Meas | Samp |
| Alabama | 2.98 | 0.30 | 2.68 | -0.09 | 0.02 | -0.10 | 0.08 | 0.05 | 0.03 | 0.07 | -0.01 | 0.08 | 0.14 | 0.08 | 0.06 |
| Alaska | 2.67 | 0.20 | 2.48 | 0.11 | 0.17 | -0.06 | -0.04 | 0.02 | -0.06 | 0.05 | 0.01 | 0.04 | 0.15 | 0.08 | 0.08 |
| Arizona | 1.55 | 0.08 | 1.47 | 0.09 | 0.02 | 0.06 | -0.02 | -0.05 | 0.03 | -0.05 | -0.03 | -0.02 | -0.01 | 0.06 | -0.07 |
| Arkansas | 1.91 | 0.07 | 1.84 | 0.04 | 0.06 | -0.02 | -0.12 | 0.00 | -0.13 | 0.00 | 0.04 | -0.04 | -0.06 | 0.02 | -0.08 |
| California | 1.55 | 0.20 | 1.35 | 0.09 | 0.11 | -0.02 | 0.00 | 0.10 | -0.09 | 0.05 | 0.06 | -0.01 | -0.07 | 0.00 | -0.08 |
| Colorado | 1.49 | 0.24 | 1.25 | -0.02 | -0.02 | 0.00 | 0.11 | 0.14 | -0.04 | 0.05 | 0.09 | -0.04 | 0.11 | 0.02 | 0.09 |
| Connecticut | 1.20 | 0.13 | 1.07 | -0.08 | -0.06 | -0.02 | 0.07 | 0.08 | -0.01 | 0.07 | 0.07 | 0.00 | -0.05 | 0.01 | -0.06 |
| Delaware | 0.43 | 0.11 | 0.32 | 0.02 | 0.03 | -0.01 | 0.00 | 0.02 | -0.03 | 0.04 | 0.03 | 0.00 | 0.03 | -0.03 | 0.06 |
| Florida | 1.31 | 0.01 | 1.30 | 0.00 | 0.01 | -0.01 | -0.02 | -0.01 | -0.02 | -0.10 | 0.00 | -0.11 | 0.04 | 0.04 | 0.01 |
| Georgia | 1.56 | 0.04 | 1.53 | 0.01 | 0.04 | -0.02 | 0.01 | -0.01 | 0.01 | 0.04 | 0.03 | 0.01 | 0.05 | 0.01 | 0.03 |
| Hawaii | 1.87 | 0.28 | 1.59 | -0.01 | -0.08 | 0.07 | 0.06 | 0.00 | 0.06 | -0.05 | 0.01 | -0.06 | -0.07 | -0.03 | -0.04 |
| Idaho | 1.02 | 0.09 | 0.93 | -0.04 | 0.03 | -0.07 | 0.02 | 0.00 | 0.02 | 0.00 | 0.04 | -0.04 | 0.06 | 0.05 | 0.01 |
| Illinois | 2.48 | 0.03 | 2.44 | -0.17 | 0.00 | -0.17 | -0.17 | 0.03 | -0.19 | -0.15 | 0.02 | -0.17 | -0.09 | 0.05 | -0.14 |
| Indiana | 0.95 | 0.04 | 0.91 | 0.02 | 0.02 | 0.00 | -0.03 | -0.01 | -0.02 | 0.05 | 0.06 | -0.01 | -0.04 | 0.01 | -0.06 |
| Kansas | 1.41 | 0.20 | 1.22 | 0.05 | 0.06 | -0.01 | -0.05 | -0.02 | -0.03 | -0.08 | -0.04 | -0.04 | 0.05 | 0.07 | -0.02 |
| Louisiana | 1.98 | 0.01 | 1.97 | -0.02 | 0.05 | -0.07 | -0.04 | 0.02 | -0.07 | -0.01 | 0.02 | -0.03 | -0.02 | 0.04 | -0.07 |
| Maine | 0.85 | 0.17 | 0.68 | 0.00 | 0.00 | -0.01 | 0.07 | 0.12 | -0.04 | 0.08 | 0.06 | 0.02 | -0.04 | 0.02 | -0.06 |
| Maryland | 1.98 | 0.19 | 1.79 | -0.04 | 0.12 | -0.16 | -0.15 | -0.02 | -0.13 | -0.02 | 0.05 | -0.07 | -0.05 | 0.08 | -0.14 |
| Massachusetts | 1.49 | 0.18 | 1.32 | -0.13 | -0.07 | -0.06 | 0.11 | 0.01 | 0.10 | 0.01 | 0.03 | -0.02 | -0.05 | -0.04 | -0.01 |
| Michigan | 1.40 | 0.11 | 1.29 | -0.01 | 0.01 | -0.02 | 0.05 | 0.02 | 0.03 | 0.04 | 0.05 | -0.01 | -0.09 | -0.04 | -0.05 |
| Minnesota | 1.21 | 0.23 | 0.98 | -0.05 | -0.06 | 0.01 | 0.04 | -0.01 | 0.06 | 0.02 | -0.01 | 0.02 | 0.04 | 0.07 | -0.03 |
| Mississippi | 1.82 | 0.15 | 1.66 | 0.03 | 0.07 | -0.04 | -0.03 | -0.04 | 0.01 | 0.07 | 0.06 | 0.00 | 0.02 | 0.04 | -0.02 |
| Missouri | 1.37 | 0.05 | 1.32 | -0.03 | 0.05 | -0.07 | -0.04 | 0.00 | -0.04 | -0.07 | -0.03 | -0.04 | 0.00 | 0.04 | -0.04 |

*(Table continues)*

Table 19 (continued)

| State[a] | Target variance | | | [FPE variance components – target variance components] | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Replicate 1 | | | Replicate 2 | | | Replicate 3 | | | Replicate 4 | | |
| | Total | Meas | Samp | Total | Meas | Samp | Total | Meas | Samp | Total | Meas | Samp | Total | Meas | Samp |
| Nevada | 1.54 | 0.06 | 1.47 | 0.08 | 0.05 | 0.03 | -0.16 | -0.01 | -0.15 | 0.11 | 0.09 | 0.01 | 0.19 | 0.10 | 0.09 |
| New Hampshire | 0.97 | 0.11 | 0.86 | -0.07 | -0.04 | -0.03 | 0.00 | 0.02 | -0.02 | -0.05 | -0.02 | -0.03 | -0.03 | 0.02 | -0.05 |
| New Jersey | 1.38 | 0.01 | 1.37 | 0.00 | 0.01 | -0.02 | 0.00 | 0.01 | -0.01 | -0.01 | 0.01 | -0.02 | -0.06 | 0.00 | -0.06 |
| New Mexico | 2.34 | 0.37 | 1.97 | -0.10 | 0.13 | -0.23 | -0.36 | -0.17 | -0.18 | 0.22 | 0.42 | -0.20 | -0.33 | -0.10 | -0.22 |
| New York | 1.19 | 0.02 | 1.17 | 0.04 | 0.00 | 0.04 | -0.02 | 0.02 | -0.04 | 0.00 | 0.01 | -0.01 | 0.09 | 0.03 | 0.06 |
| North Carolina | 1.04 | 0.11 | 0.93 | 0.03 | 0.04 | -0.01 | 0.09 | 0.05 | 0.04 | 0.00 | -0.04 | 0.04 | 0.07 | 0.05 | 0.02 |
| North Dakota | 0.72 | 0.04 | 0.68 | 0.00 | -0.02 | 0.02 | 0.05 | 0.03 | 0.02 | 0.11 | 0.07 | 0.04 | 0.02 | 0.03 | -0.01 |
| Ohio | 1.33 | 0.13 | 1.20 | -0.08 | 0.01 | -0.10 | -0.01 | 0.07 | -0.07 | 0.01 | 0.11 | -0.10 | -0.10 | -0.02 | -0.08 |
| Oregon | 1.69 | 0.21 | 1.47 | 0.07 | -0.02 | 0.10 | 0.10 | 0.07 | 0.03 | 0.11 | 0.09 | 0.02 | 0.27 | 0.21 | 0.06 |
| Rhode Island | 1.74 | 0.06 | 1.67 | -0.07 | 0.07 | -0.15 | -0.13 | -0.03 | -0.10 | 0.03 | 0.15 | -0.12 | -0.11 | 0.05 | -0.16 |
| South Carolina | 1.65 | 0.11 | 1.54 | -0.01 | 0.09 | -0.10 | 0.06 | 0.07 | -0.01 | -0.06 | 0.06 | -0.12 | 0.00 | 0.10 | -0.10 |
| Tennessee | 2.56 | 0.04 | 2.52 | 0.04 | 0.00 | 0.04 | -0.05 | 0.05 | -0.10 | -0.11 | 0.02 | -0.13 | -0.02 | -0.01 | -0.01 |
| Texas | 1.09 | 0.10 | 0.99 | -0.03 | 0.03 | -0.06 | -0.04 | -0.03 | 0.00 | 0.08 | 0.04 | 0.04 | 0.06 | 0.14 | -0.08 |
| Utah | 1.04 | 0.10 | 0.94 | 0.04 | 0.06 | -0.03 | 0.07 | 0.00 | 0.07 | -0.09 | -0.05 | -0.04 | 0.10 | 0.00 | 0.10 |
| Vermont | 0.83 | 0.14 | 0.69 | -0.05 | -0.07 | 0.02 | -0.02 | -0.01 | -0.01 | 0.04 | 0.12 | -0.08 | 0.16 | 0.15 | 0.01 |
| Virginia | 2.24 | 0.06 | 2.18 | 0.04 | 0.01 | 0.03 | -0.01 | 0.03 | -0.04 | -0.04 | 0.09 | -0.13 | -0.05 | 0.02 | -0.07 |
| Washington | 1.26 | 0.09 | 1.17 | -0.01 | 0.04 | -0.05 | -0.05 | 0.05 | -0.10 | 0.01 | 0.04 | -0.04 | -0.02 | -0.03 | 0.01 |
| Wisconsin | 0.72 | 0.03 | 0.68 | -0.04 | 0.00 | -0.05 | 0.00 | 0.03 | -0.03 | 0.01 | 0.03 | -0.02 | 0.00 | -0.02 | 0.02 |
| Wyoming | 0.71 | 0.04 | 0.67 | -0.02 | -0.02 | 0.00 | 0.05 | 0.03 | 0.02 | 0.08 | 0.03 | 0.05 | -0.01 | 0.00 | -0.01 |

*Note.* FPE = full-population estimation. Meas = measurement variance. Samp = sampling variance. Detail may not sum to totals because of rounding. SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment. Authors' calculations.

[a] Forty-two states with the state achievement test score as a school-level sampling variable were included in the study.

**Table 20**

***Differences Between FPE MSE and Target Variance and Differences Between NAEP-Like MSE and Target Variance, for All Students, Reading, Grade 4, by State: 2003***

| State[a] | Target variance | $N$ | [MSE (FPE) – variance (target)] | | | | | [MSE (NAEP-like) – variance (target)] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Average | Rep 1 | Rep 2 | Rep 3 | Rep 4 | Average | Rep 1 | Rep 2 | Rep 3 | Rep 4 |
| Average | † | † | 0.02 | 0.01 | 0.00 | 0.03 | 0.02 | 1.57 | 1.56 | 1.57 | 1.59 | 1.55 |
| Alabama | 2.98 | 3,495 | 0.06 | -0.08 | 0.08 | 0.07 | 0.16 | 0.84 | 0.64 | 0.83 | 1.00 | 0.91 |
| Alaska | 2.67 | 2,712 | 0.08 | 0.11 | -0.04 | 0.07 | 0.19 | 0.81 | 0.92 | 0.82 | 0.76 | 0.74 |
| Arizona | 1.55 | 3,776 | 0.01 | 0.10 | 0.00 | -0.04 | -0.01 | 2.52 | 2.17 | 2.29 | 2.52 | 3.10 |
| Arkansas | 1.91 | 3,162 | -0.03 | 0.05 | -0.12 | 0.03 | -0.06 | 2.36 | 2.57 | 2.27 | 2.27 | 2.32 |
| California | 1.55 | 8,297 | 0.03 | 0.09 | 0.04 | 0.05 | -0.05 | 1.27 | 0.99 | 1.47 | 1.23 | 1.38 |
| Colorado | 1.49 | 3,466 | 0.07 | -0.02 | 0.11 | 0.05 | 0.15 | 1.07 | 1.13 | 0.98 | 1.30 | 0.87 |
| Connecticut | 1.20 | 3,207 | 0.02 | -0.08 | 0.07 | 0.10 | 0.00 | 1.47 | 1.81 | 1.75 | 1.23 | 1.10 |
| Delaware | 0.43 | 2,959 | 0.08 | 0.17 | 0.00 | 0.12 | 0.03 | 0.81 | 1.29 | 0.78 | 0.57 | 0.60 |
| Florida | 1.31 | 3,502 | -0.01 | 0.02 | -0.02 | -0.09 | 0.04 | 1.36 | 1.19 | 1.58 | 1.48 | 1.17 |
| Georgia | 1.56 | 5,353 | 0.04 | 0.04 | 0.01 | 0.06 | 0.05 | 0.94 | 0.71 | 1.07 | 1.09 | 0.88 |
| Hawaii | 1.87 | 3,493 | 0.00 | -0.01 | 0.09 | -0.05 | -0.04 | 1.58 | 1.70 | 1.43 | 1.82 | 1.37 |
| Idaho | 1.02 | 3,262 | 0.03 | -0.04 | 0.07 | 0.01 | 0.06 | 1.40 | 1.46 | 1.22 | 1.62 | 1.29 |
| Illinois | 2.48 | 4,864 | -0.14 | -0.16 | -0.14 | -0.15 | -0.09 | 2.87 | 3.26 | 2.61 | 3.34 | 2.24 |
| Indiana | 0.95 | 3,624 | 0.00 | 0.03 | -0.03 | 0.05 | -0.04 | 1.04 | 1.07 | 0.87 | 1.24 | 0.97 |
| Kansas | 1.41 | 3,020 | -0.01 | 0.06 | -0.05 | -0.08 | 0.05 | 0.85 | 0.96 | 0.75 | 0.78 | 0.93 |
| Louisiana | 1.98 | 2,864 | -0.01 | -0.01 | -0.04 | 0.02 | -0.02 | 1.81 | 1.66 | 2.01 | 2.12 | 1.43 |
| Maine | 0.85 | 2,735 | 0.07 | 0.09 | 0.11 | 0.13 | -0.04 | 1.76 | 2.53 | 1.66 | 1.20 | 1.64 |
| Maryland | 1.98 | 3,431 | -0.03 | 0.01 | -0.07 | 0.00 | -0.05 | 1.21 | 0.95 | 0.82 | 1.67 | 1.40 |
| Massachusetts | 1.49 | 4,396 | 0.00 | -0.10 | 0.11 | 0.02 | -0.04 | 1.30 | 1.63 | 1.22 | 1.17 | 1.19 |
| Michigan | 1.40 | 3,675 | 0.01 | 0.00 | 0.09 | 0.04 | -0.09 | 0.94 | 0.85 | 0.75 | 1.22 | 0.94 |
| Minnesota | 1.21 | 3,407 | 0.03 | -0.04 | 0.05 | 0.03 | 0.06 | 1.63 | 1.68 | 1.55 | 1.44 | 1.86 |
| Mississippi | 1.82 | 3,269 | 0.04 | 0.06 | 0.00 | 0.08 | 0.04 | 0.22 | 0.27 | 0.29 | 0.17 | 0.14 |
| Missouri | 1.37 | 3,347 | -0.02 | 0.01 | -0.03 | -0.07 | 0.00 | 1.19 | 0.95 | 1.10 | 1.42 | 1.29 |

*(Table continues)*

Table 19 (continued)

| State[a] | Target variance | *N* | [MSE (FPE) – variance (target)] | | | | | [MSE (NAEP-like) – variance (target)] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | verage | Rep 1 | Rep 2 | Rep 3 | Rep 4 | Average | Rep 1 | Rep 2 | Rep 3 | Rep 4 |
| Nevada | 1.54 | 3,108 | 0.07 | 0.09 | -0.12 | 0.11 | 0.19 | 4.37 | 4.45 | 4.04 | 4.47 | 4.52 |
| New Hampshire | 0.97 | 3,182 | -0.04 | -0.07 | 0.00 | -0.05 | -0.03 | 1.19 | 1.43 | 1.11 | 1.01 | 1.20 |
| New Jersey | 1.38 | 3,497 | 0.00 | 0.00 | 0.01 | 0.00 | -0.01 | 1.13 | 1.12 | 1.40 | 1.06 | 0.92 |
| New Mexico | 2.34 | 2,787 | -0.11 | -0.04 | -0.34 | 0.22 | -0.30 | 2.08 | 1.72 | 2.67 | 2.20 | 1.74 |
| New York | 1.19 | 4,325 | 0.03 | 0.06 | -0.02 | 0.00 | 0.09 | 1.48 | 1.17 | 1.72 | 1.68 | 1.35 |
| North Carolina | 1.04 | 4,810 | 0.05 | 0.03 | 0.09 | 0.01 | 0.08 | 1.28 | 1.33 | 1.42 | 1.23 | 1.14 |
| North Dakota | 0.72 | 2,922 | 0.06 | 0.01 | 0.05 | 0.14 | 0.03 | 0.90 | 0.86 | 0.94 | 0.78 | 1.00 |
| Ohio | 1.33 | 4,631 | -0.02 | -0.07 | 0.00 | 0.01 | -0.03 | 2.64 | 2.60 | 2.25 | 2.57 | 3.12 |
| Oregon | 1.69 | 3,176 | 0.15 | 0.11 | 0.10 | 0.12 | 0.28 | 3.34 | 2.99 | 3.60 | 3.46 | 3.32 |
| Rhode Island | 1.74 | 3,162 | -0.06 | -0.06 | -0.13 | 0.08 | -0.11 | 1.49 | 1.50 | 1.68 | 1.27 | 1.50 |
| South Carolina | 1.65 | 3,403 | 0.02 | 0.04 | 0.06 | -0.03 | 0.01 | 0.96 | 1.00 | 1.10 | 0.82 | 0.92 |
| Tennessee | 2.56 | 3,533 | -0.02 | 0.06 | -0.03 | -0.09 | -0.01 | 0.82 | 0.79 | 0.67 | 1.03 | 0.80 |
| Texas | 1.09 | 5,067 | 0.07 | 0.02 | -0.04 | 0.12 | 0.17 | 3.98 | 4.24 | 3.67 | 3.48 | 4.52 |
| Utah | 1.04 | 3,668 | 0.04 | 0.04 | 0.09 | -0.08 | 0.11 | 1.90 | 1.76 | 1.94 | 1.96 | 1.93 |
| Vermont | 0.83 | 2,734 | 0.04 | -0.05 | -0.02 | 0.04 | 0.16 | 1.05 | 0.99 | 1.18 | 0.89 | 1.12 |
| Virginia | 2.24 | 3,308 | 0.00 | 0.04 | 0.00 | -0.01 | -0.05 | 1.34 | 1.61 | 1.43 | 1.01 | 1.32 |
| Washington | 1.26 | 3,635 | 0.01 | 0.05 | -0.01 | 0.01 | -0.02 | 1.94 | 1.52 | 2.30 | 2.40 | 1.54 |
| Wisconsin | 0.72 | 3,048 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 2.18 | 1.80 | 2.10 | 2.27 | 2.55 |
| Wyoming | 0.71 | 2,716 | 0.03 | 0.00 | 0.05 | 0.09 | -0.01 | 0.60 | 0.36 | 0.71 | 0.67 | 0.65 |

*Note* .† = not applicable. FPE = full-population estimation. MSE = mean square error. *N* = number of students. Detail may not sum to totals because of rounding. SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2003 Reading Assessment. Authors' calculations.

[a] Forty-two states with the state achievement test score as a school-level sampling variable were included in the study.

## 4. Discussion

The analyses presented in this report contribute to our understanding of the problems with employing the current method for estimating state NAEP statistics, when the target shifts from all students who could be assessed by NAEP to all students. They also shed some light on the advantages and disadvantages of possible remedies. Specifically, we noted that McLaughlin's comparisons of state test scores for classified students who were excluded from NAEP with those of classified students who were not excluded, indicated that the former performed more poorly (on average). Since NAEP scores and state test scores are positively correlated at the school level, a plausible inference is that the test scores of excluded classified students are not missing completely at random (MCAR); rather, the more poorly a student would perform on NAEP, the more likely that student is to be excluded.

We approached the issue somewhat differently. For 42 states, we categorized all classified students by a pair of characteristics derived from the questionnaire that is filled out for each such student. For each state, we found that there were substantial differences in exclusion rates among the different categories of the resulting matrix and, moreover, that these rates were strongly negatively correlated with the mean scores of the assessed students. Again, the implication is that classified students' scores are not MCAR. Obviously, our findings are consistent with and, in a sense, account for those of McLaughlin referenced above.

We took the argument a step further by calculating indirectly standardized exclusion rates and found that they were substantially less variable than the observed exclusion rates. We concluded that the differences among states in aggregate exclusion rates for both SD and ELL students could not be explained simply by differences in the characteristics of these students. Together, these results support the assertion that straightforward comparisons of NAEP results among some states are subject to bias.

It is important to note that even if the indirectly standardized exclusion rates had tracked the observed exclusion rates, the problem of bias remains. The difficulty is that even if students are excluded at random conditional on their characteristics (i.e., missing at random or MAR), then they are not MCAR, which is necessary for the current NAEP procedure to perform well with a different target population. Thus, in a situation in which students are excluded according to a MAR process, we should expect to see some bias in the estimates derived from a NAEP-like procedure. Indeed, this is what was observed in Condition 1 of the HumRRO simulation and in our simulation, as well.

How serious is the problem of bias? A simple but indirect approach to answering this question is to look at the difference between the actual number of excluded students and the number that would have been excluded had states experienced uniform category-specific exclusion rates. The results obtained in section 2 suggest that for many states the differences are relatively large.

A more direct approach employs simulations. The report by Wise et al. (2006) documented the improvement in MSE realized by using FPEs rather than the current NAEP procedure when the data are MAR. Again, this is consistent with our results. When student scores are not even MAR (Conditions 2 and 3 of the HumRRO simulation), both the FPE and current NAEP methods yield biased estimates—as expected. However, the comparative advantage (with respect to the MSE criterion) of FPEs is even greater in these conditions.
In view of these findings, several issues remain for further consideration. We believe there are three categories of issues: data, methodology, and policy.

### *4.1 Data Issues*

The generation of PPV for excluded students is done separately for SD and ELL students. Those students who are classified as both SD and ELL can be combined with either group. In earlier work, McLaughlin (2000, 2001, 2003) included those students with the SD group. We adopted that choice for our simulation. More recently (McLaughlin, 2005), the recommendation is made to include them with the ELL group because that group tends to be smaller. The decision is arbitrary, but not inconsequential. Using our simulation, we obtained PPV for this SD/ELL group with both choices. The differences in mean PPV by state are substantial; indeed, for most states the squared difference in the means is larger than all but one of the variance components used to generate the PPV. Although the number of students in this group is quite small for most states, it is necessary to make a principled and defensible decision on how to treat this group of students.

The FPE methodology proposed by McLaughlin and the one developed in this report differ in their treatment of missing data on the predictors and the differences should be resolved before further empirical studies are carried out.

In the present approach, a school-level variable that reflects the average score of the school on the state test is used as a predictor. This variable is used in the selection of the NAEP school sample. (For states in which this variable is not available, a variable related to the median income of the school's ZIP code can be employed instead.) McLaughlin does not make use of this variable, with the rationale that use of state test data might be regarded as contaminating NAEP results. Again, this difference should be resolved.

### *4.2 Technical Issues*

*Prediction.* The choice of the regression model is critical to the generation of PPV. The methodology for predictor selection currently favored by McLaughlin (2005) and the one presented in this report are similar but not identical. As pointed out earlier, both methods suffer from the possibility that the estimated regression coefficients are biased. One remedy would be to incorporate the background data directly into the conditioning model, as has been done for the NAEP 2007 assessments. This change reduces the effort required to produce FPEs in real time. However, this can result in substantial regression to the mean for those groups without cognitive data.

*Variance estimation.* In the HumRRO simulations, the variances of McLaughlin's FPEs were typically only slightly greater than the corresponding variances of the estimates based on the complete data. One might expect that the lack of cognitive data would manifest itself in a greater price paid in terms of variance. A concern, then, is whether the variances used to generate the PPV are sufficiently large; that is, whether all sources of uncertainty have been properly taken into account. This is critical because once the PPV are generated, the standard NAEP variance estimation machinery is applied, so that PPV are treated as if they were PV. Accordingly, the uncertainty must be built into the process that generates the PPV.

In this regard, there are at least two questions deserving further study.

1. The formula for the variance of a point on a regression plane is based on the assumption of a fixed matrix of predictors. In this setting, the matrix is actually a realization from a distribution of such matrices, which is induced by the sampling of students and schools. It is not clear whether that variability is somehow already accounted for and, if not, whether it is of sufficient magnitude to affect the results.
2. A close analysis of the results of the simulation reveals that, for excluded students, the average variance between PPV within students is larger than the average variance

between PV within students. The direction of the relationship is reasonable in view of the fact that the latter are derived from a model that includes cognitive data. (There is no basis to judge whether the magnitude of the difference is plausible.) However, the jackknife estimate of the variance due to sampling for the FPE is typically smaller than the corresponding estimate for the estimate based on the complete data. At this juncture, it is not clear whether this is a reasonable result and, if not, what would be an appropriate remedy. One suggestion is that this is due to the fact that imputing PV in the manner described here reduces the clustering in the sample.

### *4.3 Policy Issues*

An approach based on FPEs attempts to level the playing field by imputing PV for all excluded students. This constitutes a material change in the target population, which could not be adopted without extensive discussion—and eventual acceptance by the National Assessment Governing Board. However, conceptual analysis and empirical results together indicate that neither the current NAEP procedure nor the FPE constitutes an ideal solution. The former because it assumes that all excluded students could not meaningfully participate in NAEP and the latter because it implicitly assumes that all students could obtain a meaningful NAEP score.[13] In a sense, these two approaches are located at opposite ends of a continuum of possible procedures and one can surmise that a strategy superior to either can be found somewhere along that continuum. That strategy may well produce estimates that are closer to those of the FPE than those of the current procedure.

What might such an approach look like? One alternative would be to generate the PPV for all excluded students, but allow each state to exclude a fixed percentage, say 10%. Another, more complex alternative would recognize that the populations of classified students do differ from state to state with respect to the prevalence of characteristics associated with their ability to meaningfully participate in NAEP. Consequently, another alternative would be to allow states to exclude a certain percentage of students having a particular combination of characteristics, with the percentage varying by combination.

For example, in the simulation presented in the previous section, students were placed in 1 of 10 categories. In general, the categories differed both in exclusion rates and the average score for included students. In principle, it would be possible to set a maximum exclusion rate for each category based, in part, on the observed distribution of exclusion rates for that category across states. That would be relatively easy to do if the category-specific exclusion rates did not vary greatly. Unfortunately, they do, and so setting such maximums would require making choices that could be regarded as somewhat arbitrary. Thus, even if most observers were to agree that the current NAEP estimates are problematic, finding a consensus alternative is far from automatic. Ultimately, the difficulty is that it is well-nigh impossible to devise a solution that would be regarded as fair by all stakeholders.

Any attempt to change the target population must reckon with a number of challenges: (a) Communicating change can be difficult and there is bound to be confusion, as well as charges from some quarters that it is politically motivated; (b) since it is impossible to identify an optimal procedure, the choice of an alternative will involve both technical considerations and value judgments, each of which can be criticized on some basis; (c) a change will elicit a variety of reactions from the jurisdictions. How they respond can materially affect the integrity of the new procedure. Our hope is that this report will provide a foundation for further work in this area that will lead eventually to a new approach to reporting NAEP results.

## References


Cohen, J. E. (1986). An uncertainty principle in demography. *The American Statistician*, *40*, 32–39.

Ford, B. L. (1983). An overview of hot-deck procedures. In W. G. Madow, I. Olkin, & D. B. Rubin (Eds.), *Incomplete data in sample surveys* (pp. 185–207). New York: Academic Press.

McLaughlin, D. (2000, June). *Protecting state NAEP trends from changes in SD/ELL inclusion rates*. Unpublished manuscript, American Institutes for Research, Palo Alto, CA.

McLaughlin, D. (2001, November). *Exclusions and accommodations affect state NAEP gain statistics: Math grade 4, 1996 to 2000*. Unpublished manuscript, American Institutes for Research, Palo Alto, CA.

McLaughlin, D. (2003, April). *Full-population estimates of reading achievement gains between 1998 and 2002*. Unpublished manuscript, American Institutes for Research, Palo Alto, CA.

McLaughlin, D. (2005, July). *Properties of NAEP full-population estimates*. Unpublished manuscript, American Institutes for Research, Palo Alto, CA.

Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, *56*, 177–196.

Mosteller, F., & Tukey, J. W. (1977). *Data analysis and regression: A second course in statistics*. Reading, MA: Addison-Wesley.

Wise, L. L., Hoffman, R. G., & Becker, D. E. (2006, April). *Testing NAEP full-population estimates for sensitivity to violation of assumptions: Phase II*. Unpublished manuscript, HumRRO, Alexandria, VA.

## Notes

[1] This research was carried out while Henry Braun was a distinguished presidential appointee at ETS.

[2] McLaughlin (2005) modified this classification in his most recent analysis. Students who are both SD and ELL are now included with the ELL group, called *ELL all*. Thus, the two new groups are SD only and ELL all.

[3] The term *hot-deck procedure* refers to a class of stochastic mechanisms for generating missing data (Ford, 1983)

[4] Since the conditioning model that generates the PV does not include some the student characteristics derived from the questionnaires filled out for the SD and ELL students, there is a possibility of bias in the estimates of the regression coefficients of the variables based on those characteristics (Mislevy, 1991). This issue is considered further in section 4.

[5] Although one can surmise how differences in practices at the school and district levels may arise, the causes of these putative systematic differences are not particularly germane at this juncture.

[6] The characteristics are the grade level of instruction and the severity of the disability. These characteristics were selected on the basis of their strong association with exclusion rates and NAEP performance.

[7] The characteristics are the grade level of instruction and the number of years of instruction in English. Again, these characteristics were selected on the basis of their strong association with exclusion rates and NAEP performance.

[8] Since the distributions of the exclusion probabilities are non-normal, we employ the interquartile range rather than the standard deviation as a measure of dispersion.

[9] The homogeneity is with reference to the two characteristics used to classify the SD and ELL students.

[10] It is possible, but unlikely, that one would reach a different finding with other pairs characteristic variables, or by further subdividing the sample. See Cohen (1986) for a relevant analysis.

[11] For Condition 1, a propensity score model was estimated based on the original NAEP data and that model was then applied to the completed data in order to select those records marked for deletion of the cognitive data.

[12] The variance component $V_k^{(1)}$, which was employed in the present simulation, was not used for the version of this method evaluated in the HumRRO simulation.

[13] To the extent that there are students enrolled in public schools that cannot meaningfully participate in NAEP, the imputation of PV for those students based on the relationships between NAEP performance and student characteristics for assessed students (SD and/or ELL) is based on a counter-factual.