The Effects of Finite Sampling Corrections on State Assessment Sample Requirements

James R. Chromy
Research Triangle Institute

Commissioned by the NAEP Validity Studies (NVS) Panel
August 1998

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

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

For Information:

NAEP Validty Studies (NVS)
American Institutes for Research
1791 Arastradero Road
PO Box 1113
Palo Alto, CA 94302
Phone: 650/ 493-3550
Fax: 650/ 858-0958

Contents

List of Tables

Problem Statement

Current Practices

States participating in the National Assessment of Educational Progress State Assessment program (state NAEP) are required to sample at least 2,500 students selected from at least 100 schools per subject assessed. In this ideal situation, 25 students are assessed for a subject in each school selected for that subject. If more than one subject is being assessed for the given state and grade, say k subjects (with k usually 2), then as many as 25k students may be assessed at a single school for the target grade if the grade enrollment is sufficiently large.

Two problems have arisen in implementing the required design: (1) some states have too few schools—sometimes fewer than 100 within a target grade—and (2) some states have small schools, thereby requiring many more than 100 schools to obtain a sample of 2,500 students per subject. Specific policies have been developed to exempt states from the stringent sample design requirement above for each case (*e.g.*, Forsyth *et al.* 1996, Freund and Carlson 1997, Rust 1997).

The "partial sample option" addresses the first problem and allows states to negotiate a smaller sample, but not less than 1,250 students per subject assessed. The plan generally requires at least one session per subject per eligible and cooperating school and may require two, three, or four sessions per subject per grade in some schools. Per policy implemented in 1994, there must be 100 schools in the sample or the total number of schools eligible if less than 100.

A "sparse state sample option" has been proposed to deal with the second problem for the 1998 state NAEP. If the required number of schools exceeds 120 when applying the required design, a sampling plan may be negotiated that meets the following: (1) at least 115 schools per grade, (2) at least 80 schools per subject within each grade, and (3) no school selected with less than half the probability required in the initial requirements.

Both options yield a minimum sample size of 1,250 students per subject per grade and in practice would yield sample sizes closer to 2,500.

Theoretical Basis for Using or Ignoring the Finite Population Correction Factor

The sample size requirements for states participating in the state assessment program have been developed to allow adequate precision for estimates of various reporting domains commonly used in state assessment reports. Domains include the total population and also smaller groups defined by gender, education of parents, race or ethnicity, and other factors.

Requiring adequately high precision is equivalent to requiring that the variance of estimates be bounded above by some maximum value. Often relative standard error is used as a standard, but this can be converted to a variance requirement for each specific estimate.

In discussing simple random sampling, Cochran (1973, p.25) suggests that in practice, the finite population correction factor can be ignored (i.e., treated as a multiplicative factor of one) if the sampling fraction does not exceed five percent. A more general reason for ignoring the finite population correction factor, however, is stated in his discussion of comparisons between domain means (p. 39). For this type of analytic purpose, one wishes to test whether two (domain) means could have been drawn from the same infinite population. Since such a test relates to an infinite population, Cochran states it is not appropriate to apply a finite population correction to the variance formulation. Cochran does not re-address these issues with regard to stratified or cluster sampling, both of which apply to the state assessment samples. The usual variance formulation for stratified sampling is equivalent to applying finite population correction factor of zero to the variance component associated with between-stratum differences. When all schools are sampled, schools are treated as strata rather than as primary sampling units. When schools are treated as strata, there is no contribution of between-school differences to the variance of estimates. When a high fraction of schools is sampled, it seems natural to take an intermediate position: namely, to apply a finite population correction factor between zero and one to the between-school component of variance.

Approach

In this paper, we explore the application of finite population correction factors to the between-school component of variance and examine how this might effect sample size requirements in the types of states that currently require exemptions from the minimum sample requirements for the state NAEP. We also explore how we might preserve the infinite population assumptions for hypothesis testing relating to comparisons between domain means. For this preliminary exploration, we develop hypothetical school and student population structures and hypothetical variance component distributions. For each variance component distribution, we determine the effective sample size resulting from the minimum state NAEP sample size requirements when the infinite population assumptions are a good approximation to reality (i.e., when the finite population correction factor can be ignored at all stages of sampling). We then examine how this effective sample size can be maintained with an alternate sample design which recognizes the population structure and applies the finite population correction factor at the school stage of sampling.

Modeling Assumptions

The Ideal Population Structure

For variance modeling purposes, we assume that the ideal population structure has the following components:

- Fifty geographic strata each with a large number of schools
- Large and equal numbers of eligible schools in each stratum
- A large number of eligible students in each school

From such a population, the sample design for a grade assessment would consist of two schools per stratum and 25 students per subject assessed from each sample school, or a total of 100 schools and 2,500 students per subject.

Variance Component Distributions

We will investigate three variance component distributions corresponding to three levels of intra-school correlation coefficients: 0.10, 0.15, and 0.20. Preliminary results from an ongoing study of National Assessment designs support estimates near 0.15.[1] The

1 Estimates of the variance component distributions for strata, schools, and students for states in the 1992 Trial State Assessment of mathematics varied widely from state to state. Their simple means were 0.082, 0.082, and 0.837, respectively, for strata, schools, and students. The mean intra-school correlation coefficient was 0.163. Median values for the variance components were 0.073, 0.077, and 0.849, respectively. The median intra-school correlation coefficient was 0.15. Two states were omitted from the analyses due to problems with the stratification variable on the file.

preliminary results also support splitting the remaining variance approximately equally between strata and schools. Table 1 shows how these assumptions partition the total variance. Variance component distribution B is the best match to average preliminary data. Distributions A and C are included in recognition of the variability among states as well as the possibility that the variance component distribution for other measures (other than mathematics composite scores) may differ from the empirical results. The table also shows the effective sample size that would be achieved if the population and sample were as described above. Variance component distribution A would correspond to a design effect of 2.15 for the estimates relating to the aggregate population. Variance component distributions B and C lead to design effects of 2.73 and 3.30.[2] Smaller design effects would be expected for some subgroups of the total population, particularly subgroups which tend to partition all or most sessions. Variance component distributions for subgroup estimates are discussed later in this paper.

The variance of an estimated mean for the ideal setting can be modeled as:

$$V_1 = \frac{\sigma_2^2}{100} + \frac{\sigma_3^2}{2500} .$$

The first component of variance, σ_1^2, drops out since all strata are included with certainty, and implicitly, a finite population correction factor of zero is applied. The remaining two components of variance correspond to schools and students, respectively. The variance under simple random sampling (assuming a finite population correction factor of 1) would be simply:

$$V_{SRS} = \frac{\sigma^2}{2500} .$$

The design effect can be computed as a ratio of the design-based variance to the variance under simple random sampling, or:

$$Deff_1 = \frac{V_1}{V_{SRS}} .$$

2 The design effect can be computed as the ratio of the actual sample size to the effective sample size; *e.g.*, for model A the design effect is 2,500/1,163 or 2.15.

With a little manipulation, this can be expressed for the ideal design as:

$$Deff_1 = 2500\left[\frac{\sigma_2^2/\sigma^2}{100} + \frac{\sigma_3^2/\sigma^2}{2500}\right].$$

The inverse of the term in brackets can also be interpreted as the effective sample size which is shown in the final column of Table 1.

Table 1. Variance Component Distributions

Distribution	Intra-School Correlation	Proportion of Population Variance Associated with:			Design Effect	Effective Sample Size*
		Strata σ_1^2/σ^2	Schools σ_2^2/σ^2	Students σ_3^2/σ^2		
A	0.10	0.050	0.050	0.900	2.15	1,163
B	0.15	0.075	0.075	0.850	2.73	917
C	0.20	0.100	0.100	0.800	3.30	758

* The effective sample sizes shown are for a sample of 100 schools and 2,500 students per subject selected from an ideal hypothetically infinite population.

Finite Population Variance Models

When we consider incorporating the finite population correction factor at the school and/or student level, there are at least two possible options:

- Variance model 1: Apply the finite population correction factor at the school level and ignore it (set to 1) at the student level

- Variance model 2: Apply the finite population correction factor at both levels and add back a variance component for the finite population

Both of these models are presented in oversimplified form—particularly with regard to all schools having about the same enrollment, or schools within major strata having about the same enrollment. Even though this is unrealistic, it should still help to develop an understanding of the options being considered without developing empirically-based formulae for real state populations.

The finite variance model 1 can be written as:

$$V_{f1} = \frac{\sigma_2^2}{n_2}\left[\frac{N_2 - n_2}{N_2}\right] + \frac{\sigma_3^2}{n_2 n_3}$$

where N_2 refers to the number of schools in the population, n_2 refers to the number of schools in the sample, and n_3 refers to the number of students selected per school. Note that if all schools are selected in the sample, this effectively treats schools as strata with all the variance contributions coming from sampling within schools. Also, no finite population correction factor is applied within schools, so that hypothesis tests about differences between domain means can reasonably be based on infinite population assumptions.

Another way to recognize infinite population assumptions would be based on model 2 which adds back a component of variance related to treating the finite population as a simple random sample from an infinite population. This variance can be modeled as:

$$V_{f2} = \frac{\sigma_2^2}{n_2}\left[\frac{N_2 - n_2}{N_2}\right] + \frac{\sigma_3^2}{n_2 n_3}\left[\frac{N_3 - n_3}{N_3}\right] + \frac{\sigma^2}{N_2 N_3}$$

where σ^2 is the overall population variance, N_3 refers to the average number of eligible students per school.

The choice of appropriate variance models depends on how one views the finite real world arising from or being generated by some infinite process or super-population model. Model 1 treats schools and their served communities as being fixed and finite; it views the students as arising from individual super-populations. The characteristics of the communities served and educational environments of the schools are unique and can vary considerably from school to school. Student performance in model 1 can be viewed as being the combined outcome of community and school environments. To allow us to perform analytical studies and to use standard infinite population-based tests (e.g., those based on Normal distribution theory), the student performance measures are viewed as arising from the individual super-population models specific to the community and school environments of each individual school. The variance component for schools is a function of the differences among the basic processes occurring in each school's educational and community environment.

In contrast to model 1, model 2 treats the state's finite population of students and their performance measures as arising from a single infinite process or super-population model. The finite sample arising from this single model is then arranged into schools and their served communities. The arrangement process is not random and results in a nonzero school variance component. To represent the true variance under this process, the finite population variance (finite at all stages) is based on the particular super-population outcome and the particular arrangement of students into schools (the first two terms in V_{f2}). Then, an estimate of variance arising from the generation of the finite population from the super-population model is added (the final term in V_{f2}).[3]

In the remainder of this report, we limit consideration to model 1 because we believe it provides a better representation of the relationship of the finite population to the underlying infinite processes we wish to evaluate in the analysis of assessment data. We view the structure of schools and their served communities as fundamental contributors to the characteristics of students and the performance measures of the student population rather than viewing all states' students as arising from a single process and then simply being partitioned in some arbitrary manner among the schools in a state. Note that stratification and its effectiveness in controlling the variance of estimates is largely influenced by the craft of the study designers and not by any random process; this view of effective stratification is intuitively more consistent with model 1.

In order to study populations with a mixture of large and small schools, we model a partition of the population into two major strata based on size of school. This more general model reflects a mix of large and small schools and is written as:

$$V_{f1,M} = W_1^2\left[\frac{\sigma_2^2}{n_{12}}\left(\frac{N_{12} - n_{12}}{N_{12}}\right) + \frac{\sigma_3^2}{n_{12}n_{13}}\right] + W_2^2\left[\frac{\sigma_2^2}{n_{22}}\left(\frac{N_{22} - n_{22}}{N_{22}}\right) + \frac{\sigma_3^2}{n_{22}n_{23}}\right]$$

where W_1 is the proportion of eligible students in large schools and W_2 $(= 1 - W_1)$ is the proportion of eligible students in small schools. Population and sample sizes are appropriately defined with an additional leading subscript. Note that common variance components are assumed across size of school strata.

3 Kendall and Stuart (1966, pp. 190-191) show how the variance of a statistic can be partitioned based on any condition c into a component which is the expected value over all c of the variance of the statistic given c, and a component which is the variance over all c of the expected value of the statistic given c. In this case the condition c is defined as a particular super-population outcome of a particular state's finite population of students.

Other Anomalous Population Structures

Table 2 shows some distributions of schools by size for selected states and grades; data are extracted from the National Center for Education Statistics Common Core of Data 1993 (Sierra Systems Consultants, Inc. 1994). No state fits the ideal population model; these states were selected because they were believed to exhibit some of the problems encountered in the strict interpretation of state NAEP sample size guidelines. For the purposes of summarization, small schools are defined as those with 1 to 49 students in the target grade. Large schools have 50 or more students in the target grade. In actual practice, probability-proportional-to-size samples of schools would be selected; probabilities could be adjusted so that approximately equal probability samples of students could be selected after considering the feasible number of sessions at each school. We use artificial examples to examine the impacts of sample allocation on effective sample sizes both to simplify the calculations and to enhance interpretation.

Table 2. 1993 School Distributions for Selected States and Grades

	Large School Stratum (50+ Students)			Small School Stratum (1-49 Students)		
State and Grade	Schools	Total Students	Students/ School	Schools	Total Students	Students/ School
Delaware, Grade 4	51	8,113	159.1	17	172	10.1
Delaware, Grade 12	29	5,721	197.3	19	161	8.5
Rhode Island, Grade 4	115	9,153	79.6	65	2,450	37.7
Rhode Island, Grade 12	39	8,039	206.1	4	55	13.8
Nebraska, Grade 4	158	11,748	74.4	695	10,643	15.3
Nebraska, Grade 12	77	13,424	174.3	261	5,154	19.7
Texas, Grade 4	2,511	266,357	106.1	750	18,018	24.1
Texas, Grade 12	734	164,642	224.3	729	14,239	19.5
Alaska, Grade 4	97	7,643	78.8	253	2,512	9.9
Alaska, Grade 12	30	5,439	181.3	195	1,529	7.8

In Table 3, we show the hypothetical ideal (population 1) as well as five other hypothetical but finite populations which illustrate some of the situations that prevent strict application of the state NAEP guidelines and which roughly match some of the cases shown in Table 2. Populations 2 and 3 illustrate the few schools problem where the "partial sample option" has been applied; these are similar to Delaware grades 4 and 12 or Rhode Island grade 12. Population 4 illustrates the problem with many small schools for which a "sparse sample option" has been proposed; this population resembles the

Nebraska grade 4 population. Population 5 illustrates a mix of large and small schools in a large state; population 5 most closely resembles the situation in Texas for grade 4. Population 6 illustrates a mix of large and small schools in a small state and is similar to Alaska grade 12.

Table 3. Assumed Population Structures

	Large School Stratum			Small School Stratum		
Population Type	**Schools**	**Total Students**	**Students/ School**	**Schools**	**Total Students**	**Students/ School**
Population 1: Hypothetical Ideal	Infinite	Infinite	Infinite	0	0	0
Population 2: 80 Schools	80	10,000	125	0	0	0
Population 3: 50 Schools	50	8,000	160	0	0	0
Population 4: Sparse	158	11,580	75	700	10,500	15
Population5: Mixed, Large	2,500	267,500	107	1,000	18,000	18
Population5: *Mixed, Small*	30	6,000	200	200	1,600	8

Sample Design Options

For each of the five hypothetical finite populations (populations 2, 3, 4, 5, and 6 in Table 3, we consider four alternate sample designs. The first design considers attempts to follow the state NAEP guidelines as closely as possible with first priority given to achieving a sample of 2,500 students per subject and second priority given to achieving a sample of at least 100 schools. All designs use proportional allocation to the two sizes of school strata; i.e., the student sample size is approximately proportional (within rounding of design parameters) to the student population in each size stratum. The alternate designs proposed for each hypothetical finite population involve a reduced sample size based on achieving the same precision as the population 1 ideal would achieve under each of the three variance component distributions shown in Table 1; the precision standard is based on achieving or exceeding the effective sample sizes shown in the rightmost column of Table 1 when the particular variance component distribution is assumed. For example, if we assume variance component distribution A, then we require an effective sample size of at least 1,163 students. Similarly, for assumed variance component distributions B and C, we require effective sample sizes of at least 917 and 758, respectively.

The Few Schools Problem

Hypothetical populations 2 and 3 illustrate the “too few schools” problem. State population 2 has only 80 large schools and no small schools (Table 4). The design that most closely meets state NAEP guidelines would include all 80 schools with an average of 31.25 students per school. Effective sample sizes based on finite variance model 1 naturally exceed those that would be required under the infinite population model since they apply a finite correction factor of zero to the first stage variance component. Since this variance model completely removes the variance associated with schools when all schools are selected (i.e., it treats schools as a stratification variable), the effective sample size actually increases as the intra-school correlation coefficient increases from 0.10 for variance component distribution A to 0.20 for variance component distribution C. Alternate designs are presented for each variance component distribution; these designs meet or exceed the comparable infinite population effective sample sizes shown in Table 1. The alternate designs for all three variance component distributions reduce the student sample per school to 25 and scale down the number of schools to below 80; note that the effective sample sizes decrease as we move from variance component distribution A to C when schools are sampled and contribute to the variance.

Table 4. Sample Design Options and Effective Sample Sizes for Population 2

	Total		Large School Stratum		Small School Stratum		Effective Sample Size Under Variance Component Distribution		
	Schools	Students	Schools	Students/ School	Schools	Students/ School	A	B	C
Population 2	80	10,000	80	125	0	0			
State NAEP	80	2,500	80	31.25	0	0	2,778	2,941	3,125
Design for Distribution A*	58	1,450	58	25	0	0	**1,166**	1,062	975
Design for Distribution B*	54	1,350	54	25	0	0	1,033	**925**	837
Design for Distribution C*	54	1,350	54	25	0	0	1,033	925	**837**

* Alternate sampling designs employing finite sampling correction assumptions that, for the named variance component distribution, meet or exceed the effective sample sizes obtained under the current state NAEP design (with no finite sampling corrections).

Table 5 shows similar results when a state has only 50 large schools and no small schools.

Table 5. Sample Design Options and Effective Sample Sizes for Population 3

	Total		Large School Stratum		Small School Stratum		Effective Sample Size Under Variance Component Distribution		
	Schools	Students	Schools	Students/ School	Schools	Students/ School	A	B	C
Population 3	50	8,000	50	160	0	0			
State NAEP	50	2,500	50	50	0	0	2,778	2,941	3,125
Design for Distribution A*	47	1,175	47	25	0	0	**1,205**	1,221	1,237
Design for Distribution B*	43	1,075	43	25	0	0	1,000	**966**	935
Design for Distribution C*	40	1,000	40	25	0	0	870	816	**769**

* Alternate sampling designs employing finite sampling correction assumptions that, for the named variance component distribution, meet or exceed the effective sample sizes obtained under the current state NAEP design (with no finite sampling corrections).

In general, these two artificial populations lend support to a policy which allows states with few schools to reduce their sample size below those usually specified for participation in the state NAEP.

The Many Small Schools Problem

Population 4 (Table 6) illustrates the problem encountered when states have many small schools. About one-half of the state's student population is enrolled in small schools. In this example, the state NAEP requirements could be met with a sample of 151 schools. Depending on the variance component distribution assumed, school sample sizes for optional sample designs range from 105 to 115 schools to achieve effective sample sizes equivalent to those obtained using the strictly infinite population model. One of the factors operating in determining effective sample size for this case is the reduced clustering effect associated with the small schools in a large portion of the sample.

Table 6. Sample Design Options and Effective Sample Sizes for Population 4

	Total		Large School Stratum		Small School Stratum		Effective Sample Size Under Variance Component Distribution		
	Schools	Students	Schools	Students/ School	Schools	Students/ School	A	B	C
Population 4	858	22,350	158	75	700	15			
State NAEP	151	2,501	53	25	98	12	1,580	1,302	1,155
Design for Distribution A*	115	1,913	41	25	74	12	**1,166**	952	838
Design for Distribution B*	111	1,852	40	25	71	12	1,126	**919**	809
Design for Distribution C*	105	1,754	38	25	67	12	1,060	864	**760**

* Alternate sampling designs employing finite sampling correction assumptions that, for the named variance component distribution, meet or exceed the effective sample sizes obtained under the current state NAEP design (with no finite sampling corrections).

Population 5 (Table 7) was chosen to illustrate a case where both types of limitations influence the sample allocation; in fact, this population does not exhibit either problem since the state NAEP standard design requires only 104 schools. Because the number of schools is large in both school size strata, the finite population correction has only minimal influence on effective sample size. Required sample sizes for both schools and total students are reduced only moderately by taking account of the finite population correction.

Table 7. Sample Design Options and Effective Sample Sizes for Population 5

	Total		Large School Stratum		Small School Stratum		Effective Sample Size Under Variance Component Distribution		
	Schools	Students	Schools	Students/ School	Schools	Students/ School	A	B	C
Population 5	3,500	285,500	2,500	107	1,000	18			
State NAEP	104	2,500	94	25	10	15	1,206	956	795
Design for Distribution A*	101	2,425	91	25	10	15	**1,169**	926	770
Design for Distribution B*	100	2,410	91	25	9	15	1,162	**921**	766
Design for Distribution C*	99	2,385	90	25	9	15	1,150	911	**758**

* Alternate sampling designs employing finite sampling correction assumptions that, for the named variance component distribution, meet or exceed the effective sample sizes obtained under the current state NAEP design (with no finite sampling corrections).

Population 6 (shown in Table 8) was also selected to exhibit the many schools problem. This represents a much smaller state, and application of the finite population correction factor does materially reduce the required sample size from the standard state NAEP model. The state NAEP requirement is not excessive in magnitude, but it does represent a high proportion of the school population. The alternate design options allow for reduction of the proportion of schools sampled.

Table 8. Sample Design Options and Effective Sample Sizes for Population 6

	Total		Large School Stratum		Small School Stratum		Effective Sample Size Under Variance Component Distribution		
	Schools	Students	Schools	Students/ School	Schools	Students/ School	A	B	C
Population 6	230	7,600	30	200	200	8			
State NAEP	117	2,502	30	66	87	6	2,673	2,704	2,869
Design for Distribution A*	70	1,110	30	29	40	6	**1,169**	1,175	1,235
Design for Distribution B*	61	917	29	25	32	6	930	**917**	943
Design for Distribution C*	55	843	27	25	28	6	799	763	**759**

* Alternate sampling designs employing finite sampling correction assumptions that, for the named variance component distribution, meet or exceed the effective sample sizes obtained under the current state NAEP design (with no finite sampling corrections).

Design Effects and Subgroup Estimates

Table 9 shows the design effects for total population estimates, for males, and for a region constituting one third of the state assuming the six hypothetical populations studied above; all results are shown for variance component distribution A and the sample allocation chosen for that distribution in Tables 4 through 9. We use variance model 1 and define the design effect as:

$$Deff_{f1} = \frac{V_{f1}}{V_{SRS}} \ .$$

Note that the numerator in this expression incorporates the finite population correction[4] for the school component of variance and that any contribution from the component of variance associated with within-state strata is completely eliminated. No finite population correction is assumed for the simple random sample variance in the denominator.

Table 9. Design Effects Assuming Variance Component Distribution A

	Total Population			All Males			Region (1/3 of State)		
Pop.	Actual Sample Size	Effective Sample Size	Design Effect	Actual Sample Size	Effective Sample Size	Design Effect	Actual Sample Size	Effective Sample Size	Design Effect
1	2,500	1,163	2.15	1,250	820	1.52	833	388	2.15
2	1,450	1,166	1.24	725	676	1.07	483	389	1.24
3	1,175	1,205	0.98	588	627	0.94	392	402	0.98
4	1,913	1,166	1.64	957	753	1.27	638	389	1.64
5	2,425	1,169	2.07	1,213	815	1.49	808	390	2.07
6	1,110	1,169	0.95	555	600	0.93	370	390	0.95

Design effects for subgroups which partition all (or nearly all) sessions tend to reduce the design effect since the effect of clustering is reduced; the design effects for males exhibit this behavior. Subgroups which are formed in terms of entire schools, such as regional estimate, maintain the same clustering effects, and assuming constant variance components across subgroups, the same design effects.

4 For population structure 1, the hypothetical infinite population, the finite correction factor is exactly 1.

More complex changes in design effects are associated with subgroups such as ethnic groups which may exhibit varying concentrations over schools. Table 10 shows design effects for three populations based on their distribution across schools in selected states. Design effects were computed assuming stratification by ethnic subgroup concentration and averaging clustering effects over strata. The model used to approximate these design effects in Table10 took account of the distribution of schools by concentration of the population subgroup. It also took account of the differing finite population correction factors in different size of school strata for populations 5 and 6.

Table 10. Design Effects for Selected Ethnic Subgroups Assuming Variance Component Distribution A

Population Structure	Population Subgroup	Distribution Based on	Actual Sample size	Effective Sample Size	Design Effect
3	Blacks	Delaware	334	363	0.92
5	Hispanics	Texas	858	517	1.66
6	Indians/Alaskan natives	Alaska	343	346	0.99

To gain a better intuitive understanding of the design effects for subpopulations when using a finite population correction factor, it may be useful to examine a simpler approximation to the design effect for subgroups. When there is no stratification and no finite population correction, the design effect is often expressed as $Deff = 1 + \rho(\bar{n}_3 - 1)$ where we carry over the notation from the previous section on finite population variance models and define the intra-school correlation coefficient as $\rho = 1 - \sigma_3^2/\sigma^2$. If we wish to incorporate stratification, subgroup estimation, and a finite population correction at the first stage as assumed in variance model 1, the simple model can be enhanced as $Deff = 1 + \rho(\gamma d \bar{n}_3 fpc_2 - 1)$ where $\gamma = \sigma_2^2/(\sigma_1^2 + \sigma_2^2)$, or the portion of the remaining variance associated with schools, d represents the proportion of the population that belongs to a particular subgroup or domain, and fpc_2 represents the finite population correction factor for the school stage of sample selection. Several observations can be made from this model. If the finite population correction, fpc_2, goes to zero, the design effect becomes simply $1 - \rho$ taking full advantage of the initial strata and schools as stratifying variables to totally eliminate contributions from the first two components of variance. If the product, $\gamma d \bar{n}_3 fpc_2$, becomes less than 1 for any reason, the design effect will also be less than 1. For the three populations and sample allocations considered in Table 10, this model projects design effects of 0.92, 1.31, and 1.03, respectively. The differences occur because the calculations in Table 10 for models 3 and 5 utilized the partitioned or mixed variance model, $V_{f1,M}$, and the finite population correction factor is zero in the first partition. The simpler model uses an average finite population correction factor which does not recognize the mix of zero and nonzero corrections.

Many of the sample size requirements for state NAEP stem from the need to support subgroup level estimates. One must examine both expected yield and projected design effects for population subgroups. The main problem for some subgroup estimates may be that proportional allocation will not yield adequate sample sizes to support the planned analyses. Without adequate stratification controls in a proportional allocation design, the subgroup sample size may behave as an uncontrolled random variable, leaving the adequacy of the sample size to chance. Stratification can be applied to reduce this type of variability in achieved sample sizes; drawing larger overall samples can also reduce the sample-to-sample variability in achieved subgroup sample sizes, thus reducing the risk of an unusually small subgroup sample occurring by chance.

Conclusions and Recommendations

The investigation above lends support to alternate sample size specifications in both states with few schools and states with many small schools. Some additional changes in optional designs should also be considered. The alternate allocations used above assumed proportional or near proportional allocations to the large school and small school strata. Appropriate cost modeling and the application of optimum allocation theory could yield less costly designs which would most likely favor more sampling from the larger schools. The precision requirements for each design were based on maintaining the effective sample size that would pertain to the ideal hypothetical population case (or, equivalently, to the case in which the state had a sufficiently large number of schools such that finite population corrections would have very little impact on the variance). A better approach would be to base the required effective sample size on the needs of data users.

Better estimates of variance components associated with strata, schools, and students for a variety of assessment measures would also be of great assistance in investigations such as this one. These estimates should be combined with the development of variable survey cost models so that optimal sample design decisions could be based on minimizing overall cost that is subject to meeting specified precision (or, equivalently, effective sample size) requirements.

Allowing states to implement design options other than the standard requirement of at least 100 schools and at least 2,500 students per subject could actually help to reduce the variable costs associated with administration, field quality control, scoring, and data processing. Analyzing a state's school and student population distributions in order to fully develop an alternative which incorporates the reduced sample size and still meets analytic requirements for selected subgroups as well as for the total population would add to the total cost of state assessments. If this process could be routinized, the increased cost should be controllable. Some initial experience would quickly disqualify a large number of states from consideration so that these additional costs should also only apply to a few special cases. In general, analysis of the adequacy of sample and overall survey designs with respect to users' analytic needs should be viewed as essential to the ongoing success of any longitudinal data series such as the National and State assessments of education; the review of state assessment designs certainly fits within this general concept.

References

Cochran, W. G. (1977). *Sampling techniques*. New York: John Wiley & Sons.

Forsyth, R., Hambleton, R., Linn, R., Mislevy, R., & Yen, W. (1966). Design/feasibility team report to the National Assessment Governing Board.

Freund, D., & Carlson, J. (1997). Memo to NAEP Staff on the Effect on State NAEP of reduced no. of schools *Revised*. Educational Testing Service.

Kendall, M. G., & Stuart, A. (1966). *The advanced theory of statistics*. London: Charles Griffin & Company Limited.

Rust, K. (1997). Reduced sample option for states with a high proportion of small schools. Draft report. Westat, Inc.

Sierra Systems Consultants, Inc. (1994). *CD-ROM installation and technical user guide: National Center for Education Statistics Common Core of Data (CCD93 CD-ROM DISC 1988/1989 - 1993/94)*. Arlington, VA.