Federal Sample Sizes for Confirmation of State Tests in the No Child Left Behind Act

Paul Mosquin
RTI International

James Chromy
RTI International

Commissioned by the NAEP Validity Studies (NVS) Panel
May 2004

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

The NAEP Validity Studies Panel was formed by the American Institutes for Research under contract with the National Center for education Statistics. Points of view or opinions expressed in this paper do not necessarily represent the official positions of the U.S. Department of Education or the American Institutes for Research.

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

Panel Members:

Albert E. Beaton
Boston College

Peter Behuniak
Connecticut State Department of Education

George W. Bohrnstedt
American Institutes for Research

James R. Chromy
Research Triangle Institute

Phil Daro
East Bay Community Foundation

Lizanne DeStefano
University of Illinois

Richard P. Durán
University of California

David Grissmer
RAND

Larry Hedges
University of Chicago

Gerunda Hughes
Howard University

Robert Linn
University of Colorado

Donald M. McLaughlin
American Institutes for Research

Ina V.S. Mullis
Boston College

Jeffrey Nellhaus
Massachusetts State Department of Education

P. David Pearson
Michigan State University

Lorrie Shepard
University of Colorado

David Thiessen
University of North Carolina-Chapel Hill

Project Director:

Frances B. Stancavage
American Institutes for Research

Project Officer:

Patricia Dabbs
National Center for Education Statistics

The authors would like to thank Beth Scarloss for her help in preparing this manuscript.

For Information:

NAEP Validity Studies (NVS)
American Institutes for Research
1791 Arastradero Road
Palo Alto, CA 94304-1337
Phone: 650/ 493-3550
Fax: 650/ 858-0958

Table of Contents

1. Introduction

This paper addresses statistical aspects of the No Child Left Behind (NCLB) Act (Bush, 2001), with the following goals: to further the discussion on how gaps in performance might be defined and to offer candidate gap estimators, to evaluate candidate gap estimators with respect to three separate student performance measures, to provide state-level distributions of major racial and ethnic groups, and to use the obtained state-level race and ethnicity distributions to calculate minimum sample sizes for state-level sampling on federal confirmation tests for each candidate gap estimator and performance measure.

The concept of gaps in student performance appears in many places throughout the NCLB Act, especially with respect to gaps in achievement between groups of students considered disadvantaged and not disadvantaged. Unfortunately, the legislation does not provide a statistical definition of a gap, so definition and implementation remains an open question. Notable efforts to clarify the situation have been made (Holland, 2002), but so far the issue remains unresolved. Here we will discuss in general what a gap might be, provide some additional approaches to estimating gaps, and discuss advantages and disadvantages of each.

Each of the candidate definitions of gap given will be evaluated with respect to three quantitative measures of student performance derived from the National Assessment of Educational Progress (NAEP): mean scale scores, percentage at or above basic achievement level, and percentage at or above proficient achievement level. This evaluation will be done both through a discussion of the statistical properties of the various gap estimators and through comparison of state-level sample sizes computed for each.

In the process, we will attempt to identify the disadvantaged racial and ethnic groups within states that might be expected to be adequately sampled under a proportional allocation sampling plan (i.e., no targeted sampling). Besides choice of gap statistic and performance measure, this determination will depend on a variety of factors. At the state level, the varying population sizes of the advantaged and disadvantaged groups will play a role, as will the varying current levels of state performance: states with large gaps will require smaller sample sizes in order to track progress. It may also be desirable to set sample sizes independent of current state performance; this approach will also be studied here.

The paper is divided into six main sections. In Section 2 we provide a brief introduction to the Act and describe the statistical aspects of the legislation. How gaps are defined in the Act is described, as well as the related statistic, “adequate yearly progress”

(AYP). Section 3 describes how a gap might be defined operationally, gives suggested candidate gap estimators together with their variances, and evaluates their performance in terms of margin of error. Section 4 defines the racial and ethnic groups most likely covered by the Act and uses existing databases to find state-level distributions of these groups. Section 5 identifies state-level sample sizes for the 4th grade NAEP mathematics assessment when the improvement targets can vary by state. Section 6 identifies state-level sample sizes for fixed improvement targets across states. Finally, Section 7 gives conclusions and further recommendations.

2. Review of the No Child Left Behind Act

The No Child Left Behind Act of 2001 mandates state-level assessments, given yearly to at least 95 percent of the student body at schools receiving public funding. These state-level assessments began in the 2001-2002 school year for reading and mathematics and must also be given in science beginning in 2005-2006. They are to be designed to be consistent with currently accepted educational standards. Until 2004, each child must be tested at least once in grades 3 through 5, grades 6 through 9, and grades 10 through 12. Beginning in 2005, the testing will be expanded to every grade from 3 through 8.

The results of the state assessments are to be used to evaluate improvement in academic performance overall and among various groups of disadvantaged students in each state. There are four groups of disadvantaged students specified in the Act:

1. Economically disadvantaged
2. Major racial and ethnic groups
3. Disabled
4. Limited English proficiency

Improvement in each group is expected to occur both by decreasing the number of students in the group considered to have only basic proficiency, and by reducing the observed gap in performance between the disadvantaged group and their more advantaged peers. The first type of improvement occurs when “adequate yearly progress” is made, while the second occurs through reduction of the observed performance gap between the two groups. The Act requires states to largely develop their own approaches to implementation.

To calibrate, or “confirm,” the different state testing methodologies, federal tests will be given to a sample of students in each state. Federal reading and mathematics tests will be given every two years in grades 4 and 8, starting with the 2002-2003 school year. Calibration and confirmation of state results will be through implementation of federal gap and AYP definitions. These federal implementations may differ in detail from those adopted by specific states, but they should have enough general applicability to confirm or contradict the state results. Adequate yearly progress and gaps are now discussed in turn, with a focus on performance gaps.

Adequate yearly progress

Adequate yearly progress is measured both overall and for each of the four disadvantaged groups specified by the Act. A group makes adequate progress if one of the following is true:

1. The proportion of students at the basic proficiency level decreases linearly over time, until after 12 years there are no more students in this category.1
2. The percentage of students at the basic proficiency level decreases by more than 10 percent of the previous year's value, and the group advances in either graduation rate or another state-defined indicator of progress.2

Note that a linear decrease to no students at the basic proficiency level and a 10 percent decrease per year are very different standards. In the latter case, over the 12 years the total number of basic proficient students is only reduced to $0.9^{12}*100\% = 28.2\%$ of its initial level.

The flexibility of the Act appears to leave both the exact statistic used in measuring AYP and the definition of "basic proficiency"[3] up to the individual states. However, the Act does define how an initial baseline level of student performance might be obtained, and these baselines presumably determine how AYP is computed in later years. The starting year percentage of students who are at basic proficiency is found by taking the larger of the following:[4]

1. The highest percentage of basic proficiency students among the four disadvantaged groups.
2. The percentage of basic-proficiency students in the school at the 20th percentile. This is calculated by ranking the schools in ascending order of overall percentage proficient, then summing the percentage of state enrollment over ranks, and taking the proficiency percentage of the school where the cumulative sum is 20 percent.[5]

[1] The NCLB Act of 2001, Section 1111 (b)(2)(F).
[2] The NCLB Act of 2001, Section 1111 (b)(2)(I)(i).
[3] The act describes proficiency categories of basic, proficient, and advanced. These state-defined categories may not be the same as the similarly named NAEP classifications. We presume that basic proficiency as defined in the act corresponds roughly to the "below basic" achievement level on the NAEP classifications.
[4] The NCLB Act of 2001, Section 1001, (b)(2)(E).
[5] This is described in the act as "the school in the 20th percentile in the state, based on enrollment, among all schools ranked by percentage of students at the proficient level."

Apparently these are meant to apply for finding starting points for both disadvantaged groups and all students combined; however, it is difficult to see how the first approach could be used for all students combined, and how the second would be used for disadvantaged groups. The first approach also appears to start each of the disadvantaged groups at the percentage proficiency of the worst among them; the motivation for this is unclear.

Performance gaps

In addition to AYP, states are evaluated under the Act by their ability to reduce or eliminate the performance gap between disadvantaged and other students. Performance gaps are not as clearly defined by the legislation as adequate yearly progress; in this section we review sections of the legislation that specifically refer to gaps in an attempt to find guidance about how a gap might be defined. The Act requires that these gaps be reduced at the state, local education agency, and school level.

> The first mention of a gap is in the lead phrase describing the Act's purpose:
> *To close the achievement gap with accountability, flexibility, and choice, so that no child is left behind.*

This is soon followed by an enumeration of the goals of the Act, among which is goal (3):[6]

> *(3) Closing the achievement gap between high- and low-performing children, especially the achievement gaps between minority and nonminority students, and between disadvantaged children and their more advantaged peers.*

Gaps are referred to only once in the important Section 1111, apparently as a form of adequate yearly progress:

> *(B) ADEQUATE YEARLY PROGRESS.—Each State plan shall demonstrate, based on academic assessments described in paragraph (3), and in accordance with this paragraph, what constitutes adequate yearly progress of the State, and of all public elementary schools, secondary schools, and local educational agencies in the State, toward enabling all public elementary school and secondary school students to meet the State's student academic achievement standards, while working toward the goal of narrowing the achievement gaps in the State, local educational agencies, and schools.*[7]

Reduction of gap size qualifies a state for performance recognition if they have

[6] The NCLB Act of 2001, Section 1001, (3).
[7] The NCLB Act of 2001, Section 1111 (b)(2)(B). Emphasis added.

> *i) Significantly closed the achievement gap between the groups of students described in Section 1111(b)(2);*[8]

where part ii) relates good performance to adequate yearly progress.

Later in the Act, the description of the groups for which performance gaps should be estimated is expanded on:

> *… to eliminate the achievement gap that separates low-income and minority students from other students.*[9]

And in a funding section, schools are to be rewarded for

> *(B) Closing the academic achievement gap for those groups of students farthest away from the proficient level on the academic assessments administered by the State under Section 1111.*[10]

Later, the State and Local Flexibility Section lists as a goal:

> *(7) To narrow achievement gaps between the lowest and highest achieving groups of students so that no child is left behind.*[11]

The remainder of the Act refers to "narrowing" or "reducing" of the "achievement gap" in a general sense, often with reference to Section 1111, which presumably means as applied to the four main groups of disadvantaged students defined in that section.

In summary then, although no statistical definition is given, the Act suggests that a gap represents a performance difference between two groups of students, one of which is disadvantaged. Disadvantaged apparently means having lower performance as a group on the assessments. The Act also suggests that among disadvantaged groups, those of lower proficiency have more important gaps than others. The Act is unclear in specifying the extent to which gaps are to be reduced: at times it says they should be reduced, and at times it says they should be eliminated.

The Act identifies the group(s) to which the disadvantaged group should be compared as "highest achieving groups," "other students," or "more advantaged peers." When multiple ethnic groups are present, it is unclear whether there is a single more advantaged group or multiple advantaged groups, and whether they should be treated separately or collapsed. This may be a state-by-state issue to resolve; in cases where multiple advantaged groups can be identified, our preference would be to collapse them as a group (for example, Asians and whites in a number of states).

[8] The NCLB Act of 2001, Section 1117. (b)(1)(B)(i).
[9] The NCLB Act of 2001, Section 2122 (b)(1)(B)(2).
[10] The NCLB Act of 2001, Section 5411 (b)(1)(B)(2).
[11] The NCLB Act of 2001, Section 6132 (7).

3. Defining a performance gap

Two statistical measures are mentioned in the Act: one to determine if adequate yearly progress occurs, and the other to determine if performance gaps decrease. As described in Section 2, the AYP statistic is reasonably well defined within the Act, while the gap statistic is not. In this section we consider what a gap statistic might measure, offer candidate gap statistics, provide variances for the difference or change between two gaps, and compare the margins of error of the candidate statistics.

What is a gap?

As the previous section demonstrates, the definition of a gap is left vague within the Act. The Act does make clear, however, that gaps describe a performance difference between two groups of students at a given time, or perhaps a performance difference between a group of students and a constant standard.

One approach to arriving at a definition of gap would be to compare observed test score distributions, as illustrated in Holland (2002), using cumulative distribution functions (cdfs). Holland's paper describes graphical methods for portraying distributions of the same group at two points in time (differences), of different groups at the same point in time (gaps), or of different groups at two points in time (differences in gaps). This approach provides a very sensitive method for visualizing performance differences, as the observed distributions contain a great deal more information concerning test scores than typically contained in a few summary statistics.

In this paper, however, we will consider approaches based on sample means and proportions. The sampling theory for these has been extensively developed, and this choice will thus allow us to focus immediately on necessary sample sizes for the federal confirmation of state results. The approach adopted here should be seen as complementary to Holland's distribution function approach. Because the statistics we use are simple functions of observed distribution functions, tests of these statistics are in fact tests of aspects of the distribution functions.

That said, we are unsure if gaps should ultimately be defined solely in terms of differences in distribution functions. A "gap" seems to necessarily imply a difference in location, and distribution functions provide far more information than just this. For example, with a distribution-function definition of gap, we might conclude that two groups of students show a performance gap, even though they both share the same mean or median. In this case, a gap has been found to exist, but which group would we consider advantaged? Can there be a gap in performance with neither group performing better according to a measure of location? For this reason, it would seem that slightly less information than that contained in a distribution function should be used to define a gap.

As demonstrated in Holland (2002), distribution functions provide an excellent tool for the *visualization* of performance differences between groups of students. Used for either exploratory data analysis or for follow-up analysis of gaps deemed significant through hypothesis testing, comparison of distribution functions allows a rapid understanding of important differences between the two groups.

We now consider definitions of gap based on the two statistics mentioned above: sample means and proportions.

Gaps based on sample means and proportions

An initial, somewhat intuitive definition of gap would be a difference in measures of location between two distributions, possibly a difference in sample means. This captures the idea that gaps are "distances" in performance between groups of individuals. Sample means benefit statistically from a well-developed theory, as well as their easy interpretation as averages. For sample means of continuous variates (such as standardized scores), the gap at time t is then $\bar{y}_t - \bar{x}_t$, where $\bar{y}_t$ is the mean in the advantaged group, and $\bar{x}_t$ the mean in the disadvantaged group. For sample proportions (i.e., sample means of zero-one valued variables), the gap at time t would be written $\hat{q}_t - \hat{p}_t$ where $\hat{q}_t$ is the sample proportion at or above the target proficiency level in the advantaged group, and $\hat{p}_t$ is the sample proportion at or above the target proficiency level in the disadvantaged group. If gaps were defined as differences from fixed values, then the gap statistics would be $\bar{y}_0 - \bar{x}_t$ or $\hat{q}_0 - \hat{p}_t$ with the advantaged group performance fixed, perhaps at a baseline year value.

Whether to use continuous scale scores or a discretized proficiency level is left as a decision for later, although each has its own advantages and disadvantages. Scale scores are obtained directly from the item response model fit, and they do not suffer a potentially information-reducing transformation, as is the case for proportions. They may for this reason require relatively lower sample sizes. Gaps measured by proportion of students at or above a given proficiency level may complement the AYP statistic (based on decreases in the proportion at each state's basic proficiency level), allowing a common framework for interpretation.

Gap statistics based on differences in sample means of standardized scores relate to a cdf approach in that the difference between two means is the area between the two cdfs.[12] The hypothesis test of difference in means then can be interpreted as a test that the area between cdfs is non-zero, and so closely tests a gap defined as the space between two distribution functions (Nettles et al., 2002).

A gap statistic based on a difference in proportions relates to a cdf-based approach in that the difference between two proportions is the vertical distance between the two cdfs at the test score cut-point separating the proficiency categories. The hypothesis test of difference in proportions then can be interpreted as a test that the two cdfs are not equal at a specific test-score value.

Comparing two gaps

For gap statistics defined as a difference in sample means (or sample proportions), an immediate question is how to compare two gaps that are measured at different points in time. An obvious approach is to take their difference, $(\bar{y}_t - \bar{x}_t) - (\bar{y}_{t-1} - \bar{x}_{t-1})$, and to say that there has been a reduction or improvement in the gap if this value is negative. We believe, however, that direct application of this approach has certain drawbacks that suggest the value of exploring alternative approaches.

Drawbacks of initial approach

A serious drawback of using only the difference in gaps to determine if performance improvement has occurred is that reductions in the gap do not necessarily mean that either group has improved individually. Gap reduction emphasizes equality of performance among the various advantaged and disadvantaged groups regardless of the change in performance in each group separately. A reduction can occur when both groups improve, one group improves and the other worsens, or both groups worsen. Although a gap reduced by worsening performance in both groups could be considered a "performance improvement" under the Act, it would not seem to be an improvement in any objective sense.

The separation of the change in the gap according to contributions from each of the two groups can be seen algebraically as a simple re-expression of the difference in gaps:

$$(\bar{y}_t - \bar{x}_t) - (\bar{y}_{t-1} - \bar{x}_{t-1}) = (\bar{y}_t - \bar{y}_{t-1}) - (\bar{x}_t - \bar{x}_{t-1}) = \Delta\bar{y}_t - \Delta\bar{x}_t$$

where $\Delta\bar{x}_t = \bar{x}_t - \bar{x}_{t-1}$ and $\Delta\bar{y}_t$ is defined similarly. The gap is reduced if this difference is negative. However, as indicated above, this can be accomplished in three different ways: the

[12] Note that if the cdfs cross, then some of the areas must be negative.

advantaged group's performance improves, but the disadvantaged group's improves even more; the advantaged group's performance deteriorates, while the disadvantaged group's improves; and finally, both groups deteriorate in performance with the disadvantaged group deteriorating less. Only one of these outcomes represents a clear improvement in performance.

These effects are depicted graphically in Figure 3.1. In this figure, which plots advantaged versus disadvantaged group performance, we consider the area above the 45-degree line that passes through the origin. This line represents all possible values of advantaged and disadvantaged group performance of zero gap, while the area above it represents all possible values at which advantaged group performance exceeds disadvantaged group performance. We divide this area into six separate regions, indicated by Roman numerals and discussed below, which represent different scenarios for change in the relative performance of the two groups.

A second line, parallel to the zero-gap line, is fixed by point $(\bar{x}_0, \bar{y}_0)$ representing mean disadvantaged and advantaged group performance in the base year (or the previous assessment). If the current assessment value $(\bar{x}_t, \bar{y}_t)$ lies along this second line, then the gap will not have changed. If the current value lies closer to the zero gap line (regions I, II, and III), then the gap will have decreased, while if it lies further away (regions IV, V, and VI), an increase in gap will have occurred. Similarly, the initial assessment values $(\bar{x}_0, \bar{y}_0)$ divide the plot into quadrants, depending on whether disadvantaged group performance has increased or decreased, and whether advantaged group performance has increased or decreased.

Figure 3.1.

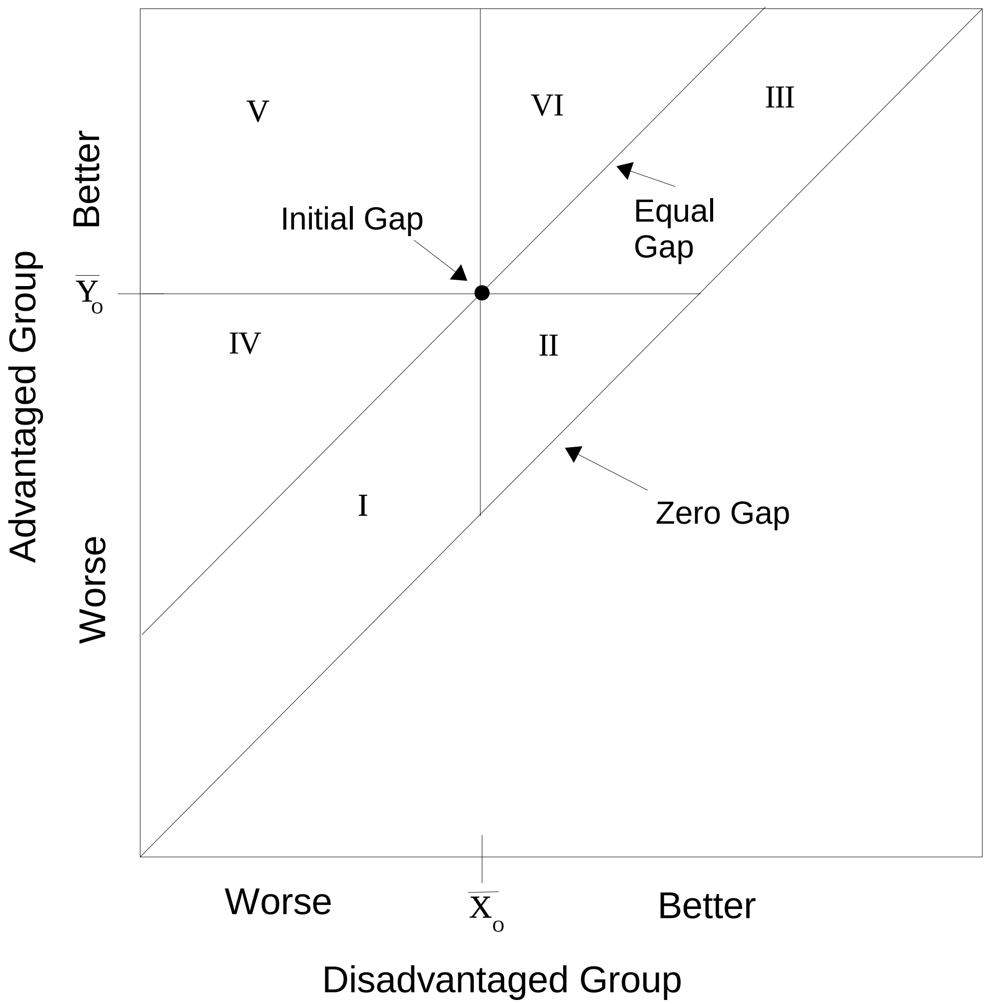


Figure 3.1: Six regions in which a change in gaps might occur for advantaged and disadvantaged groups. The plotted point represents the baseline year gap-value. The diagonal lines represent equal gap-values. The vertical line separates decreased performance by the disadvantaged group from improved performance, and the horizontal line separates decreased performance by the advantaged group from improved performance.

Thus we see that, among the three regions representing a decrease in gap, an actual improvement in performance for the disadvantaged group only occurs in regions II and III. Similarly, the three regions representing an increase in gap are also of varied desirability. In region IV both groups deteriorate in performance with the disadvantaged group deteriorating more. In region V the advantaged improves while the disadvantaged deteriorates, and in region VI both improve in performance, with the advantaged improving more than the disadvantaged.

Suggested approaches

Our two approaches to measuring gaps and their improvement can then be illustrated with reference to Figure 3.1. Two of the regions represent absolute improvement in test scores for both groups (the upper right quadrant), while three of the regions represent an increased equality of performance, as defined by a reduction in the gap.

The first approach restricts improvement to region III, where both the disadvantaged group improves and the performance becomes more equal. In this case, the gap is defined as $\bar{y}_t - \bar{x}_t$, the change in gap is estimated as $\Delta\bar{y}_t - \Delta\bar{x}_t$, and there is said to be an improvement in the gap if

1. $\Delta\bar{y}_t - \Delta\bar{x}_t < 0$, and
2. $\Delta\bar{y}_t \geq 0$.

The first condition requires that there be more equality between the groups, while the second requires that the advantaged group not deteriorate. This approach both requires equality and individual improvement of the two groups.

The second approach restricts improvement to the region where both groups improve, corresponding to regions III and VI. In this case the gap is defined as $\bar{y}_0 - \bar{x}_t$, a difference between the base-year mean of the advantaged group and the current mean of the disadvantaged group at time *t*, and the change in gaps is estimated as $-\Delta\bar{x}_t$. An improvement in the gap would occur if

1. $-\Delta\bar{x}_t < 0_t$, and
2. $\Delta\bar{y}_t \geq 0$.

The first condition requires that the disadvantaged group improve, while the second requires that the advantaged group not deteriorate. This approach does not require increasing equality, but does require individual improvement within the two groups.

The above is a refinement of what an “improvement in gap” represents, but not of the gap statistics themselves. For either approach, conditions 1 and 2 should be tested simultaneously, using a multivariate test, in order to determine if the gap has decreased.

Adequate yearly progress

Unlike gaps, adequate yearly progress is defined in terms of a single group, identified as disadvantaged. The Act mandates that the percentage of students who exhibit only basic proficiency be reduced to zero after 12 years, or that it be reduced by 10 percent per year (for a total reduction of about 72 percent).

Figure 3.1 can also be used to depict AYP, which is achieved if the new value of the percentage of scores at or above proficient for disadvantaged students lies to the right of the previous value, regardless of performance in the advantaged (or any other) group. The goal in this case is not to reach the diagonal line through the origin, but instead, to reach a specific value (either a 72 percent or 100 percent reduction of the starting value). "Improvement" in adequate yearly progress can reward states in which disadvantaged group performance improves, but advantaged group performance deteriorates. In that case, change falling in region II (and its extension below the diagonal line) represents "improvement." Therefore one of the regions not considered improvement when figuring gap statistics is considered an improvement for AYP.

The figure also illustrates the correlation between gap statistic and AYP statistic. For example, if gap statistics are measured using the proportion of scores at or above the proficient level, and if either of the gap statistics proposed here shows an improvement, then we would know that an increase in AYP has also occurred. Statistically, this correlation is represented as $Corr(\Delta\bar{y}_t - \Delta\bar{x}_t, \Delta\hat{p}_t)$, where $\Delta\hat{p}_t$ is the adequate yearly progress statistic. In the case where percentage at or above proficient is used for computing gaps, this becomes $Corr(\Delta\hat{q}_t - \Delta\hat{p}_t, \Delta\hat{p}_t)$ for the first gap statistic, and $Corr(-\Delta\hat{p}_t, \Delta\hat{p}_t)$ for the second, which is complete correlation. One might want to avoid excessive correlation. However, gaps and AYP are inherently correlated; therefore it is not possible, or even desirable, to avoid correlation entirely.

In selecting a gap performance measure, comparability with the AYP statistic is more important than correlation. Adequate yearly progress is already defined within the Act based on the percentage of scores exceeding the basic proficiency level. The basic proficiency level corresponds roughly to the percentage *below basic* on the NAEP scale. Therefore, of the various statistics that might be used for measuring a gap on the NAEP scale—proportion at or above the basic, proficient, or advanced achievement level, or mean standardized score—the proportion at or above the basic achievement level will both have the greatest correlation with the adequate yearly progress statistic and also be the most directly comparable. Since gaps and AYP measure different performance objectives (equality vs. absolute improvement), it follows that using the same basic statistic to measure each would simplify both interpretation and the presentation of results (for example, both could be depicted together on a plot such as that in Figure 3.1). If a choice is made to measure gaps and AYP by different statistics, the benefits to the overall analysis should be identified. For example, perhaps other aspects of the performance distribution besides the chosen cut-point are important, or perhaps sample sizes can be smaller.

What are the gap improvement targets?

To select sample sizes for the biennial state NAEP assessments it is important to first have a clear understanding of the magnitude of differences to be detected. These differences will vary according to whether gaps or adequate yearly progress are considered.

For adequate yearly progress, the legislation is reasonably clear: a reduction of basic-level proficiency (below basic on NAEP) of either at least 72 percent or 100 percent must be achieved. This can be either a progression of 10 percent decreases from each previous year's level, or a linear decrease to zero. Not mentioned in the legislation, but perhaps worth consideration, is decreasing the proportion of basic-level proficiency in the disadvantaged group to equal the proportion observed in the advantaged group.

With performance gaps, a target goal is not as clear. The Act requires either a reduction or an elimination of gaps. Guidelines for the level of reduction might be obtained from historical patterns; however, these suggest that a very small change is to be expected. A study by Yen (2002) on how gaps perform over time found that, in general, both advantaged and disadvantaged groups improve together, so the change in gap can be negligible. Small changes cannot be detected given current sample sizes on NAEP; we will therefore instead consider the goal to be *elimination* of the performance gap.

With the goal of eliminating the performance gap over the 12-year period, states could be held to one of two biennial targets for improvement: 1) the current amount remaining to be improved, divided by the number years remaining, or 2) the total amount of improvement necessary from the baseline year divided by the total number of years covered by the initiative (i.e., 12 years). The first approach maintains a specific schedule, but falling behind early could quickly lead to unattainable goals. This approach could also fail to reward improved performance in later years. The second approach does not maintain a schedule, but provides fixed targets for each year. We will use the second method to determine sample sizes.

Note that the first approach to establishing a target is most compatible with the gap statistic $\bar{y}_t - \bar{x}_t$, in which the gap is defined as the difference in performance measured in the current time period. The gap improvement target is then reset each year at the current value of $\bar{y}_t - \bar{x}_t$ divided by the number of years remaining. The second approach is more consistent with the gap statistic $\bar{y}_0 - \bar{x}_t$, as in this case the annual target is always 1/12 of the distance between the two groups at baseline.

The above discusses alternatives for setting state-specific targets. It is also possible to develop targets that are constant across all states, which will be considered separately in Section 6.

Variance of difference-in-gap statistics

We now provide simple variance expressions for the difference-in-gap statistics mentioned above: $\Delta\bar{y}_t - \Delta\bar{x}_t$, and $-\Delta\bar{x}_t$ for sample means, and $\Delta\hat{q}_t - \Delta\hat{p}_t$, and $-\Delta\hat{p}_t$ for sample proportions. These variance expressions will be used to determine sample sizes for state NAEP if it is to be used as the federal confirmation test. In order to reduce the dimension of the problem to a manageable level, simplifying approximations will be given. Note that $-\Delta\hat{p}_t$ has the same variance as AYP statistic $\Delta\hat{p}_t$, allowing sample sizes to be found for $\Delta\hat{p}_t$ as well.

Among the statistics based on sample means, variances can be given as

$$Var(\Delta\bar{y}_t - \Delta\bar{x}_t) = \frac{2\sigma^2 deff}{n}\left(\frac{\tau_1 + \tau_2}{\tau_1\tau_2}\right)$$

$$= \frac{2\sigma^2 deff}{n}\left(\frac{1}{\tau_1} + \frac{1}{\tau_2}\right)$$

and

$$Var(-\Delta\bar{x}_t) = \frac{2\sigma^2 deff}{n}\left(\frac{1}{\tau_1}\right)$$

where σ^2 is the variance of the standardized score distribution, *deff* is the design effect,[13] *n* is the state-level effective sample size, τ_1 is the proportion of the student body within the disadvantaged group, and τ_2 is the proportion of the student body within the advantaged group. Note that, as expected, the variance of $-\Delta\bar{x}_t$ is less than that of $\Delta\bar{y}_t - \Delta\bar{x}_t$.

For the difference-in-gap statistics based on sample proportions, variances can be given as

$$Var(\Delta\hat{q}_t - \Delta\hat{p}_t) = \left(\frac{p_{t-1}(1-p_{t-1}) + p_t(1-p_t)}{\tau_1} + \frac{q_{t-1}(1-q_{t-1}) + q_t(1-q_t)}{\tau_2} \right) \frac{deff}{n}$$

and

$$Var(-\Delta\hat{p}_t) = \left(\frac{p_{t-1}(1-p_{t-1}) + p_t(1-p_t)}{\tau_1} \right) \frac{deff}{n}$$

where p_t is the proportion of the disadvantaged group at or above the NAEP basic/proficient level at time *t*, and q_t is the proportion of the advantaged group at or above the NAEP basic/proficient level at time *t*. As with sample means, $-\Delta\hat{p}_t$ can be seen to have lower variance than $\Delta\hat{q}_t - \Delta\hat{p}_t$.

[13] The design effect (*deff*) allows for statements to be made about the variance of a statistic measured using a complex survey design by using the variance expression for a simple random sample in combination with prior knowledge of the performance of similar complex designs. Deff is defined as the ratio of the variance for the statistic of interest under the complex survey design to the variance of the same statistic under a simple random sample. That is,

$$deff = \frac{Var_{design}(\hat{\theta})}{Var_{SRS}(\hat{\theta})}.$$

This definition allows the known (and typically simple) variance expression for the simple random sample to be used in conjunction with an approximate range of values for *deff* from similar or previous studies to provide insight on the variances associated with a proposed complex design.

In addition to the design effect, the *effective sample size* can provide another way to illustrate the effect of a complex sample design. The effective sample size is the simple random sample size that would give the same *standard error* as that seen in the complex design. That is,

$$\frac{Var_{SRS}(\hat{\theta})}{n_{eff}(\hat{\theta})} = \frac{Var_{design}(\hat{\theta})}{n_{design}},$$

where in this paper we will write n instead of $n_{eff}(\hat{\theta})$. Note that by the definition of design effect

$$n_{design} = n_{eff}(\hat{\theta})\, deff$$

the design sample size n_{design} may also be referred to as the nominal sample size.

The variances for proportions allow a simpler approximation when, as expected, the change in percentage proficiency between times 1 and 2 is small relative to its level at time 1. In this case

$$Var(\Delta\hat{q}_t - \Delta\hat{p}_t) \cong 2\left(\frac{p_t(1-p_t)}{\tau_1} + \frac{q_t(1-q_t)}{\tau_2}\right)\frac{deff}{n}$$

and

$$Var(-\Delta\hat{p}_t) \cong \frac{2p_t(1-p_t)deff}{\tau_1 n}.$$

Furthermore, these variance expressions are maximized at $p_t = q_t = 0.5$. The dependence on the unknown quantities p_t and q_t can therefore be removed for values close to one half. That is:

$$Var(\Delta\hat{q}_t - \Delta\hat{p}_t) \le \frac{deff}{2n}\left(\frac{1}{\tau_1} + \frac{1}{\tau_2}\right)$$

and

$$Var(-\Delta\hat{p}_t) \le \frac{deff}{2\tau_1 n}.$$

This upper bound can serve as a reasonable, conservative approximation for the required *n* for intermediate values of p_t and q_t (between, for our purposes, .25 and .75). For example, under the approximation, $p_t(1-p_t)$=0.5*0.5=.25, while even at p_t=0.25 we have 0.25*0.75≈ 0.19.

Assumptions required to arrive at the above variance expressions include the following: fixed-size samples of $\tau_1 n$ individuals of the disadvantaged group and $\tau_2 n$ individuals of the advantaged group, design effects equal across both of these samples, independently and identically distributed observations (according to a model-based approach to sampling), design effects equal across the two time periods, population standard deviations of standardized scores equal across the two time periods, and equal sample sizes at time 1 and time 2.

The most important of these assumptions are the first three listed: two separate samples, equal design effects in each sample, and independent, identically distributed test scores. From a model-based perspective, the assumption of independent, identically distributed test scores is not strictly true, as the test scores were obtained through the fitting of an item response model and so contain some model-induced dependence. It is not expected that this dependence will be overly large, but if it were accounted for, a covariance term would appear in the above expressions. The assumption of two separate samples limits the problem to the consideration of fixed sample sizes. Under the realistic condition of random sample sizes, variances are likely to be somewhat larger, although the effect is not expected to be large. Perhaps the most important assumption is that of equal design effects within both samples.[14] This assumption is unlikely to hold in practice, so we might want to use the larger of the design effects in the two separate samples. The sample allocation can be controlled by the survey design, and in this paper we assume proportional allocation (i.e., no oversampling of any targeted groups).

Margin of error

Recommended sample sizes for state-level racial and ethnic group sampling will be those required in order to establish margins of error for the various difference-in-gap statistics at less than or equal to a fixed amount.[15] The biennial target for a given state, in turn, determines the fixed amount. We now provide a brief review of the meaning and use of margins of error, a graphical analysis to illustrate the behavior of margins of error for the difference-in-gap statistics given above, and a brief review of how margins of error are used to find the effective sample size.

A 95 percent one-sided confidence level margin of error for each of the difference-in-gap statistics is equal to 1.65 times the square root of the variance, according to the large sample limiting normal distribution implied by the above assumptions.[16] That is,

$$ME = 1.65\sqrt{\frac{2\sigma^2 deff}{n}\left(\frac{1}{\tau_1}+\frac{1}{\tau_2}\right)}$$

for $\Delta\bar{y}_t - \Delta\bar{x}_t$, with margins of error for other difference-in-gap statistics obtained similarly. The sample point estimate of a difference-in-gap statistic plus or minus its margin of error provides an approximate 95 percent confidence interval for the true value of the difference-in-gap statistic over the population as a whole.

---

[14] This assumption only applies to $\Delta\bar{y}_t - \Delta\bar{x}_t$ and $\Delta\hat{q}_t - \Delta\hat{p}_t$.

[15] The approach relies on the theoretical limiting normal distributions of difference-in-gap statistics based on the stated assumptions and simple random sampling (where design effects are equal to 1). Under other sampling designs the conclusions may be viewed as an approximation.

[16] For proportions, there are additional assumptions that $n\tau_1 p \geq 5$, $n\tau_1(1-p) \geq 5$, $n\tau_2 q \geq 5$, and $n\tau_2(1-q) \geq 5$.

Margins of error can also provide some insight into hypothesis testing. Recall that in a hypothesis test, we wish to determine whether the sample supports a specific "null hypothesis" population value for the statistic of interest, at a specified α level of significance. This significance level is the probability of the null value being rejected when it is in fact the true value. A hypothesis-test interpretation of a confidence interval is that it contains all population values of the statistic of interest that, if tested against the value observed in the sample, would not have led to a rejection of the null hypothesis.[17]

However, we note that the hypothesis-test interpretation of margins of error applies to the difference-in-gap statistics without introduction of the constraint $\Delta\bar{y}_t \geq 0$, as described previously. For this reason the sample sizes provided here are often a lower bound on required sample sizes for the purposes of hypothesis testing. The requirement $\Delta\bar{y}_t \geq 0$, if accepted, makes the hypothesis tests multivariate, and the univariate approach implied by difference-in-gap margin of error analysis does not apply. If the $\Delta\bar{y}_t \geq 0$ requirement is dropped, then the confidence interval interpretation does apply, and sample sizes given here are the required sample sizes. Further, since the $\Delta\bar{y}_t \geq 0$ requirement does not exist for AYP, the sample sizes given here apply directly to that case.

Our primary use of margins of error will be to obtain sample sizes. To illustrate the use of margins of error for this purpose, consider the variance for $\Delta\bar{y}_t - \Delta\bar{x}_t$. Its margin of error is

$$ME = 1.65\sqrt{\frac{2\sigma^2 deff}{n}\left(\frac{1}{\tau_1}+\frac{1}{\tau_2}\right)}.$$

Given a specified target margin of error, a minimum sample size would be obtained by squaring both sides and expressing in terms of *n*:

$$n = 1.65^2\,\frac{2\sigma^2 deff}{ME^2}\left(\frac{1}{\tau_1}+\frac{1}{\tau_2}\right).$$

Obtaining *n* then requires values for σ^2, τ_1, τ_2, and *deff*, in addition to the margin or error, and these are often set equal to values observed in previous studies. In Section 4, such estimates will be obtained for the various racial and ethnic groups in each of the states.

[17] For a two-sided test.

Equal margin of error plots

For the difference-in-gap statistics discussed in this paper, contour plots of equal margin of error give insight into conditions under which adequate sampling is possible. These plots are given for difference-in-gap statistics based on both sample means and proportions.

The equal margin of error plot for effective sample size versus the proportion of disadvantaged students is given in Figure 3.2(a) for the difference-in-gap statistic $\Delta\bar{y}_t - \Delta\bar{x}_t$. Each contour line gives the margin of error in standard deviation units for a design effect of 1. For simplicity, it is assumed that $\tau_1 + \tau_2 = 1$ (i.e., that only the advantaged and disadvantaged groups are present). If $\tau_1 + \tau_2 < 1$, then the margin of error will be larger. It is clear from this plot that as the percentage of disadvantaged students becomes small, the required sample increases rapidly. Without resorting to oversampling, groups that represent a small percentage of the population benefit very little from an increased overall sample size.

Figure 3.2

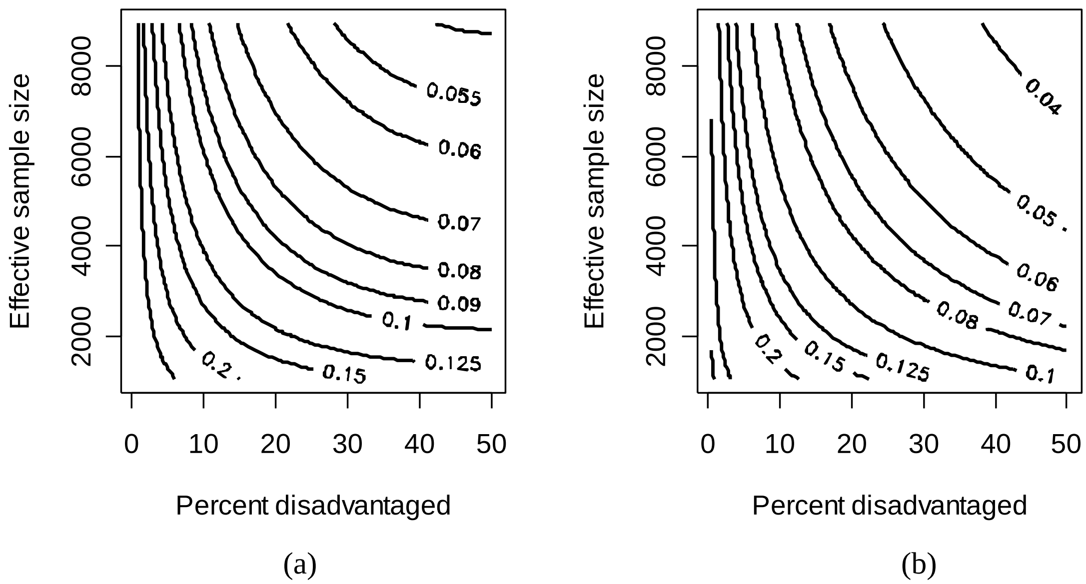


Figure 3.2: Equal margin of error curves for (a) difference-in-gap statistic $\Delta\bar{y}_t - \Delta\bar{x}_t$ and (b) difference-in-gap statistic $-\Delta\bar{x}_t$ at various sample sizes and percentages of the population disadvantaged. Plotted values are

$$\text{(a) } ME = 1.65\sqrt{\frac{2}{\tau_1(1-\tau_1)n}}$$

and

$$\text{(b) } ME = 1.65\sqrt{\frac{2}{\tau_1 n}}.$$

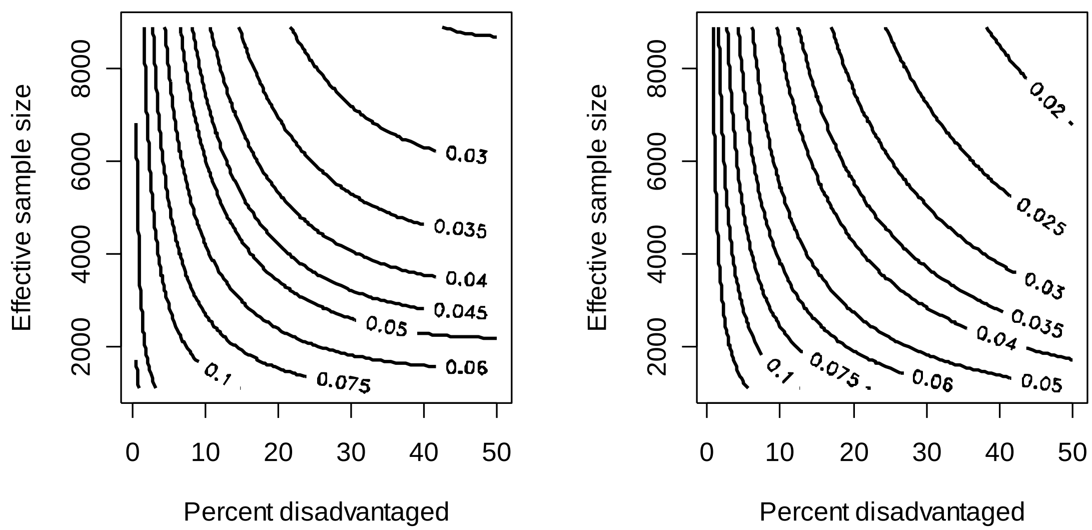
Design effect is equal to one, and units are in standard deviations.

Plots of this sort can also be used to obtain margins of error for designs with non-unit standard deviations or other design effects. For non-unit σ, the contour line values are multiplied by the value of σ, while for other design effects they are multiplied by $\sqrt{deff}$. If, for example, the standard deviation of standardized test scores was 35 and the design effect was 2, contour line values would be multiplied by $35\sqrt{2} \approx 49$.

Figure 3.2(b) shows a margin of error contour plot for difference-in-gap statistic $-\Delta\bar{x}_t$. The plot is similar to that of Figure 3.2(a), as the variance expressions differ only by a factor of $1/(1-\tau_1)$, which is close to 1 for small proportions of the disadvantaged group. Again, the greatly increasing margin of error for smaller proportions disadvantaged is apparent. However, as τ_1 increases, margins of error continually decrease, unlike in Figure 3.2(a) where they reach a minimum at τ_1=0.5. As expected from the variance expressions, margins of error are smaller for a given sample size and τ_1 as compared to Figure 3.2(a).

Figure 3.3 gives margin-of-error contour plots for (a) difference-in-gap statistic $\Delta\hat{q}_t - \Delta\hat{p}_t$, and (b) difference-in-gap statistic $-\Delta\hat{p}_t$. These plots bear a strong similarity to those of Figure 3.2 because, under the simplifying assumptions, the margins of error differ only by a multiplicative constant.

Figure 3.3



(a) (b)

Figure 3.3: Equal margin of error curves for (a) difference-in-gap statistic $\Delta\hat{q}_t - \Delta\hat{p}_t$ and (b) difference-in-gap statistic $-\Delta\hat{p}_t$ at various sample sizes and percentages of the population disadvantaged. Plotted values are

(a) $$ME = 1.65\sqrt{\frac{1}{2\tau_1(1-\tau_1)n}}$$

and

(b) $$ME = 1.65\sqrt{\frac{1}{2\tau_1 n}}.$$

Design effect is equal to one and units are in standard deviations.

The simplifying assumptions used to create Figure 3.3 are that 1) only advantaged and disadvantaged groups are present in the population, and 2) that the proportion at each achievement level is 0.5. These assumptions affect the estimated margins of error in different ways. On the one hand, if more than one disadvantaged group is present, the margins of error will be larger than estimated since the proportion advantaged, τ_2, is now less than $1-\tau_1$. On the other hand, setting the proportion at each achievement level to 0.5 leads to the largest possible margin of error. The choice of this approximation is motivated by the improvement in disadvantaged group performance expected under the NCLB Act. As the proportion of the disadvantaged group at or above the basic or proficient achievement levels changes (as is expected) over the 12-year period of the Act, the sample size required to detect change will also change. In the case of students at or above the basic achievement level, most states' disadvantaged group mathematics scores (see Table 5.4 in the appendix) are expected to improve through P $\geq$ Basic = 0.5 towards the advantaged group's performance level of about P $\geq$ Basic = 0.7 or 0.8. Margins of error seen in Figure 3.3 might then be expected over many of the 12 years. The approximation is more reasonable for the P $\geq$ Basic achievement level than for the P $\geq$ Proficient achievement level, as in the latter case, the advantaged group currently has around 0.4 at or above the proficient level, and the disadvantaged group much less (see Table 5.6 in the appendix). In this case, the approximation may be more reasonable towards the end of the 12-year period when ideally both advantaged and disadvantaged groups will have P $\geq$ Proficient close to 0.4 or 0.5.

4. State-level distributions of race and ethnicity

Obtaining state-level distributions of race and ethnicity requires that the groups be identified and defined, that variables from existing data sets corresponding to the definitions be found, and that these variables then be used to produce the state-level distributions.

A variety of approaches are used to record race and ethnicity. For example, the older U.S. Census race/ethnicity variables required individuals to assign themselves to a single category, while the current variables allow multiple races. Since the NCLB Act itself does not specify any one approach, we will look for a data set that provides the most precise estimates.

The NCLB Act suggests that the disadvantaged racial and ethnic groups to be studied include American Indian, black, and Hispanic, and these correspond well with categories commonly recorded in most major data sets. Modifications seen in some data sets include categories such as "Asian or Pacific Islander" or "black not Hispanic," which for our purposes will be considered the same as the primary category (i.e., Asian, black) except in states where the difference is substantial (e.g., Hawaii).

Three data sets were used to obtain state-level population distributions: the 1998 National Assessment of Educational Progress (NAEP),[18] the 2000-01 Common of Core Data (CCD) Public School Universe Survey,[19] and 2000 U.S. Census Bureau data.

The 2000-01 CCD Public Universe Survey provides a complete census listing of all public elementary and secondary schools in the states and other U.S. administrative regions.[20] With exceptions, it provides basic information on each school in the data set, including student counts by grade, gender, race, and ethnicity. The 1998 NAEP data set is from a sample survey of around 448,000 students in grades 4, 8, and 12 in 40 states as well as the District of Columbia, Department of Defense schools, and the Virgin Islands. Information was collected on race/ethnicity, English proficiency status, and disability status, among other variables. The 2000 U.S. Department of Census data provides population level estimates for all U.S. inhabitants.

Of the NAEP and CCD data sets, the CCD data set is more recent and furthermore, it is a census covering all institutions of interest to the NCLB Act. For many states, it also has district-level and school-level counts, which can facilitate the development of more sophisticated sampling approaches, if necessary. For these reasons, estimates of state-level population counts have been obtained from the CCD data set whenever possible. If CCD estimates were not available, the NAEP data set was used. U.S. Census values were used as a last resort when neither the CCD nor the NAEP data set could be used.

Both the CCD and the NAEP data sets contain the desired race/ethnicity categories with minor differences. CCD records five categories: American Indian, Asian or Pacific Islander, Hispanic, black not Hispanic, and white. NAEP records the same five categories and in addition a sixth “other” category. CCD ethnicity data is available for all states with the exception of Idaho, Tennessee, and Washington. For these states NAEP data and/or Census data have been used to obtain required population estimates.

Table A-1 (see appendix) gives the percentages of 4th grade students in the various CCD race/ethnicity categories by state, and Table A-2 gives a similar table for the 8th grade data. The tables show widespread variation in racial and ethnic distributions across the states. In contrast, across grades within a state, the distributions are very similar.

---

[18] Data obtained from tables in Allen, Donoghue, & Schoeps (2001)
[19] Available at http://nces.ed.gov/ccd/pubschuniv.html
[20] These latter include the District of Columbia, the Bureau of Indian affairs, the Department of Defense, and overseas territories.

A cross-classification table of the distributions of state-level race/ethnicity percentages according to various cut-points is given in Table A-3. For example, in the first row of the table we see that there is only one state in which Asian/Pacific Islanders represent more than 50 percent of the student population, but there are 44 states in which whites represent more than 50 percent of the student population. From this table it is clear that American Indians and Asian/Pacific Islanders will be particularly hard to target due to small populations in most states. In contrast, there are notably more states in which blacks and Hispanics are a moderate to large percentage of the population.

Table A-4 provides counts of the total number of students by grade and by state. These counts allow the identification of states where small student populations may limit the feasible sample sizes.

5. State-dependent performance targets

This section provides state-level sample sizes for the difference-in-gap estimators given earlier using NAEP standardized scale scores, the proportion at or above the basic achievement level, and the proportion at or above the proficient achievement level.

The 2000 NAEP mathematics assessment for 4th grade students (Braswell et al., 2001) was used to provide estimates of expected achievement values by state. It is expected that general trends will largely hold for the other relevant NAEP assessments (8th grade mathematics and 4th and 8th grade reading), although separate computations will be needed to calculate specific sample sizes. A total of 40 states have standardized test data from this assessment.[21] The remaining states covered by the NCLB Act, however, did not participate in the assessment, and we are therefore unable to provide sample size estimates for them based on expected achievement values.

Standardized test scores

NAEP standardized tests scores are obtained from an item response model fitted to the raw test score data. The scores themselves take values from 0 to 300 or 0 to 500, depending on the assessment (reading, mathematics, etc.)

Considering the test scores as continuous data, the difference-in-gap statistics of interest are $\Delta\bar{y} - \Delta\bar{x}_t$, and $-\Delta\bar{x}_t$. The sample size expression based on margin of error depends on the desired margin of error, the standard deviation of the test scores σ, the study

[21] The ten missing states are Alaska, Connecticut, Delaware, Florida, New Hampshire, New Jersey, Pennsylvania, South Dakota, Washington, and Wisconsin. The District of Columbia is also missing.

design effect, the proportion of disadvantaged students τ_1, and, in the case of $\Delta\bar{y} - \Delta\bar{x}_t$, the proportion of advantaged students τ_2. Parameters τ_1 and τ_2 can be estimated from the state-level distributions described in Section 4, while the design effect will be set equal to one, corresponding to the effective sample size for a simple random sample. Sample sizes for studies with other sampling designs can be obtained by multiplying the sample size by the study design effect.

The standard deviation of the test scores for the 2000 Grade 4 NAEP mathematics assessment was set equal to 30. This value was not estimated directly from the year 2000 standardized scores, but was selected based on two considerations. First, that the base year mathematics scores for grades 4, 8, 12 have a standard deviation set equal to 50 for all grades combined, which implies a within-grade standard deviation of less than 50, and second, that estimates of standard deviations from the 1996 mathematics assessment results support this value. Standard deviations are estimated as $\hat{\sigma} = SE\sqrt{n/deff}$, where *SE* is the standard error, and the 1996 estimates are given in Table A-5 (in the appendix) by race/ethnicity category. The estimated standard deviations for 1996 also suggest that the assumption of equal standard deviations across groups is reasonable. Note that these estimates may be biased upward as the design effects are based on both public and private school scores, whereas the NCLB Act is only concerned with public schools.

Mean scale scores for each advantaged (Asian, white) and disadvantaged group (American Indian, black, Hispanic) are given in Table A-6. Also given is a mean score for the advantaged group as a whole, as well as the size of the gap for each of the disadvantaged groups. The advantaged group score was computed as a weighted average of white and Asian scores, with the exception of Hawaii, where only white was considered advantaged. Since Asians are generally present at low frequency in the population, the advantaged group score is very close to the white score. Gaps are often around one standard deviation $(\hat{\sigma} = 30)$ in size. More specifically, the average gap is 17.9 for American Indians (with a range of 8 to 30), 27.7 for blacks (with a range of 20 to 38), and 22.4 for Hispanics (with a range of 14 to 35).

Minimum effective sample sizes for the estimation of differences in gap for each of the disadvantaged groups using both difference-in-gap statistics are given in Table A-7. Desired margins of error were set equal to the observed gap divided by six, assuming that the target was a linear reduction of the states' observed year 2000 gaps over the six biennial tests covered by the Act. The large variation in required sample sizes across states and disadvantaged groups, as well as the smaller sample sizes for difference-in-gap statistic $-\Delta\bar{x}_t$, are readily apparent in this table.

A visual representation of the effective sample sizes of Table A-7 is given in Figure 5.1, which plots the natural logarithm of the required sample size against the percentage of the disadvantaged group as given in Table A-1. Horizontal lines indicating log sample sizes of 1,000, 5,000, and 10,000 are included for reference. As expected, all of the plots show general trends of decreasing sample sizes as the percentage of the disadvantaged population increases. The departure from a smoothly decreasing curve is due to the differing observed gaps in each of the states. States with smaller year 2000 gaps have smaller yearly targets for improvement, and these, in turn, require larger sample sizes to detect. American Indians in Oklahoma, for example, have an observed gap of 12, well below the average gap of 17.9 for this racial group, and for this reason require very large sample sizes even though Oklahoma has the largest percentage American Indian population among the states with data. Also apparent from these plots are the many states that require effective sample sizes below 5,000 and the fair number that require effective sample sizes below 1,000. Larger black and Hispanic populations in a number of states lead to greatly reduced sample size requirements.

Figure 5.1

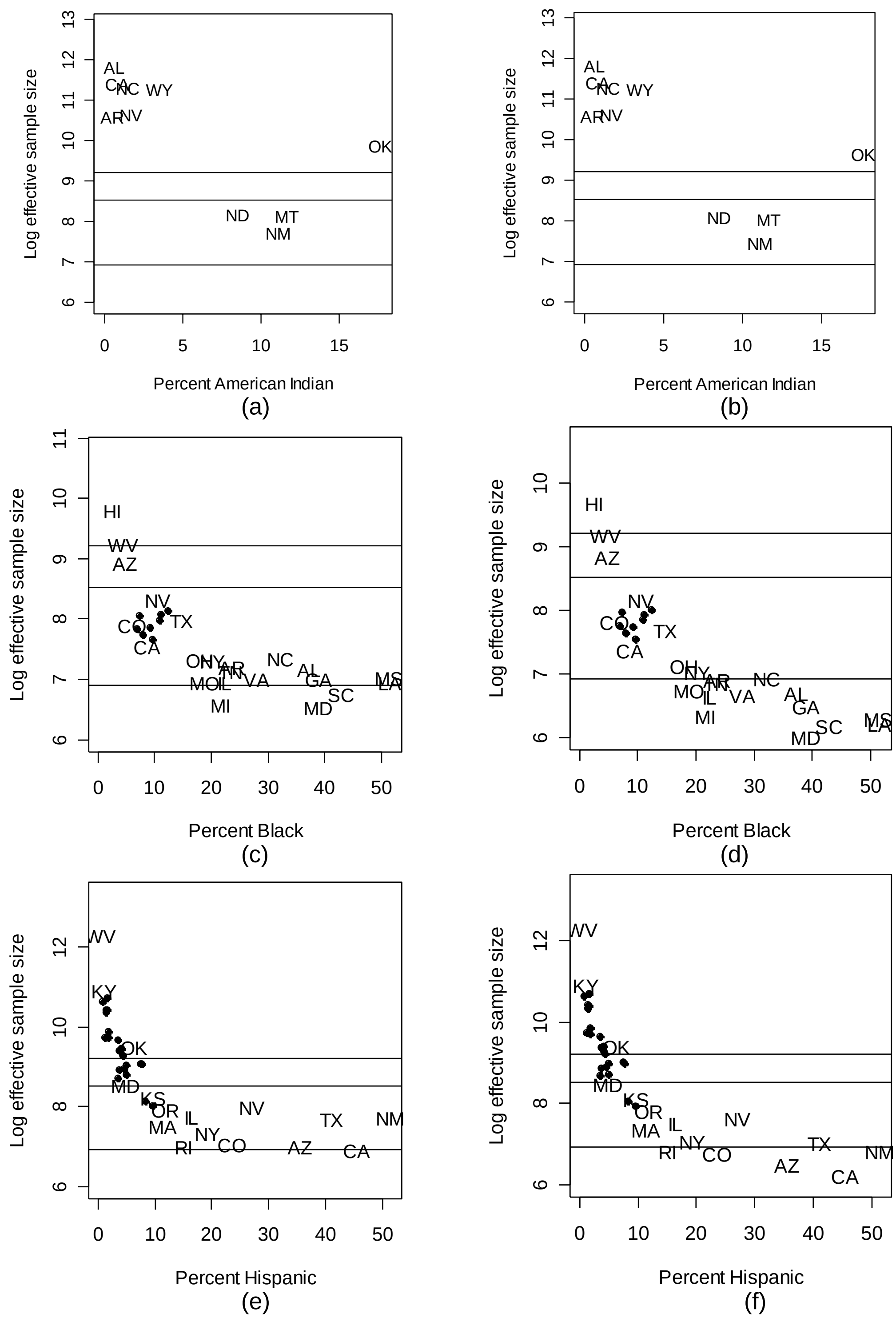


Figure 5.1: Log effective sample size versus percentage of disadvantaged group for difference-in-gap statistic $\Delta\bar{y}_t - \Delta\bar{x}_t$ (a, c, & e), and $-\Delta\bar{x}_t$, (b, d, & f) using mean standardized test score for NAEP grade 4 mathematics. Horizontal lines are log(1000), log(5000), and log(10000). Design effect equals 1, and standard deviation of standardized test scores is set equal to 30. Performance target is zero gap after twelve years. (Source: 2000 NAEP fourth grade mathematics assessment)

Effective sample size contour plots of margin of error versus disadvantaged group percentages are provided in Figure 5.2. These plots illustrate the effect that the specified margin of error has on sample sizes. Changes in margin of error simply shift the sample size up or down vertically. The effect of margin of error on the American Indian sample size in Oklahoma, for example, is again clear. Not all states with data are present on plots a, c, and e for the difference-in-gap statistic $\Delta\bar{y}_t - \Delta\bar{x}_t$. Constructing these plots requires an assumption that $(1-\tau_1) \approx \tau_2$, that is, that the percentage of advantaged students is approximately equal to 100 percent minus the percentage disadvantaged students. This is not true in states with more than one disadvantaged group of significant population size. States where the sample size under this assumption differs by more than 10 percent from that given in Table A-7 are not included in the plots.

Figure 5.2

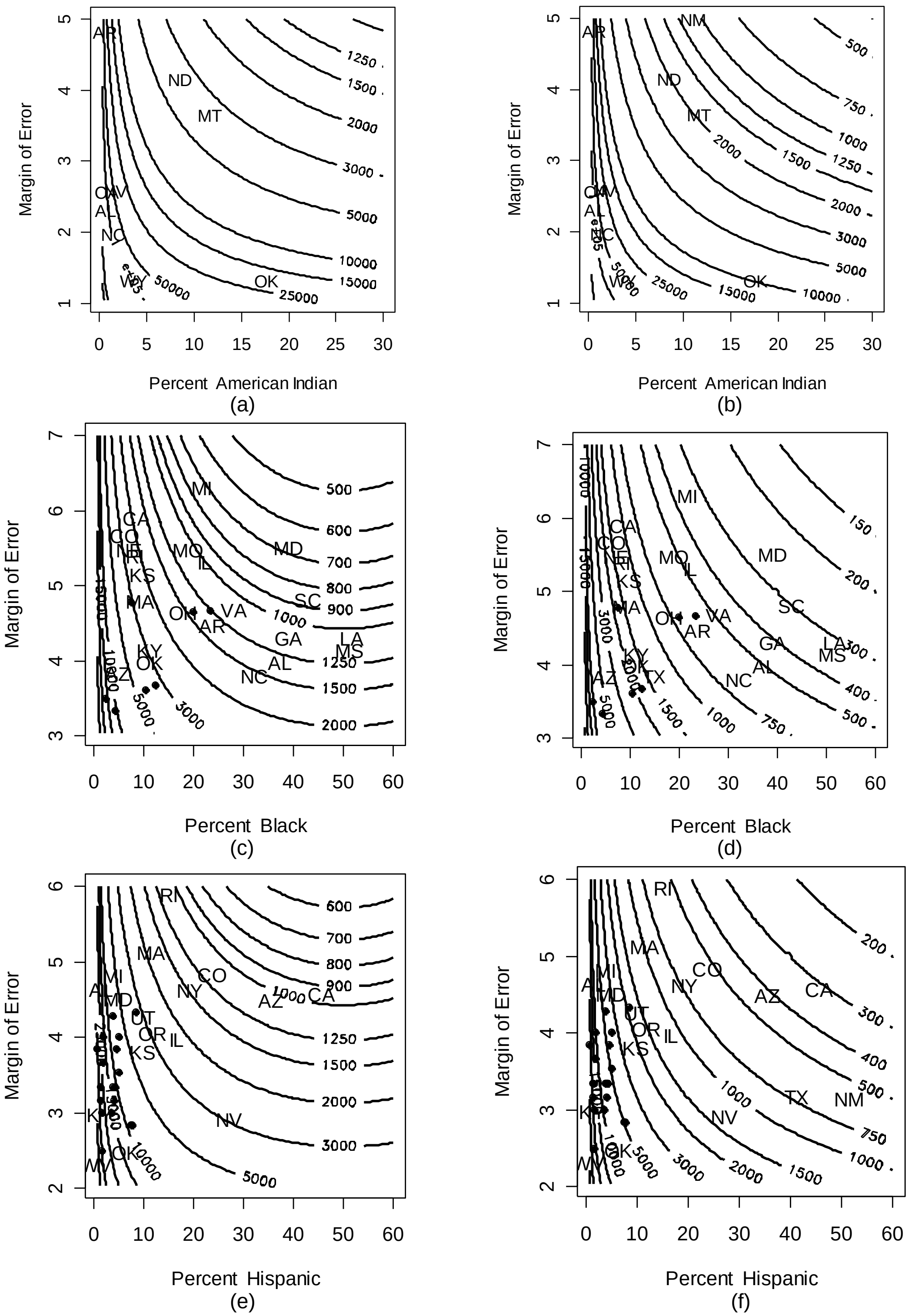


Figure 5.2: Effective sample size contour plots of margin of error versus percentage disadvantaged for difference-in-gap statistic $\Delta\bar{y}_t - \Delta\bar{x}_t$ (a, c, & e), and $-\Delta\bar{x}_t$, (b, d, & f). Sample sizes for states with effective sample sizes reasonably approximated by the margin of error formula are plotted.

(Source: 2000 NAEP fourth grade mathematics assessment)

Percentage at or above basic

The NAEP category of percentage at or above basic represents the percentage of students whose standardized scores were equal to or exceeded the cut-point for the basic achievement level.

The difference-in-gap statistics of interest are $\Delta\hat{q}_t - \Delta\hat{p}_t$ and $-\Delta\hat{p}_t$, and the sample size expression based on margin of error depends on the study design effect, the proportion disadvantaged τ_1, and in the case of $\Delta\hat{q}_t - \Delta\hat{p}_t$, the proportion advantaged τ_2. Parameters τ_1 and τ_2 can be estimated from state-level distributions discussed in Section 4, while the design effect will be set equal to one, corresponding to a simple random sample. Sample sizes for studies with other sampling designs can be obtained by multiplying the sample size by the study design effect. The margin of error also depends on parameters p_{t-1}, p_t, q_{t-1}, and q_t; however, we make the simplifying assumptions that they are all close enough in value to 0.5 to use this as an approximating upper bound, as described in Section 3.

The percentage at or above the basic level for the NAEP year 2000 mathematics assessment for advantaged (Asian, white) and disadvantaged groups (American Indian, black, Hispanic) are given in Table A-8. Also provided is the percentage of students at or above the basic level for the advantaged group as a whole, and the size of the gap for each of the disadvantaged groups. As before, the advantaged group score is computed as a weighted average of white and Asian scores, with the exception of Hawaii, where only white was considered advantaged. As was observed for the mean scale scores, the advantaged group percentages are very close to the white percentages due to the relatively small number of Asian students in the state populations. The distribution of gaps shows large variation across states and disadvantaged groups: American Indians have an average gap of 25.0 (with a range of 8 to 51), blacks have an average gap of 39.5 (with a range of 28 to 54), and Hispanics have an average gap of 29.5 (with a range of 15 to 45). Many of the percentages are close to 50 percent, suggesting that use of the variance upper bound is reasonable.

The minimum effective sample sizes for the estimation of differences in gap for each of the disadvantaged groups using the two difference-in-gap statistics are given in Table A-9. Desired margins of error were set equal to the observed gap divided by six, assuming that the target was a linear reduction of the states' observed year 2000 gaps over the six biennial tests covered by the Act. Again, a large variation in required sample sizes across states and disadvantaged groups is observed, with smaller sample sizes shown for difference-in-gap statistic $-\Delta\hat{p}_t$.

Plots of the natural logarithm of the effective sample size against disadvantaged group percentages are given in Figure 5.3. All of the plots show the general trend of decreasing sample sizes as the percentage disadvantaged population increases. As with mean scale scores, many states require total sample sizes below 5,000 and some require sample sizes below 1,000. Blacks and Hispanics again have smaller sample requirements in those states where they have larger population sizes.

Figure 5.3

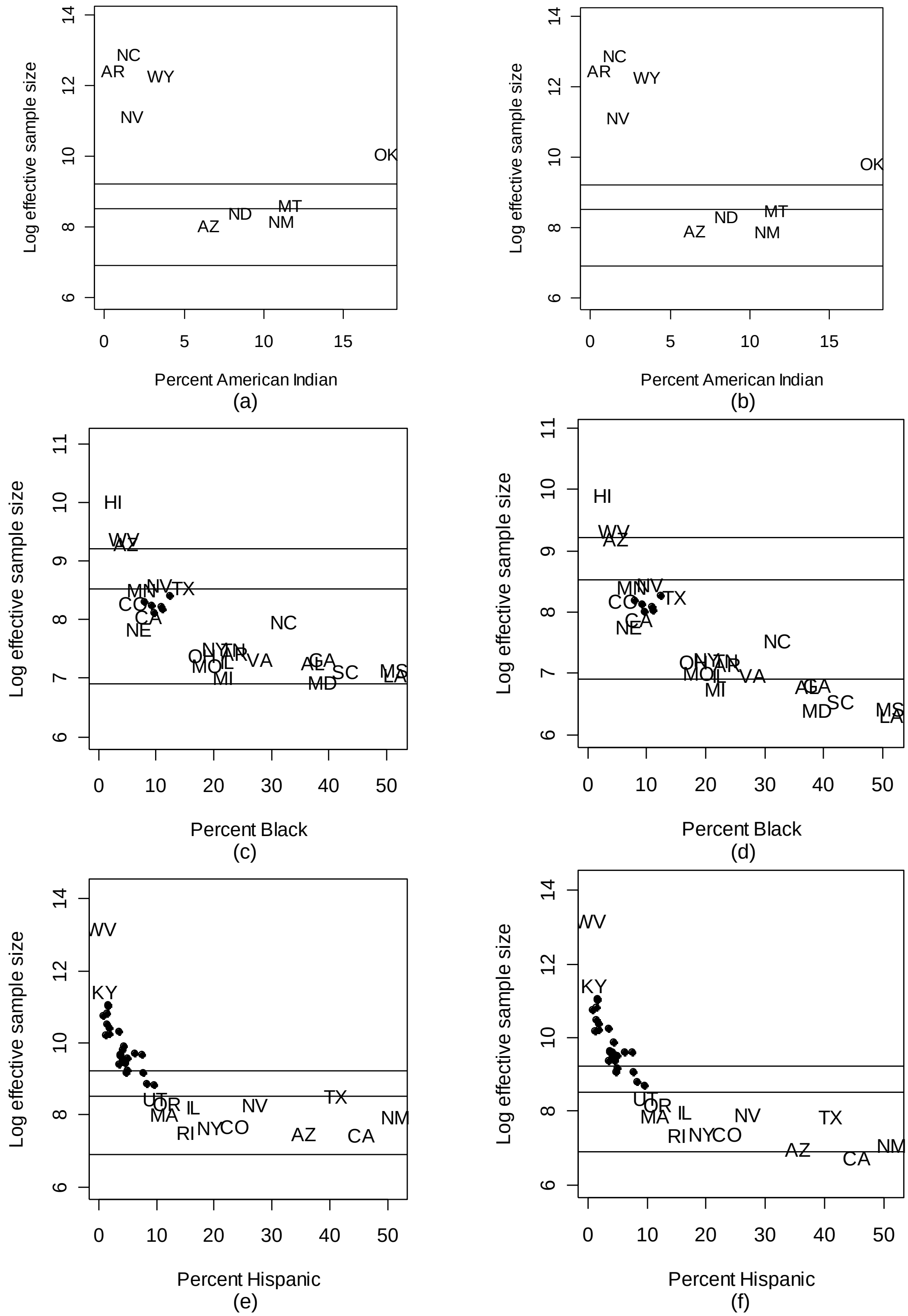


Figure 5.3: Log effective sample size versus percentage disadvantaged group for difference-in-gap statistic $\Delta\hat{q}_t - \Delta\hat{p}_t$ (a, c, & e), and $-\Delta\hat{p}_t$, (b, d, & f) using percentage of students at or above NAEP basic achievement level. Horizontal lines are log(1000), log(5000), and log(10000). Design effect equals 1. Performance target is zero gap after twelve years.

(Source: 2000 NAEP fourth grade mathematics assessment)

Contour plots of margin of error versus disadvantaged group percentages are provided in Figure 5.4 to illustrate the effect the specified margin of error has on sample sizes. As in the case of mean scale scores, not all states are present on the difference-in-gap statistic $\Delta\hat{q}_t - \Delta\hat{p}_t$ plots a, c, & e, due to failure to meet the assumption $(1-\tau_1) \approx \tau_2$. States where the sample size under this assumption differed by more than 10 percent from that given in Table A-9 are not included in the plots.

Percentage at or above proficient

The NAEP category of percentage at or above proficient represents the percentage of students whose standardized scores were equal to or exceeded the cut-point for the proficient achievement level.

As with percentage at or above basic, the difference-in-gap statistics of interest are $\Delta\hat{q}_t - \Delta\hat{p}_t$, and $-\Delta\hat{p}_t$, and the sample size expression based on margin of error then depends on the study design effect, the proportion disadvantaged τ_1, and in the case of $\Delta\hat{q}_t - \Delta\hat{p}_t$, the proportion advantaged τ_2. Parameters τ_1 and τ_2 can be estimated from state-level distributions described in Section 4, and the design effect is set equal to one. Sample sizes for studies with other sampling designs can be obtained by multiplying the sample size by the study design effect. Since the margin of error also depends on parameters p_{t-1}, p_t, q_{t-1}, and q_t, we make the simplifying assumptions that they can be approximated by the average of the proportion at or above proficient for the advantaged and disadvantaged groups as given in Table A-10. This is considered a slightly better approximation than setting the overall maximum at 0.5, as both proportions are generally both below 0.5.

The percentages of students at or above the proficient level for the NAEP 2000 mathematics assessment for advantaged (Asian, white) and disadvantaged groups (American Indian, black, Hispanic) are given in Table 5.6 (see appendix). Also provided is the percentage of students at or above the proficient level for the advantaged group as a whole, and the size of the gap for each of the disadvantaged groups. Again, percentages for the advantaged group are very close to the white percentages due to the relatively small number of Asian students in state populations.

Figure 5.4

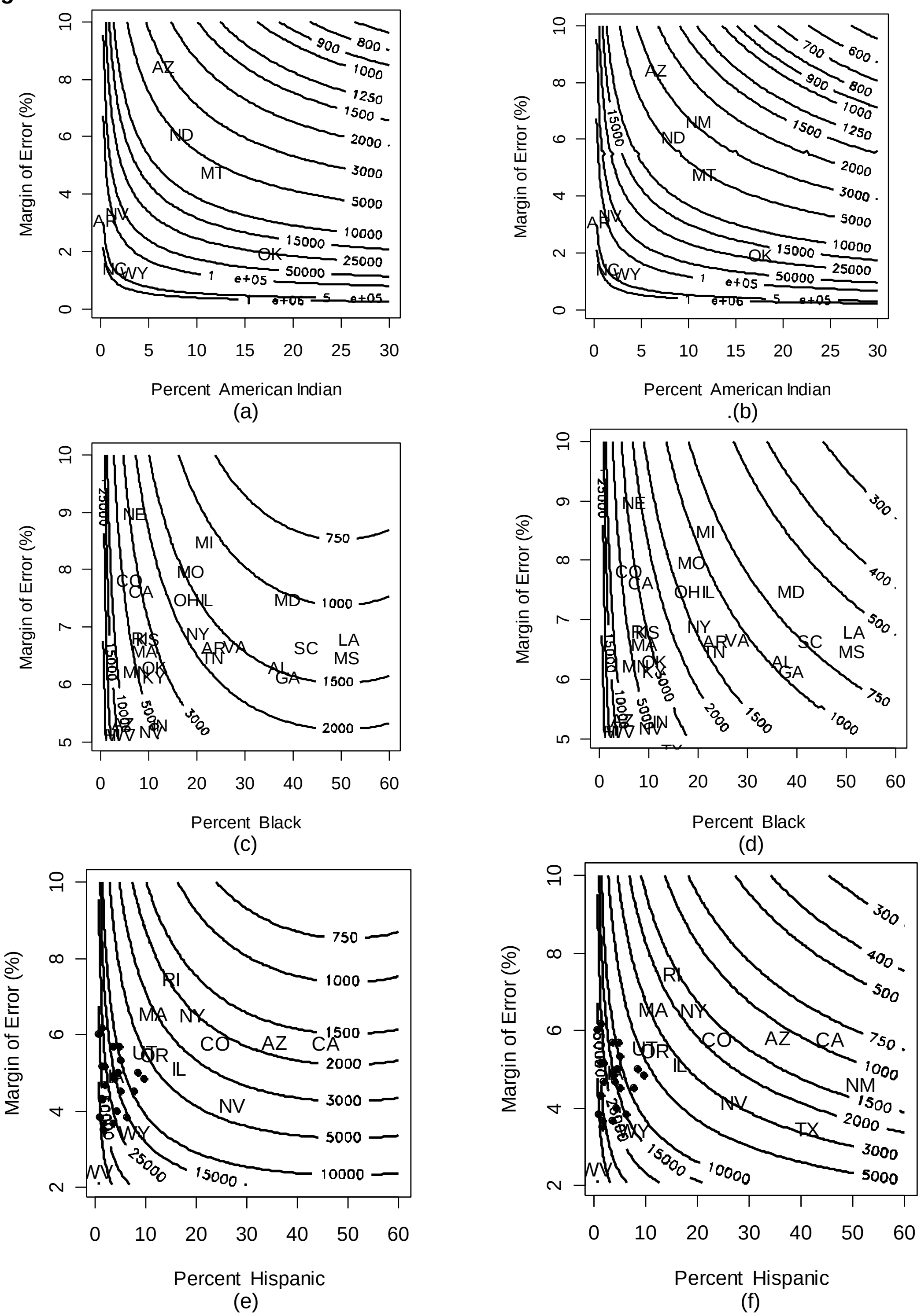


Figure 5.4: Effective sample size contour plots for margin of error versus percentage disadvantaged for difference-in-gap statistic $\Delta\hat{q}_t - \Delta\hat{p}_t$ (a, c, & e), and $-\Delta\hat{p}_t$, (b, d, & f) using percentage of students at or above basic achievement level. Sample sizes for states with effective sample sizes reasonably approximated by the margin of error formula are plotted.

(Source: 2000 NAEP fourth grade mathematics assessment)

The distribution of gaps shows large variation across states and disadvantaged groups, although not as much as the distribution of gaps for students at or above the basic level, since with the proficient level as the criterion, all percentages are smaller than 50 percent. American Indians have an average gap of 15.4 with a range of 8 to 22, blacks have an average gap of 23.8 with a range of 13 to 35, and Hispanics have an average gap of 19.1 with a range of 6 to 32.

The minimum effective sample sizes for the estimation of differences in gaps for each of the disadvantaged groups for both difference-in-gap statistics are given in Table A-11. Desired margins of error were set equal to the observed gap divided by six, assuming a target of linearly reducing states observed year 2000 gaps over the six biennial tests covered by the Act. Again, a large variation in required sample sizes across states and disadvantaged groups is observed, with smaller sample sizes shown for difference-in-gap statistic $-\Delta\hat{p}_t$. Sample sizes are much larger than for either the percentage of students at or above basic achievement level, or mean scale score difference-in-gap statistics.

Plots of the natural logarithm of the required effective sample size against disadvantaged group percentages are given in Figure 5.5. All of the plots show the trend observed for other statistics of decreasing sample sizes as the percentage disadvantaged population increases. In this case, only one student group in one state falls below an effective sample size of 1,000 (Maryland, for black), although there are still a number below 5,000. All of the American Indian sample sizes exceed 5,000.

Figure 5.5

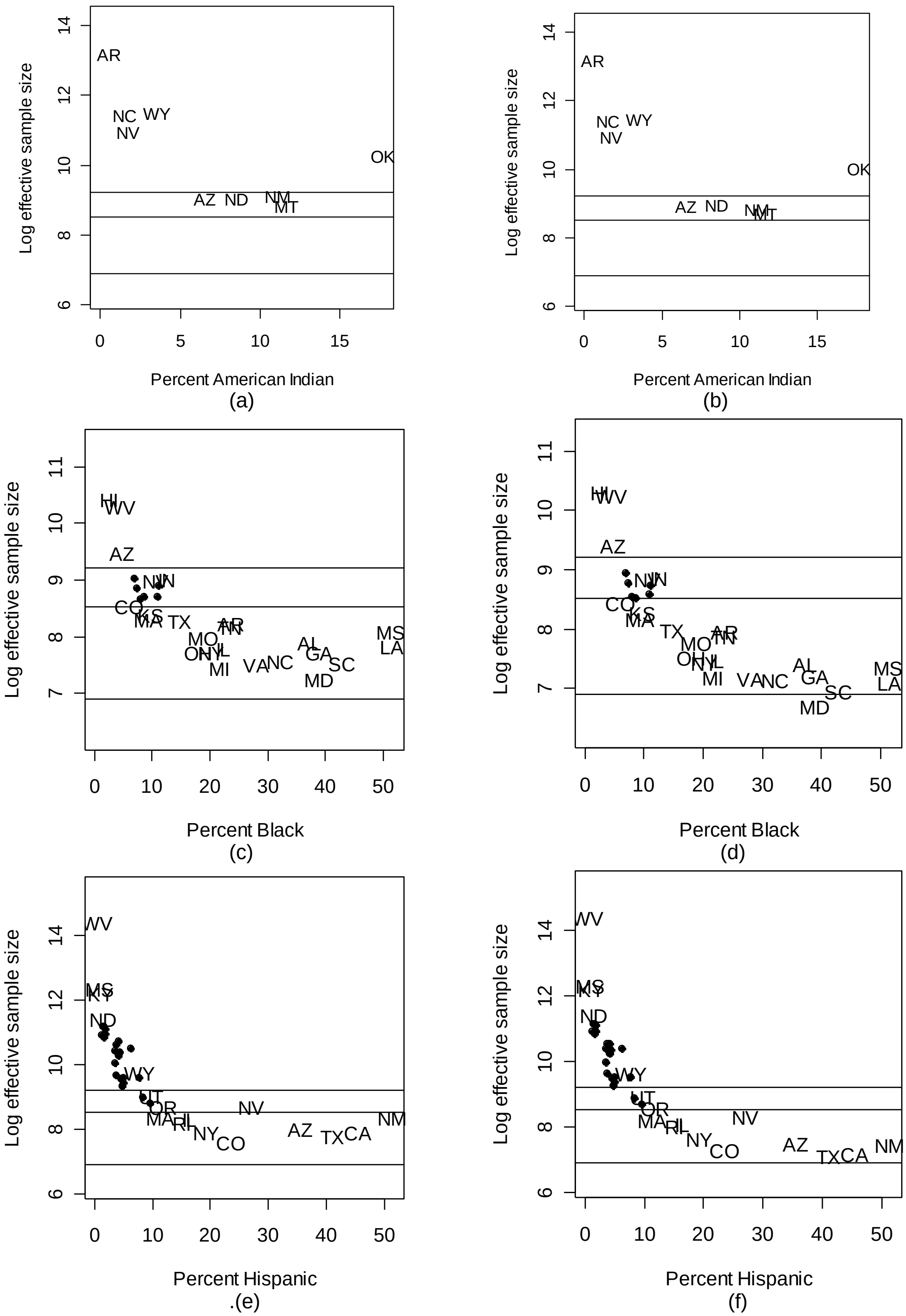


Figure 5.5: Log effective sample size versus percentage disadvantaged group for difference-in-gap statistic $\Delta\hat{q}_t - \Delta\hat{p}_t$ (a, c, & e), and $-\Delta\hat{p}_t$, (b, d, & f) using percentage of students at or above NAEP proficient achievement level. Horizontal lines are log(1000), log(5000), and log(10000). Design effect equals 1. Performance target is zero gap after twelve years.

Contour plots of margin of error versus percentage disadvantaged group are not provided for percentage at or above proficient level because setting the approximation p_{t-1}, p_t, q_{t-1}, and q_t all equal to .5 was not used for sample sizes in this case. The margin of error expression differs too much from that used to produce the plots.

6. Fixed performance targets

In contrast to the previous section, which included state-specific goals to be used in finding state-level sample sizes, this section will describe an approach to finding sample sizes based upon a fixed, common goal across all states. A fixed performance goal can avoid complications introduced by specific state targets, such as the fact that small, less important differences in gaps can require huge sample sizes to detect. It also allows sample sizes to be obtained for all states, both those with previous NAEP data and those without.

Specifying fixed performance targets

The first step in setting fixed goals is to find an acceptable common goal for all states. The two approaches taken here are first, to study the legislation for guidance, and second, to observe the precision obtained in previous NAEP samples (as provided in Carlson, 2003) in order to suggest the levels of precision attainable in practice.

The legislative targets of the Act were described in Section 2. Achieving adequate yearly progress requires that over 12 years, the percentage of students scoring at the basic proficiency level be driven to zero, or be decreased by 10 percent per year. Gaps are either to be reduced or eliminated. As an example, using the 4th grade mathematics assessment and equating the basic proficiency level, as given in the legislation, with below basic on NAEP, consider a mean scale score of 235 for the advantaged group and a score of 205 for the disadvantaged group, and a percentage at or above the basic achievement level of 75 percent for the advantaged group and 25 percent for the disadvantaged group. These numbers would suggest targets of a mean scale score gap reduction of 5 points per two years [(235-205)/6=5.0], and a reduction of 8.3 percentage points per two years [(75-25)/6=8.3] for gaps measured on the basic achievement level. For AYP for the advantaged group, the proficiency targets would be 25/6=4.2% per two years in order to eliminate below basic performance, and about $(25-.9^{12}*25)/6=3.0\%$ per two years using the 10 percent rule. It is not clear how AYP legislative targets would be found using mean scale scores.

A study of previously observed levels of NAEP precision is given in Carlson (2003). For each performance measure, a fixed standard error was chosen by inspecting plots of observed NAEP standard error versus NAEP sample size for all of the disadvantaged and advantaged groups. For example, for the year 2000 4th grade mathematics assessment, a standard error of approximately 3.0 could be expected at a NAEP sample size of 200 (regardless of race or ethnicity).[22] In Carlson's analysis, the typical sample size of 200 was then used to ask whether each state's racial/ethnic composition would allow the state to achieve that sample size for each of its racial and ethnic groups without any oversampling.

Table A-12 contrasts the margins of error required to detect the legislative targets with the observed NAEP margins of error for both mean scale score and percentage at or above the basic achievement level in the 4th grade mathematics assessment. It considers both AYP (gaps based on differences from a constant: $-\Delta\hat{p}_t, -\Delta\hat{x}_t$), and standard performance gaps (gaps based on differences in current means: $\Delta\hat{q}_t - \Delta\hat{p}_t, \Delta\bar{y}_t - \Delta\bar{x}_t$).[23] It can be seen that the legislative targets are smaller than the observed margins of error, and therefore detecting them should require somewhat larger sample sizes than those in previous NAEP measurements.

In our analysis, we will assume Carlson's typical standard error, and ask what sample size it implies for the racial/ethnic minority, as well as for the state as a whole (without oversampling). This analysis can identify the approximate sample sizes required to detect changes in performance of at least the size of the current typical NAEP margin of error for the total population.

Sample sizes for fixed performance targets

Tables A-13 through A-24 give effective and nominal sample sizes by state for AYP, and changes in gap statistics for both mean scale scores (MSS) and percentage at or above the basic achievement level based on the NAEP 4th grade mathematics assessment. Tables are provided separately for each of the three disadvantaged groups.

[22] Note that the NAEP sample size of 200 is not directly comparable to the NAEP sample sizes given later, because the sample size of 200 is for the standard error of a single mean whereas the statistics of interest are functions of either two (adequate yearly progress) or four (changes in gap) means.

[23] This section of the report was written later, after recommendations from NAEP led to restricting analysis to gap statistics based on current performance differences. It was also decided to no longer consider the percentage of students at or above the proficient achievement level as a performance measurement. Support for these decisions is provided in Section 7. Because adequate yearly progress is still of interest, the analysis here refers to gaps based on differences from a constant ($-\Delta\hat{p}_t, -\Delta\hat{x}_t$) as "adequate yearly progress," and gaps based on differences in current means ($\Delta\hat{q}_t - \Delta\hat{p}_t, \Delta\bar{y}_t - \Delta\bar{x}_t$) as "performance gaps."

Effective sample sizes were obtained by substituting typical NAEP margins of error from Table A-12 into the expressions detailed in Section 3. That is, a margin of error expression under simple random sampling was solved for the required effective sample size. Margins of error were specified according to $\alpha = .05$ one-sided confidence intervals, and effective sample sizes were obtained through solving for n, where n represents an overall effective sample size. Note that the sum of the effective sample sizes for the disadvantaged and advantaged groups does not, in general, equal the total effective sample size due to the presence of other groups in the sample.

Nominal sample sizes were then obtained by multiplying the effective sample sizes by the design effect, which here is taken to equal 3.[24] Total number of students per state has been provided for comparison.

Tables A-13 through A-18 give the effective and nominal sample sizes for AYP, sorted in descending order by the percentage of disadvantaged students. The nominal sample size for the total sample is also given, along with the actual number of students in the state or the target grade level. From these tables it is apparent that tracking AYP based on mean scale scores generally allow smaller nominal sample sizes than tracking AYP based on the percentage at or above basic. The standard NAEP state sample size per subject per grade is about 2,500 students; therefore, without oversampling, detectable differences in mean scale scores at about the current level of NAEP precision could currently be obtained for American Indians in 3 states, blacks in 23 states and Hispanics in 12 states. For percentage at or above the basic achievement level these counts are 1, 17, and 8 states respectively.

Tables A-19 through A-24 give the effective and nominal sample sizes for performance gaps, sorted in descending order by percentage of disadvantaged students. Again, it is apparent that the gap statistics based on mean scale scores generally allow smaller nominal sample sizes than those based on the percentage at or above basic. Detectable differences in mean scale scores at about the current level of NAEP precision (with an approximately 2,500 student sample) could currently be obtained for American Indians in 4 states, blacks in 31 states and Hispanics in 14 states. For the percentage at or above the basic achievement level, these counts are 1, 10, and 5 states respectively.

7. Conclusions and recommendations

This paper has addressed a variety of issues relating to federal confirmation of state assessments for the No Child Left Behind Act. Principal questions have concerned how gaps

[24] NAEP design effects have an approximate range of 2 to 4.

and improvement through reducing gaps should be defined, which statistics are most appropriate for measuring gap improvement, which performance measures might be used, and what state NAEP sample sizes would be required for the different targeted minority groups.

Choice of gap statistic

We have described two approaches for measuring a gap: the difference in the current year performance between the disadvantaged group and the advantaged group, and the difference between the disadvantaged group's performance in the current year and that of the advantaged group in a baseline year.

Of the two approaches, the difference in performance between groups in the current year is the natural choice because it incorporates the most essential aspect of equality. However, this approach has the drawback of having a somewhat larger variance due to the contribution from the advantaged group. Furthermore, if a zero gap after a certain amount of time is the target, this method might produce more challenging improvement goals.

Gaps measured according to difference in performance between the disadvantaged group in the current year and the advantaged group in the baseline year eliminate the advantaged group's contribution to the variance and therefore require smaller sample sizes. The influence of the advantaged group remains only through its role in providing a target against which improvement in the disadvantaged group is measured. A disadvantage of this method is that scores can be recorded as 'improving' (see Figure 3.1) when in fact the absolute inequality between the two groups may be constant or increasing. For example, it allows gaps to "improve" when the advantaged group and the disadvantaged group are improving together. Historically such a situation is not unexpected (Yen, 2002).

For both approaches, the possibility of deteriorating performance in the advantaged group introduces what appears to be a necessary requirement: that "improvement" be defined to only occur if the advantaged group does not deteriorate in performance. This requirement prevents a change in a gap being recorded as "improvement" when objectively the results are unclear or perhaps tend in the opposite direction (again, see Figure 3.1).

Such a requirement does introduce some additional complications, however. Testing procedures must be multivariate, and the sample size determination is more complex. In this paper we have provided sample sizes based on target margins of error for a decrease in the gap without considering the possibility of decreasing advantaged-group performance. In terms of the illustration given in Figure 3.1, the sample sizes used here refer to a decrease in gaps of such a size as to eliminate the gap after 12 years for any point on the zero gap line. If consideration of declining advantaged-group performance were introduced, the sample sizes would likely be larger. Precisely how to determine sample sizes in this case is a topic for further study.

This paper has also discussed adequate yearly progress, whose variance and sample sizes are the same as gap statistic $-\Delta\hat{p}_t$.[25] Adequate yearly progress does not require comparisons across groups, and so the sample sizes in Table A-9 apply directly. For these reasons, it appears that, given gaps and AYP, adequate yearly progress is the easier measure to confirm.

Choice of performance measure

In addition to the two approaches to defining a "gap," three measures of performance have been considered: mean scale score, the proportion of students at or above the NAEP basic achievement level, and the proportion of students at or above the NAEP proficient achievement level.

Mean scale scores required the smallest sample sizes of the three, and allowed simplified computation through variance expressions that do not depend on the mean. This may offer some advantages for computing sample sizes, since mean scores are expected to change dramatically over the lifetime of the Act. Thus, appropriate sample sizes could be set once for the duration of the Act.

Using the proportion of students at or above the basic achievement level required sample sizes larger than using the mean scale score, but smaller than when using the proportion of students at or above the proficient level. Thus the proportion of students at or above the basic achievement level appears to be the more usable of the two achievement level performance measures. It also appears to be most compatible with the AYP statistic, providing a consistent quantitative measure for both gaps and adequate yearly progress. When gaps are measured relative to the advantaged-group performance at baseline, however, the gap statistics based on the proportion of students at or above basic is completely correlated with AYP, so the two measures become redundant. Another consideration is that, like the measure of the proportion of students at or above the proficient achievement level, the proportion at or above basic has a variance that depends on its mean. Sample sizes might therefore have to be recomputed each year to maintain specific margins of error. Refining the approach given in this paper to account for current and projected values of these proportions is a topic for future work.

[25] For example, when gap is defined relative to the advantaged group's baseline level.

Gap statistics based on the proportion of students at or above the proficient level required the largest sample sizes of the three performance measures. In large part, this may have been due to the smaller margins of error since there was simply less of a gap to reduce. The variance approximation for this gap statistic (based on the average of advantaged and disadvantaged groups) was smaller, although this was apparently not enough to offset the smaller margins of error.

State-level sample sizes

Selection of state-level sample sizes depends on a variety of considerations described in this paper. Assuming that a difference in mean scale scores is used to measure gaps and changes in the percentage at or above the basic achievement level are used to measure annual yearly progress, then another important consideration is whether improvement targets are fixed or depend on existing performance differences and legislative goals.

If the improvement targets depend on legislative goals and existing differences (Tables A-7, A-9, A-12, Figures 5.3, 5.4), then the requirements for setting sample size are much more demanding. To illustrate this difference, consider changes in performance gaps for the American Indians. From Figure 5.1(a), it can be seen that four of the NAEP states (MT, ND, NM, OK) have substantial American Indian populations. For these states, effective sample size requirements are (from Table A-7) 3,570, 3,642, 2,333, and 19,877, respectively. Note that Oklahoma, the state with the largest American Indian population, is also the state with the largest required sample size — due to the strong performance of that group on the NAEP assessments. Nominal sample sizes, assuming a design effect of 3, are 10,710, 10,926, 6,999, and 59,631. In contrast, nominal sample sizes for these same four states assuming a fixed typical NAEP precision (from Table A-19) are 2,352, 2,856, 2,856, and 1,731. To see why this is the case, examine Table A-12. This table shows that a typical NAEP precision is 7.8, which, in Figure 5.2(a), is outside the range of the y-axis. As described earlier, detecting the relatively small changes necessary to eliminate the gap for a disadvantaged group with relatively high baseline performance requires a considerably larger sample size than current NAEP assessments typically provide, at least for this disadvantaged group.

Note also that larger samples are necessary if the legislative target require that gaps be reduced by some fraction of their current levels, compared to samples for targets requiring that gaps be reduced to zero. From a sampling perspective, therefore, more realistic decreases in gap (i.e., not to zero) require larger sample sizes to detect.

In states where the nominal sample size for a disadvantaged group is only slightly larger than what might realistically be attempted with proportional allocation, oversampling might be introduced. Whether or not to oversample would be considered on a state-by-state basis. Efforts to increase the representation of one group can, of course, lead to decreased representation of other groups if overall sample sizes remain fixed. Therefore, any effects on the precision of estimates for other groups should be considered.

Other issues

We have used the margin of error to develop sample sizes. Another common approach is to use power analysis. In a power analysis, the problem is specified in terms of a hypothesis test, and a minimum sample size to detect a certain level of difference between null and alternative hypothesis is found. A power analysis would generally lead to larger sample size requirements than the analyses described here. Such an analysis is a topic for further study.

In this study, we limit our analyses to the NAEP data from the 4th grade mathematics assessments. For other assessments the general conclusions of this paper should also apply, although specific sample sizes will of course require separate analyses.

Only states included in the NAEP assessments allowed for sample size determination by all three margin of error calculation approaches. If there were a need to obtain margins of error for non-NAEP states, information on the ethnic and racial makeup of those states might be used in conjunction with performance information from other sources (such as state tests).

The sample size results of this study depend on various assumptions and approximations. Were these reasonable? We believe that our estimates of sample size reveal the important trends, and the relative magnitude of the effects. Some analysis of robustness is warranted, however, and could increase the precision of the sample size estimates given here.

Under the NCLB Act, one of the goals of the NAEP assessments is to confirm the results of state-administered assessments. However the state assessments have much larger sample sizes than NAEP. Therefore, it is clear from the results of this study, that NAEP samples will not allow direct confirmation of all state assessment results concerning gaps and adequate yearly progress. Ideally, NAEP samples will at least suggest no inconsistency. The ability of the NAEP samples to confirm state results might be improved by changing or expanding the approach to confirmation (perhaps by restricting confirmation to adequate yearly progress), or by altering the sample collection methodology (for example, pairing student level results on both the NAEP and state-level assessments).

References

Allen, N.L., Carlson, J.E., & Zelenak, C.A. (1999). *The NAEP 1996 Technical Report* (NCES 1999-452). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Allen, N.L., Donoghue, J.R., & Schoeps, T.L. (2001). *The NAEP 1998 Technical Report* (NCES 2001-509). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Braswell, J.S., Lutkus, A.D., Grigg, W.S., Santapau, S.L., Tay-Lim, B.S.-H., & Johnson, M.S. (2001). *The Nations Report Card: Mathematics 2000* (NCES 2001-517). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Bush, G.W. (2001). *The no child left behind act*. Public Law 107-110, U.S. Congress.

Carlson J. (2003). *Data on grades 4 and 8 subgroup compositions of states participating in state NAEP*. Working paper, unpublished manuscript.

Holland, P. (2002). Measuring progress in student achievement: Changes in scores and score-gaps over time. In *Report of the Ad Hoc Committee on Confirming Test Results* (Appendix B). Washington, DC: National Assessment Governing Board.

National Center for Education Statistics. (2001). *Common Core of Data 2000-01: Information on Public Schools and School Districts in the United States.* Available on-line at http://nces.ed.gov/ccd/pubschuniv.html

Nettles, M., Domenech, D., Haertel, D., Kopp, N., Paulson, D., Ravitch, D., Ward, M., Whirry, M., & Palmer Wolf, D. (2002). Using the National Assessment of Educational Progress to confirm state test results. In *Report of the Ad Hoc Committee on Confirming Test Results*. Washington, DC: National Assessment Governing Board.

United States Census Bureau. (2000). *United States 2000 Census*. Available on-line at http://www.census.gov/main/www/cen2000.html

Yen, W. (2002). The use of NAEP results in a confirming role for ESEA: A thought experiment with historical data from state A. In *Report of the Ad Hoc Committee on Confirming Test Results* (Appendix C). Washington, DC: National Assessment Governing Board.

Appendix

Table A-1: Race/ethnicity distributions for the fourth grade by state or district.

State or District	Percentage American Indian	Percentage Asian or Pacific Islander	Percentage Hispanic	Percentage Black not Hispanic	Percentage White
Alabama	0.65	0.71	1.37	37.15	60.12
Alaska	26.20	5.18	3.45	4.99	60.19
Arizona	6.53	1.96	35.35	4.67	51.49
Arkansas	0.51	0.74	3.87	23.55	71.33
California	0.80	10.47	45.40	8.60	34.74
Colorado	1.28	2.72	23.56	5.96	66.48
Connecticut	0.21	2.73	13.59	14.01	69.47
Delaware	0.21	2.37	6.63	32.61	58.17
District of Columbia	0.03	1.29	9.07	85.01	4.60
Florida	0.29	1.74	19.55	25.23	53.19
Georgia	0.15	2.03	5.12	38.96	53.74
Hawaii	0.39	72.81	4.20	2.46	20.15
Idaho	1.38[3]	0.89[3]	7.78[3]	0.39[3]	89.57[3]
Illinois	0.16	3.20	16.45	22.22	57.97
Indiana	0.18	0.88	3.58	12.36	83.00
Iowa	0.52	1.65	4.15	4.53	89.15
Kansas	1.40	2.23	9.76	9.82	76.79
Kentucky	0.16	0.60	0.97	11.01	87.26
Louisiana	0.68	1.13	1.45	51.55	45.19
Maine	0.30	1.14	0.65	1.26	96.64
Maryland	0.29	4.09	4.88	38.91	51.83
Massachusetts	0.26	4.47	11.32	9.24	74.71
Michigan	0.99	1.82	3.67	21.51	72.01
Minnesota	2.29	5.22	3.80	7.46	81.22
Mississippi	0.16	0.61	0.82	51.24	47.16
Missouri	0.30	1.16	1.95	18.75	77.84
Montana	11.59	0.82	1.80	0.79	85.00
Nebraska	1.73	1.30	8.42	7.04	81.51
Nevada	1.71	5.39	27.12	10.54	55.24
New Hampshire	0.18	1.29	1.90	1.10	95.54
New Jersey	0.17	6.39	15.44	17.77	60.23
New Mexico	11.05	0.93	51.34	2.47	34.22
New York	0.40	5.89	19.02	19.98	54.71
North Carolina	1.50	1.76	4.66	32.23	59.86
North Dakota	8.49	0.86	1.49	1.32	87.84
Ohio	0.12	1.12	1.69	17.80	79.28
Oklahoma	17.64	1.36	6.27	11.20	63.53
Oregon	2.23	3.75	11.68	3.04	79.29
Pennsylvania	0.11	1.90	4.82	16.27	76.89
Rhode Island	0.43	3.12	14.92	8.12	73.40
South Carolina	0.25	0.84	1.92	42.88	54.12
South Dakota	12.68	0.94	1.58	1.46	83.35
Tennessee	0.07[2]	2.21[2]	4.46[1]	23.37[1]	69.89[1]
Texas	0.30	2.53	41.10	14.75	41.33
Utah	1.60	2.76	9.78	1.07	84.79
Vermont	0.49	1.10	0.57	1.25	96.59
Virginia	0.27	3.93	5.03	28.04	62.73
Washington	2.57[2]	8.83[2]	10.71[1]	4.44[1]	73.45[1]
West Virginia	0.08	0.48	0.40	4.42	94.61
Wisconsin	1.54	3.61	4.84	11.18	78.83
Wyoming	3.49	1.02	7.63	1.38	86.48

(Source: CCD 2000-2001)

[1]NAEP 1998 non-response adjusted estimates, 4th grade reading
[2]Adjusted Census 2000 counts for all inhabitants after allowing for NAEP 1998 estimates
[3]Adjusted Census 2000 counts for all inhabitants.

Table A-2: Race/ethnicity distributions for the eighth grade by state or district.

State or District	Percentage American Indian	Percentage Asian or Pacific Islander	Percentage Hispanic	Percentage Black not Hispanic	Percentage White
Alabama	0.81	0.69	1.10	35.53	61.88
Alaska	24.45	5.71	3.28	4.18	62.38
Arizona	6.88	1.96	31.75	4.42	54.98
Arkansas	0.47	0.93	3.25	22.23	73.12
California	0.90	11.49	40.65	8.66	38.30
Colorado	1.09	2.62	20.48	5.38	70.42
Connecticut	0.24	2.54	12.58	13.02	71.62
Delaware	0.43	2.19	5.31	31.76	60.31
District of Columbia	0.00	2.02	9.08	84.46	4.45
Florida	0.26	1.91	18.87	24.05	54.91
Georgia	0.14	2.16	4.04	37.36	56.30
Hawaii	0.28	72.82	4.76	2.03	20.10
Idaho	1.38[3]	0.89[3]	7.78[3]	0.39[3]	89.57[3]
Illinois	0.21	3.36	14.14	20.09	62.20
Indiana	0.22	0.92	3.09	11.01	84.75
Iowa	0.53	1.69	3.21	3.62	90.95
Kansas	1.31	2.05	7.98	8.46	80.21
Kentucky	0.22	0.73	0.76	9.54	88.76
Louisiana	0.67	1.22	1.28	50.83	45.99
Maine	0.49	0.91	0.57	0.96	97.07
Maryland	0.34	4.42	4.17	36.06	55.00
Massachusetts	0.31	4.25	10.10	8.31	77.04
Michigan	1.12	1.83	3.18	16.95	76.92
Minnesota	2.12	4.83	2.77	5.97	84.32
Mississippi	0.13	0.70	0.58	48.38	50.21
Missouri	0.33	1.06	1.68	15.68	81.25
Montana	10.50	1.09	1.88	0.50	86.04
Nebraska	1.63	1.44	6.33	6.35	84.24
Nevada	1.83	5.59	23.25	10.06	59.26
New Hampshire	0.19	1.16	1.50	0.99	96.16
New Jersey	0.20	6.30	14.07	16.32	63.11
New Mexico	10.57	0.99	50.04	2.12	36.28
New York	0.36	5.81	16.19	18.34	59.29
North Carolina	1.56	1.72	3.78	30.61	62.33
North Dakota	7.99	0.84	1.11	0.72	89.34
Ohio	0.13	1.13	1.56	15.48	81.71
Oklahoma	17.38	1.36	5.47	10.30	65.49
Oregon	2.19	3.79	8.96	2.50	82.56
Pennsylvania	0.14	1.94	4.29	14.56	79.07
Rhode Island	0.43	3.17	11.12	7.27	78.01
South Carolina	0.22	0.94	1.65	41.79	55.40
South Dakota	9.50	0.93	1.12	0.88	87.57
Tennessee	0.06[2]	2.03[2]	3.63[1]	20.79[1]	73.49[1]
Texas	0.28	2.58	38.38	14.40	44.37
Utah	1.58	2.73	8.24	0.84	86.60
Vermont	0.59	1.21	0.49	0.81	96.90
Vermont	0.24	4.00	4.30	26.12	65.35
Washington	2.51[2]	8.64[2]	10.41[1]	3.52[1]	74.92[1]
West Virginia	0.09	0.54	0.32	4.22	94.83
Wisconsin	1.47	3.03	4.12	9.69	81.69
Wyoming	3.06	0.81	6.80	0.80	88.54

(Source: CCD 2000-2001)

[1]NAEP 1998 non-response adjusted estimates, 4th grade reading
[2]Adjusted Census 2000 counts for all inhabitants after allowing for NAEP 1998 estimates
[3]Adjusted Census 2000 counts for all inhabitants.

Table A-3: Numbers of states in which the various race/ethnicity categories represent specified percentages of the state's student population[1]

Percentage of student population	Race or Ethnicity Category					
	American Indian	Asian or Pacific Islander	Hispanic	Black not Hispanic	White	Total
50+ to 100%	0	1	1	2	44	48
25+ to 50%	1	0	4	8	5	18
10+ to 25%	4	1	10	17	1	33
5+ to 10%	2	6	8	7	0	25
1+ to 5%	13	31	23	14	1	82
0+ to 1%	31	12	5	3	0	51

(Source: CCD 2000-2001, NAEP 1998, and Census 2000 as described in text)

[1]Percentages are averages of 4th grade and 8th grade estimates.

Table A-4: Number of students by state or district for fourth and eighth grades.

State or District	Total 4th Grade Students	Total 8th Grade Students
Alabama	59,735	56,922
Alaska	10,646	10,377
Arizona	72,295	65,526
Arkansas	35,724	34,873
California	489,043	441,877
Colorado	57,055	55,371
Connecticut	44,687	42,597
Delaware	8,848	9,075
District of Columbia	5,830	3,371
Florida	194,292	185,657
Georgia	116,678	109,124
Hawaii	15,291	13,424
Idaho	18,949	19,003
Illinois	160,495	149,045
Indiana	79,738	73,882
Iowa	36,448	36,458
Kansas	34,975	35,785
Kentucky	50,181	47,707
Louisiana	63,874	61,992
Maine	16,077	17,000
Maryland	69,279	64,647
Massachusetts	78,287	74,527
Michigan	133,612	128,453
Minnesota	63,334	66,254
Mississippi	40,177	36,588
Missouri	71,222	68,728
Montana	11,682	12,517
Nebraska	21,357	21,864
Nevada	28,616	25,327
New Hampshire	16,852	17,209
New Jersey	100,622	92,094
New Mexico	25,493	24,870
New York	217,997	203,429
North Carolina	105,105	99,295
North Dakota	7,982	8,651
Ohio	143,116	141,777
Oklahoma	47,064	46,276
Oregon	42,661	41,497
Pennsylvania	142,366	143,638
Rhode Island	12,490	11,750
South Carolina	54,468	53,259
South Dakota	9,583	10,303
Tennessee	73,373	66,188
Texas	313,731	304,419
Utah	35,910	34,579
Vermont	7,736	8,005
Virginia	92,073	87,440
Washington	78,418	77,059
West Virginia	21,995	21,902
Wisconsin	64,455	67,950
Wyoming	6,736	7,284

(Sources: CCD 2000-2001, NAEP 1998, and Census 2000 as described in text)

Table A-5: Estimated standardized test score standard deviations $\hat{\sigma}$ for four racial/ethnic groups in the 1996 NAEP 4th grade mathematics assessment.

Group	Design effect[1]	Sample Size[2]	Standard Error[2]	$\hat{\sigma}$
Asian/Pacific Islander	3.86	157	4.6	29.2
Black	6.32	782	2.4	26.7
Hispanic	4.04	730	2.2	29.6
White	4.74	3,442	1.1	29.6

(Source: 1996 NAEP fourth grade mathematics assessments)

[1]From NAEP 1996 Technical Report and Allen, Carlson, & Zelenak (1999)
[2]From NAEP Data Tool v2.0.

Table A-6: Mean standardized test scores for racial and ethnic groups and estimated gaps for disadvantaged groups by state.

State	Advantaged Group Mean Score			Disadvantaged Group Mean Score			Gap		
	Asian	White	Weighted Average	American Indian	Black	Hispanic	American Indian	Black	Hispanic
Alabama	234	229	229	215	205	201	14	24	28
Arizona	****	231	231.1	****	208	204	****	23.1	27.1
Arkansas	227	225	225	196	198	205	29	27	20
California	246	229	228.5	213	193	201	15.5	35.5	27.5
Colorado	****	243	243.1	****	209	214	****	34.1	29.1
Georgia	216	232	232	****	206	208	****	26	24
Hawaii	****	225	225	****	204	205	****	21	20
Idaho	****	230	230	****	****	213	****	****	17
Illinois	****	237	237	****	205	213	****	32	24
Indiana	****	238	238	****	216	220	****	22	18
Iowa	****	235	235	****	****	216	****	****	19
Kansas	****	238	238	****	207	215	****	31	23
Kentucky	****	225	225	****	200	207	****	25	18
Louisiana	****	230	230	****	204	210	****	26	20
Maine	240	231	231	****	****	****	****	****	****
Maryland	239	237	237.2	****	204	210	****	33.2	27.2
Massachusetts	****	241	240.9	****	212	210	****	28.9	30.9
Michigan	235	239	239	****	201	210	****	38	29
Minnesota	****	240	239.7	****	211	214	****	28.7	25.7
Mississippi	****	224	224	****	199	201	****	25	23
Missouri	****	235	235	****	202	213	****	33	22
Montana	****	234	234	212	****	219	22	****	15
Nebraska	224	232	232	****	199	206	****	33	26
Nevada	****	228	227.6	212	206	210	15.6	21.6	17.6
New Mexico	247	227	227	197	****	208	30	****	19
New York	****	238	238.9	****	211	211	****	27.9	27.9
North Carolina	****	241	241	229	218	218	12	23	23
North Dakota	****	233	233	208	****	214	25	****	19
Ohio	****	236	236	****	208	218	****	28	18
Oklahoma	240	230	230	222	206	215	8	24	15
Oregon	221	230	230.5	****	****	206	****	****	24.5
Rhode Island	****	234	233.5	****	201	198	****	32.5	35.5
South Carolina	****	233	233	****	204	209	****	29	24
Tennessee	247	227	227	****	199	207	****	28	20
Texas	222	243	243.2	****	220	224	****	23.2	19.2
Utah	****	232	231.7	****	****	206	****	****	25.7
Vermont	243	233	233	****	****	****	****	****	****
Virginia	****	240	240.2	****	212	219	****	28.2	21.2
West Virginia	****	227	227	****	207	213	****	20	14
Wyoming	234	232	232	224	****	215	8	****	17

(Source: 2000 NAEP fourth grade mathematics assessment)

**** Indicates either sample size too low for reliable estimate, or jurisdiction did not participate, or special analyses raised concerns about accuracy (see Braswell et al. 2001 for further details).

Table A-7: Effective sample sizes for two difference-in-gap statistics based on a performance target of zero performance gap after 12 years.

State	$\Delta\bar{y}_t - \Delta\bar{x}_t$			$-\Delta\bar{x}_t$		
	American Indian	Black	Hispanic	American Indian	Black	Hispanic
Alabama	140,598	1,328	16,772	139,118	825	16,402
Arizona	****	7,690	1,129	****	7,072	680
Arkansas	41,263	1,364	12,021	40,972	1,028	11,409
California	92,856	1,935	1,028	91,239	1,626	513
Colorado	****	2,764	1,184	****	2,545	884
Georgia	****	1,138	6,534	****	670	5,985
Hawaii	****	18,254	12,693	****	16,269	10,504
Idaho	****	****	8,526	****	****	7,851
Illinois	****	1,057	2,363	****	776	1,863
Indiana	****	3,384	15,868	****	2,949	15,218
Iowa	****	****	12,312	****	****	11,773
Kansas	****	2,103	3,840	****	1,870	3,418
Kentucky	****	2,886	56,870	****	2,564	56,250
Louisiana	****	1,070	31,433	****	507	30,480
Maine	****	****	****	****	****	****
Maryland	****	697	5,306	****	411	4,880
Massachusetts	****	2,557	1,867	****	2,290	1,634
Michigan	****	734	6,005	****	568	5,721
Minnesota	****	3,120	7,340	****	2,872	7,031
Mississippi	****	1,142	41,171	****	551	40,473
Missouri	****	1,070	19,135	****	864	18,674
Montana	3,570	****	44,523	3,145	****	43,609
Nebraska	****	2,498	3,415	****	2,302	3,100
Nevada	43,368	4,196	3,025	42,179	3,575	2,090
New Mexico	2,333	****	2,343	1,775	****	952
New York	****	1,511	1,569	****	1,137	1,194
North Carolina	83,610	1,576	7,706	81,621	1,035	7,165
North Dakota	3,642	****	33,328	3,324	****	32,777
Ohio	****	1,545	32,821	****	1,265	32,143
Oklahoma	19,877	3,208	13,706	15,629	2,736	12,498
Oregon	****	****	2,882	****	****	2,526
Rhode Island	****	2,280	1,123	****	2,062	940
South Carolina	****	871	16,535	****	490	15,978
Tennessee	****	1,275	10,501	****	963	9,889
Texas	****	2,963	2,249	****	2,217	1,161
Utah	****	****	3,040	****	****	2,735
Vermont	****	****	****	****	****	****
Virginia	****	1,126	8,413	****	793	7,823
West Virginia	****	10,434	225,970	****	9,970	225,023
Wyoming	82,157	****	8,698	79,007	****	8,000

(Source: 2000 NAEP fourth grade mathematics assessment)

**** Indicates either sample size too low for reliable estimate, or jurisdiction did not participate, or special analyses raised concerns about accuracy (see Braswell et al. 2001 for further details).

Table A-8: Percentage at or above basic achievement level for racial and ethnic groups and estimated gaps for disadvantaged groups by state.

State	Advantaged Group Percentage			Disadvantaged Group Percentage			Gap		
	Asian	White	Weighted Average	American Indian	Black	Hispanic	American Indian	Black	Hispanic
Alabama	****	74	74	****	36	37	****	38	37
Arizona	77	75	75.1	24	43	40	51.1	32.1	35.1
Arkansas	****	68	68	49	28	39	19	40	29
California	71	71	71	****	25	36	****	46	35
Colorado	89	88	88	****	41	53	****	47	35
Georgia	****	75	75	****	38	43	****	37	32
Hawaii	56	68	68	****	37	40	****	31	28
Idaho	****	76	76	****	****	49	****	****	27
Illinois	****	82	82	****	37	51	****	45	31
Indiana	****	83	83	****	51	61	****	32	22
Iowa	****	81	81	****	****	51	****	****	30
Kansas	****	83	83	****	42	54	****	41	29
Kentucky	****	66	66	****	29	43	****	37	23
Louisiana	****	76	76	****	35	45	****	41	31
Maine	****	75	75	****	****	****	****	****	****
Maryland	82	81	81.1	****	36	47	****	45.1	34.1
Massachusetts	81	87	86.7	****	47	47	****	39.7	39.7
Michigan	****	83	83	****	32	49	****	51	34
Minnesota	77	84	83.6	****	46	54	****	37.6	29.6
Mississippi	****	66	66	****	27	30	****	39	36
Missouri	****	82	82	****	34	54	****	48	28
Montana	****	78	78	49	****	57	29	****	21
Nebraska	****	75	75	****	21	45	****	54	30
Nevada	64	72	71.3	51	40	46	20.3	31.3	25.3
New Mexico	****	70	70	30	****	42	40	****	28
New York	90	85	85.5	****	44	46	****	41.5	39.5
North Carolina	****	86	86	77	58	56	9	28	30
North Dakota	****	79	79	42	****	53	37	****	26
Ohio	****	82	82	****	37	60	****	45	22
Oklahoma	****	77	77	65	39	54	12	38	23
Oregon	77	73	73.2	****	****	40	****	****	33.2
Rhode Island	55	79	78	****	37	33	****	41	45
South Carolina	****	77	77	****	37	46	****	40	31
Tennessee	****	70	70	****	31	46	****	39	24
Texas	90	89	89.1	****	60	68	****	29.1	21.1
Utah	61	76	75.5	****	****	42	****	****	33.5
Vermont	****	75	75	****	****	****	****	****	****
Virginia	88	86	86.1	****	46	59	****	40.1	27.1
West Virginia	****	70	70	****	39	55	****	31	15
Wyoming	****	77	77	69	****	56	8	****	21

(Source: 2000 NAEP fourth grade mathematics assessment)

**** Indicates either sample size too low for reliable estimate, or jurisdiction did not participate, or special analyses raised concerns about accuracy (see Braswell et al. 2001 for further details).

Table A-9: Effective sample sizes for two difference-in-gap statistics based on percentage at or above basic achievement level and a performance target of zero performance gap after 12 years.

State	$\Delta\hat{q}_t - \Delta\hat{p}_t$			$-\Delta\hat{p}_t$		
	American Indian	Black	Hispanic	American Indian	Black	Hispanic
Alabama	****	1,472	26,680	****	914	26,091
Arizona	3,229	11,090	1,873	2,877	10,199	1,127
Arkansas	267,017	1,726	15,881	265,133	1,301	15,073
California	****	3,207	1,767	****	2,695	882
Colorado	****	4,039	2,271	****	3,719	1,694
Georgia	****	1,561	10,209	****	919	9,351
Hawaii	****	23,269	17,989	****	20,738	14,887
Illinois	****	****	9,389	****	****	8,645
Indiana	****	1,485	3,935	****	1,090	3,101
Iowa	****	4,442	29,505	****	3,872	28,298
Indiana	****	****	13,718	****	****	13,118
Kansas	****	3,339	6,709	****	2,970	5,972
Kentucky	****	3,659	96,754	****	3,252	95,700
Louisiana	****	1,195	36,342	****	566	35,242
Maine	****	****	****	****	****	****
Maryland	****	1,052	9,405	****	621	8,650
Massachusetts	****	3,767	3,146	****	3,374	2,752
Michigan	****	1,132	12,135	****	876	11,561
Minnesota	****	5,055	15,390	****	4,653	14,742
Mississippi	****	1,304	46,681	****	629	45,889
Missouri	****	1,404	32,813	****	1,135	32,022
Montana	5,707	****	63,099	5,028	****	61,804
Nebraska	****	2,591	7,126	****	2,388	6,468
Nevada	71,626	5,577	4,090	69,662	4,752	2,826
New Mexico	3,645	****	2,997	2,773	****	1,218
New York	****	1,895	2,171	****	1,425	1,653
North Carolina	412,885	2,954	12,581	403,065	1,940	11,698
North Dakota	4,618	****	49,438	4,215	****	48,621
Ohio	****	1,661	61,030	****	1,360	59,770
Oklahoma	24,539	3,554	16,193	19,295	3,031	14,766
Oregon	****	****	4,346	****	****	3,810
Rhode Island	****	3,968	1,936	****	3,588	1,621
South Carolina	****	1,272	27,529	****	715	26,601
Tennessee	****	1,826	20,256	****	1,379	19,076
Texas	****	5,260	5,209	****	3,936	2,689
Utah	****	****	4,956	****	****	4,458
Vermont	****	****	****	****	****	****
Virginia	****	1,543	14,251	****	1,086	13,251
West Virginia	****	12,063	546,791	****	11,527	544,500
Wyoming	228,213	****	15,832	219,463	****	14,562

(Source: 2000 NAEP fourth grade mathematics assessment)

**** Indicates either sample size too low for reliable estimate, or jurisdiction did not participate, or special analyses raised concerns about accuracy (see Braswell et al. 2001 for further details).

Table A-10: Percentage at or above proficient achievement level for state racial and ethnic groups of NAEP year 2000 4th grade mathematics assessment and estimated gaps among advantaged and disadvantaged groups.

State	Advantaged Group Percentage			Disadvantaged Group Percentage			Gap		
	Asian	White	Weighted Average	American Indian	Black	Hispanic	American Indian	Black	Hispanic
Alabama	****	23	23	****	4	5	****	19	18
Arizona	28	26	26.1	4	5	6	22.1	21.1	20.1
Arkansas	****	18	18	9	2	6	9	16	12
California	25	25	25	****	2	5	****	23	20
Colorado	45	41	41.2	****	6	9	****	35.2	32.2
Georgia	****	29	29	****	6	8	****	23	21
Hawaii	14	19	19	****	3	7	****	16	12
Idaho	****	24	24	****	****	8	****	****	16
Illinois	****	32	32	****	5	8	****	27	24
Indiana	****	34	34	****	14	16	****	20	18
Iowa	****	30	30	****	****	13	****	****	17
Kansas	****	36	36	****	7	11	****	29	25
Kentucky	****	20	20	****	2	9	****	18	11
Louisiana	****	23	23	****	4	7	****	19	16
Maine	****	25	25	****	****	****	****	****	****
Maryland	40	36	36.3	****	5	10	****	31.3	26.3
Massachusetts	41	39	39.1	****	7	10	****	32.1	29.1
Michigan	****	37	37	****	4	15	****	33	22
Minnesota	32	39	38.6	****	11	13	****	27.6	25.6
Mississippi	****	16	16	****	2	6	****	14	10
Missouri	****	28	28	****	4	11	****	24	17
Montana	****	28	28	8	****	12	20	****	16
Nebraska	****	29	29	****	6	7	****	23	22
Nevada	21	23	22.8	7	5	8	15.8	17.8	14.8
New Mexico	****	22	22	5	****	6	17	****	16
New York	47	34	35.3	****	5	7	****	30.3	28.3
North Carolina	****	38	38	21	9	13	17	29	25
North Dakota	****	27	27	7	****	12	20	****	15
Ohio	****	32	32	****	3	12	****	29	20
Oklahoma	****	20	20	12	3	9	8	17	11
Oregon	36	26	26.5	****	****	6	****	****	20.5
Rhode Island	21	30	29.6	****	4	5	****	25.6	24.6
South Carolina	****	28	28	****	4	12	****	24	16
Tennessee	****	23	23	****	4	9	****	19	14
Texas	48	41	41.4	****	12	14	****	29.4	27.4
Utah	16	28	27.6	****	****	8	****	****	19.6
Vermont	****	31	31	****	****	****	****	****	****
Virginia	45	35	35.6	****	6	11	****	29.6	24.6
West Virginia	****	19	19	****	6	13	****	13	6
Wyoming	****	28	28	18	****	12	10	****	16

(Source: 2000 NAEP fourth grade mathematics assessment)

**** Indicates either sample size too low for reliable estimate, or jurisdiction did not participate, or special analyses raised concerns about accuracy (see Braswell et al. 2001 for further details).

Table A-11: Effective sample sizes for two difference-in-gap statistics based on percentage at or above proficient achievement level and a performance target of zero performance gap after 12 years.

State	$\Delta\hat{q}_t - \Delta\hat{p}_t$			$-\Delta\hat{p}_t$		
	American Indian	Black	Hispanic	American Indian	Black	Hispanic
Alabama	****	2,750	54,290	****	1,707	53,092
Arizona	8,833	13,485	3,079	7,872	12,401	1,854
Arkansas	555,866	3,883	39,178	551,945	2,926	37,183
California	****	5,992	2,759	****	5,034	1,377
Colorado	****	5,211	2,027	****	4,798	1,512
Georgia	****	2,333	14,297	****	1,374	13,095
Hawaii	****	34,206	44,307	****	30,485	36,666
Idaho	****	****	14,373	****	****	13,235
Illinois	****	2,488	4,201	****	1,825	3,311
Indiana	****	8,297	33,057	****	7,231	31,705
Iowa	****	****	28,839	****	****	27,578
Kansas	****	4,505	6,492	****	4,007	5,779
Kentucky	****	6,054	209,765	****	5,380	207,479
Louisiana	****	2,599	69,577	****	1,231	67,469
Maine	****	****	****	****	****	****
Maryland	****	1,430	11,239	****	843	10,337
Massachusetts	****	4,078	4,326	****	3,652	3,785
Michigan	****	1,762	22,305	****	1,364	21,250
Minnesota	****	6,999	15,755	****	6,443	15,091
Mississippi	****	3,314	236,910	****	1,599	232,893
Missouri	****	3,019	55,893	****	2,440	54,545
Montana	7,084	****	69,566	6,241	****	68,139
Nebraska	****	8,248	7,823	****	7,602	7,101
Nevada	59,770	8,234	6,207	58,132	7,015	4,289
New Mexico	9,425	****	4,419	7,171	****	1,796
New York	****	2,291	2,825	****	1,723	2,150
North Carolina	96,270	1,981	13,767	93,980	1,301	12,800
North Dakota	8,921	****	93,264	8,141	****	91,722
Ohio	****	2,310	50,688	****	1,892	49,642
Oklahoma	29,682	7,229	35,107	23,339	6,165	32,012
Oregon	****	****	6,220	****	****	5,453
Rhode Island	****	5,687	3,704	****	5,141	3,100
South Carolina	****	1,899	66,138	****	1,067	63,909
Tennessee	****	3,593	32,002	****	2,714	30,138
Texas	****	4,022	2,465	****	3,010	1,272
Utah	****	****	8,472	****	****	7,621
Vermont	****	****	****	****	****	****
Virginia	****	1,869	12,388	****	1,316	11,519
West Virginia	****	30,010	1,837,216	****	28,676	1,829,520
Wyoming	103,467	****	17,455	99,499	****	16,055

(Source: 2000 NAEP fourth grade mathematics assessment)

**** Indicates either sample size too low for reliable estimate, or jurisdiction did not participate, or special analyses raised concerns about accuracy (see Braswell et al. 2001 for further details).

Table A-12: Typical state margins of error required to detect AYP and gap legislative targets based on the NAEP 2000 4th grade mathematics assessment and typical state margins of error observed under current NAEP precision.

	Legislative targets		Observed NAEP precision[1]	
	AYP	Change in gaps	AYP	Change in gaps
Mean scale score	NA[2]	5.0	7.0	7.8
Percentage at or above NAEP basic achievement level	4.1 or 3.0	8.3	8.1	9.4

[1]Computed using disadvantaged $SE(\overline{x}_d) = 3.0$, advantaged $SE(\overline{x}_a) = 1.5$, disadvantaged $SE(\hat{p}_d) = .035$, and advantaged $SE(\hat{p}_a) = .02$. For advantaged standard errors, see figures 1 and 3 of Carlson(2003).

[2]Adequate yearly progress has legislative targets given with respect to percentage proficient (equated here with percent at or above the NAEP basic achievement level), and it is not clear how these targets would be interpreted using mean scale scores. There is an observed NAEP precision for this statistic, however.

Table A-13: American Indian effective and nominal sample sizes for adequate yearly progress in NAEP 4th grade mathematics mean scale scores. Margin of error set according to observed NAEP 2000 4th grade mathematics precision.

State	Percentage disadvantaged	Effective disadvantaged sample size	Nominal disadvantaged sample size	Nominal total sample size	Number of grade 4 students in state
Alaska	26.2	100	300	1,146	10,646
Oklahoma	17.6	100	300	1,705	47,064
South Dakota	12.7	100	300	2,363	9,583
Montana	11.6	100	300	2,587	11,682
New Mexico	11.0	100	300	2,728	25,493
North Dakota	8.5	100	300	3,530	7,982
Arizona	6.5	100	300	4,616	72,295
Wyoming	3.5	100	300	8,572	6,736
Washington	2.6	100	300	11,539	78,418
Minnesota	2.3	100	300	13,044	63,334
Oregon	2.2	100	300	13,637	42,661
Nebraska	1.7	100	300	17,648	21,357
Nevada	1.7	100	300	17,648	28,616
Utah	1.6	100	300	18,750	35,910
North Carolina	1.5	100	300	20,000	105,105
Wisconsin	1.5	100	300	20,000	64,455
Idaho	1.4	100	300	21,429	18,949
Kansas	1.4	100	300	21,429	34,975
Colorado	1.3	100	300	23,077	57,055
Michigan	1.0	100	300	30,000	133,612
California	0.8	100	300	37,500	489,043
Louisiana	0.7	100	300	42,858	63,874
Alabama	0.6	100	300	50,000	59,735
Arkansas	0.5	100	300	60,000	35,724
Iowa	0.5	100	300	60,000	36,448
Vermont	0.5	100	300	60,000	7,736
Hawaii	0.4	100	300	75,000	15,291
New York	0.4	100	300	75,000	217,997
Rhode Island	0.4	100	300	75,000	12,490
Florida	0.3	100	300	100,000	194,292
Maine	0.3	100	300	100,000	16,077
Maryland	0.3	100	300	100,000	69,279
Massachusetts	0.3	100	300	100,000	78,287
Missouri	0.3	100	300	100,000	71,222
Texas	0.3	100	300	100,000	313,731
Virginia	0.3	100	300	100,000	92,073
Connecticut	0.2	100	300	150,000	44,687
Delaware	0.2	100	300	150,000	8,848
Georgia	0.2	100	300	150,000	116,678
Illinois	0.2	100	300	150,000	160,495
Indiana	0.2	100	300	150,000	79,738
Kentucky	0.2	100	300	150,000	50,181
Mississippi	0.2	100	300	150,000	40,177
New Hampshire	0.2	100	300	150,000	16,852
New Jersey	0.2	100	300	150,000	100,622
South Carolina	0.2	100	300	150,000	54,468
Ohio	0.1	100	300	300,000	143,116
Pennsylvania	0.1	100	300	300,000	142,366
Tennessee	0.1	100	300	300,000	73,373
West Virginia	0.1	100	300	300,000	21,995
District of Columbia	0.0	100	300	>300,000	5,830

Table A-14: Black effective and nominal sample sizes for adequate yearly progress in NAEP 4th grade mathematics mean scale scores. Margin of error set according to observed NAEP 2000 4th grade mathematics precision.

State	Percentage disadvantaged	Effective disadvantaged sample size	Nominal disadvantaged sample size	Nominal total sample size	Number of grade 4 students in state
District of Columbia	85.0	100	300	353	5,830
Louisiana	51.5	100	300	583	63,874
Mississippi	51.2	100	300	586	40,177
South Carolina	42.9	100	300	700	54,468
Georgia	39.0	100	300	770	116,678
Maryland	38.9	100	300	772	69,279
Alabama	37.2	100	300	807	59,735
Delaware	32.6	100	300	921	8,848
North Carolina	32.2	100	300	932	105,105
Virginia	28.0	100	300	1,072	92,073
Florida	25.2	100	300	1,191	194,292
Arkansas	23.6	100	300	1,272	35,724
Tennessee	23.4	100	300	1,283	73,373
Illinois	22.2	100	300	1,352	160,495
Michigan	21.5	100	300	1,396	133,612
New York	20.0	100	300	1,500	217,997
Missouri	18.8	100	300	1,596	71,222
New Jersey	17.8	100	300	1,686	100,622
Ohio	17.8	100	300	1,686	143,116
Pennsylvania	16.3	100	300	1,841	142,366
Texas	14.7	100	300	2,041	313,731
Connecticut	14.0	100	300	2,143	44,687
Indiana	12.4	100	300	2,420	79,738
Oklahoma	11.2	100	300	2,679	47,064
Wisconsin	11.2	100	300	2,679	64,455
Kentucky	11.0	100	300	2,728	50,181
Nevada	10.5	100	300	2,858	28,616
Kansas	9.8	100	300	3,062	34,975
Massachusetts	9.2	100	300	3,261	78,287
California	8.6	100	300	3,489	489,043
Rhode Island	8.1	100	300	3,704	12,490
Minnesota	7.5	100	300	4,000	63,334
Nebraska	7.0	100	300	4,286	21,357
Colorado	6.0	100	300	5,000	57,055
Alaska	5.0	100	300	6,000	10,646
Arizona	4.7	100	300	6,383	72,295
Iowa	4.5	100	300	6,667	36,448
Washington	4.4	100	300	6,819	78,418
West Virginia	4.4	100	300	6,819	21,995
Oregon	3.0	100	300	10,000	42,661
Hawaii	2.5	100	300	12,000	15,291
New Mexico	2.5	100	300	12,000	25,493
South Dakota	1.5	100	300	20,000	9,583
Wyoming	1.4	100	300	21,429	6,736
Maine	1.3	100	300	23,077	16,077
North Dakota	1.3	100	300	23,077	7,982
Vermont	1.3	100	300	23,077	7,736
New Hampshire	1.1	100	300	27,273	16,852
Utah	1.1	100	300	27,273	35,910
Montana	0.8	100	300	37,500	11,682
Idaho	0.4	100	300	75,000	18,949

Table A-15: Hispanic effective and nominal sample sizes for adequate yearly progress in NAEP 4th grade mathematics mean scale scores. Margin of error set according to observed NAEP 2000 4th grade mathematics precision.

State	Percentage disadvantaged	Effective disadvantaged sample size	Nominal disadvantaged sample size	Nominal total sample size	Number of grade 4 students in state
New Mexico	51.3	100	300	585	25,493
California	45.4	100	300	661	489,043
Texas	41.1	100	300	730	313,731
Arizona	35.3	100	300	850	72,295
Nevada	27.1	100	300	1,108	28,616
Colorado	23.6	100	300	1,272	57,055
Florida	19.5	100	300	1,539	194,292
New York	19.0	100	300	1,579	217,997
Illinois	16.4	100	300	1,830	160,495
New Jersey	15.4	100	300	1,949	100,622
Rhode Island	14.9	100	300	2,014	12,490
Connecticut	13.6	100	300	2,206	44,687
Oregon	11.7	100	300	2,565	42,661
Massachusetts	11.3	100	300	2,655	78,287
Washington	10.7	100	300	2,804	78,418
Kansas	9.8	100	300	3,062	34,975
Utah	9.8	100	300	3,062	35,910
District of Columbia	9.1	100	300	3,297	5,830
Nebraska	8.4	100	300	3,572	21,357
Idaho	7.8	100	300	3,847	18,949
Wyoming	7.6	100	300	3,948	6,736
Delaware	6.6	100	300	4,546	8,848
Oklahoma	6.3	100	300	4,762	47,064
Georgia	5.1	100	300	5,883	116,678
Virginia	5.0	100	300	6,000	92,073
Maryland	4.9	100	300	6,123	69,279
Pennsylvania	4.8	100	300	6,250	142,366
Wisconsin	4.8	100	300	6,250	64,455
North Carolina	4.7	100	300	6,383	105,105
Tennessee	4.5	100	300	6,667	73,373
Hawaii	4.2	100	300	7,143	15,291
Iowa	4.2	100	300	7,143	36,448
Arkansas	3.9	100	300	7,693	35,724
Minnesota	3.8	100	300	7,895	63,334
Michigan	3.7	100	300	8,109	133,612
Indiana	3.6	100	300	8,334	79,738
Alaska	3.4	100	300	8,824	10,646
Missouri	2.0	100	300	15,000	71,222
New Hampshire	1.9	100	300	15,790	16,852
South Carolina	1.9	100	300	15,790	54,468
Montana	1.8	100	300	16,667	11,682
Ohio	1.7	100	300	17,648	143,116
South Dakota	1.6	100	300	18,750	9,583
North Dakota	1.5	100	300	20,000	7,982
Alabama	1.4	100	300	21,429	59,735
Louisiana	1.4	100	300	21,429	63,874
Kentucky	1.0	100	300	30,000	50,181
Mississippi	0.8	100	300	37,500	40,177
Maine	0.7	100	300	42,858	16,077
Vermont	0.6	100	300	50,000	7,736
West Virginia	0.4	100	300	75,000	21,995

Table A-16: American Indian effective and nominal sample sizes for adequate yearly progress in NAEP 4th grade mathematics percentage at or above the basic achievement level. Margin of error set according to observed NAEP 2000 4th grade mathematics precision.

State	Percentage disadvantaged	Effective disadvantaged sample size	Nominal disadvantaged sample size	Nominal total sample size	Number of grade 4 students in state
Alaska	26.2	154	462	1,764	10,646
Oklahoma	17.6	154	462	2,625	47,064
South Dakota	12.7	154	462	3,638	9,583
Montana	11.6	154	462	3,983	11,682
New Mexico	11.0	154	462	4,200	25,493
North Dakota	8.5	154	462	5,436	7,982
Arizona	6.5	154	462	7,108	72,295
Wyoming	3.5	154	462	13,200	6,736
Washington	2.6	154	462	17,770	78,418
Minnesota	2.3	154	462	20,087	63,334
Oregon	2.2	154	462	21,000	42,661
Nebraska	1.7	154	462	27,177	21,357
Nevada	1.7	154	462	27,177	28,616
Utah	1.6	154	462	28,875	35,910
North Carolina	1.5	154	462	30,800	105,105
Wisconsin	1.5	154	462	30,800	64,455
Idaho	1.4	154	462	33,000	18,949
Kansas	1.4	154	462	33,000	34,975
Colorado	1.3	154	462	35,539	57,055
Michigan	1.0	154	462	46,200	133,612
California	0.8	154	462	57,750	489,043
Louisiana	0.7	154	462	66,000	63,874
Alabama	0.6	154	462	77,000	59,735
Arkansas	0.5	154	462	92,400	35,724
Iowa	0.5	154	462	92,400	36,448
Vermont	0.5	154	462	92,400	7,736
Hawaii	0.4	154	462	115,500	15,291
New York	0.4	154	462	115,500	217,997
Rhode Island	0.4	154	462	115,500	12,490
Florida	0.3	154	462	154,000	194,292
Maine	0.3	154	462	154,000	16,077
Maryland	0.3	154	462	154,000	69,279
Massachusetts	0.3	154	462	154,000	78,287
Missouri	0.3	154	462	154,000	71,222
Texas	0.3	154	462	154,000	313,731
Virginia	0.3	154	462	154,000	92,073
Connecticut	0.2	154	462	231,000	44,687
Delaware	0.2	154	462	231,000	8,848
Georgia	0.2	154	462	231,000	116,678
Illinois	0.2	154	462	231,000	160,495
Indiana	0.2	154	462	231,000	79,738
Kentucky	0.2	154	462	231,000	50,181
Mississippi	0.2	154	462	231,000	40,177
New Hampshire	0.2	154	462	231,000	16,852
New Jersey	0.2	154	462	231,000	100,622
South Carolina	0.2	154	462	231,000	54,468
Ohio	0.1	154	462	462,000	143,116
Pennsylvania	0.1	154	462	462,000	142,366
Tennessee	0.1	154	462	462,000	73,373
West Virginia	0.1	154	462	462,000	21,995
District of Columbia	0.0	154	462	>462,000	5,830

Table A-17: Black effective and nominal sample sizes for adequate yearly progress percentage in NAEP 4th grade mathematics at or above the basic achievement level. Margin of error set according to observed NAEP 2000 4th grade mathematics precision.

State	Percentage disadvantaged	Effective disadvantaged sample size	Nominal disadvantaged sample size	Nominal total sample size	Number of grade 4 students in state
District of Columbia	85.0	154	462	544	5,830
Louisiana	51.5	154	462	898	63,874
Mississippi	51.2	154	462	903	40,177
South Carolina	42.9	154	462	1,077	54,468
Georgia	39.0	154	462	1,185	116,678
Maryland	38.9	154	462	1,188	69,279
Alabama	37.2	154	462	1,242	59,735
Delaware	32.6	154	462	1,418	8,848
North Carolina	32.2	154	462	1,435	105,105
Virginia	28.0	154	462	1,650	92,073
Florida	25.2	154	462	1,834	194,292
Arkansas	23.6	154	462	1,958	35,724
Tennessee	23.4	154	462	1,975	73,373
Illinois	22.2	154	462	2,082	160,495
Michigan	21.5	154	462	2,149	133,612
New York	20.0	154	462	2,310	217,997
Missouri	18.8	154	462	2,458	71,222
New Jersey	17.8	154	462	2,596	100,622
Ohio	17.8	154	462	2,596	143,116
Pennsylvania	16.3	154	462	2,835	142,366
Texas	14.7	154	462	3,143	313,731
Connecticut	14.0	154	462	3,300	44,687
Indiana	12.4	154	462	3,726	79,738
Oklahoma	11.2	154	462	4,125	47,064
Wisconsin	11.2	154	462	4,125	64,455
Kentucky	11.0	154	462	4,200	50,181
Nevada	10.5	154	462	4,400	28,616
Kansas	9.8	154	462	4,715	34,975
Massachusetts	9.2	154	462	5,022	78,287
California	8.6	154	462	5,373	489,043
Rhode Island	8.1	154	462	5,704	12,490
Minnesota	7.5	154	462	6,160	63,334
Nebraska	7.0	154	462	6,600	21,357
Colorado	6.0	154	462	7,700	57,055
Alaska	5.0	154	462	9,240	10,646
Arizona	4.7	154	462	9,830	72,295
Iowa	4.5	154	462	10,267	36,448
Washington	4.4	154	462	10,500	78,418
West Virginia	4.4	154	462	10,500	21,995
Oregon	3.0	154	462	15,400	42,661
Hawaii	2.5	154	462	18,480	15,291
New Mexico	2.5	154	462	18,480	25,493
South Dakota	1.5	154	462	30,800	9,583
Wyoming	1.4	154	462	33,000	6,736
Maine	1.3	154	462	35,539	16,077
North Dakota	1.3	154	462	35,539	7,982
Vermont	1.3	154	462	35,539	7,736
New Hampshire	1.1	154	462	42,000	16,852
Utah	1.1	154	462	42,000	35,910
Montana	0.8	154	462	57,750	11,682
Idaho	0.4	154	462	115,500	18,949

Table A-18: Hispanic effective and nominal sample sizes for adequate yearly progress in NAEP 4th grade mathematics percentage at or above the basic achievement level. Margin of error set according to observed NAEP 2000 4th grade mathematics precision.

State	Percentage disadvant.	Effective disadvant. sample size	Nominal disadvant. sample size	Nominal total sample size	Number of grade 4 students in state
New Mexico	51.3	154	462	901	25,493
California	45.4	154	462	1,018	489,043
Texas	41.1	154	462	1,125	313,731
Arizona	35.3	154	462	1,309	72,295
Nevada	27.1	154	462	1,705	28,616
Colorado	23.6	154	462	1,958	57,055
Florida	19.5	154	462	2,370	194,292
New York	19.0	154	462	2,432	217,997
Illinois	16.4	154	462	2,818	160,495
New Jersey	15.4	154	462	3,000	100,622
Rhode Island	14.9	154	462	3,101	12,490
Connecticut	13.6	154	462	3,398	44,687
Oregon	11.7	154	462	3,949	42,661
Massachusetts	11.3	154	462	4,089	78,287
Washington	10.7	154	462	4,318	78,418
Kansas	9.8	154	462	4,715	34,975
Utah	9.8	154	462	4,715	35,910
District of Columbia	9.1	154	462	5,077	5,830
Nebraska	8.4	154	462	5,500	21,357
Idaho	7.8	154	462	5,924	18,949
Wyoming	7.6	154	462	6,079	6,736
Delaware	6.6	154	462	7,000	8,848
Oklahoma	6.3	154	462	7,334	47,064
Georgia	5.1	154	462	9,059	116,678
Virginia	5.0	154	462	9,240	92,073
Maryland	4.9	154	462	9,429	69,279
Pennsylvania	4.8	154	462	9,625	142,366
Wisconsin	4.8	154	462	9,625	64,455
North Carolina	4.7	154	462	9,830	105,105
Tennessee	4.5	154	462	10,267	73,373
Hawaii	4.2	154	462	11,000	15,291
Iowa	4.2	154	462	11,000	36,448
Arkansas	3.9	154	462	11,847	35,724
Minnesota	3.8	154	462	12,158	63,334
Michigan	3.7	154	462	12,487	133,612
Indiana	3.6	154	462	12,834	79,738
Alaska	3.4	154	462	13,589	10,646
Missouri	2.0	154	462	23,100	71,222
New Hampshire	1.9	154	462	24,316	16,852
South Carolina	1.9	154	462	24,316	54,468
Montana	1.8	154	462	25,667	11,682
Ohio	1.7	154	462	27,177	143,116
South Dakota	1.6	154	462	28,875	9,583
North Dakota	1.5	154	462	30,800	7,982
Alabama	1.4	154	462	33,000	59,735
Louisiana	1.4	154	462	33,000	63,874
Kentucky	1.0	154	462	46,200	50,181
Mississippi	0.8	154	462	57,750	40,177
Maine	0.7	154	462	66,000	16,077
Vermont	0.6	154	462	77,000	7,736
West Virginia	0.4	154	462	115,500	21,995

Table A-19: American Indian effective and nominal sample sizes for changes in gaps for NAEP 4th grade mathematics mean scale scores. Margin of error set according to observed NAEP 2000 4th grade mathematics precision.

State	Percentage disadvant.	Percentage advant.	Effective disadvant. sample size	Effective advantaged sample size	Effective total sample size	Nominal disadvant. sample size	Nominal advant. sample size	Nominal total sample size	Number of grade 4 students in state
Alaska	26.2	65.4	113	280	428	339	840	1,284	10,646
Oklahoma	17.6	64.9	102	375	577	306	1,125	1,731	47,064
South Dakota	12.7	84.3	93	612	726	279	1,836	2,178	9,583
Montana	11.6	85.8	91	673	784	273	2,019	2,352	11,682
New Mexico	11.0	35.1	106	335	952	318	1,005	2,856	25,493
North Dakota	8.5	88.7	88	916	1,033	264	2,748	3,099	7,982
Arizona	6.5	53.5	90	735	1,375	270	2,205	4,125	72,295
Wyoming	3.5	87.5	84	2,087	2,385	252	6,261	7,155	6,736
Washington	2.6	82.3	83	2,643	3,212	249	7,929	9,636	78,418
Minnesota	2.3	86.4	83	3,095	3,580	249	9,285	10,740	63,334
Oregon	2.2	83.0	83	3,054	3,678	249	9,162	11,034	42,661
Nebraska	1.7	82.8	82	3,905	4,716	246	11,715	14,148	21,357
Nevada	1.7	60.6	83	2,919	4,814	249	8,757	14,442	28,616
Utah	1.6	87.6	82	4,447	5,079	246	13,341	15,237	35,910
North Carolina	1.5	61.6	82	3,364	5,460	246	10,092	16,380	105,105
Wisconsin	1.5	82.4	82	4,368	5,299	246	13,104	15,897	64,455
Idaho	1.4	90.5	82	5,332	5,894	246	15,996	17,682	18,949
Kansas	1.4	79.0	82	4,593	5,812	246	13,779	17,436	34,975
Colorado	1.3	69.2	82	4,409	6,371	246	13,227	19,113	57,055
Michigan	1.0	73.8	82	6,053	8,198	246	18,159	24,594	133,612
California	0.8	45.2	82	4,595	10,165	246	13,785	30,495	489,043
Louisiana	0.7	46.3	82	5,499	11,869	246	16,497	35,607	63,874
Alabama	0.6	60.8	81	7,602	12,497	243	22,806	37,491	59,735
Arkansas	0.5	72.1	81	11,341	15,737	243	34,023	47,211	35,724
Iowa	0.5	90.8	81	13,943	15,356	243	41,829	46,068	36,448
Vermont	0.5	97.7	81	15,997	16,376	243	47,991	49,128	7,736
Hawaii	0.4	93.0	81	19,346	20,812	243	58,038	62,436	15,291
New York	0.4	60.6	81	12,323	20,335	243	36,969	61,005	217,997
Rhode Island	0.4	76.5	81	14,252	18,624	243	42,756	55,872	12,490
Florida	0.3	54.9	81	15,339	27,924	243	46,017	83,772	194,292
Maine	0.3	97.8	81	25,727	26,312	243	77,181	78,936	16,077
Maryland	0.3	55.9	81	15,402	27,541	243	46,206	82,623	69,279
Massachusetts	0.3	79.2	81	24,165	30,520	243	72,495	91,560	78,287
Missouri	0.3	79.0	81	20,937	26,504	243	62,811	79,512	71,222
Texas	0.3	43.9	81	11,814	26,939	243	35,442	80,817	313,731
Virginia	0.3	66.7	81	19,543	29,318	243	58,629	87,954	92,073
Connecticut	0.2	72.2	81	27,584	38,207	243	82,752	114,621	44,687
Delaware	0.2	60.5	81	22,608	37,342	243	67,824	112,026	8,848
Georgia	0.2	55.8	81	29,243	52,432	243	87,729	157,296	116,678
Illinois	0.2	61.2	81	30,476	49,821	243	91,428	149,463	160,495
Indiana	0.2	83.9	81	37,994	45,294	243	113,982	135,882	79,738
Kentucky	0.2	87.9	81	43,202	49,171	243	129,606	147,513	50,181
Mississippi	0.2	47.8	81	23,383	48,948	243	70,149	146,844	40,177
New Hampshire	0.2	96.8	81	43,598	45,027	243	130,794	135,081	16,852
New Jersey	0.2	66.6	81	30,709	46,098	243	92,127	138,294	100,622
South Carolina	0.2	55.0	81	17,953	32,666	243	53,859	97,998	54,468
Ohio	0.1	80.4	81	54,584	67,897	243	163,752	203,691	143,116
Pennsylvania	0.1	78.8	81	56,868	72,174	243	170,604	216,522	142,366
Tennessee	0.1	72.1	81	87,479	121,324	243	262,437	363,972	73,373
West Virginia	0.1	95.1	81	92,855	97,646	243	278,565	292,938	21,995
District of Columbia	0.0	5.9	81	13,923	236,654	243	41,769	709,962	5,830

Table A-20: Black effective and nominal sample sizes for changes in gaps for NAEP 4th grade mathematics mean scale scores. Margin of error set according to observed NAEP 2000 4th grade mathematics precision.

State	Percentage disadvant.	Percentage advantaged	Effective disadvant. sample size	Effective advantaged sample size	Effective total sample size	Nominal disadvant. sample size	Nominal advantaged sample size	Nominal total sample size	Number of grade 4 students in state
District of Columbia	85.0	5.9	1236	86	1,454	3,708	258	4,362	5,830
Louisiana	51.5	46.3	170	152	328	510	456	984	63,874
Mississippi	51.2	47.8	166	155	324	498	465	972	40,177
South Carolina	42.9	55.0	143	183	333	429	549	999	54,468
Georgia	39.0	55.8	136	195	349	408	585	1,047	116,678
Maryland	38.9	55.9	136	195	349	408	585	1,047	69,279
Alabama	37.2	60.8	129	211	347	387	633	1,041	59,735
Delaware	32.6	60.5	124	229	378	372	687	1,134	8,848
North Carolina	32.2	61.6	122	233	379	366	699	1,137	105,105
Virginia	28.0	66.7	114	271	406	342	813	1,218	92,073
Florida	25.2	54.9	117	255	463	351	765	1,389	194,292
Arkansas	23.6	72.1	107	325	451	321	975	1,353	35,724
Tennessee	23.4	72.1	106	327	454	318	981	1,362	73,373
Illinois	22.2	61.2	110	301	491	330	903	1,473	160,495
Michigan	21.5	73.8	104	355	481	312	1,065	1,443	133,612
New York	20.0	60.6	107	323	533	321	969	1,599	217,997
Missouri	18.8	79.0	99	418	528	297	1,254	1,584	71,222
New Jersey	17.8	66.6	102	380	571	306	1,140	1,713	100,622
Ohio	17.8	80.4	98	442	550	294	1,326	1,650	143,116
Pennsylvania	16.3	78.8	97	468	594	291	1,404	1,782	142,366
Texas	14.7	43.9	107	318	725	321	954	2,175	313,731
Connecticut	14.0	72.2	96	493	682	288	1,479	2,046	44,687
Indiana	12.4	83.9	92	623	743	276	1,869	2,229	79,738
Oklahoma	11.2	64.9	94	544	838	282	1,632	2,514	47,064
Wisconsin	11.2	82.4	91	670	813	273	2,010	2,439	64,455
Kentucky	11.0	87.9	91	719	818	273	2,157	2,454	50,181
Nevada	10.5	60.6	94	541	892	282	1,623	2,676	28,616
Kansas	9.8	79.0	90	724	917	270	2,172	2,751	34,975
Massachusetts	9.2	79.2	90	766	968	270	2,298	2,904	78,287
California	8.6	45.2	96	501	1,108	288	1,503	3,324	489,043
Rhode Island	8.1	76.5	89	835	1,091	267	2,505	3,273	12,490
Minnesota	7.5	86.4	87	1,008	1,166	261	3,024	3,498	63,334
Nebraska	7.0	82.8	87	1,022	1,234	261	3,066	3,702	21,357
Colorado	6.0	69.2	87	1,010	1,459	261	3,030	4,377	57,055
Alaska	5.0	65.4	87	1,129	1,727	261	3,387	5,181	10,646
Arizona	4.7	53.5	87	996	1,863	261	2,988	5,589	72,295
Iowa	4.5	90.8	84	1,685	1,856	252	5,055	5,568	36,448
Washington	4.4	82.3	85	1,563	1,900	255	4,689	5,700	78,418
West Virginia	4.4	95.1	84	1,800	1,893	252	5,400	5,679	21,995
Oregon	3.0	83.0	83	2,262	2,724	249	6,786	8,172	42,661
Hawaii	2.5	93.0	83	3,105	3,340	249	9,315	10,020	15,291
New Mexico	2.5	35.1	86	1,220	3,471	258	3,660	10,413	25,493
South Dakota	1.5	84.3	82	4,696	5,571	246	14,088	16,713	9,583
Wyoming	1.4	87.5	82	5,149	5,885	246	15,447	17,655	6,736
Maine	1.3	97.8	82	6,274	6,416	246	18,822	19,248	16,077
North Dakota	1.3	88.7	82	5,477	6,174	246	16,431	18,522	7,982
Vermont	1.3	97.7	82	6,312	6,462	246	18,936	19,386	7,736
New Hampshire	1.1	96.8	81	7,135	7,369	243	21,405	22,107	16,852
Utah	1.1	87.6	81	6,645	7,590	243	19,935	22,770	35,910
Montana	0.8	85.8	81	8,794	10,246	243	26,382	30,738	11,682
Idaho	0.4	90.5	81	18,447	20,394	243	55,341	61,182	18,949

Table A-21: Hispanic effective and nominal sample sizes for changes in gaps for NAEP 4th grade mathematics mean scale scores. Margin of error set according to observed NAEP 2000 4th grade mathematics precision.

State	Percentage disadvant.	Percentage advantaged	Effective disadvant. sample size	Effective advantaged sample size	Effective total sample size	Nominal disadvant. sample size	Nominal advantaged sample size	Nominal total sample size	Number of grade 4 students in state
New Mexico	51.3	35.1	197	135	384	591	405	1,152	25,493
California	45.4	45.2	161	160	354	483	480	1,062	489,043
Texas	41.1	43.9	155	166	378	465	498	1,134	313,731
Arizona	35.3	53.5	133	201	376	399	603	1,128	72,295
Nevada	27.1	60.6	116	259	427	348	777	1,281	28,616
Colorado	23.6	69.2	108	315	456	324	945	1,368	57,055
Florida	19.5	54.9	109	305	555	327	915	1,665	194,292
New York	19.0	60.6	106	335	553	318	1,005	1,659	217,997
Illinois	16.4	61.2	102	378	618	306	1,134	1,854	160,495
New Jersey	15.4	66.6	99	426	639	297	1,278	1,917	100,622
Rhode Island	14.9	76.5	96	491	641	288	1,473	1,923	12,490
Connecticut	13.6	72.2	96	506	700	288	1,518	2,100	44,687
Oregon	11.7	83.0	92	649	782	276	1,947	2,346	42,661
Massachusetts	11.3	79.2	92	640	808	276	1,920	2,424	78,287
Washington	10.7	82.3	91	695	845	273	2,085	2,535	78,418
Kansas	9.8	79.0	90	728	922	270	2,184	2,766	34,975
Utah	9.8	87.6	89	797	910	267	2,391	2,730	35,910
District of Columbia	9.1	5.9	204	132	2,242	612	396	6,726	5,830
Nebraska	8.4	82.8	89	867	1,047	267	2,601	3,141	21,357
Idaho	7.8	90.5	87	1,011	1,118	261	3,033	3,354	18,949
Wyoming	7.6	87.5	87	998	1,140	261	2,994	3,420	6,736
Delaware	6.6	60.5	89	811	1,339	267	2,433	4,017	8,848
Oklahoma	6.3	64.9	88	908	1,399	264	2,724	4,197	47,064
Georgia	5.1	55.8	88	952	1,707	264	2,856	5,121	116,678
Virginia	5.0	66.7	87	1,141	1,711	261	3,423	5,133	92,073
Maryland	4.9	55.9	87	997	1,783	261	2,991	5,349	69,279
Pennsylvania	4.8	78.8	85	1,387	1,761	255	4,161	5,283	142,366
Wisconsin	4.8	82.4	85	1,442	1,749	255	4,326	5,247	64,455
North Carolina	4.7	61.6	87	1,139	1,849	261	3,417	5,547	105,105
Tennessee	4.5	72.1	85	1,374	1,905	255	4,122	5,715	73,373
Hawaii	4.2	93.0	84	1,852	1,992	252	5,556	5,976	15,291
Iowa	4.2	90.8	84	1,830	2,016	252	5,490	6,048	36,448
Arkansas	3.9	72.1	85	1,572	2,181	255	4,716	6,543	35,724
Minnesota	3.8	86.4	84	1,900	2,198	252	5,700	6,594	63,334
Michigan	3.7	73.8	84	1,691	2,290	252	5,073	6,870	133,612
Indiana	3.6	83.9	84	1,956	2,332	252	5,868	6,996	79,738
Alaska	3.4	65.4	85	1,598	2,444	255	4,794	7,332	10,646
Missouri	2.0	79.0	82	3,318	4,200	246	9,954	12,600	71,222
New Hampshire	1.9	96.8	82	4,160	4,296	246	12,480	12,888	16,852
South Carolina	1.9	55.0	83	2,374	4,319	249	7,122	12,957	54,468
Montana	1.8	85.8	82	3,899	4,543	246	11,697	13,629	11,682
Ohio	1.7	80.4	82	3,877	4,823	246	11,631	14,469	143,116
South Dakota	1.6	84.3	82	4,359	5,172	246	13,077	15,516	9,583
North Dakota	1.5	88.7	82	4,840	5,456	246	14,520	16,368	7,982
Alabama	1.4	60.8	82	3,627	5,963	246	10,881	17,889	59,735
Louisiana	1.4	46.3	83	2,642	5,702	249	7,926	17,106	63,874
Kentucky	1.0	87.9	81	7,342	8,356	243	22,026	25,068	50,181
Mississippi	0.8	47.8	82	4,718	9,877	246	14,154	29,631	40,177
Maine	0.7	97.8	81	12,060	12,333	243	36,180	36,999	16,077
Vermont	0.6	97.7	81	13,815	14,142	243	41,445	42,426	7,736
West Virginia	0.4	95.1	81	19,099	20,085	243	57,297	60,255	21,995

Table A-22: American Indian effective and nominal sample sizes for changes in gaps for NAEP 4th grade mathematics percentage at or above the basic achievement level. Margin of error set according to observed NAEP 2000 4th grade mathematics precision.

State	Percentage disadvant.	Percentage advantaged	Effective disadvant. sample size	Effective advantaged sample size	Effective total sample size	Nominal disadvant. sample size	Nominal advantaged sample size	Nominal total sample size	Number of grade 4 students in state
Alaska	26.2	65.4	216	538	823	648	1,614	2,469	10,646
Oklahoma	17.6	64.9	196	720	1,110	588	2,160	3,330	47,064
South Dakota	12.7	84.3	177	1,177	1,396	531	3,531	4,188	9,583
Montana	11.6	85.8	175	1,294	1,507	525	3,882	4,521	11,682
New Mexico	11.0	35.1	203	644	1,831	609	1,932	5,493	25,493
North Dakota	8.5	88.7	169	1,761	1,985	507	5,283	5,955	7,982
Arizona	6.5	53.5	173	1,414	2,644	519	4,242	7,932	72,295
Wyoming	3.5	87.5	160	4,013	4,586	480	12,039	13,758	6,736
Washington	2.6	82.3	159	5,082	6,176	477	15,246	18,528	78,418
Minnesota	2.3	86.4	158	5,952	6,885	474	17,856	20,655	63,334
Oregon	2.2	83.0	158	5,873	7,072	474	17,619	21,216	42,661
Nebraska	1.7	82.8	158	7,510	9,069	474	22,530	27,207	21,357
Nevada	1.7	60.6	159	5,613	9,256	477	16,839	27,768	28,616
Utah	1.6	87.6	157	8,552	9,768	471	25,656	29,304	35,910
North Carolina	1.5	61.6	158	6,469	10,500	474	19,407	31,500	105,105
Wisconsin	1.5	82.4	157	8,400	10,190	471	25,200	30,570	64,455
Idaho	1.4	90.5	157	10,253	11,335	471	30,759	34,005	18,949
Kansas	1.4	79.0	157	8,832	11,176	471	26,496	33,528	34,975
Colorado	1.3	69.2	157	8,478	12,251	471	25,434	36,753	57,055
Michigan	1.0	73.8	156	11,640	15,765	468	34,920	47,295	133,612
California	0.8	45.2	157	8,836	19,548	471	26,508	58,644	489,043
Louisiana	0.7	46.3	157	10,574	22,825	471	31,722	68,475	63,874
Alabama	0.6	60.8	156	14,618	24,032	468	43,854	72,096	59,735
Arkansas	0.5	72.1	155	21,810	30,262	465	65,430	90,786	35,724
Iowa	0.5	90.8	155	26,813	29,530	465	80,439	88,590	36,448
Vermont	0.5	97.7	155	30,763	31,491	465	92,289	94,473	7,736
Hawaii	0.4	93.0	155	37,204	40,023	465	111,612	120,069	15,291
New York	0.4	60.6	155	23,697	39,104	465	71,091	117,312	217,997
Rhode Island	0.4	76.5	155	27,407	35,814	465	82,221	107,442	12,490
Florida	0.3	54.9	155	29,498	53,699	465	88,494	161,097	194,292
Maine	0.3	97.8	155	49,475	50,599	465	148,425	151,797	16,077
Maryland	0.3	55.9	155	29,619	52,963	465	88,857	158,889	69,279
Massachusetts	0.3	79.2	155	46,471	58,691	465	139,413	176,073	78,287
Missouri	0.3	79.0	155	40,264	50,970	465	120,792	152,910	71,222
Texas	0.3	43.9	155	22,718	51,805	465	68,154	155,415	313,731
Virginia	0.3	66.7	155	37,582	56,380	465	112,746	169,140	92,073
Connecticut	0.2	72.2	155	53,047	73,474	465	159,141	220,422	44,687
Delaware	0.2	60.5	155	43,477	71,811	465	130,431	215,433	8,848
Georgia	0.2	55.8	155	56,237	100,829	465	168,711	302,487	116,678
Illinois	0.2	61.2	155	58,607	95,809	465	175,821	287,427	160,495
Indiana	0.2	83.9	155	73,064	87,103	465	219,192	261,309	79,738
Kentucky	0.2	87.9	155	83,080	94,560	465	249,240	283,680	50,181
Mississippi	0.2	47.8	155	44,968	94,131	465	134,904	282,393	40,177
New Hampshire	0.2	96.8	155	83,841	86,590	465	251,523	259,770	16,852
New Jersey	0.2	66.6	155	59,055	88,649	465	177,165	265,947	100,622
South Carolina	0.2	55.0	155	34,524	62,820	465	103,572	188,460	54,468
Ohio	0.1	80.4	155	104,968	130,570	465	314,904	391,710	143,116
Pennsylvania	0.1	78.8	155	109,362	138,796	465	328,086	416,388	142,366
Tennessee	0.1	72.1	154	168,229	233,314	462	504,687	699,942	73,373
West Virginia	0.1	95.1	154	178,567	187,780	462	535,701	563,340	21,995
District of Columbia	0.0	5.9	155	26,774	455,104	465	80,322	1,365,312	5,830

Table A-23: Black effective and nominal sample sizes for changes in gaps for NAEP 4th grade mathematics percentage at or above the basic achievement level. Margin of error set according to observed NAEP 2000 4th grade mathematics precision.

State	Percentage disadvant.	Percentage advantaged	Effective disadvant. sample size	Effective advantaged sample size	Effective total sample size	Nominal disadvant. sample size	Nominal advantaged sample size	Nominal total sample size	Number of grade 4 students in state
District of Columbia	85.0	5.9	2,377	165	2,797	7,131	495	8,391	5,830
Louisiana	51.5	46.3	326	293	631	978	879	1,893	63,874
Mississippi	51.2	47.8	319	298	623	957	894	1,869	40,177
South Carolina	42.9	55.0	274	352	639	822	1,056	1,917	54,468
Georgia	39.0	55.8	262	375	671	786	1,125	2,013	116,678
Maryland	38.9	55.9	261	375	671	783	1,125	2,013	69,279
Alabama	37.2	60.8	248	406	668	744	1,218	2,004	59,735
Delaware	32.6	60.5	237	440	726	711	1,320	2,178	8,848
North Carolina	32.2	61.6	235	448	728	705	1,344	2,184	105,105
Virginia	28.0	66.7	219	520	780	657	1,560	2,340	92,073
Florida	25.2	54.9	225	489	890	675	1,467	2,670	194,292
Arkansas	23.6	72.1	205	625	867	615	1,875	2,601	35,724
Tennessee	23.4	72.1	204	629	872	612	1,887	2,616	73,373
Illinois	22.2	61.2	210	578	944	630	1,734	2,832	160,495
Michigan	21.5	73.8	199	682	924	597	2,046	2,772	133,612
New York	20.0	60.6	205	621	1,024	615	1,863	3,072	217,997
Missouri	18.8	79.0	191	803	1,016	573	2,409	3,048	71,222
New Jersey	17.8	66.6	195	731	1,097	585	2,193	3,291	100,622
Ohio	17.8	80.4	188	849	1,056	564	2,547	3,168	143,116
Pennsylvania	16.3	78.8	186	899	1,141	558	2,697	3,423	142,366
Texas	14.7	43.9	206	612	1,395	618	1,836	4,185	313,731
Connecticut	14.0	72.2	184	947	1,312	552	2,841	3,936	44,687
Indiana	12.4	83.9	177	1,198	1,428	531	3,594	4,284	79,738
Oklahoma	11.2	64.9	181	1,046	1,611	543	3,138	4,833	47,064
Wisconsin	11.2	82.4	175	1,288	1,563	525	3,864	4,689	64,455
Kentucky	11.0	87.9	174	1,382	1,573	522	4,146	4,719	50,181
Nevada	10.5	60.6	181	1,040	1,714	543	3,120	5,142	28,616
Kansas	9.8	79.0	173	1,393	1,762	519	4,179	5,286	34,975
Massachusetts	9.2	79.2	172	1,473	1,861	516	4,419	5,583	78,287
California	8.6	45.2	184	963	2,131	552	2,889	6,393	489,043
Rhode Island	8.1	76.5	171	1,605	2,097	513	4,815	6,291	12,490
Minnesota	7.5	86.4	168	1,937	2,241	504	5,811	6,723	63,334
Nebraska	7.0	82.8	167	1,965	2,372	501	5,895	7,116	21,357
Colorado	6.0	69.2	168	1,942	2,806	504	5,826	8,418	57,055
Alaska	5.0	65.4	166	2,171	3,320	498	6,513	9,960	10,646
Arizona	4.7	53.5	168	1,915	3,582	504	5,745	10,746	72,295
Iowa	4.5	90.8	162	3,240	3,568	486	9,720	10,704	36,448
Washington	4.4	82.3	163	3,005	3,652	489	9,015	10,956	78,418
West Virginia	4.4	95.1	162	3,461	3,640	486	10,383	10,920	21,995
Oregon	3.0	83.0	160	4,350	5,238	480	13,050	15,714	42,661
Hawaii	2.5	93.0	158	5,970	6,422	474	17,910	19,266	15,291
New Mexico	2.5	35.1	165	2,346	6,674	495	7,038	20,022	25,493
South Dakota	1.5	84.3	157	9,030	10,713	471	27,090	32,139	9,583
Wyoming	1.4	87.5	157	9,902	11,317	471	29,706	33,951	6,736
Maine	1.3	97.8	156	12,065	12,339	468	36,195	37,017	16,077
North Dakota	1.3	88.7	157	10,532	11,873	471	31,596	35,619	7,982
Vermont	1.3	97.7	156	12,139	12,426	468	36,417	37,278	7,736
New Hampshire	1.1	96.8	156	13,721	14,171	468	41,163	42,513	16,852
Utah	1.1	87.6	156	12,778	14,595	468	38,334	43,785	35,910
Montana	0.8	85.8	156	16,910	19,703	468	50,730	59,109	11,682
Idaho	0.4	90.5	155	35,474	39,218	465	106,422	117,654	18,949

Table A-24: Hispanic effective and nominal sample sizes for changes in gaps for NAEP 4th grade mathematics percentage at or above the basic achievement level. Margin of error set according to observed NAEP 2000 4th grade mathematics precision.

State	Percentage disadvant.	Percentage advantaged	Effective disadvant. sample size	Effective advantaged sample size	Effective total sample size	Nominal disadvant. sample size	Nominal advantaged sample size	Nominal total sample size	Number of grade 4 students in state
New Mexico	51.3	35.1	379	260	738	1,137	780	2,214	25,493
California	45.4	45.2	309	308	680	927	924	2,040	486,527
Texas	41.1	43.9	299	318	726	897	954	2,178	313,731
Arizona	35.3	53.5	256	387	724	768	1,161	2,172	72,295
Nevada	27.1	60.6	223	498	821	669	1,494	2,463	28,616
Colorado	23.6	69.2	207	606	876	621	1,818	2,628	57,056
Florida	19.5	54.9	209	587	1,068	627	1,761	3,204	194,320
New York	19.0	60.6	203	644	1,063	609	1,932	3,189	217,881
Illinois	16.4	61.2	196	727	1,187	588	2,181	3,561	160,495
New Jersey	15.4	66.6	190	818	1,228	570	2,454	3,684	100,622
Rhode Island	14.9	76.5	184	943	1,232	552	2,829	3,696	12,490
Connecticut	13.6	72.2	183	972	1,346	549	2,916	4,038	44,682
Oregon	11.7	83.0	176	1,248	1,503	528	3,744	4,509	42,810
Massachusetts	11.3	79.2	176	1,230	1,554	528	3,690	4,662	78,287
Washington	10.7	82.3	174	1,336	1,624	522	4,008	4,872	78,505
Kansas	9.8	79.0	173	1,400	1,772	519	4,200	5,316	35,036
Utah	9.8	87.6	172	1,532	1,749	516	4,596	5,247	35,910
District of Columbia	9.1	5.9	392	254	4,311	1,176	762	12,933	5,830
Nebraska	8.4	82.8	170	1,668	2,014	510	5,004	6,042	21,357
Idaho	7.8	90.5	168	1,944	2,149	504	5,832	6,447	13,501
Wyoming	7.6	87.5	168	1,918	2,192	504	5,754	6,576	6,736
Delaware	6.6	60.5	171	1,558	2,574	513	4,674	7,722	8,850
Oklahoma	6.3	64.9	169	1,746	2,690	507	5,238	8,070	47,064
Georgia	5.1	55.8	168	1,831	3,282	504	5,493	9,846	116,678
Virginia	5.0	66.7	166	2,194	3,290	498	6,582	9,870	92,073
Maryland	4.9	55.9	168	1,917	3,428	504	5,751	10,284	69,279
Pennsylvania	4.8	78.8	164	2,668	3,386	492	8,004	10,158	142,366
Wisconsin	4.8	82.4	163	2,772	3,363	489	8,316	10,089	64,455
North Carolina	4.7	61.6	166	2,191	3,555	498	6,573	10,665	105,105
Tennessee	4.5	72.1	164	2,642	3,663	492	7,926	10,989	73,412
Hawaii	4.2	93.0	161	3,560	3,830	483	10,680	11,490	15,291
Iowa	4.2	90.8	161	3,520	3,876	483	10,560	11,628	36,448
Arkansas	3.9	72.1	163	3,022	4,193	489	9,066	12,579	35,724
Minnesota	3.8	86.4	161	3,654	4,227	483	10,962	12,681	63,334
Michigan	3.7	73.8	162	3,252	4,404	486	9,756	13,212	134,163
Indiana	3.6	83.9	161	3,761	4,484	483	11,283	13,452	79,738
Alaska	3.4	65.4	162	3,072	4,699	486	9,216	14,097	10,646
Missouri	2.0	79.0	158	6,380	8,077	474	19,140	24,231	71,208
New Hampshire	1.9	96.8	157	7,999	8,261	471	23,997	24,783	16,852
South Carolina	1.9	55.0	160	4,565	8,306	480	13,695	24,918	54,463
Montana	1.8	85.8	158	7,498	8,736	474	22,494	26,208	11,682
Ohio	1.7	80.4	158	7,455	9,274	474	22,365	27,822	143,373
South Dakota	1.6	84.3	157	8,382	9,945	471	25,146	29,835	9,583
North Dakota	1.5	88.7	157	9,307	10,492	471	27,921	31,476	7,982
Alabama	1.4	60.8	158	6,975	11,467	474	20,925	34,401	59,692
Louisiana	1.4	46.3	159	5,080	10,965	477	15,240	32,895	63,884
Kentucky	1.0	87.9	156	14,118	16,069	468	42,354	48,207	49,837
Mississippi	0.8	47.8	157	9,074	18,993	471	27,222	56,979	40,177
Maine	0.7	97.8	155	23,191	23,718	465	69,573	71,154	16,121
Vermont	0.6	97.7	155	26,567	27,196	465	79,701	81,588	7,736
West Virginia	0.4	95.1	155	36,729	38,624	465	110,187	115,872	21,995