DOCUMENT RESUME

ED 206 709 TM 810 612

AUTHOR Williams, Rick L.: And Others
TITLE NAEP Year 11 Design Efficiency Study. Final Report.
INSTITUTION Research Triangle Inst., Research Triangle Park, N.C.
SPONS AGENCY Education Commission of the States, Denver, Colo. National Assessment of Educational Progress.: National Center for Education Statistics (ED), Washington, D.C.: National Inst. of Education (ED), Washington, D.C.
REPORT NO RTI-1969-01-01-F
PUB DATE Jun 81
CONTRACT OEC-0-74-0506.
GRANT NIE-G-80-0003
NOTE 50p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Educational Assessment: Elementary Secondary Education: Error of Measurement: National Competency Tests: *Research Design: *Research Methodology: *Sampling: Testing Programs
IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

The National Assessment of Educational Progress in-school sampling design is a three-stage stratified design. Stratification variables include region, size of community and socioeconomic status. The three levels of sample selection are Primary Sampling Units (PSUs), schools and students. In general, two and sometimes three PSUs are selected from each stratum for variance estimation. The stratification variables are assumed fixed and not subject to change; therefore, the problem of finding the optimal design is reduced to finding the number of PSUs, schools and students per stratum that will minimize cost for a given variance. Following a brief overview of the sample drawn for Year 11, presented in Section 2, the cost model developed for the purpose of the present study is outlined in Section 3. Section 4 describes the statistics which were selected for analysis, and Section 5 derives the corresponding variance and covariance component models. Finally, Section 6 describes the optimization procedure used, and Section 7 provides a summary of the results. Primary type of information provided by report: Procedures (Sampling). (Author/BW)

RESEARCH TRIANGLE INSTITUTE

RTI/1969/01-01 F

Final Report

NAEP Year 11 Design Efficiency Study

by

Rick L. Williams
David V. Budescu
James R. Chromy

Sampling Research and Design Center

Prepared for

National Assessment of Educational Progress

June 1981

RESEARCH TRIANGLE PARK, NORTH CAROLINA 27709

TABLE OF CONTENTS

	Page
LIST OF TABLES	iii
1. INTRODUCTION	1
2. SAMPLE OVERVIEW	4
3. COST MODELS	7
4. STATISTICS ANALYZED	17
5. VARIANCE AND COVARIANCE MODELS	21
6. OPTIMIZATION PROBLEM	25
7. NUMERICAL RESULTS	26
8. CONCLUSIONS	40
REFERENCES	41
APPENDIX A: COVARIANCE COMPONENTS FOR TWO ESTIMATED TOTALS	42

LIST OF TABLES

Table		Page
3-1	Percentage of Time Allocated to Different Activities by DSs and EAs	8
3-2	Variable Cost Components Summary: 1980 Levels	12
3-3	Estimated Cost Parameters	16
4-1	Within-Package Score Definitions	19
4-2	Cross-Package Score Definitions	20
4-3	Analysis Populations	20
7-1	NAEP Design Optimization for Within-Package Means	30
7-2	NAEP Design Optimization for Within-Package Means	30
7-3	Projected Percent RSE's for Within-Package Means Assuming 10 PSU's/Stratum, 2 Rep's/PSU and 15 Student/Rep	32
7-4	Percent RSE Constraints for Within-Package Means	32
7-5	NAEP Design Optimization for Within-Package Means	33
7-6	NAEP Design Optimization for Within-Package Means	33
7-7	NAEP Design Optimization for Within-Package Means	34
7-8	NAEP Design Optimizations for Cross-Package Means	36
7-9	Projected Percent RSE's for Cross-Package Means Assuming 50 PSU's, 2 Rep's/PSU, and 15 Students/Rep	37
7-10	Percent RSE Constraints for Cross-Package Means	38

1. INTRODUCTION

Design efficiency studies are conducted to determine whether sampling procedures have been effective and what sample design provides minimum cost for a given variance. The results from the design efficiency studies are used to plan future assessments.

The National Assessment in-school sampling design is a three-stage stratified design. Stratification variables include region, size of community (SOC) and socioeconomic status (SES). The three levels of sample selection are PSUs, schools and students. In general, two and sometimes three PSUs are selected from each stratum for variance estimation. The stratification variables are assumed fixed and not subject to change; therefore, the problem of finding the optimal design is reduced to finding the number of PSUs, schools and students per stratum that will minimize cost for a given variance. One of the objectives of the design efficiency study has been to determine the "optimal" values of these parameters.

If only one statistic is used, the solution for the optimal design is well known (Kish, 1974). It is rare, however, in any survey to have only one statistic of interest. In the past, the best optimality criterion for many statistics was not obvious. Some possibilities that have been considered are: 1) the design that has minimum average variance at the given cost; 2) the design with maximum average efficiency relative to the separate statistic optima; 3) the design with minimum average loss (inverse efficiency); and 4) the design that minimizes cost subject to variance constraints for each of the separate statistics.

The average of several quantities is meaningful only if all the quantities are measured on the same scale with similar units. The variances of different statistics would be measured on different scales; hence, the minimum average variance does not seem to be a meaningful criterion. To avoid this problem, the efficiency for a particular statistic is defined. This efficiency is a ratio with numerator equal to the minimum variance that can be achieved by the optimal design for that statistic and the denominator equal to the variance of the statistic for the given design. The problem now is what criterion can be formulated for optimality, based on the average efficiency. A possibility is to find the design with large average efficiency at the given cost with small variance of efficiencies over all statistics for this design. The trade-off between the maximum, mean and minimum variance of efficiencies is not easy to define. This criterion was used in the Year 3 efficiency study (Shah, Folsom, Clayton, 1973).

Kish (1974) has advocated inverting the efficiency to form what he calls the loss function of a particular design. The advantage of Kish's loss function approach is that a simple analytic solution exists for the minimum average loss where the averaging may be weighted. This is the optimal criterion used in the Year 7 efficiency study (Sherdon, Folsom, Clemmer, 1977).

While minimizing cost subject to a set of variance constraints for key statistics provides an appealing solution, efficient computational algorithms for obtaining such solutions have not generally been available. Such an algorithm was recently derived and software for its implementation was developed by Dr. James R. Chromy. This method will be used in the present study.

To determine optimal designs estimates of the variance and cost component associated with each stage of sampling and statistic of interest are necessary. Following a brief overview of the sample drawn for Year 11, which will be presented in Section 2, we outline the cost model developed for the purpose of the present study in Section 3. Section 4 describes the statistics which were selected for analysis, and Section 5 derives the corresponding variance and covariance component models. Finally, Section 6 describes the optimization procedure used, and Section 7 provides a summary of the results.

2. SAMPLE OVERVIEW

The National Assessment sampling design is a three-stage stratified probability sample. Stratification variables include region, community size, and socioeconomic status. An overview of the general sampling and weighting process is included here for completeness and reference.

The National Assessment sample is designed to be representative of students in three age classes, 9-, 13-, and 17-year-olds, in all schools and communities in the nation. It is also designed to produce, for a variety of subpopulations, performance estimates which are relatively unbiased and which meet certain precision requirements.

Primary Sampling Units (PSUs) are geopraphic land areas consisting of a single county or several counties. Each year approximately 83 PSUs are randomly selected on a probability basis so that every county and every state in the United States has a positive chance of being included in the sample.

At the second stage of sampling, a list of all schools, both public and private, within each of the selected PSUs is developed and a probability sample of these schools is selected for each of the three age classes. The number of schools selected in each PSU is determined by the approximate number of students in the eligible age group attending each school. Schools are selected in such a way that any given school will not appear in the sample more than once in a four-year period. In most years, about 1,600 schools are selected; the number selected in a particular year depends upon the number of distinct packages.

The third and final stage of sampling is the selection of a random sample of students from the eligible age group at each selected school. A total of approximately 2,600 respondents is obtained for each National

Assessment package. Generally, the students are selected from one to eight schools within each selected PSU for each of the three age groups being assessed.

Selected students who do not show up for assessment are termed nonrespondents. Response rates for 9- and 13-year-olds tend to average about 85 percent, whereas the response rate for 17-year-olds averages 75 percent. Seventeen-year-olds who miss their appointments are followed up in school the day after the assessment. Seventeen-year-old dropouts and early graduates are located in their homes and administered packages. According to census data, about 10 percent of the 17-year-olds are not enrolled in school. Including these out-of-school individuals in the target population enables National Assessment to apply its results to the entire population of 17-year-olds rather than only to those enrolled in school. The assessment of dropouts and early graduates is termed the Supplementary Frame Assessment.

Sample weights adjusted for nonresponse are computed for each age class. The weights are calculated as the reciprocal of the appropriate selection probabilities. Sample weights are used to calculate ratio estimates of the proportions of population members who respond in alternative ways to assessment exercises. So that the proportion of population members who respond in alternative ways can be calculated based on community location and occupation of parents, the assessment data are postclassified into seven size and type of community (STOC) categories.

Each year from 75,000 to 100,000 persons are assessed in one or more learning areas normally taught in schools. In the past, NAEP has conducted major assessments in art, career and occupational development, citizenship, literature, mathematics, music, reading, science, social

studies, and writing, and six of these areas have already been reassessed. Year 11 of the project (1979-80) was the third assessment in the areas of reading and literature.

3. COST MODELS

Cost models are used in design efficiency studies to show how total survey costs would be affected by changing some of the design parameters. Experience in any given survey at a fixed level of the design parameters does not usually provide any directly useable data base for estimating the parameters of a useful cost model.

For NAEP, the budget prepared for any given year was allocated to the following five categories:

1. Packages;
2. Travel points;
3. Schools;
4. Administrative sessions; and
5. Students.

In addition, a non-allocated or fixed cost component was identified.

The cost model parameters estimates are only applicable within certain limits and assumptions based on current NAEP operating practices. A few of the major assumptions are listed below:

1. The current schedule of 13-year-old assessment in the fall, 9-year-old assessment in January and February, and 17-year-old assessment in the spring is followed;
2. Any travel point in the sample for one age class assessment is also included in the other two age class assessment at the same level (same number of replications of each package administration);
3. No more that ten group package sessions may normally be assigned to a single school;
4. No more than 25 students (expected response) may be assigned to a package session; and
5. Each package will consist of no more that 45 minutes of paced tape exercises.

Any departure from these assumptions may require revision of cost model parameters.

Allocation of 1978 Budgets.

Administration. The major portion of the variable costs of an in-school assessment is associated with the field work required to gain cooperation and collect the NAEP package data. Furthermore, a large portion of this budget is associated directly or indirectly with the off-site staff of District Supervisors (DSs) and temporary Exercise Administrators (EAs).

Table 3-1 shows the assumed allocation of DS and EA time. The allocations are based on:

1. The technical proposal for the period;
2. The usual operating practice as defined in the Year 09 DS Manual; and
3. Updated District Supervisor (DS) expense reports.

Five variable cost categories (packages, travel points, schools, administrative sessions, and students) are identified in the overview. In analyzing the budget, it was noted that many cost items relate directly to the number of District Supervisors required to conduct the assessment.

Table 3-1. Percentage of Time Allocated to Different Activities by DSs and EAs

	Activity				
	Total	Travel Point	Schools	Sessions	Students
DS	100	22.8	56.6	7.7	12.9
EA	100	12.6	20.1	37.5	29.8

A special category called "DS-related" was used to accumulate all such costs. The total in this category was then allocated to variable cost categories according to the percentage distribution shown in Table 3-1.

Costs not allocated to any of the variable cost categories belong in the fixed cost or setup category. The general approach to allocating budgeted costs is discussed by category in the following paragraphs.

Costs placed initially in the "DS-related" category included:

(1) Fifty percent of labor costs for the associated project director for administration, the field director, and the administrative secretarial support;

(2) Eighty-five percent of the labor costs for the regional supervisors, the administrative coordinator, and the survey assistant;

(3) All DS labor costs;

(4) Duplication of cassette group tapes;

(5) Help wanted advertising costs;

(6) DS relocation expenses;

(7) Supplies including ring-binders, jiffy bags, date books, and corrugated boxes;

(8) Shipping and communication costs of DSs including mailings to and from DSs, shipment of reports, and postal cards;

(9) Central staff travel to supervise DSs, to conduct quality checks, to recruit new DSs, and to train DSs (excluding travel to the annual training session);

(10) DS travel to training sessions and to debriefing sessions; and

(11) Seventy-five percent of printing costs for DS manuals and demonstration packages.

Costs allocated to packages included:

(1) Ninety percent of the labor costs for the proofing and tape coordinator; and

(2) Production costs for magnetic tapes.

Costs allocated to travel points included:

(1) 22.8 percent of items initially allocated to DS-related expenses;

(2) DS travel to and from PSUs to conduct assessment;

(3) DS travel to and from PSUs to hold introductory meetings; and

(4) 12.6 percent of EA services.

Costs allocated to schools included:

(1) 56.6 percent of items initially allocated to DS-related expenses;

(2) Fifty percent of computer programmer labor costs;

(3) Supplies expenses including envelopes, mailing tapes, mailing labels, stationery, portfolios, and SLF storage envelopes;

(4) Computer costs including usage charges, magnetic tapes, and print ribbons;

(5) Shipping and communications costs for school mailings and for toll calls to schools;

(6) Central staff travel expenses for large city contacts;

(7) DS travel within PSUs to conduct assessment;

(8) DS travel within PSUs to conduct introductory meetings;

(9) Seventy-five percent of printing costs for introductory materials, memoranda, school official questionnaires, and school workers;

(10) 20.5 percent of EA services; and

(11) E[illegible]l expenses.

Costs allocated to administrative sessions included:

(1) 7.7 percent of DS-related expenses;

(2) Tape recorder repair expenses;

(3) Depreciation of cassette recorders (approximated by twenty percent of cost budgeted for 1979);

(4) Seventy-five percent printing costs for administration schedules, EA manuals, and EA administrative instructions; and

(5) 37.5 percent of EA services.

Costs allocated to students included:

(1) 12.9 percent of DS-related expenses;

(2) Supplies expenses for pencils and art materials;

(3) Shipping costs for bus and freight shipments;

(4) Excess baggage charges for DS travel;

(5) Seventy-five percent of printing costs for student listing forms (SLFs) and parental permission forms; and

(6) 29.8 percent of EA services.

Based on these allocations, the 1980 variable administration costs were determined as follows (numbers are rounded):

Variable cost associated with	Amount
Package setup	$1,200.00
Travel points	2,816.00
Schools	292.00
Administrative sessions (35 min.)	13.41
Students	1.57

Sampling. Sampling costs constitute a much smaller part of the total budget. Variable sampling costs were associated with primary sampling units and with schools. Budgets for 1980 were examined under various design configurations actually considered prior to implementation. Results of sampling cost component estimation for the Year 07 design efficiency studies (Sherdon, Folsom, Clemmer, 1977) were also examined. Based on these considerations, the following variable costs at 1980 levels were developed:

Task	PSU component	School component
Select school sample	$140.00	$11.64
Package assignment	---	43.05
Weights	---	9.31
Total	140.00	64.00

Print and Scoring. Only approximate variable costs for the printing and scoring components are included in the model: for package setup, $10,000 and for students, $1.25. Scoring costs can vary widely depending on the mix of hand scoring and direct optical scanning.

Variable cost components are summarized in Table 3-2.

Table 3-2 Variable Cost Components Summary: 1980 Levels

Name	Symbol	Amount
Fixed	C_o	$271,000.00
Package setup	C_p	1,200.00
Travel points	C_1	3,056.00
Schools	C_s	356.00
Administrative sessions	C_a	13.41
Students (edit & score)	C_e	2.82

Relation of cost and variance models. In order to seek optimum design configurations, cost and variance models must be stated in terms of the same design parameters. Basic design parameters include the number of primary sampling units (PSUs), the number of replicates per PSU, and the number of students sampled per replicate. The term "replicate" is used to denote the number of group administrations planned for each group package with a PSU. Since all packages cannot be administered in each sample school, the number of schools usually exceeds the number of replicates.

The cost function has been found to be very sensitive to the number of schools. The variance funtion (for group packages) is directly affected only by the number of replicates. The number of schools which must be selected, on the average, in each replicate depends on a number of factors:

(1) The number and type of packages assigned to each age class;

(2) The number of students selected per package;

(3) The distribution of school sizes in terms of age eligibles; and

(4) Any administrative restrictions employed to limit the burden placed on individual schools.

A number of models relating the number of schools per replicate to package configurations and sample sizes were studied using Year 01 through 09 data. It was noted that the number of schools per replicate should be at least one even if only one package was used. The following model was selected for its fit to the data and for its intuitive appeal:

$$s_a = \mathrm{Max}[1, b_{1a} G_a] \tag{3.1}$$

where

s_a = the average number of schools per replicate for age class-a, and

G_a = the number of group packages assigned for age class-a.

The value b_{1a} was obtained for each age class by ordinary least squares fitting of the model

$$s_a = b_{1a} G_a \tag{3.2}$$

to the Year 01 through 09 data. The model ignores the sample size for group package sessions; this assumption would be unrealistic if the group session size could vary without limit. In practice, scheduled sessions for more than 25 students have been difficult to manage. Allowing

for nonresponse and some variability in assigned sample sizes to achieve weight stability, an average achieved sample size of 16 students per group session is considered near the feasible maximum. Year 01 through 06 assessments were targeted for 12 respondents per group package session; Year 07 through 09 assessments were targeted for 16 respondents per group package session.

Since the clustering effect for individual packages is less than that for group packages under the sample allocation schemes normally employed, the sample size per replicate for individual packages should be less than that for group packages under equivalent precision requirements.

Estimated values for b_{1a} based on ordinary least squares fits were as follows (standard errors are indicated in parenthesis):

Age class	$\hat{b}_{1a}$
9 (a=1)	.419 (.030)
13 (a=2)	.312 (.024)
17 (a=3)	.212 (.013)

Cost model parameters. A number of cost models suitable for studying alternative design configurations can be developed from the data in Table 3-2. If the assumptions outlined above hold, a cost model can be stated in terms of the number of PSUs, replicates, and students sampled per package as follows:

$$C = C_o + n_1C_1 + n_1n_2C_2 + n_1n_2n_3 \sum_{a=1}^{3} G_aC_{3G_a}$$

$$= C_o + n_1C_1 + n_1n_2C_2 + n_1n_2n_3C_3 \qquad (3.3)$$

where

C = total

C_o = fixed cost component (may be a function of number of packages);

C_1 = cost associated with adding one PSU;

C_2 = cost associated with adding one replicate;

C_{3G_a} = cost associated with adding one respondent to a group session at age class a;

$$C_3 = \sum_{a=1}^{3} G_a C_{3Ga}$$

n_1 = number of PSUs;

n_2 = number of replicates per PSU;

n_3 = number of student respondents per replicate per group package.

The value of C_1 is stated in Table 3-2 directly; C_2 and C_{3G_a} can be determined from the values in Table 3-2 and certain assumed relationships of cost model and variance model parameters.

The value of C_2, costs associated with adding one replicate, can be stated as functions of the school cost component, C_s; the estimated regression parameter, $\hat{b}_{1a}$, relating numbers of schools to numbers of packages; the number of group packages, G_a, for each age group a; and the administration session cost, C_a. Symbolically, C_2 can be expressed as

$$C_2 = C_a \sum_{a=1}^{3} G_a + C_s \sum_{a=1}^{3} \hat{b}_{1a} G_a \tag{3.4}$$

The student cost component, C_{3G_a}, for group sessions can be expressed in terms of student editing and scoring costs (C_e). Assuming cost structures are approximately the same for all three ages,

$$C_{3G_a} = C_e \tag{3.5}$$

Finally this yields

$$C_2 = \$13.41(G_1 + G_2 + G_3) + \$356.00[.419G_1 + .312G_2 + .212G_3] \tag{3.6}$$

$$C_{3G_a} = \$2.82 \quad (a=1,2,3) \tag{3.7}$$

$$C_3 = \$2.82\ [G_1 + G_2 + G_3] \tag{3.8}$$

Three cost models were entertained. The first one used assumed a 40 package assessment with 11, 15, and 14 group packages for the three ordered age groups, respectively. This is the assignment used in Year 11. The two other cost models assumed 3 packages per age group and 6 packages per age, respectively. The estimated cost parameters for these three models are shown in Table 3-3.

Table 3-3. Estimated Cost Parameters

Number of Packages	C_1	C_2	C_3
11 for age 9 15 for age 13 14 for age 17	3056.00	4899.29	112.80
6 per age group	3056.00	2255.63	50.76
3 per age group	3056.00	1188.64	25.38

4. STATISTICS ANALYZED

It was emphasized in the introduction that the present study will seek to optimize the design simultaneously for several statistics. Specifically, 58 items, intended to measure two NAEP subobjectives, were selected. From these items, 21 linear combinations were defined for analysis. All the items are pertinent to objective IV of the 1979-80 Assessment--"Application of study skills in reading." The following two subobjectives were specifically addressed:

A. "Obtains information from nonprose reading facilitators"

B. "Obtains information from materials commonly found in libraries or resource centers."

Subobjective (a) attempts to evaluate whether the students use visual aids when reading and whether they can correctly interpret information given in charts, maps and graphs. The second subobjective is directed at measuring the extent to which students use various reference materials and whether they can find specific information in these materials (e.g., dictionaries, encyclopedias, etc). All the items selected for analysis are multiple choice in format with a single correct response.

The scores analyzed in this study fall into two categories--within-package scores and cross-package scores. Within-package scores are those defined from items taken entirely from a single package. Conversely, cross-package scored involve items taken from multiple packages.

The 15 within-package scores shown in Table 4-1 were considered. For each student taking one of the indicated packages, the score was defined to be the proportion of items involved that the student answered correctly. The statistics of ultimate interest were the means of the scores over students or, in other words, the mean proportions answered correctly.

Cross-package scores were constructed by averaging related within-package estimates. That is, if $\hat{R}_1, \hat{R}_2, \ldots, \hat{R}_G$ are related within-package means, then their associated cross-package mean is

$$\hat{R} = [\hat{R}_1 + \hat{R}_2 + \ldots + \hat{R}_G]/G. \tag{4.1}$$

A cross-package mean was defined ~~for~~ both subobjectives within each age group. These are presented in Table 4-2.

Finally analyses were conducted for the six populations listed in Table 4-3.

Table 4-1. Within-Package Score Definitions

Within-Package Score Number	Age	Package	Item Numbers	Associated Subobjectives*
1	9	4	7a-7d	b
2	9	5	8a-8d	b
3	9	5	9a-9c	a
4	9	8	5a-5d,9a-9d	a
5	13	2	15a,15b	a
6	13	4	7a-7d	b
7	13	6	7a-7c	a
8	13	6	9a-9e	b
9	13	8	9a-9d	a
10	17	1	7a,7b	a
11	17	1	10a-10c	a
12	17	4	7a-7d	b
13	17	6	9a-9c	b
14	17	13	8a-8d	b
15	17	13	9a-9c	a

*Subobjectives:

a. Obtains information from nonprose reading facilitators.

b. Obtains information from materials commonly found in libraries or resource centers.

Table 4-2. Cross-Package Score Definitions

Age	Defining Within-Package Score Numbers From Table 4-1 Subobjectives (a)*	(b)**
9	3,4	1,2
13	5,7,9	6,8
17	10,11,15	12,13,14

*Obtains information from nonprose reading facilitators.

**Obtains information from materials commonly found in libraries or resource centers.

Table 4-3. Analysis Populations

All students (National)
Non-whites
Males
Females
Students with parents' education less than high school
Students with parents' education at least high school

5. VARIANCE AND COVARIANCE MODELS

Variance models are developed to demonstrate how the precision of the estimates would be affected by changes in the design parameters. In this section two general variance models are derived which can be applied to (1) scores drawn from a single package and (2) mean scores combined across packages.

For Year 11 assume a with replacement three stage design of

1) PSUs
2) Schools
3) Students

and that the sample is selected with probabilities proportional to size (PPS) at the first two stages and with equal probabilities at the last stage. Let

y_{gijk} = response for student-k from school-j from PSU-i for package-g

n_1 = number of sample PSUs

n_2 = number of replicates per PSU

n_3 = number of sample students per school

G = number of packages of interest

N = number of PSUs in population

N_i = number of schools in PSU-i population

N_{ij} = number of age eligible students in school-ij

A_{ij} = size measure for school-ij

$P_i = A_i/A_+$ = single draw probability for PSU-i

$P_{j(i)} = A_{ij}/A_{i+}$ = single draw probability for school-ij given PSU-i selected.

The estimate of a population total, say Y, from a single package is

$$\hat{y}_g = \sum_{i=1}^{n_1} [n_1 P_i]^{-1} \sum_{j=1}^{n_2} [n_1 P_{j(i)}]^{-1} \sum_{k=1}^{n_3} y_{gijk} N_{jk}/n_3 \qquad (5.1)$$

Now, consider a separate ratio estimator of a cross-package mean score for items taken from G distinct packages. That is,

$$\hat{R} = [\hat{R}_1+\hat{R}_2+\ldots+\hat{R}_G]/G \tag{5.2}$$

where

$\hat{R}_g = \hat{y}_g/\hat{x}_g$

= the within package mean score for package-g

$\hat{y}_g$ = the estimated total for the score from the package-g sample (g=1,2,..., G)

and

$\hat{x}_g$ = the total number students estimated from the package-g sample (g = 1,2,...,G).

The variance of $\hat{R}$ is

$$V(\hat{R}) = [\sum_{g=1}^{G} V(\hat{R}_g) + 2 \sum_{g=1}^{G-1} \sum_{g'=g+1}^{G} Cov(\hat{R}_g,\hat{R}_{g'})]/G^2 \tag{5.3}$$

It is commonly known and has been used in previous NAEP efficiency studies that the $V(\hat{R}_g)$ can be decomposed into

$$V(\hat{R}_g) = \sigma_g^2(1)/n_1 + \sigma_g^2(2)/n_1n_2 + \sigma_g^2(3)n_1n_2n_3 \tag{5.4}$$

where

$\sigma_g^2(1)$ = the between PSU contribution to variance,

$\sigma_g^2(2)$ = the between school within PSU contribution to variance,

$\sigma_g^2(3)$ = the between student within school contribution to variance.

Software is available at RTI to estimate these three components (Shah, 1979). Model (5-4) is exactly that required for scores drawn from a single package.

Now, the Taylor Series approximation to the $Cov(\hat{R}_g,\hat{R}_{g'})$ is

$$Cov(\hat{R}_g,\hat{R}_{g'}) = Cov(\hat{y}_g/\hat{x}_g,\hat{y}_{g'}/\hat{x}_{g'})$$

$$\doteq \frac{1}{X_g X_{g'}} [Cov(\hat{y}_g,\hat{y}_{g'}) - R_g Cov(\hat{y}_{g'},\hat{x}_g)$$

$$- R_{g'} Cov(\hat{y}_g,\hat{x}_{g'}) + R_g R_{g'} Cov(\hat{x}_g,\hat{x}_{g'})]$$

$$= \mathrm{Cov}[(\hat{y}_g - R_g\hat{x}_g)/X_g,\ (\hat{y}_{g'} - R_{g'}\hat{x}_{g'})/X_{g'}]$$

$$= \mathrm{Cov}(\hat{z}_g, \hat{z}_{g'}) \tag{5.5}$$

where x_g and R_g are the corresponding population values of $\hat{x}_g$ and $\hat{R}_g$, respectively. The form of the covariance term is explicitly derived in Appendix A. It is shown that

$$\mathrm{Cov}(\hat{R}_g, \hat{R}_{g'}) = \mathrm{Cov}(\hat{z}_g, \hat{z}_{g'})$$

$$= \sigma_{gg'}(1)/n_1 + t\sigma_{gg'}(2)/n_1n_2 \tag{5.6}$$

where $\sigma_{gg'}(1)$ and $\sigma_{gg'}(2)$ are components of covariance for the first and second stages, respectively, analogous to the variance components in (5.4). Also, t is the proportion of schools where both package-g and -g' are administered.

Next, combining (5.3), (5.4) and (5.6)

$$V(\hat{R}) = G^{-2} \sum_{g=1}^{G} [\sigma_g^2(1)/n_1 + \sigma_g^2(2)/n_1n_2 + \sigma_g^2(3)/n_1n_2n_3$$

$$+ 2 \sum_{g=1}^{G-1} \sum_{g'=g+1}^{G} [\sigma_{gg'}(1)/n_1 + t\sigma_{gg'}(2)/n_1n_2]$$

$$= G^{-2} \{\sigma^2(1)/n_1 + \sigma^2(2)[1 + 2t\rho]/n_1n_2 + \sigma^2(3)/n_1n_2n_3\} \tag{5.7}$$

where

$$\sigma^2(1) = \sum_{g=1}^{G} \sigma_g^2(1) + 2 \sum_{g=1}^{G-1} \sum_{g'=g+1}^{G} \sigma_{gg'}(1) \tag{5.8}$$

$$\sigma^2(2) = \sum_{g=1}^{G} \sigma_g^2(2) \tag{5.9}$$

$$\rho = [\sum_{g=1}^{G-1} \sum_{g'=g+1}^{G} \sigma_{gg'}(2)]/\sigma^2(2) \tag{5.10}$$

$$\sigma^2(3) = \sum_{g=1}^{G} \sigma_g^2(3) \tag{5.11}$$

The above model for the variance of $\hat{R}$, equation (5.7), contains a parameter (t) for the proportion of schools where a given pair of packages were jointly administered. This is a function of the number of schools per replicate since every package is administered within a replicate. Chromy, Clemmer, and Jones (1980) have shown that the number of schools per replicate can be approximated by a function of the total number of packages. This implies that t probably does not vary with the sample design parameters (i.e., n_1, n_2, and n_3). Thus the value of t observed in the data will be substituted into the model.

The model outlined above was developed at the stratum level. However, in determining the optimal allocation, the population will be stratified in eight strata defined by the cross-classification of the geographical region (West, Central, Northeast, and Southeast), and the community size (rural = no place with 25,000 or more population in 1970 Census, Urban = otherwise). Variance and covariance components will be estimated for each of the strata.

Finally, variance and covariance components will be estimated for each of six subpopulations (domains) judged to be of interest: National, Non-white males, Females, students with parents' education less than high school, and students with parents' education at least at the high school level.

6. OPTIMIZATION PROBLEM

Thus far, a linear cost model has been developed of the general form

$$C = \sum_{h=1}^{H} C(h)x(h) \tag{6.1}$$

where C(h) is the cost of adding an additional unit to the hth stage of the sample and x(h) is the sample size for stage-h. In addition, a variance model has also been developed of the form

$$V(k) = \sum_{h=1}^{H} V(kh)/x(h) \tag{6.2}$$

where V(kh) is the component of variance associated with the hth stage of sampling for statistic-k and x(h) is as before.

Combining these two models, the problem at hand is to find the values of x(h) (h=1,2,..., H) which minimize C subject to:

(a) $V(k) \leqq V^*(k)$ (k=1,2,..., K)

and

(b) $x(h) \geqq 0$ (h=1,2,..., H)

where V*(h) is a positive constraint on the variance of statistic-k.

Several approximate solution methods for this problem are described by Cochran (1977). In addition, numerical solution methods have been given by Hartley and Hocking (1963). Chatterjee (1966) Zukhovitsky and Adeyeva (1966), and Huddleston et al (1970). The solution method used in this report is a numerical solution developed by Chromy (1970). The algorithm is written in BASIC and operates interactively on the HP2000 computer.

7. NUMERICAL RESULTS

Estimates of the variance model parameters were obtained using data from the NAEP Year 11 Assessment. For package level statistics, separate components were estimated for each of the eight strata. This led to a stratified variance model which allows for a separate allocation of resources to each stratum. The data could not support the estimation of stratified models for the cross-package means. For these statistics, a single set of national level components were estimated.

The remainder of this section will be broken into two parts--one for within-package statistics and the other for cross-package statistics.

Within-Package Analyses.

Variance models were estimated for each of the six domains in Table 4-3 for the 15 scores in Table 4-1. This yielded 90 estimated models each with 24 components (3 levels by 8 strata).

The variance components in these models are estimated from the NAEP sample data and, hence, are subject to sampling variation. In an attempt to smooth out this variability, groups of components which were expected to be of comparable size were identified and smoothed estimates obtained. The components were grouped by age, stratum and domain. Within an age-stratum-domain group let $\sigma_i^2(j)$ be the variance component for the ith score for the jth level of the design (j=1,2,3 for PSUs, replicates and students, respectively) and let n be the number of scores in the group. Define

$$\delta_i(j) = \sigma_i^2(j)/\sigma_i^2(+) \tag{7.1}$$

where

$$\sigma_i^2(+) = \sum_j \sigma_i^2(j). \tag{7.2}$$

Also define

$$\bar{\delta}_+(j) = \sum_i \delta_i(j)/n \tag{7.3}$$

and

$$\bar{\sigma}^2_+(+) = \sum_i \sigma^2_i(+)/n \tag{7.4}$$

Finally, the smoothed components were calculated as

$$\tilde{\sigma}^2(j) = \bar{\delta}_+(j)\bar{\sigma}^2_+(+) \tag{7.5}$$

Within an age-stratum-domain group, this process estimates an average total variation ($\bar{\sigma}^2_+(+)$) and an average proportion of variation attributable to the jth stage of sampling ($\bar{\delta}_+(j)$). The product of these two estimates estimates the stage-j variance component for the group. This approach was taken, rather than directly averaging the components, since the total variation and the proportions of variation were expected to exhibit greater stability.

The smoothing process resulted in a separate variance model for each domain within each age group for a total of 18 estimated variance models each with 24 levels (8 strata by 3 sampling stages).

Once all the models have been parameterized, it becomes necessary to select appropriate variance constraints. In this situation it is often convenient to work in terms of relative variances and standard errors. To see this, recall that a 95 percent normal theory confidence interval (C.I.) for a mean $\bar{Y}$ is approximately

$$\text{C.I.} = \bar{Y} \pm 2SE(\bar{Y}) \tag{7.6}$$

where $SE(\bar{Y})$ is the standard error of the mean. A common precision constraint is to require that the half-width of the confidence interval

be less than some multiple of the mean, say $\alpha\bar{Y}$. Thus, it is required that

$$C.I. = \bar{Y} \pm \alpha\bar{Y}. \tag{7.7}$$

This implies that

$$\alpha\bar{Y} = 2SE(\bar{Y}) \tag{7.8}$$

or

$$\alpha = 2SE(\bar{Y})/\bar{Y} \tag{7.9}$$

$$= 2RSE(\bar{Y}).$$

where $RSE(\bar{Y})$ is the relative standard error of the mean. Hence, the expected half-width of a 95 percent confidence interval relative to the mean is twice the relative standard error of the mean.

To take advantage of the relationship between the relative standard error (RSE) and the expected width of a confidence interval in setting variance constraints, the general variance model shown in (6.2) can be recast by dividing through by the squared mean to yield

$$V(k)/\bar{Y}^2(k) = \sum_{h=1}^{H} V(kh)/\bar{Y}^2(k)x(h) \tag{7.10}$$

or

$$RV(k) = \sum_{h=1}^{H} RV(kh)/x(h) \qquad (k=1,2,\ldots, K) \tag{7.11}$$

where

$$RV(k) = V(k)/\bar{Y}^2(k) \tag{7.12}$$

= the relative variance of statistic-k

and

$$RV(kh) = V(kh)/\bar{Y}^2(k) \tag{7.13}$$

= the relative variance component for stratum-h statistic-k.

Thus, the constraints take the form

$$RV(k) \leq [RSE^*(k)]^2 \tag{7.14}$$

or

$$\sum_{h=1}^{H} RV(kh)/x(h) \leq [RSE^*(k)]^2 \quad (k=1,2,\ldots, K) \tag{7.15}$$

where RSE*(k) is the relative standard error constraint on statistic-k. This transformation was applied to all the variance models in this study:

Optimal NAEP sample designs were obtained assuming the cost model shown in Table 3-3 with package assignments of 11, 15, and 14 to the three ordered age groups (the Year 11 cost model) and assuming the 18 smoothed within-package variance models. Table 7-1 presents the optimal design under global 10 percent relative standard error constraints. This table can be contrasted with Table 7-2 which presents the optimal design for five percent RSE constraints. The main body of each table presents, for each stratum, the number of PSU's, replicates (Rep's) per PSU and students per replicate. Note the substantial difference in the resources required for these two designs induced by the change in the constraints.

In an attempt to identify reasonable precision constraints, the percent RSE's for within-package means projected by the variance models are presented in Table 7-3 for a hypothetical sample of 10 PSU's per stratum, two replicates per PSU and 15 students per replicate. This implies a total sample size of 2,400 students per package. Table 7-3 presents the best precision possible under the above hypothetical design. Notice, that all of the projected RSE's are less than the 10 percent constraints used to prepare Table 7-1, while several are greater than

Table 7-1. NAEP Design Optimization for Within-Package Means

Region	Urban PSU's	Urban Rep's	Urban Students	Rural PSU's	Rural Rep's	Rural Students
NE	2.54	0.61	127.54	0.78	1.37	52.03
S	1.45	2.23	53.78	3.15	1.85	38.46
NC	4.67	2.18	22.48	2.05	0.66	55.57
W	3.91	0.86	82.58	2.21	1.74	28.69

20.76 Total PSU's

30.41 Total Rep's

1,343 Total Students/Package

$363,987 Variable Cost

Notes:
Year 11 cost model
10% RSE constraints

Table 7-2. NAEP Design Optimization for Within-Package Means

Region	Urban PSU's	Urban Rep's	Urban Students	Rural PSU's	Rural Rep's	Rural Students
NE	10.17	0.61	127.62	3.11	1.38	51.91
S	5.81	2.23	53.82	12.61	1.85	38.44
NC	18.70	2.18	22.47	8.20	0.66	55.46
W	15.63	0.86	82.52	8.85	1.73	28.71

83.08 Total PSU's

121.68 Total Rep's

5,371 Total Students/Package

$1,455,950 Variable Cost

Notes:
Year 11 cost model
5% RSE constraints

the Table 7-2 five percent RSE constraints. Review of Table 7-3 lead to the formation of the specially selected set of RSE constrints exhibited in Table 7-4.

A third optimization for within-package statistics is reported on in Table 7-5. This design was derived from the Year 11 cost model and the selected precision constraints in Table 7-4. Notice that this design is very similar in the total numbers of PSU's, replicates, and students to the hypothetical sample design used to generate the precision constraints in Table 7-4. On the other hand, the optimal design differs markedly in its allocation of resources to the various strata and stages from that of the hypothetical design. The design indicated in Table 7-5 demonstrates how the available resources should be allocated to minimize the cost of the survey while still meeting the designated requirements.

Two additional optimations were calculated and are presented in Tables 7-6 and 7-7. Both of these optimizations were constrainted as shown in Table 7-4. The difference between them is that the former assumed the six packages per age group cost model and the latter the three package per age group cost models (see Table 3-3). Comparison of Tables 7-5, 7-6, and 7-7 indicates that reducing the number of package reduces the cost of the optimal sample design. In addition, most of the cost saving comes about through a reduction in the number of PSU's with the sample sizes of the other two stages increasing. The reason for this becomes readily apparent when the three cost models in Table 3-3 are compared. Notice that reducing the number of packages per age group leaves the cost per PSU unchanged while substantially reducing the cost per replicate and per student. Hence, reducing the number of packages

Table 7-3. Projected Percent RSE's for Within-Package Means Assuming 10 PSU's/Stratum, 2 Rep's/PSU and 15 Students/Rep.

Domain	Age 9	Age 13	Age 17
National	2.13	2.42	1.36
Non-White	5.42	5.04	3.67
Male	2.99	2.96	1.75
Female	2.41	2.77	1.58
Parents Education < High School	7.03	6.10	3.33
Parents Education ≧ High School	2.27	2.20	1.28

Table 7-4. Percent RSE Constraints for Within-Package Means

Domain	Age 9	Age 13	Age 17
National	2.00	2.50	2.00
Non-White	5.50	5.00	4.00
Male	3.00	3.00	2.00
Female	2.50	2.50	2.00
Parents Education < High School	7.00	6.00	3.50
Parents Education ≧ High School	2.25	2.00	2.00

Table 7-5. NAEP Design Optimization for Within-Package Means

Region	Urban PSU's	Urban Rep's	Urban Students	Rural PSU's	Rural Rep's	Rural Students
NE	17.70	1.30	23.72	3.19	0.64	45.53
S	10.87	1.36	26.07	7.17	1.35	28.43
NC	9.92	2.56	21.75	2.84	3.89	18.72
W	19.37	1.29	25.75	2.11	5.23	23.42

73.77 Total PSU's

122.34 Total Rep's

2,977 Total Students/Package

$1,160,640 Variable Cost

Notes:
Year 11 cost model
Table 7-4 constraint set

Table 7-6. NAEP Design Optimization for Within-Package Means

Region	Urban PSU's	Urban Rep's	Urban Students	Rural PSU's	Rural Rep's	Rural Students
NE	14.51	1.64	25.50	2.70	0.99	46.45
S	7.70	2.05	27.53	6.04	1.57	31.31
NC	7.51	3.62	23.09	2.19	6.00	17.70
W	16.97	1.40	29.05	1.21	10.50	23.45

58.83 Total PSU's

128.59 Total Rep's

3,312 Total Students/Package

$637,925 Variable Cost

Notes:
Cost model assuming 6 packages/age group
Table 7-4 constraint set

Table 7-7. NAEP Design Optimization for Within-Package Means

Region	Urban PSU's	Urban Rep's	Urban Students	Rural PSU's	Rural Rep's	Rural Students
NE	11.94	2.07	27.89	2.05	1.59	42.63
S	5.89	2.86	29.19	5.17	2.11	31.02
NC	5.74	5.15	24.10	1.78	8.33	17.67
W	14.59	1.69	31.68	1.11	12.53	24.70

48.27 Total PSU's

138.70 Total Rep's

3,758 Total Students/Package

$407,736 Variable Cost

Notes:

Cost model assuming 3 packages/age group
Table 7-4 constraint set

makes it less expensive to select more replicates, more students, and fewer PSU's to meet the same variance constraints.

Cross-Package Analyses.

As was noted previously, it was not possible to estimate all of the parameters in the fully stratified variance model for cross-package means. For this reason, a non-stratified national three stage variance model will be used here. The three stages correspond to PSU's, replicates within PSU's and students within replicates. As was done for the within-package analyses, the variance models were placed on a relative scale by dividing through by the squared mean.

All of the design optimizations for cross-package statistics are presented in Table 7-8. The sequence of optimizations proceeds similarly to the within-package analyses. First, two design optimizations were performed assuming the Year 11 cost model (see Table 3-3) and either global ten percent relative standard error (RSE) constraints or five percent RSE constraints. These two constraint sets led to designs that differed marked in cost and number of PSU's. However, the number of replicates per PSU and students per replicate were left unchanged.

The shape of the constraint space was explored by considering the projected RSE's in Table 7-9 for a hypothetical sample design consisting of 50 PSU's, 2 replicates per PSU, and 15 students per replicate. This table indicates that the ten percent RSE constraints were generally too loose, while the five percent RSE constraints were often too tight. This led to the constraint set presented in Table 7-10.

The remaining three optimizations presented in Table 7-8 were calculated using the constraints in Table 7-10. These three designs were derived from the three cost models in Table 3-3. Surprisingly,

Table 7-8. NAEP Design Optimizations for Cross-Package Means

Cost Model & Constraints	PSU's	Rep's/ PSU	Students/ Rep's	Total Rep's	Total Students/ Package	Variable Cost
Year 11 cost model						
10% RSE's	26.67	4.03	9.84	107.47	1,058	$727;364
5% RSE's	106.69	4.03	9.84	429.88	4,231	2,909,460
Table 7-9 Constraints	48.39	2.37	16.02	114.90	1,841	918,514
Six packages/age group cost model						
Table 7-9 Constraints	48.19	2.37	16.47	114.30	1,883	500,682
Three packages/age group cost model						
Table 7-9 Constraints	47.50	2.37	18.19	112.46	2,045	330,759

Table 7-9. Projected Percent RSE's for Cross-Package Means Assuming 50 PSU's, 2 Rep's/PSU, and 15 Students/Rep.

Domain	Age 9 Subobjective a	Age 9 Subobjective b	Age 13 Subobjective a	Age 13 Subobjective b	Age 17 Subobjective a	Age 17 Subobjective b
National	3.14	2.81	2.45	1.36	1.24	1.08
Non-White	10.07	7.58	6.53	3.05	2.88	2.11
Male	1.81	7.06	2.92	2.51	1.68	1.39
Female	4.30	2.32	2.38	1.29	1.32	1.14
Parents Education < High School	7.50	4.49	4.35	3.31	2.63	2.27
Parents Education ≧ High School	3.73	2.41	1.96	1.22	1.08	0.98

Table 7-10. Percent RSE Constraints for Cross-Package Means

Domain	Age 9 Subobjective a	Age 9 Subobjective b	Age 13 Subobjective a	Age 13 Subobjective b	Age 17 Subobjective a	Age 17 Subobjective b
National	3.00	3.00	2.50	1.50	1.25	1.00
Non-white	10.00	7.50	6.50	3.00	3.00	2.25
Male	2.00	7.00	3.00	2.50	1.75	1.50
Female	4.50	2.50	2.50	1.50	1.50	1.25
Parents Educ. < High School	7.50	4.50	4.50	3.50	2.75	2.25
Parents Educ. ≧ High School	3.75	2.50	2.00	1.25	1.00	1.00

these three optimal designs are virtually the same despite the fact that the three cost models differ substantially.

8. CONCLUSIONS

In general this study produced results consistent with earlier studies and tends to confirm the current NAEP design.

Numerical solutions for within-package means tended to produce student session sizes which were much larger than the range covered by the linear cost model. It may also be noted that reducing the number of packages has the effect of reducing the optimal number of PSU's. This may be difficult to implement while still retaining the ability to produce reliable data for geographical and type of community domains.

The analysis for cross-package means was based on a fairly small number of means and bears repeating on a larger scale. Analytical solutions for group session sizes conform more closely to those of the present design.

REFERENCES

Chatterjee, S. (1966). A programming algorithm and its statistical applications. O.N.R. Tech. Rept. 1, Department of Statistics, Harvard University, Cambridge.

Chromy, J. R., "Cost Minimization With Multiple Constraints." RTI internal documentation, 1978.

Cochran, W. G., Sampling Techniques, third edition. John Wiley and Sons, 1977.

Hartley, H. O., and Hocking, R. (1963). Convexing programming by tangential approcimation. Manaagement Science, 9, 600-612.

Huddleston, H. F., Claypool, P. L., and Hocking, R. R. (1970). Optimum sample allocation to strata using convex programming. App. Stat. 19, 273-278.

Shah, B. V., R. E. Folsom, C. A. Clayton, Efficiency Study of Year-03 In-School Design. RTI final report prepared for National Assessment of Educational Programs, November 1973.

Sherdon, A. W., R. E. Folsom, A. F. Clemmer, A Study of Optimum Designs for an In-School Assessment. RTI final report prepared for National Assessment of Educational Progress, December 1977.

Zukhovitsky, S. I., and Avdeyeva, L. I. (1966). Linear and Convex Programming. W. B. Saunders, Philadelphia.

APPENDIX A

COVARIANCE COMPONENTS FOR TWO ESTIMATED TOTALS

APPENDIX A

For the derivation of the covariance between two package total estimates, consider two totals, say $\hat{u}$ and $\hat{v}$, from two distinct packages. Explicitly,

$$\hat{u} = \sum_{i=1}^{n_1} [n_1 P_i]^{-1} \sum_{j=1}^{n_2} [n_2 P_{j(i)}]^{-1} \sum_{k=1}^{n_3} u_{ijk} N_{ij}/n_3 \qquad (A.1)$$

likewise for $\hat{v}$. Recall that

$$\mathrm{Cov}(\hat{u},\hat{v}) = E_1[\mathrm{Cov}(\hat{u},\hat{v})|1)] + \mathrm{Cov}_1[E(\hat{u}|1,2),E(\hat{v}|1,2)|1] \qquad (A.2)$$

and

$$\mathrm{Cov}(\hat{u},\hat{v})|1) = E_2[\mathrm{Cov}(\hat{u},\hat{v}|1,2)|1] + \mathrm{Cov}_2[E(\hat{u}|1,2),E(\hat{v}|1,2)|1] \qquad (A.3)$$

where the numerals 1 and 2 indicate which stage of sampling the expectation is being taken over or conditioned on. Note that

$$\mathrm{Cov}(\hat{u},\hat{v}|1,2) \doteq 0 \qquad (A.6)$$

since non-overlapping simple random student samples are selected within schools for each package. Thus,

$$E_2[\mathrm{Cov}(\hat{u},\hat{v}|1,2|1] \doteq 0 \qquad (A.5)$$

Now consider,

$$E(\hat{u}|1,2) = \sum_{i=1}^{n_1} [n_1 P_i]^{-1} \sum_{j=1}^{n_2} [n_2 P_{j(i)}]^{-1} E\left(\sum_{k=1}^{n_3} u_{ijk} N_{ij}/n_3 | 1,2\right)$$

$$= \sum_{i=1}^{n_1} [n_1 P_i]^{-1} \sum_{j=1}^{n_2} [n_2 P_{j(i)}]^{-1} U_{ij+} \qquad (A.6)$$

where

$$U_{ij+} = \sum_{k=1}^{N_{ij}} U_{ijk}$$

= the population total for school-ij

Likewise,

$$E(\hat{v}|1,2) = \sum_{i=1}^{n_1} [n_1P_i]^{-1} \sum_{j=1}^{n_2} [n_2P_{j(i)}]^{-1} V_{ij+} \tag{A.7}$$

Note that,

$$E(\hat{u}|1,2) = \sum_{i=1}^{n_1} [n_1P_i]^{-1} \sum_{j=1}^{tn_2} [n_2P_{j(i)}]^{-1} U_{ij+}$$

$$+ \sum_{j=tn_2+1}^{n_2} [n_2P_{j(i)}]^{-1} U_{ij+} \tag{A.8}$$

and

$$E(\hat{v}|1,2) = \sum_{i=1}^{n_1} [n_1P_i]^{-1} \sum_{j=1}^{tn_2} [n_2P_{j(i)}]^{-1} V_{ij+}$$

$$+ \sum_{j=tn_2+1}^{n_2} [n_2P_{j(i)}]^{-1} V_{ij+} \tag{A.9}$$

where t is the proportion of sample schools where the two packages are jointly administered. Also, the summations over schools in equations (A.8) and (A.9) are likewise decomposed. Assuming that the schools where both packages are administered constitute a simple random subsample of the schools leads to

$$\mathrm{Cov}_2[E(\hat{u}|1,2),E(\hat{v}|1,2)|1]$$

$$= \sum_{i=1}^{n_1} [n_1P_i]^{-2} \mathrm{Cov}_2\left[\sum_{j=1}^{tn_2} [n_2P_{j(i)}]^{-1} U_{ij+}, \sum_{j=1}^{tn_2} [n_2P_{j(i)}]^{-1} V_{ij+}\right]$$

$$= \sum_{i=1}^{n_1} [n_1P_i]^{-2} (t/n_2) \sum_{j=1}^{N_i} P_{j(i)} \{[U_{ij+}/P_{j(i)}] - U_{i++}\}$$

$$\{[V_{ij+}/P_{j(i)}] - V_{i++}\}$$

$$= \sum_{i=1}^{n_1} [n_1 P_i]^{-1} (t/n_2) \sigma_{uv_i}(2) \tag{A.10}$$

where

$$U_{i++} = \sum_{j=1}^{N_i} U_{ij+}$$

$$V_{i++} = \sum_{j=1}^{N_i} V_{i++}$$

and

$$\sigma_{uv_i}(2) = \sum_{j=1}^{N_i} P_{j(i)} \{[U_{ij+}/P_{j(i)}] - U_{i++}\} \{[V_{ij+}/P_{j(i)}] - V_{i++}\} \tag{A.11}$$

Combining (A.3), (A.6) and (A.10) yields

$$\mathrm{Cov}(\hat{u},\hat{v}|1) = \sum_{i=1}^{n_1} [n_1 P_i]^{-2} (t/n_2) \sigma_{uv_i}(2) \tag{A.12}$$

Thus,

$$E_1[\mathrm{Cov}(\hat{u},\hat{v}|1) = (t/n_2) E_1[\sum_{i=1}^{n_1} (n_1 P_i)^{-2} \sigma_{uv_i}(2)]$$

$$= (t/n_1 n_2) \sum_{i=1}^{N} \sigma_{uv_i}(2)/P_i$$

$$= t\sigma_{uv}(2)/n_1 n_2 \tag{A.13}$$

where

$$\sigma_{uv}(2) = \sum_{i=1}^{N} \sigma_{uv_i}(2)/P_i \tag{A.14}$$

Next, note that

$$E(\hat{u}|1) = \sum_{i=1}^{n_1} [n_1 P_i]^{-1} U_{i++} \tag{A.15}$$

and

$$E(\hat{v}|1) = \sum_{i=1}^{n_1} [n_1 P_i]^{-1} V_{i++}$$

Thus,

$$\mathrm{Cov}_1[E(\hat{u}|1), E(\hat{v}|1)] = \mathrm{Cov}_1\{\sum_{i=1}^{n_1} [n_1 P_i]^{-1} U_{i++}, \sum_{i=1}^{n_1} [n_1 P_i]^{-1} V_{i++}\}$$

$$= \sum_{i=1}^{N} P_i[(U_{i++}/P_i) - U_{+++}]\,[(V_{i++}/P_i) - V_{+++}]/n_1$$

$$= \sigma_{uv}(1)/n_1 \qquad (A.16)$$

Combining (A.2), (A.13) and (A.16) yields

$$\mathrm{Cov}(\hat{u},\hat{v}) = \sigma_{uv}(1)/n_1 + t\sigma_{uv}(2)/n_1 n_2 \qquad (A.17)$$

which is the final desired representation.