NISS

Exploring Statistical Adjustment of Results from the National Assessment of Educational Progress

Valerie S.L. Williams

Technical Report Number 102
October 1999

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

Exploring Statistical Adjustment of Results from the National Assessment of Educational Progress

Valerie S. L. Williams
North Carolina Central University

ABSTRACT

Motivated by a concern for improper comparisons of states based on published test score aggregates, a series of statistical adjustments was applied to data from the 1990 and 1992 Trial State Assessments (TSAs) of the National Assessment of Educational Progress (NAEP) Using the method of direct standardization, the analyses adjusted for demographic factors, home background variables, and school-related variables. The differences between unadjusted and adjusted NAEP averages were small, the mean overall change for states being a one-point increase on the NAEP mathematics score scale. The largest single difference was an increase of ten NAEP scale points for California for adjustment based on three student-reported variables related to student's home background (number of parents living in the household, responses to the NAEP background question "How often is another language spoken in your home?," and a home environment composite variable describing the number of reading materials available in the home). The states of the Southeastern US showed the greatest average increases across all adjustments, followed by the Western states, the states in the Northeast, and finally the Central US states. The adjustment analyses resulted in a generally homogenizing effect, that is, differences among states were reduced. Although the results of this exploration are preliminary and the continued investigation of plausible adjustment variables is advised, it is recommended that reports containing unadjusted state NAEP means be augmented with adjustment results based on select variables that are generally considered beyond the control of the school and school systems.

This research was supported in part by a grant from the National Science Foundation (#RED-9452861) and the National Center for Education Statistics (U.S. Department of Education) through the American Educational Research Association. The opinions presented reflect those of the author and do not necessarily represent those of the granting agencies. I thank Daniel G. Horvitz, Lyle V. Jones, Nada Ballator, John Dossey, Karol Krotki, and Alex Sedlacek for their very helpful suggestions in the course of this research.

Correspondence concerning this paper should be addressed to: Valerie S. L. Williams, North Carolina Central University, Department of Psychology, P.O. Box 19675, Durham NC 27707; electronic mail may be sent to *vslw@wpo.nccu.edu*.

Exploring Statistical Adjustment of Results from the National Assessment of Educational Progress

Education indicators are often compared across school districts, states, and even countries. However, to the extent that participating jurisdictions are not comparable, conclusions based on comparisons of educational achievement data may be subject to question (Glaser & Linn 1992). Precisely what inferences are to be made determines the nature of comparability required of sampled jurisdictions. Some classes of inferences — for example, about differences in average achievement for extant population subgroups — simply require that certain methodological issues (e.g., varying test administration procedures or differential participation rates) be considered. Possible differences in the demographic composition of populations may be incidental when the intent of the measurement is merely to describe those populations.

However, one primary motive for conducting comparative educational studies is to allow participating jurisdictions to identify policies and practices that account for observed differences in the achievement levels of students. For a variety of reasons, Rotberg (1995) considers test score comparisons between nations, states, or schools highly misleading measures of the quality of education. Inferences about the effects of alternative policies or treatments may require substantial adjustment of the data to help unconfound the variables that are related to achievement but not influenced by the educational system. It is necessary to control these nuisance variables because they exert some impact on the measure of achievement but are not distributed in a balanced way over treatment or policy groups.

The degree to which the comparability of the populations influences results depends on the extent to which the distributions of variables correlated with educational achievement differ across the populations. Comparison groups may differ with respect to the distributions of background measures not under the control of their educational systems. These background factors include demographics, such as socioeconomic status and language spoken in the home, as well as student behavioral and attitudinal variables. Behavioral variables are student activities and habits outside of school that contribute to learning; attitudinal variables are educationally-relevant student and parent values and attitudes. In addition, educational jurisdictions can differ with respect to the availability of certain resources, such as materials and equipment and highly qualified teachers. Furthermore, education policy and process variables — including curriculum and classroom instructional practices — play an important role in student learning. The latter two classes of variables are to some degree regulated by school system policy and are critical in the evaluation of the quality of an educational program (Dossey 1993).

Adjustments to education-related statistics have had only limited application in educational surveys (Selden 1990; Tukey, Abelson, Coffman, Gilbert, Jones, & Mosteller 1971; Wainer 1994; Wainer & Kulick 1997). Wainer (1994) used direct standardization to adjust mathematics scores from the 1992 Trial State Assessment (TSA) to reflect the population distribution of the nation with respect to three demographic characteristics known to be related to educational achievement: Race/ethnicity, Community type, and classification of Limited English Proficiency. His analyses demonstrated a profound disagreement between inferences from comparisons of unadjusted state averages as opposed to inferences based on adjusted state averages.

However, some commentators have criticized the notion of correcting observed achievement levels for differential variation in factors correlated with achievement (cf. Finn 1994). An important argument against adjusting test score statistics is that these adjustments may imply an expectation of fundamental differences in ability across subgroups, or may somehow rationalize lower expectations and standards for different groups of students or jurisdictions. However, the fact remains that there are variables beyond the control of the school system which are highly correlated with achievement test scores and which are also differentially distributed across states or other reporting jurisdictions.

Comparisons of crude achievement test scores are confounded with differences in these intervening variables and, therefore, will not necessarily reflect the relative quality or efficacy of the education systems.

This study explores different sets of adjustment variables and their effects on state NAEP mathematics means. The objectives are to advance an acceptable level of comparability in the educational outcomes of the NAEP state assessment and facilitate informed comparisons among the diverse student populations in participating states. In epidemiology, death rates are often adjusted for age and gender; however, in educational measurement there is no commonly-accepted set of adjustment variables. Particularly sensitive is the use of race/ethnicity which is often employed in observational research as a proxy for underlying environmental factors that may actually be *causally* related to educational achievement. This study utilizes race/ethnicity for statistical adjustment only as a benchmark for the reader to compare the effects of other adjustment variables; the primary aim is to use variables that can logically explain differences in mathematics performance.

METHOD

Data

The restricted student-level data from the 1992 National Assessment of Educational Progress (NAEP) eighth-grade mathematics TSA were used to conduct exploratory analyses; data from the 1990 eighth-grade mathematics TSA were used in a limited way for confirmatory analyses. Since 1969, NAEP has assessed American schoolchildren, and has described educational achievement and trends in various subject areas. In the 1992 TSA, representative samples of eighth graders were surveyed in 41 states and students were assessed in six mathematical content areas (i.e., Numbers and Operations, and Geometry) measuring conceptual understanding, procedural knowledge, and problem-solving (Johnson, Mazzeo, & Kline 1993; Koffler 1991). Although many have cautioned against inferences from state NAEP data involving anything more than simple descriptions and comparisons of states (Haertel 1992; Koretz 1991), the wealth of available contextual information related to the students, their teachers and schools, and the states, can provide a richer understanding and more meaningful explication of the observed differences between states in student achievement (Phillips 1991).

The analysis of NAEP data is complicated by the fact that examinees do not receive individual mathematics achievement scores; instead of traditional point estimates of proficiency, students are assigned five different *plausible values* for each composite and subscale score (Mislevy, Beaton, Kaplan, & Sheehan 1992). This unique structure of the TSA data sets requires that five separate analyses be performed, one with each set of plausible values; for estimation and hypothesis testing, the results of the five analyses are then averaged to provide a single summary statistic.

As a strategy to aid in interpretation, the standardization analyses involve distinct sets of adjustments. For example, there are factors associated with the school and school system, such as school resource variables (i.e., teacher certification), and educational policy and process variables (i.e., frequency of testing and calculator use). There are also important factors that do not fall under the control of the school system, especially demographic characteristics and home background characteristics (i.e., parents' highest level of education). The analyses conducted can be loosely classified as Demographic Adjustments, Home Background Adjustments, and School-related Adjustments.

Demographic Adjustments. There is a set of student characteristics that are considered relatively (although, perhaps, not strictly) immutable, including student race/ethnicity, sex, community of residence, the language spoken in the home, and whether the student was born in the US or has an individualized education plan or learning disability. While the rationale for adjusting state NAEP averages based on these factors does not hold any promise for policy implications, such demographic variables are what Wainer (1994) uses "to place all states on a level playing field" (pp. 4-5) necessary for meaningful comparisons.

IEP: Based on school records, IEP indicates (Yes or No) whether a student possesses an Individualized Education Plan, reflecting a student with learning disabilities.

Sex: Primarily from school records, Sex (male or female) is provided for each student.

Race: Within the common core of the NAEP General Background questionnaire, students are asked, "Which best describes you?" with six response options ("White," "Black," "Hispanic," "Asian or Pacific Islander," "American Indian or Alaskan Native," or "Other"). From this information as well as school records, an analogous set of six Race categories was created by Educational Testing Service (ETS) for NAEP analysis and reporting purposes.

LangHome: The common core of the NAEP General Background questionnaire asks students, "How often do the people in your home speak a language other than English?" with a three-point response scale ("Never," "Sometimes," or "Always").

TOC: Based on available sampling frame information and the judgments of school principals, the type of community in which a student lived was categorized as Extreme Rural, Disadvantaged Urban, Advantaged Urban, or Other. The majority of the TOCs, described as "Other," are less extreme categories such as "Main Big City," "Urban fringe," "Medium city," or "Small place." See Rogers, Kline, Mazzeo, Johnson, Mislevy, & Rust (1990) and Rogers, Kline, Mazzeo, Johnson, Mislevy, & Rust (1992) for additional details.

USBorn: Also within the common core of the NAEP General Background questionnaire, students are asked whether or not they were born in the United States: "Were you born in one of the 50 states of the United States, in the District of Columbia, or in one of the United States territories?" Students responded "Yes" or No."

Home Background Adjustments. Mullis, Jenkins, and Johnson (1994) note that "students only spend part of their day in school" (p. 35) and demonstrate that Home Background variables of this sort differentiate between top-performing and the bottom-performing schools. The present analyses are intended to investigate the possible effects of different household environments on student math achievement, especially as possible indicators of poverty; for example, socioeconomic status can be linked to the number of parents living in the household. Data on student calculator ownership and the availability of reading materials in the home are also used here in an effort to tap possible indicators of poverty available within the NAEP secondary-use data files.

ParEd: This variable was derived by ETS from responses to two questions on the common core of the NAEP General Background questionnaire, "How far in school did your mother go?" and "How far in school did your father go?" Besides "I don't know," students could respond "She [or "He"] did not finish high school," "She [or "He"] graduated from high school," "She [or "He"] had some education after high school," and "She [or "He"] graduated from college." ParEd was coded into five categories based on the highest education level attained by either the mother or the father (the Unknown category was coded as missing data in the present analyses).

HaveCalc: The NAEP Math Background questionnaire asks students, "Do you have a calculator that you can use to do mathematics schoolwork?" with students required to respond either "Yes" or "No."

HomeEn: A composite variable available in the NAEP secondary-use data files aims to characterize each student's home environment based on the availability of relatively common reading materials. HomeEn was created by ETS by summing student responses to four items on the common core of the NAEP General Background questionnaire: "Does your family get a newspaper regularly?," "Is there an encyclopedia in your home?," "Are there more than 25 books in your home?," and "Does your family get any magazines regularly?" Students reporting "0-2 types" were combined into one HomeEn category, students reporting "3 types" constituted a second category, and students reporting "4 types" constituted a third category

SingleP: This variable describes the number of parents in the student's home and is based on student self-reports to two dichotomous NAEP General Background questionnaire items: "Does either your mother or stepmother live at home with you?" and "Does either your father or stepfather live at home with you?" Four levels of SingleP were used in the standardization analyses reported here: "Mother and father," "Father only," "Mother only," and "Neither parent."

School-related Adjustments. Mullis et al. (1994) note that "the ways in which students are taught mathematics form the foundation for the inquiry, problem solving, and communication skills that students will need throughout their lives. The curriculum and its delivery can be viewed as the distribution system for the opportunity to learn mathematics" (p. 39). A set of School-related Adjustment variables perceived by educators to play an important role in student learning was selected for standardization analyses. These variables are to some extent regulated by school and school system policy, for example, education process variables such as classroom instructional practices. In addition, educational jurisdictions can differ with respect to the availability of highly qualified mathematics teachers who majored in mathematics and are certified in mathematics education.

Tmajor: The mathematics teacher of each student participating in NAEP was asked to fill out a Math Teacher Questionnaire which includes a question pertaining to the teacher's undergraduate field of study: "What were your undergraduate major fields of study?" Included in the 1992 secondary-use data files is a variable derived from this question, with the following four classifications: Education, Mathematics, Mathematics education, or Other.

Tcert: The NAEP Math Teacher Questionnaire includes a set of questions interrogating teacher certification status: "Do you have teaching certification that is recognized by the state in which you teach in any of the following areas?" Based on responses to these questions, a teacher can be classified as being certified in Mathematics, Education, or Other.

UseCalc: Calculator use was examined with the inclusion of a question within the Math Background section of the NAEP student questionnaire: "In mathematics class, how often do you do each of the following? Use a calculator." The analyses of 1992 data used student responses coded into the four multiple-choice response alternatives: "Almost every day," "Once or twice a week," "Once or twice a month," and "Never or hardly ever." For the analysis of 1990 data, the five different multiple-choice response alternatives were coded into three reasonably similar categories: "Almost every day" alone, "Several times a week" and "About once a week" combined, and "Less than once a week" and "Never" combined.

ProbText: Another question within the Math Background section of the NAEP student questionnaire asks "In mathematics class, how often do you do each of the following? Do mathematics problems from textbooks." The four multiple-choice response alternatives of 1992 were coded into three categories: "Almost every day," "Once or twice a week," and "Once or twice a month" and "Never or hardly ever" combined. The five different multiple-choice response alternatives of the 1990 data were coded into three similar categories: "Almost every day," "Several times a week" and "About once a week" combined, and "Less than once a week" and "Never" combined.

WriteRep: How often students wrote mathematics reports or did mathematics projects was the subject of another question within the Math Background section of the NAEP student questionnaire: "In mathematics class, how often do you do each of the following? Write reports or do mathematics projects." The analyses of 1992 data used student responses coded into two categories based on the four multiple-choice response options: "Almost every day," "Once or twice a week," and "Once or twice a month," were combined, and "Never or hardly ever" was used alone. The five different multiple-choice response alternatives in the 1990 data were also coded into two categories: "Almost every day," "Several times a week," "About once a week," and "Less than once a week" were combined, and "Never" was used alone.

MathTest: This variable is based on a question also within the Math Background section of the NAEP student questionnaire which asks "In mathematics class, how often do you do each of the following? Take mathematics tests." For the 1992 analyses, responses were coded into two categories based on the four multiple-choice response alternatives: "Almost every day" and "Once or twice a week" were combined, and "Once or twice a month" and "Never or hardly ever" were combined. For the 1990 analyses, the five different multiple-choice response alternatives were coded into two categories reasonably similar to the 1992 categories: "Almost every day," "Several times a week," and "About once a week" were combined, and "Less than once a week" and "Never" were combined.

Class9th: As a starting point for investigating the effect of course-taking habits on mathematics achievement via direct standardization, a question from the Math Background section of the 1992 NAEP student questionnaire was used: "What kind of mathematics class do you expect to take in 9th grade?" Analyses reported by Mullis, Jenkins, and Johnson (1994) show that schools with average NAEP mathematics scores in the top-performing one-third of all schools assessed have students who are more likely to expect to take Algebra I or Geometry in the ninth grade, whereas schools with NAEP mathematics averages in the bottom-performing one-third have students who are more likely to report that they will take Pre-algebra or basic mathematics, general mathematics, business mathematics, or consumer mathematics (or don't know, or don't plan to take mathematics). For standardization analysis purposes, one category was created from "Algebra I or elementary algebra" and "Geometry," and a second category was created from the five remaining response alternatives: "Basic, general, business, or consumer mathematics," "Pre-algebra," "Other mathematics class," "I don't know" and "I do not expect to take a mathematics class in 9th grade." Class9th is included under School-related Adjustments primarily because it was analyzed in the context of other School-related Adjustment variables, but also because it may also reflect school curricula and tracking practices.

Statistical analyses

Multiple regression analyses of aggregated state-level data have shown that readily available demographic measures are highly predictive of average scores on the TSA (Robinson & Brandon 1994; Wolf 1992). The statistical adjustment technique, direct standardization, will be applied to student-level TSA data to help correct for differences in the distributions of the selected variables.

The unadjusted mean for North Carolina NAEP scores can be represented as

$$\underline{x}_{NC} = \Sigma p_{NCi}\underline{x}_{NCi} = p_{NC1}\underline{x}_{NC1} + p_{NC2}\underline{x}_{NC2} + p_{NC3}\underline{x}_{NC3} + \ldots + p_{NCi}\underline{x}_{NCi}\ , \qquad (1)$$

where p_{NCi} is the percentage of NC students with a value of "*i*" on the adjustment variable and $\underline{x}_{NCi}$ is the average NAEP score for the "*i*" students. North Carolina's NAEP scores can be standardized to the national population distribution by applying the NC NAEP averages in those *i* cells to the corresponding national percentages. This standardized mean for North Carolina, $\underline{x}'_{NC}$, can be represented as

$$\underline{x}'_{NC} = \Sigma p_{USi}\underline{x}_{NCi} = p_{US1}\underline{x}_{NC1} + p_{US2}\underline{x}_{NC2} + p_{US3}\underline{x}_{NC3} + \ldots + p_{USi}\underline{x}_{NCi}\ , \qquad (2)$$

which provides us with the state mean that would be expected if North Carolina possessed a population distribution similar to the nation for the adjustment variable. This statistical procedure is implemented in SUDAAN® (Shah, Barnwell, & Bieler 1997), a commercially available software package for analyzing data from complex survey designs; computational details are provided in Fleiss (1981) and Mosteller and Tukey (1977), and illustrated below.

A simple one-way example. To demonstrate the technique of direct standardization in the simplest case of a single variable with two levels, the five average plausible values for North Carolina are presented in Table 1, separately for eighth-grade students with IEPs and for students without IEPs. In the state of North Carolina, 8.9% of the eighth-grade students have learning disabilities that qualify them for an IEP, whereas for the nation only 3.3% do.

Table 1. Mean plausible values for North Carolina's eighth-grade students, with and without IEPs.

	IEP = "Yes"	IEP = "No"
Mathematics Composite Plausible Value #1	239.7358	260.1265
Mathematics Composite Plausible Value #2	241.0237	260.0619
Mathematics Composite Plausible Value #3	241.7856	259.6949
Mathematics Composite Plausible Value #4	241.3167	260.4785
Mathematics Composite Plausible Value #5	240.1721	260.3051

The unadjusted average mathematics score for North Carolina is 258.4; however, to standardize the North Carolina mathematics average to the US population percentages, we apply (2) to each of the plausible values:

$$PV_1 = (0.033)239.7358 + (0.967)260.1265 = 259.4536$$
$$PV_2 = (0.033)241.0237 + (0.967)260.0619 = 259.4336$$
$$PV_3 = (0.033)241.7856 + (0.967)259.6949 = 259.1038$$
$$PV_4 = (0.033)241.3167 + (0.967)260.4785 = 259.8462$$
$$PV_5 = (0.033)240.1721 + (0.967)260.3051 = 259.6407$$

and obtain a mean mathematics score of 259.5 by averaging across the five plausible values:

$$(259.4536 + 259.4336 + 259.1038 + 259.8462 + 259.6407) \div 5 = 259.4956 .$$

The basic formulation can be generalized to the multiway case in which two or more variables, each with two or more levels, are cross-classified. The one point increase in average eighth-grade NAEP mathematics for NC is trivial considering that the standard deviation for NAEP mathematics is approximately 37 points.

More analytic details. On self-reports where student information was illegible, omitted, or ambiguous (i.e., multiple responses), or in cases where teacher or school records were unavailable, values were set to missing for the purposes of the present analyses. In situations where cross-tabulations of national data generated missing cells, direct standardization computations would fail and further analyses were abandoned.

Standard errors for each of the five standardized mean plausible values are calculated within SUDAAN® using Taylor series linearization; however, because of the complication of the five plausible values in the NAEP data sets, an additional computation is carried out in order to obtain a single estimate of the sampling variability associated with the standardized mathematics mean for a state:

$$var(\underline{x}_{NC}) = var(PV_1) + (1 + M^{-1})\, var(PV_i), \qquad (3)$$

where $var(PV_1)$ is the sampling variance associated with the first IEP-adjusted plausible value for NC, M is the number of multiply-imputed values (here, 5), and $var(PV_i)$ is the variance of the five IEP-adjusted plausible values for NC. The standard error for NC corresponding to the IEP-adjusted mean of 259 is the square root of 9.99, or 9.99.

Similar to those presented in NAEP reports containing unadjusted state means (cf. Mullis, Dossey, Owen, & Phillips 1993), standard errors of standardization-adjusted means ranged in size from somewhat less than 1 to greater than 4 for states with greater sampling variability. Standard errors are tabled in the appendix.

Inferential approach. Consistent with the exploratory nature of this research, results are presented graphically by state within region without reference to “statistically significant differences,” although relatively large differences between unstandardized and standardized means are noted. In an attempt not to over-interpret the statistical results, increases and decreases are referred to as “relatively large” or “small and unimportant,” depending only on the magnitude of the difference between the unstandardized and standardized means. The earnest reader is free to make comparisons among states at her own discretion — of course, after determining the rationale for the hypothesis test and adjusting for multiplicity.

RESULTS

A large number of standardizations were conducted using the 1992 NAEP TSA data, with state means standardized to national proportions. Each set of standardizations — Demographic Adjustments, Home Background Adjustments, and School-related Adjustments — will be discussed separately, followed by an overview of the standardization results.

Demographic Adjustments

A set of preliminary standardization analyses were based on student demographic characteristics: student IEP status (IEP), race/ethnicity and sex (Race + Sex), community type and "How often is another language spoken in your home?" (LangHome + TOC), and community type and whether the student was born in the US (USBorn + TOC). All state means, unadjusted and adjusted, are contained in Table A-1 of Appendix A.

IEP. About 3.3% students in the nation have learning disabilities qualifying them for IEPs, states vary with respect to the proportion of their students with learning disabilities. The results for the one-way standardization investigating student IEP-status classification show that there is very little difference in unadjusted state NAEP mathematics averages compared to state averages standardized to the national proportion of students with IEPs — the average change across all states was an increase of about one NAEP point. States in the Northeast increased almost two NAEP scale points on average, and states in the Central and Southeastern and Western US, about one point. Considering that the NAEP standard deviation is about 37 points, these increases are trivial.

In Figure 1 are displayed the results of all Demographic Adjustments by US region: Central (IL, IN, IA, MI, MN, MO, NE, ND, OH, and WI), Northeast (CT, DE, ME, MD, MA, NH, NJ, NY, PA, and RI), Southeast (AL, AR, FL, GA, KY, LA, MS, NS, SC, TN, VA, and WV), and West (AZ, CA, CO, HI, ID, MT, NM, OK, OR, TX, UT, and WY). Each standardization analysis is represented by a point on the horizontal axis; the NAEP mathematics scale range for state averages, around 245 to 285, is shown on the vertical axis. The results for the standardization involving student IEP status are shown above the point labeled "IEP"and can be easily compared to the average unadjusted NAEP mathematics score for each state, labeled "1992" on the abscissa in Figure 1. The fact that the increases for IEP-standardized means are small is evident in Figure 1.

Race and Sex. Because of the well-established differences in average NAEP scores between males and females and across different race/ethnicity classifications, this analysis was conducted in order to provide a simple benchmark for comparing the results of adjustment variables with actual explanatory value. In the national NAEP sample to which the state means are standardized, 34% are white females and 35% are white males, and 8% are black males and 8% are black females.

The standardized means for states in the Central geographic region generally decreased in comparison to their unadjusted NAEP means — on average about four NAEP scale points — with both MN and NE decreasing by approximately six NAEP points. With a few exceptions, the means of Southeastern and Western states increased somewhat (on average, by about two points and one point, respectively): MS gained approximately eight NAEP scale points and LA gained about six; TX gained approximately seven NAEP scale points, and CA and NM gained about six points. Results for the states in the Northeast were mixed, but scores on average decreased by about one NAEP point. Figure 1 displays the state means adjusted for Race and Sex above the point on the horizontal axis labeled "Race + Sex." These adjusted scores reflect what states' NAEP scores would be if the same proportional representation of

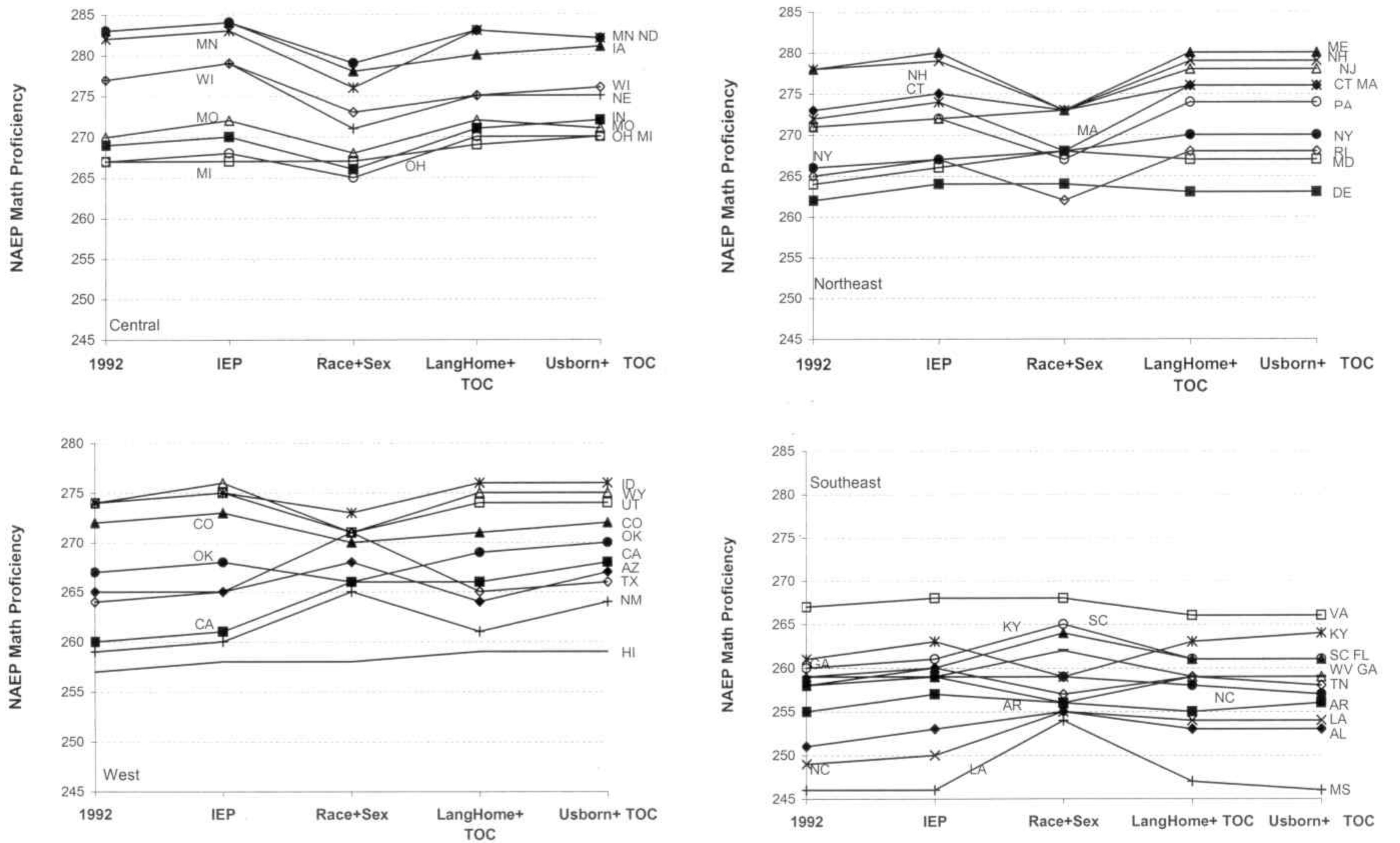

Figure 1. 1992 NAEP eighth-grade mathematics standardization results for all Demographic Adjustments by state and US region.

an adjusted mean answers the question "What would this state's average be if its demographic composition was the same as the nation?" (p. 11).

USborn and Community type. 68% of the eighth-graders in the national NAEP population live in "Other" community types and reported that they were born in the US. Figure 1 displays the state means adjusted for USborn and Community type above the point on the horizontal axis labeled "USborn + TOC." In comparison to the unstandardized NAEP means, standardized means for all states in the Northeast geographic region increased three NAEP scale points on average, with NJ increasing seven points; Western states also showed increases (the standardized mean of CA was higher by eight NAEP scale points), on average about two points; states from the Central US showed small and unimportant changes in both directions, as did the Southeastern states (the largest change was an increase of five NAEP points for LA).

"How often is another language spoken in your home?" and Community type. Adjusting NAEP means based on these two variables generally resulted in some relatively large increases for some states (NJ increased by approximately seven NAEP scale points, CA by about six, and LA by five points). Regionally, the Northeastern states averaged an increase of three NAEP points, and the Southeast and West averaged somewhat more than one point; although a few Central states showed two- or three-point decreases (IA, NE, and WI), the Central region showed no change overall because of small increases for other states (OH, IN, MI, and MO). The state means adjusted for "How often is another language spoken in your home?" and Community type are displayed in Figure 1 above the point on the horizontal axis labeled "LangHome + TOC."

Home Background Adjustments

Five standardization analyses were conducted using Home Background Adjustment variables: Parent education level and "Do you have a calculator to do math schoolwork?" (ParEd + HaveCalc), Parent education level and Home environment composite (ParEd + HomeEn), Parent education level and Home environment composite and "How often is another language spoken in your home?" (ParEd + HomeEn + LangHome), Number of parents in the household and Home environment composite (SingleP + HomeEn), and Number of parents in the household and Home environment composite and "How often is another language spoken in your home?" (SingleP + HomeEn + LangHome). Figure 2 displays the state means adjusted for the School-related factors. For comparison purposes, the average unadjusted NAEP mathematics score for each state is displayed above the point labeled "1992." Unadjusted and adjusted state means are contained in Table A-2 of Appendix A.

Parent education level and "Do you have a calculator to do math schoolwork?" This 4 × 2 standardization analysis resulted in only four states with obtained adjusted means lower than their unadjusted means — although by only a few NAEP scale points: IA (one point), MN (two points), ND (three points), and UT (two points). Standardized means for all other states showed increases, or no change. All Southeastern states showed increases of two (VA) to a maximum of six (KY) NAEP scale points with an average of nearly four points. On average states in the West and Northeast increased about two NAEP scale points overall. Figure 2 displays the state means adjusted for Parent education level and "Do you have a calculator to do math schoolwork?" above the point on the horizontal axis labeled "ParEd + HaveCalc."

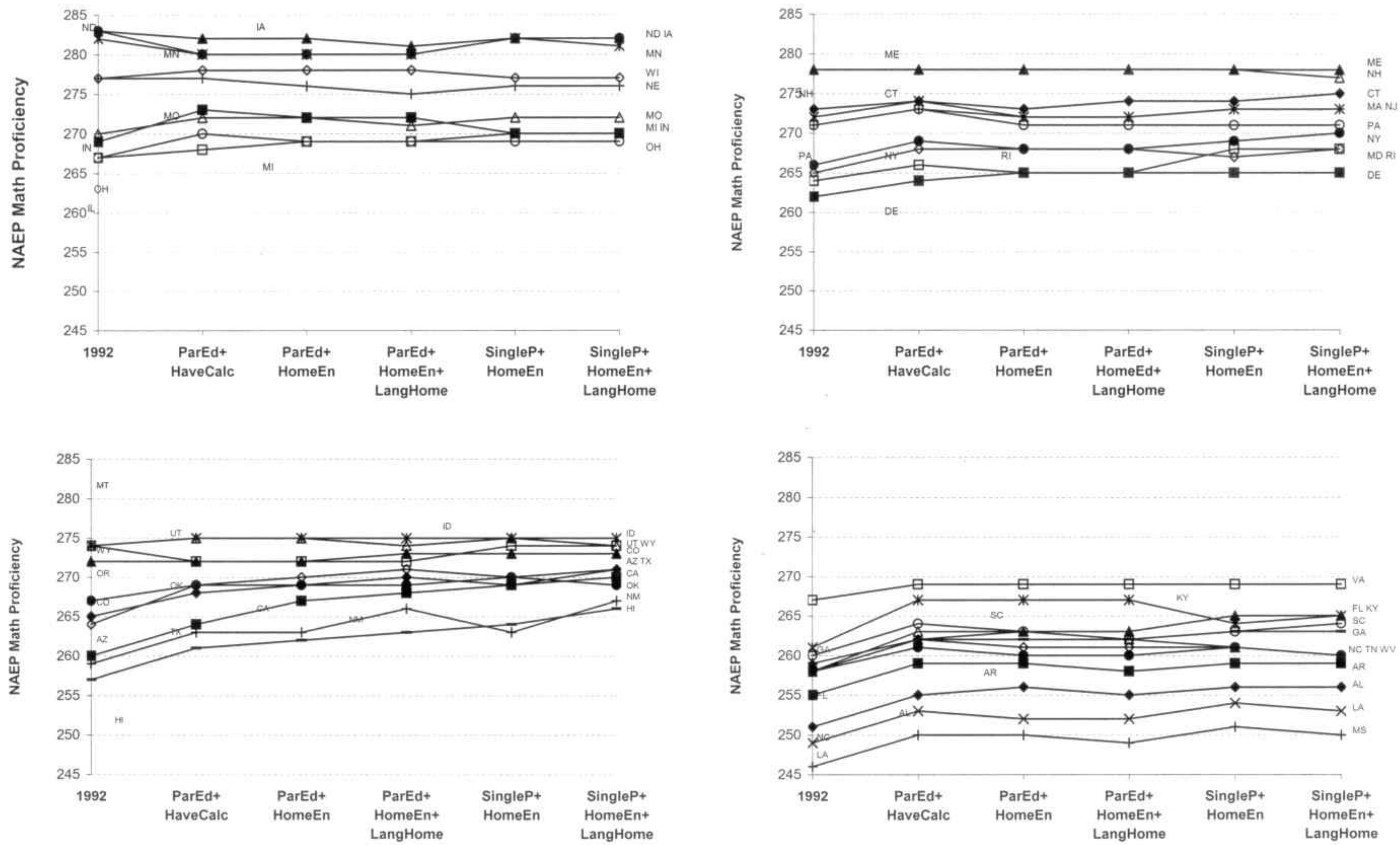

Figure 2. 1992 NAEP eighth-grade mathematics standardization results for all Home Background Adjustments by state and US region.

Parent education level and Home environment composite. The 4 × 3 standardization analyses based on these two variables show that only five states obtained adjusted means that were slightly lower than their unadjusted means, usually by only one (IA and NE), two (MN and UT), or three points (ND). The rest of the Central states demonstrated only very small increases yielding a regional average that is essentially unchanged. Adjusted means for states in the Northeast increased by about one NAEP scale point overall. However, the standardized means for all Southeastern states showed increases of at least two points (KY increased by approximately six points) resulting in a regional average increase of nearly four NAEP scale points. Three states in the Western geographical region showed standardized means that were at least five points higher than their unstandardized means: HI had an increase of about five points, TX an increase of about six points, and CA had an increase of about seven points. For the Western US, adjusted NAEP means increased by almost three points overall. The state means adjusted for Parent education level and Home environment composite are displayed in Figure 2 above the point on the horizontal axis labeled "ParEd + HomeEn."

Parent education level and Home environment composite and "How often is another language spoken in your home?" In 1992 18.3% of the nations's eighth-graders report speaking English only in their home, having all four home environment resources (newspapers, magazines, books, and encyclopedias), and a parent with a college degree; 8.5% report speaking English only in their home, having all four home environment resources, and a parent with a high school diploma.

The three-way standardization (4 × 3 × 3) of 1992 NAEP means based on these variables resulted in generally small changes for states in the Northeastern and Central US. The adjusted means of all Southeastern states increased by at least two NAEP scale points for an overall regional average of more than three points. In the West, CA increased by approximately eight points, and New Mexico and Texas each by approximately seven points; on average the Western states increased more than three NAEP scale points overall. Figure 2 displays the state means adjusted for these three variables above the point on the horizontal axis labeled "ParEd + HomeEn + LangHome."

Number of parents in the household and Home environment composite. The standardization results based on these two variables show only three standardized state means that are lower than the corresponding unadjusted means, all three in the Central region (IA, NE, and ND) but each by only about one point; the overall average increase for the Central states was less than one NAEP scale point. The standardized means in the states of the Southeast all showed increases of at least two points (AL, LA, and MS increased by approximately five points, and FL increased by about six points) for a regional average increase of nearly four NAEP scale points. Western states, too, showed a few relatively large increases in standardized NAEP means compared to unstandardized means with an average increase of almost four points: TX by about six points, HI by about seven, and CA by approximately nine points. Northeastern states increased on average somewhat less than two NAEP scale points. Figure 2 displays the state means adjusted for Number of parents in the household and the Home environment composite above the point on the horizontal axis labeled "SingleP + HomeEn."

Number of parents in the household and Home environment composite and "How often is another language spoken in your home?" 29.5% of the nations's eighth-graders reportedly live at home with both parents, speak only English in their home, and have all four home environment resources; 15.5% live in two-parent English-speaking households with three home environment resources. When state means are standardized to the national proportions for these three variables, the result is rather large (relatively speaking) differences across region. The Western geographic region increased an average of more than four NAEP scale points overall, with CA exhibiting a ten-point increase, HI a nine-point increase, NM an eight-point increase, and TX a seven-point increase; UT and WY show no change. The Southeast showed an average increase of nearly four NAEP scale points overall, the Northeast increased somewhat less than two points on average, and the Central states remained essentially the same on average. In Figure 2 are shown all state means adjusted for these three variables (Number of parents in the household and the Home environment composite and “How often is another language spoken in your home?”) above the point on the horizontal axis labeled “SingleP + HomeEn + LangHome.”

School-related Adjustments

Six standardizations used Home Background Adjustment variables: Take mathematics tests (MathTest), Math teacher undergraduate major and Certification status (Tmajor + Tcert), Math teacher undergraduate major and Certification status and Use a calculator (TMajor + TCert + UseCalc), Do problems from textbooks and Take mathematics tests and Write reports or do projects (ProbText + MathTest + Writerep), Do problems from textbooks and Take mathematics tests and Write reports or do projects and Use a calculator (ProbText + MathTest + WriteRep + UseCalc), Do problems from textbooks and Take mathematics tests and Write reports or do projects and 9th grade math class (ProbText + MathTest + WriteRep + Class9th). Figure 3 displays the state means adjusted for the School-related factors. In Figure 3 are displayed the state means adjusted for the School-related factors. Unadjusted and adjusted state means are contained in Table A-3 of Appendix A.

Take mathematics tests. Nationally, 6% of eighth-graders report that they take mathematics tests almost every day, 55% report that they take mathematics tests at least once a week, and 39% report that they take mathematics tests less than weekly. This one-way standardization analysis was intended to highlight the results of Mullis et al. (1994) which revealed that students reported more testing in the bottom-performing one-third schools compared to the top-performing one-third schools. As they note, “it may be that additional concern about the achievement of lower-performing students results in more testing, that additional time spent in testing detracts from ongoing classroom instruction, or some combination of these factors” (p. 42) that explains the association between more frequent testing and lower levels of student proficiency.

Only states in the Southeast showed changes greater than one NAEP scale point, increasing two points on average; LA increased by five NAEP scale points and AL increased by four NAEP scale points. States in the other three US geographic regions remained essentially unchanged. The results for the standardization involving Take mathematics tests are shown above the point labeled “MathTest”in Figure 3. The fact that the increases for MathTest-standardized means are generally small is evident in Figure 3.

Math teacher undergraduate major and Certification status. The possible effects of policy-governed teacher characteristics are illustrated in the standardization of NAEP means based on the undergraduate major and certification status of students’ mathematics teachers. Of the eighth-graders participating in the 1992 NAEP, approximately 44% had mathematics teachers with both an undergraduate major in mathematics and state certification to teach in mathematics.

There are a few relatively large differences between unadjusted means and means adjusted for Math teacher undergraduate major and Certification status: CA and LA increased by five NAEP scale points, PA decreased by five points, and MN decreased by seven NAEP scale points. Regionally, the West gained one NAEP scale point on average, the Central US decreased one NAEP point on average, and the Southeast and the Northeast showed no overall change. In Figure 3 are displayed the state means adjusted for Math teacher undergraduate major and Certification status, shown above the point on the horizontal

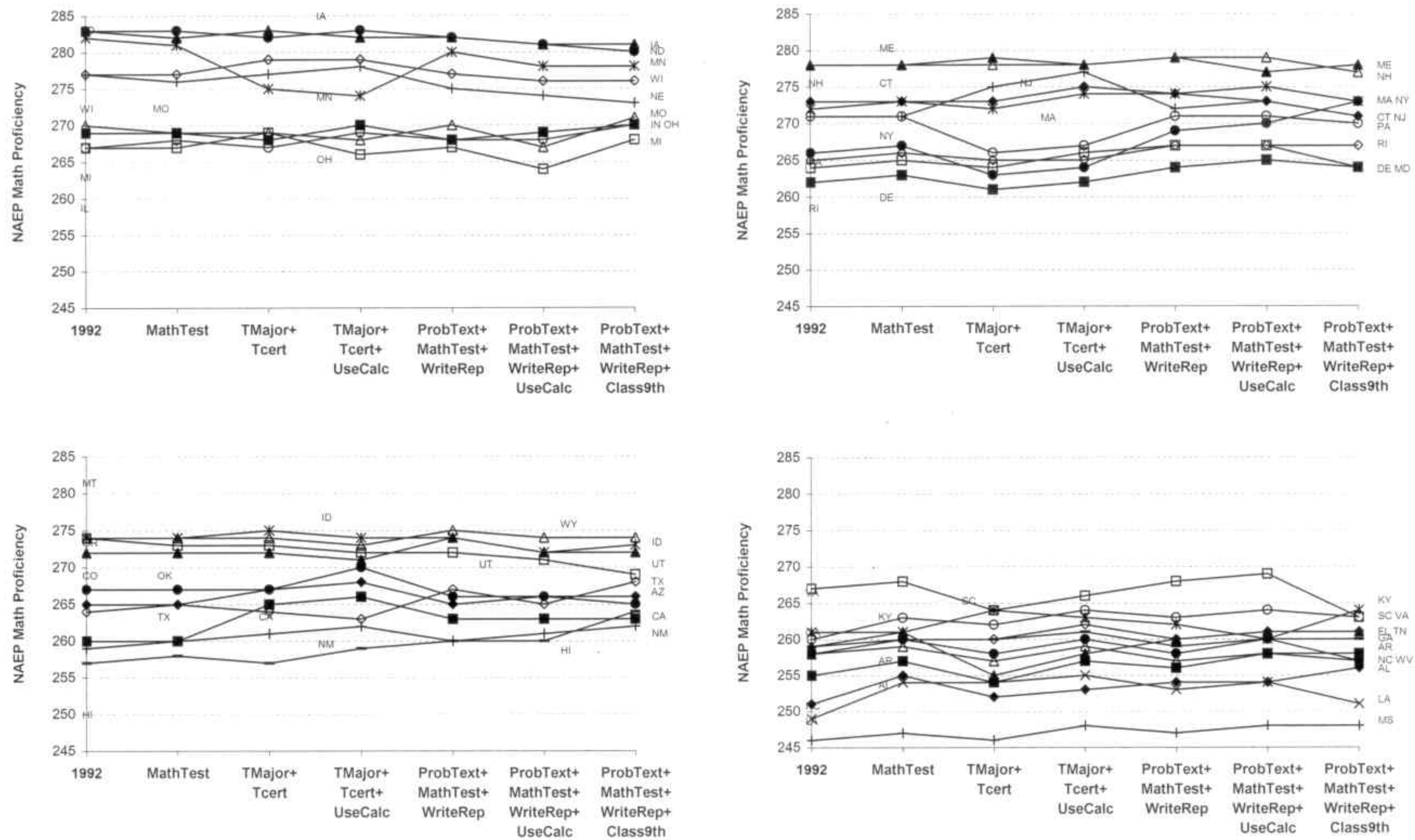

Figure 3. 1992 NAEP eighth-grade mathematics standardization results for all School-related Adjustments by state and US region.

axis labeled “TMajor + TCert” with the unadjusted NAEP mathematics averages above the point labeled “1992.”

Math teacher undergraduate major and Certification status and Use a calculator. Except for the Southeastern states, the results for this three-way standardization are very similar to the two-way standardization involving Math teacher undergraduate major and Certification status only: The West gained one NAEP scale point on average, the Central US decreased one NAEP point on average, and the Northeast showed no overall change. In the Southeast, states increased by two NAEP scale points on average, with LA gaining six NAEP points. In Figure 3 are shown all state means adjusted for these three variables (Math teacher undergraduate major and Certification status and Use a calculator) above the point on the horizontal axis labeled “TMajor + TCert + UseCalc.”

Do problems from textbooks and Take mathematics tests and Write reports or do projects. Figure 3 displays the state means adjusted for these three variables above the point on the horizontal axis labeled “ProbText + MathTest + Writerep.” The Northeastern states exhibited the greatest change overall, increasing somewhat less than two NAEP scale points; the states of the Southeast and the West increased about one NAEP scale point on average; and NAEP mathematics means in the Central states declined by about one NAEP scale point overall. No state showed change greater than four NAEP scale points.

Do problems from textbooks and Take mathematics tests and Write reports or do projects and Use a calculator. As with the standardization analysis for Do problems from textbooks and Take mathematics tests and Write reports or do projects described above, changes in state NAEP mathematics means were unremarkable, never larger than five points (LA increased by five NAEP scale points). The adjusted means for the Southeastern and the Northeastern states increased by approximately two points over their unadjusted means, while the Central states’ means decreased by about two points, and the state means in the Western US remained unchanged. The state means adjusted for these four variables (Do problems from textbooks and Take mathematics tests and Write reports or do projects and Use a calculator) are displayed in Figure 3 above the point on the horizontal axis labeled “ProbText + MathTest + WriteRep + UseCalc.”

Do problems from textbooks and Take mathematics tests and Write reports or do projects and 9th grade math class. Despite relatively large differences between a few states’ unadjusted and adjusted means (HI and NY both increased by about seven NAEP scale points), the Northeastern, Southeastern, and Western regions on average increased by about one NAEP point overall while the Central US decreased by about one NAEP point. In Figure 3 are shown all state means based on the standardization involving these four variables (Do problems from textbooks and Take mathematics tests and Write reports or do projects and 9th grade math class) above the point on the horizontal axis labeled “ProbText + MathTest + WriteRep + Class9th.”

Summary

Differences between unadjusted and adjusted NAEP averages were generally very small, the mean overall change for states being a one-point increase. Table 2 shows the frequencies of the differences between unadjusted state means and adjusted state means across the three different kinds of adjustments, Demographic, Home Background, and School-related Adjustments. Although there are a few relatively large differences, average changes are really very small. The Home Background Adjustments resulted in the greatest average increases across all states, and, in particular, the $4 \times 3 \times 3$ direct standardization based on Number of parents in the household and Home environment composite and “How often is another language spoken in your home?” yielded the largest average increase, approximately three NAEP scale points, across states.

Table 2. Frequency of differences between unadjusted and adjusted state means across all standardizations.

difference	Adjustments Demographic	Home-related	School-related
≤-7	0	0	2
-6	2	0	0
-5	3	0	2
-4	4	0	6
-3	5	3	7
-2	9	8	12
-1	10	11	34
0	17	26	56
1	43	27	43
2	35	36	39
3	16	34	27
4	6	31	8
5	5	10	5
6	4	9	3
≥7	5	10	2
total	164	205	246
mean	1.13	2.30	0.66

The states of the Southeastern US showed the greatest average increases across all adjustments ($\overline{D}$ = 3.6, s^2 = 1.3), followed by the Western states ($\overline{D}$ = 3.3, s^2 = 9.6), the states in the Northeast ($\overline{D}$ = 1.4, s^2 = 1.8), and finally the Central states ($\overline{D}$ = 0.4, s^2 = 3.5). In fact, the adjustment analyses resulted in a generally homogenizing effect, that is, differences among states were reduced. The variance of the 1992 state NAEP mathematics means is 81.2; however; the variances for the 15 sets of adjusted state means range from 41.8 (Race and sex) to 75.6 ("How often is another language spoken in your home?" and Community type) and 83.3 (IEP status).

The largest single difference was an increase of ten NAEP scale points for California for the adjustment based on three student-reported variables related to student's home environment (number of parents living in the household, "How often is another language spoken in your home?," and a composite variable describing the number of reading materials available in the home). Moreover, CA also had the greatest average change across standardization analyses, $\overline{d}$ = 5.3, s^2 = 7.9. Other states that exhibited relatively large differences between unstandardized and standardized means ($\overline{d}$) are: LA ($\overline{d}$ = 4.2, s^2 = 1.9), HI ($\overline{d}$ = 3.5, s^2 = 6.8), NM ($\overline{d}$ = 3.5, s^2 = 4.5), AL ($\overline{d}$ = 3.4, s^2 = 1.7) and TX ($\overline{d}$ = 3.3, s^2 = 7.4).

Discussion

The motivation for this exploration of statistical adjustments of educational indicators was not an attempt to attain parity in achievement for different jurisdictions, nor was it intended to obscure the disappointing academic performance of different subgroups of American students, nor identify causation at the student level. It was, instead, an effort at bringing comparability to NAEP statistics. The latter is the wholehearted objective of this study — to advance an acceptable level of comparability in the educational outcomes of the NAEP state assessment so as to facilitate informed comparisons among the diverse student populations in participating states.

As is habitually observed in both theoretical and applied treatments of statistical adjustment (cf. Mullis et al. 1994), this author verifies that it is indeed not possible to adjust for everything, and has additionally learned that it is not possible to adjust for more than a few things at a time. Direct standardization based on multiway crosstabulations behaves properly so long as cell proportions in the standard population are nonzero. It is possible to collapse across categories in order to eliminate empty cells, but this naturally degrades the fidelity of the adjustment. At any rate, despite the fact that adjustments are incomplete and imperfect, it should be recognized that this does not constitute a legitimate reason for omitting or concealing such potentially important and useful information, and doing so does a disservice to the beneficial applications of testing and score-reporting.

Based partly on the results presented in this report, I recommend the judicious reporting of state NAEP means adjusted for select variables beyond the control of schools and school systems. The purpose would not be to supplant the current reports which contain tables of unadjusted scores, but to augment the information contained in those reports. The modest wisdom of reporting unadjusted test scores alone is being challenged by various examples of state report card practices as well as the reporting of adjusted individual-level high-stakes test scores to university admissions offices. Rather than simply speculating about the effects of factors not under the control of the educational jurisdiction or dismissing altogether important kinds of inferences, educational policy and decision-making will be better served by thoughtful and relevant analyses of comparative data with statistical adjustment of student outcomes. However, recalling Wainer's (1989) error of the fifth sort — "Just because it's adjusted doesn't mean that it's helpful" (p. 131) — I encourage continued investigation of the effects and appropriateness of possible adjustment variables by all interested parties.

References

Dossey, J. (1993). What do NAEP math results mean for states? In *Analysis of State Education Indicators: State profiles and NAEP results related to state policies and practices*. Washington, DC: Council of Chief State School Officers.

Finn, Jr., C. E. (1994). Drowning in Lake Wobegon. *Education Week*, *13*(38), pp. 31, 35.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions (2nd ed.)*. New York: Wiley.

Glaser, R., & Linn, R. (Eds.). (1992). *Assessing student achievement in the states*. Stanford, CA: National Academy of Education.

Haertel, E. (1992). Reasonable inferences for the Trial State NAEP given the current design: Inferences that can and cannot be made, In Glaser, R., & Linn, R. (Eds.). *Assessing student achievement in the states*. Stanford, CA: National Academy of Education.

Johnson, E. G., Mazzeo, J., & Kline, D. L. (1993). *Technical report of the NAEP 1992 Trial State Assessment program in mathematics* (No. 23-ST05). Washington, DC: National Center for Education Statistics.

Koffler, S. L. (1991). *The technical report of NAEP's 1990 Trial State Assessment program* (No. 21-ST01). Washington, DC: National Center for Education Statistics.

Koretz, D. M. (1991). State comparisons using NAEP: Large costs, disappointing benefits. *Educational Researcher*, *20*(3), 19-21.

Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, *29*, 95-110.

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

Mullis, I., Dossey, J., Owen, E., & Phillips, G. (1993). *NAEP 1992 mathematics report card for the nation and the states*. Washington, DC: National Center for Education Statistics.

Mullis, I. V. S., Jenkins, F., & Johnson, E. G. (1994). *Effective schools in mathematics: Perspectives from the NAEP 1992 Assessment* (Report No. 23-RR-01). Washington DC: National Center for Education Statistics.

Phillips, G. W. (1991). Benefits of state-by-state comparisons. *Educational Researcher*, *20*(3), 17-19.

Robinson, G. E., & Brandon, D. P. (1994). *NAEP test scores: Should they be used to compare and rank state educational quality?* Arlington, VA: Educational Research Service.

Rogers, A. M., Kline, D. L., Mazzeo, J., Johnson, E. G., Mislevy, R. J., & Rust, K. F. (1992). *1990 Trial State Assessment: Secondary-use Data Files User Guide, revised.* Washington DC: National Center for Education Statistics.

Rogers, A. M., Kline, D. L., Mazzeo, J., Johnson, E. G., Mislevy, R. J., & Rust, K. F. (1997). *1992*

Trial State Assessment Program in Mathematics: Secondary-use Data Files User Guide, revised. Washington DC: National Center for Education Statistics.

Rotberg, I. C. (1995). Myths about test score comparisons. *Science*, *270*, 1446-1448.

Selden, R. (1990). Should test results be adjusted for socioeconomic differences? *The School Administrator*, *May 1990*, 14-18.

Shah, B. V., Barnwell, B. G., & Bieler, G. S. (1997). *SUDAAN® User's Manual, Release 7.5.* Research Triangle Park, NC: Research Triangle Institute.

Tukey, J. W., Abelson, R. P., Coffman, W. E., Gilbert, J. P., Jones, L. V., & Mosteller, F. (1971). *National Assessment of Educational Progress, Report 7, 1969-70 science: Group and balanced group results for color, parental education, and size and type of community, and balanced group results for region of the country and sex*. Denver, CO: Education Commission of the States.

Wainer, H. (1989). Eelworms, bullet holes, and Geraldine Ferraro: Some problems with statistical adjustment and some solutions. *Journal of Educational Statistics*, *14*(2), 121-140.

Wainer, H. (1994). On the academic achievement of New Jersey's public school children: Fourth and eighth grade mathematics in 1992. Research Report No. 94-29. Princeton, NJ: Educational Testing Service.

Wainer, H., & Kulick, E. (1997). A comparative study of the academic performance of Pennsylvania's public school children: Mathematics and reading between 1990 and 1996. Research Report No. 97-25. Princeton, NJ: Educational Testing Service.

Wolf, R. M. (1992). What can we learn from State NAEP? *Educational Measurement: Issues and Practice*, *11*(4), 12.

Appendix

Grade 8 Average Mathematics Proficiency
Unadjusted State Means and Means Adjusted for Various Student Background Variables

Table A-1
1992 Unadjusted State Means and Means Adjusted for Demographic Variables
IEP (I), Race+Sex (II), USBorn+TOC (III), and LangHome+TOC (IV)

	1992	I	II	III	IV
Alabama	**251**	253	255	253	253
Arizona	**265**	265	268	267	264
Arkansas	**255**	257	256	256	255
California	**260**	261	266	268	266
Colorado	**272**	273	270	272	271
Connecticut	**273**	275	273	276	276
Delaware	**262**	264	264	263	263
Florida	**259**	260	264	261	261
Georgia	**259**	259	262	259	259
Hawaii	**257**	258	258	259	259
Idaho	**274**	275	273	276	276
Indiana	**269**	270	266	272	271
Iowa	**283**	284	278	281	280
Kentucky	**261**	263	259	264	263
Louisiana	**249**	250	255	254	254
Maine	**278**	280	273	280	280
Maryland	**264**	266	268	267	267
Massachusetts	**272**	274	268	276	276
Michigan	**267**	267	267	270	269
Minnesota	**282**	283	276	282	283
Mississippi	**246**	246	254	246	247
Missouri	**270**	272	268	271	272
Nebraska	**277**	279	271	275	275
New Hampshire	**278**	279	273	279	279
New Jersey	**271**	272	273	278	278
New Mexico	**259**	260	265	264	261
New York	**266**	267	268	270	270
North Carolina	**258**	259	259	257	258
North Dakota	**283**	284	279	282	283
Ohio	**267**	268	265	270	270
Oklahoma	**267**	268	266	270	269
Pennsylvania	**271**	272	267	274	274
Rhode Island	**265**	267	262	268	268
South Carolina	**260**	261	265	261	261
Tennessee	**258**	260	257	258	259
Texas	**264**	265	271	266	265
Utah	**274**	275	271	274	274
Virginia	**267**	268	268	266	266
West Virginia	**258**	259	256	259	259
Wisconsin	**277**	279	273	276	275
Wyoming	**274**	276	271	275	275
District of Columbia	**234**	234	281	249	250

Table A-2
1992 Unadjusted State Means and Means Adjusted for Home Background Variables
ParEd+HaveCalc (I), ParEd+HomeEn (II), ParEd+HomeEn+LangHome (III), SingleP+HomeEn (IV), and SingleP+HomeEn+LangHome (V)

	1992	I	II	III	IV	V
Alabama	**251**	255	256	255	256	256
Arizona	**265**	268	269	270	269	271
Arkansas	**255**	259	259	258	259	259
California	**260**	264	267	268	269	270
Colorado	**272**	272	272	273	273	273
Connecticut	**273**	274	273	274	274	275
Delaware	**262**	264	265	265	265	265
Florida	**259**	262	263	263	265	265
Georgia	**259**	262	262	262	263	263
Hawaii	**257**	261	262	263	264	266
Idaho	**274**	275	275	275	275	275
Indiana	**269**	273	272	272	270	270
Iowa	**283**	282	282	281	282	282
Kentucky	**261**	267	267	267	264	265
Louisiana	**249**	253	252	252	254	253
Maine	**278**	278	278	278	278	278
Maryland	**264**	266	265	265	268	268
Massachusetts	**272**	274	272	272	273	273
Michigan	**267**	268	269	269	270	270
Minnesota	**282**	280	280	280	282	281
Mississippi	**246**	250	250	249	251	250
Missouri	**270**	272	272	271	272	272
Nebraska	**277**	277	276	275	276	276
New Hampshire	**278**	278	278	278	278	277
New Jersey	**271**	273	272	272	273	273
New Mexico	**259**	263	263	266	263	267
New York	**266**	269	268	268	269	270
North Carolina	**258**	261	260	260	261	260
North Dakota	**283**	280	280	280	282	282
Ohio	**267**	270	269	269	269	269
Oklahoma	**267**	269	269	269	270	269
Pennsylvania	**271**	273	271	271	271	271
Rhode Island	**265**	268	268	268	267	268
South Carolina	**260**	264	263	262	263	264
Tennessee	**258**	262	261	261	261	260
Texas	**264**	269	270	271	270	271
Utah	**274**	272	272	272	274	274
Virginia	**267**	269	269	269	269	269
West Virginia	**258**	263	263	262	261	260
Wisconsin	**277**	278	278	278	277	277
Wyoming	**274**	275	275	274	275	274
District of Columbia	**234**	238	238	238	242	242

Table A-3
1992 Unadjusted State Means and Means Adjusted for School-related Variables
MathTest (I), TMajor+TCert (II), TMajor+TCert+UseCalc (III), ProbText+MathTest+WriteRep (IV), ProbText+MathTest+WriteRep+UseCalc (V), and ProbText+MathTest+WriteRep+Class9th (VI)

	1992	I	II	III	IV	V	VI
Alabama	**251**	255	252	253	254	254	256
Arizona	**265**	265	267	268	265	266	266
Arkansas	**255**	257	254	257	256	258	258
California	**260**	260	265	266	263	263	263
Colorado	**272**	272	272	271	274	272	272
Connecticut	**273**	273	273	275	274	273	271
Delaware	**262**	263	261	262	264	265	264
Florida	**259**	261	255	258	260	261	261
Georgia	**259**	260	260	261	259	260	260
Hawaii	**257**	258	257	259	260	260	264
Idaho	**274**	274	275	274	274	272	273
Indiana	**269**	269	268	270	268	269	270
Iowa	**283**	282	283	282	282	281	281
Kentucky	**261**	261	264	263	262	260	264
Louisiana	**249**	254	254	255	253	254	251
Maine	**278**	278	279	278	279	277	278
Maryland	**264**	265	264	266	267	267	264
Massachusetts	**272**	273	272	274	274	275	273
Michigan	**267**	267	269	266	267	264	268
Minnesota	**282**	281	275	274	280	278	278
Mississippi	**246**	247	246	248	247	248	248
Missouri	**270**	269	269	268	270	267	271
Nebraska	**277**	276	277	278	275	274	273
New Hampshire	**278**	278	278	278	279	279	277
New Jersey	**271**	271	275	277	272	273	271
New Mexico	**259**	260	261	262	260	261	262
New York	**266**	267	263	264	269	270	273
North Carolina	**258**	260	258	260	258	260	257
North Dakota	**283**	283	282	283	282	281	280
Ohio	**267**	268	267	269	268	268	270
Oklahoma	**267**	267	267	270	266	266	265
Pennsylvania	**271**	271	266	267	271	271	270
Rhode Island	**265**	266	265	265	267	267	267
South Carolina	**260**	263	262	264	263	264	263
Tennessee	**258**	260	260	262	260	261	261
Texas	**264**	265	264	263	267	265	268
Utah	**274**	273	273	272	272	271	269
Virginia	**267**	268	264	266	268	269	263
West Virginia	**258**	259	257	259	257	258	257
Wisconsin	**277**	277	279	279	277	276	276
Wyoming	**274**	274	274	273	275	274	274
District of Columbia	**234**	237	233	233	239	239	238

Table A-4
1990 Unadjusted State Means and Means Adjusted for School-related Variables
SingleP+HomeEn+LangHome (I), ParEd+HomeEn+LangHome (II), and ProbText+MathTest+WriteRep (IV)

	1990	I	II	III
Alabama	**253**	255	255	251
Arizona	**260**	265	265	259
Arkansas	**256**	257	259	255
California	**256**	263	261	257
Colorado	**267**	268	267	267
Connecticut	**270**	269	269	271
Deleware	**261**	262	262	262
Florida	**255**	258	258	256
Georgia	**259**	260	261	259
Hawaii	**251**	257	256	253
Idaho	**271**	271	271	270
Indiana	**267**	267	268	265
Iowa	**278**	276	277	278
Kentucky	**257**	258	260	255
Louisiana	**246**	248	249	247
Maryland	**261**	261	260	262
Michigan	**264**	265	265	263
Minnesota	**275**	274	274	274
Montana	**280**	280	279	280
Nebraska	**276**	274	274	274
New Hampshire	**273**	272	272	274
New Jersey	**270**	269	269	270
New Mexico	**256**	263	264	256
New York	**261**	263	263	263
North Carolina	**250**	252	253	250
North Dakota	**281**	279	278	280
Ohio	**264**	264	265	264
Oklahoma	**263**	264	264	260
Oregon	**271**	271	271	270
Pennsylvania	**266**	266	267	266
Rhode Island	**260**	261	262	261
Texas	**258**	263	263	259
Virginia	**264**	264	265	264
West Virginia	**256**	256	260	253
Wisconsin	**274**	273	274	274
Wyoming	**272**	271	272	271
District of Columbia	**231**	235	234	232