NATIONAL CENTER FOR EDUCATION STATISTICS

Working Paper Series

The Working Paper Series was created in order to preserve the information contained in these documents and to promote the sharing of valuable work experience and knowledge. However, these documents were prepared under different formats and did not undergo vigorous NCES publication review and editing prior to their inclusion in the series.

U. S. Department of Education
Office of Educational Research and Improvement

NATIONAL CENTER FOR EDUCATION STATISTICS

Working Paper Series

Schools and Staffing Survey (SASS) Papers Presented at Meetings of The American Statistical Association

Working Paper No. 94-01 July 1994

Contact: Dan Kasprzyk
Special Surveys and Analysis Branch
(202) 219-1588

U. S. Department of Education
Office of Educational Research and Improvement

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Sharon P. Robinson
Assistant Secretary

National Center for Education Statistics
Emerson J. Elliott
Commissioner

National Center for Education Statistics

"The purpose of the Center shall be to collect, analyze, and disseminate statistics and other data related to education in the United States and in other nations."—Section 406(b) of the General Education Provisions Act, as amended (20 U.S.C. 1221e–1).

July 1994

Foreword

Each year a large number of written documents are generated by NCES staff and individuals commissioned by NCES which provide preliminary analyses of survey results and address technical, methodological, and evaluation issues. Even though they are not formally published, these documents reflect a tremendous amount of unique expertise, knowledge, and experience.

The ***Working Paper Series*** was created in order to preserve the valuable information contained in these documents and to promote the sharing of valuable work experience and knowledge. However, these documents were prepared under different formats and did not undergo vigorous NCES publication review and editing prior to their inclusion in the series. Consequently, we encourage users of the series to consult the individual authors for citations.

To receive information about submitting manuscripts or obtaining copies of the series, please contact Suellen Mauchamer at (202) 219-1828 or U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, 555 New Jersey Ave., N.W., Room 400, Washington, D.C. 20208-5652.

Susan W. Ahmed
Acting Associate Commissioner
Statistical Standards and
Methodology Division

Samuel S. Peng
Branch Chief
Statistical Services and
Methodological Research Branch

Table of Contents

Papers Presented at the Meeting of the American Statistical Association: Section on Survey Research Methods, August 1993

Papers Presented at the Meeting of the American Statistical Association: Social Statistics Section, August 1993

Acknowledgment

This first issue of the **Working Paper Series** is dedicated to Mr. Roger A. Herriot, the late Associate Commissioner for Statistical Standards and Methodology, who passed away in April, 1994. He initiated the discussion which led to the creation of this **Working Paper Series**. We are grateful to him for his vision, leadership, and support.

The Schools and Staffing Survey: Research Issues

Daniel Kasprzyk
National Center for Education Statistics

Introduction

Until relatively recently, data on the workplace and main workforce of our education system, schools, teachers, and administrators, have not been available on a regular basis. The Schools and Staffing Survey(SASS) was designed to meet the need for information on the characteristics and experience of teachers and administrators, to describe the essential features of the school as a place to work and a place to learn, and to provide data on aspects of teacher supply and demand, and attrition. The SASS was first fielded in the 1987-88 school year, was repeated in the 1990-91 school year, and is intended to be conducted every three years.

The SASS is a complex undertaking, both in design and in implementation. Because of its complexity and the fact that it is a regular recurring program of the National Center for Education Statistics(NCES), the SASS staff realize that opportunities exist to learn from the successes and failures of each collection cycle. These lessons provide the opportunity to make improvements to the program as well as gather information on the quality of the survey's data products.

This paper is intended to serve two purposes: 1) to provide a brief overview of the SASS program; and 2) to identify areas of research or survey evaluation in which work is ongoing or planned.

Overview of the SASS

The SASS is an integrated system of surveys of public and private schools, school districts,school principals and administrators, and teachers. The data collection program consists of seven mail-out/mail-back surveys implemented during one school year, followed by a mail survey of a subsample of teachers one year later. These surveys include:

1. a survey of public school districts(local education agencies);
2. a survey of schools, public and private;
3. a survey of school administrators/principals in the public and private sectors;
4. a survey of teachers in the public and private sectors.

In the school year following the implementation of the SASS, a subsample of teachers in the SASS teacher survey are selected to be included in the SASS Teacher Followup Survey. This mail survey, a survey of public and private school teachers, was designed to provide information on teacher attrition and retention in public and private schools.

The SASS program has also included a research sample as part of its ongoing operations, thus providing opportunities to study questionnaire design, content, operational and survey methods issues in the context of a large scale operation.

The SASS is a broad multi-dimensional program, guided by four principal objectives:

1. to provide data on the components of teacher supply and demand, shortages and turnover, and the policies and practices influencing supply and demand.
2. to provide data on the principal/administrator workforce, including demographics and economic characteristics as well as their academic background, qualifications, and experience.
3. to provide data on teachers, including demographic characteristics, academic background, qualifications and experience, teaching assignments, workloads, and compensation.
4. to provide data on school conditions and programs, including enrollments, staffing, organization, teaching load, problems and locus of control.

The SASS accomplishes these objectives through a design that allows the development of state and national estimates for public schools and affiliation and national estimates for private schools. Schools are the primary sampling unit, and a sample of teachers, on average between four and eight, is selected in each sample school selected. Public school districts are included in sample when one or more schools in the district are selected. The following are sample sizes for the 1990-91 SASS: 5400 local education agencies; 13,200 schools (9,900 public and 3,300 private); 13,200 school principals/administrators (9,900 public and 3,300 private); and 65,200 teachers (56,000 public and 9,200 private).

By the nature of its content and design, the SASS provides opportunities to address issues on education policy. The existing SASS data linkages among the levels of the education hierarchy - - teachers, principals, schools and school districts - - and the potential to add several more, namely students and parents, indicates the importance of the SASS in the elementary and secondary education statistics program in the National Center for Education Statistics. With this in mind, this paper presents the current operational and research issues of the program. The topics covered are: 1)frame and coverage issues; 2) questionnaire design; 3) data collection; 4) estimation and nonresponse; 5) measurement issues; and 6) evaluation of survey data.

Frame and Coverage Issues

The sampling frames for the school component of the 1990-91 SASS are the "public school universe" of the NCES' Common Core of Data(CCD) and the Private School Universe Survey(PSS). The CCD contains statistical information collected annually on all public elementary and secondary schools and school districts in the nation. State Education Agencies(SEA) compile and transmit data they collect from schools and school districts into formats defined by NCES. Information they provide includes school or district name, address, school type, enrollment and student characteristics, and the number of classroom teachers.

The Private School Universe Survey(PSS) is a data collection aimed at building an accurate and complete list of private schools in the U. S. The schools on the PSS come from a combination of private school lists and area frame searches. The PSS contains data on religious orientation, level and size of school, enrollment, number of graduates, and number of teachers employed.

CCD - SASS Differences

The 1990-91 SASS represented one of the first uses of the CCD for sampling purposes. School data in SASS were collected using the definitions established in CCD. However, an understanding of the relationship of these definitions to principals' and administrators' understandings of our concepts was limited. Furthermore, because of the time needed to edit several CCD variables and the time needed to draw and prepare the sample, the 1988-89 CCD was used to draw the 1990-91 SASS sample. Consequently, a number of schools in the SASS sample reported teacher counts and student enrollment counts that differed from the CCD file. Timing, school reorganization,CCD misreports and definition misunderstandings may play a role. A project is underway to understand these differences by characterizing the schools with discrepant information.

Evaluation of the Private School List Building

The sampling frame for private schools is a combination of list and area frame samples, the latter being necessary to compensate for the known undercoverage of the list frame. The private school list frame is composed of private schools contained on a commercially available list from Quality Education Data(QED), private school associations, state departments of education, and other sources listing private schools. Recognizing the large undercoverage of the list frame in the previous PSS(approximately 20%), the 1991-92 PSS made a substantial effort to acquire additional private school lists from the states. A project to evaluate this effort and the impact of these new sources is underway.

Evaluation of the Teacher Listing

The sampling of the teacher component of the SASS requires schools in the SASS sample to provide a list of teachers in their school along with the following information: whether new or experienced; race/ethnicity; bilingual/English as a Second Language(ESL); field of teaching. The issue for the SASS program is whether the school is filling the teacher list properly. SASS data have shown an inconsistency in the number of teachers listed by the schools during the listing operation early in the school year versus the numbers provided later in the year on the school questionnaire. A study is being developed that we expect will improve our insight as to how teacher estimates can be improved. This study will involve a reinterview of the person providing the teacher lists. Reconciliation of differences in the original and reinterview list will provide an approximate idea of the "true" number of teachers. Also, under consideration is a reinterview and reconciliation of responses from the teacher listing operation and the school questionnaire.

SASS Student Sample

In recent years, interest has grown in augmenting the SASS program with a sample of students. Because SASS is a national probability sample of schools, SASS is capable of providing a national probability sample of students distributed across elementary and secondary grades. Other NCES school-based surveys do not yield such samples, because they are oriented to one or two grade levels. This sample could lead to the study of equity issues: which students are taught by better/worse prepared teachers? Which students are participating in various programs? The statistical issues with the student sample are the development of procedures to draw a sample of students at the school and the ability to correctly calculate a probability of selection.

During the 1990-91 SASS, a research panel was fielded to address the issue of how a student sample should drawn - whether by an employee of the school or by the Census Bureau. Frazier(1992) reported on the results of this test and found that it was difficult for an untrained school employee to correctly draw the sample. This field test also showed that the questions aimed at determining the probability of selection did not work well. Thus, NCES and Census Bureau staff continue to work on this problem with a view to implementing another pretest in early 1993.

Expand Survey Coverage
The SASS target population is limited in two ways. First, schools offering only kindergarten and prekindergarten classes are defined to be out-of-scope in the SASS. This is a serious limitation given the current strong interest in policy issues related to early childhood education. Furthermore, NCES has no sampling frames that adequately cover the prekindergarten and kindergarten programs in the public and private sectors. During the next year, NCES plans to study alternative approaches to improving the coverage of prekindergarten and kindergarten programs. The adequacy of different sources of information on these programs will be reviewed and assessed. An exploration of the possibility of using the area search and area sampling approaches used in the PSS will also be reviewed. Second, the American Trust Territories(American Samoa, Guam, Virgin Islands, Northern Marianas, and Puerto Rico) are included in the CCD universe, but not in the SASS. Staff will explore the feasibility and cost of expanding SASS coverage to include the territories.

Questionnaire Design

SASS is a system of mail surveys with telephone follow-up. The surveys require teachers, principals, personnel officers, and administrative assistants to be able to respond to questions about complex concepts without much help. The complexity of the current forms and concepts offer many opportunities for the respondent to make mistakes. Several projects are now underway to address the shortcomings of the design and format of the current questionnaires.

Cognitive Research Program
Results from the reinterview program at the conclusion of the 1988 cycle of the SASS indicated a number of items required improvement(Bushery, Royce, and Kasprzyk, 1992). The program's response was two fold: 1) to commit a substantial amount of professional time to reviewing completed questionnaires) and 2) to conduct a cognitive research program in preparation for the 1990-91 SASS. Similarly, in preparation for the next cycle of SASS, a research program of detailed, probing interviews using the public school questionnaire was initiated in the fall of 1991. Jenkins(1992) reports on results of the latter study. This program of cognitive research will next focus on two questionnaires. First, the SASS student questionnaire will be reviewed and redesigned for testing in the spring of 1993. Second, the school questionnaire, already the focus of much review, will be redesigned, reformatted, and tested.

Computer Assisted Interviewing
The increasing availability of personal computers in schools and school districts suggests a potential application of computer assisted interviewing in the SASS. Since SASS is a mail survey, a suggestion has been made to consider sending diskette-based data collection instruments to schools and school districts as an alternative to the paper and pencil instruments. This data collection system's potential for improving the data collected in SASS is significant. Automated range checks, edits, logical edits, and skip pattern checks provide opportunities to clarify reported data at their source -- the respondent -- immediately upon reporting. Plans are being developed to design, implement, and evaluate an automated data collection instrument.

Teacher's Self-Report of Academic Background
The 1987-88 SASS provided evidence of teachers having difficulty reporting their academic backgrounds. In response, two views of collecting these data have been proposed - by asking for the number of credits earned in critical subject areas or by asking for the number of courses taken in each critical subject area. During the 1990-91 SASS, a small research panel, 200 schools, having a sample of 867 teachers was fielded to test these two approaches. As part of the test, we asked for a signed authorization from each teacher to allow the NCES to obtain transcripts from the schools he/she attended. Transcripts will be coded by subject area to allow comparisons to the self-reported data.

Data Collection

Several issues have arisen concerning the operation of the SASS data collection system. These issues bear on the quality of the reported data and the improvement of the timing of the availability of the SASS sample.

Data Collection Mode
SASS was designed to be primarily a mail-out/mail-back survey. Telephone followup was used for all sample units not returning the mail questionnaire. Because there is a substantial telephone followup(33% for the public schools and 46% for the private schools), there is concern about possible response bias due to the mode. Parmer, Shen, and Tan(1992) address the issue of possible response bias by mode.

Improving the Availability of the SASS Sample
Teacher sampling for the SASS requires the

development of a teacher list for each sampled school. These lists of teachers are requested from the schools in the SASS sample, checked in at the regional office for completeness, mailed to Jeffersonville where the total number of teachers of each type are keyed and transmitted to Washington. Washington then sends sampling instructions back to Jeffersonville where the specific teachers are selected. The data for the selected teachers are keyed and transmitted back to Washington. Washington then matches the teacher information to the school information to create the sample file. The sample file is then used to mail questionnaires to the teachers.

This cumbersome process has an obvious and direct bearing on the timeliness of the availability of the teacher sample. A working group has been established to study the potential for improving the efficiency of this operation.

Data Comparability Project

Response burden is a concern for all federal data collections. The hierarchical design of the SASS and a number of individual items, particularly as they relate to school district staffing, have proven burdensome to SASS respondents. In response to the reactions of several school districts and in pursuit of better data collection methods, NCES developed a project to test whether state education agencies have the capacity to provide data from their automated record systems that would otherwise be collected in SASS from local education agencies. How comparable are the data available from the state education agencies to the data collected in the SASS from the individual local education agencies? Blank(1992) reports on the results of this project and its direction in the future. Successful collection of district level staffing items from a state automated record system would lead to major rethinking of SASS data collection methods for the teacher demand and shortage survey.

Estimation and Nonresponse

While unit response rates in the SASS are quite good, nonresponse remains a concern because of the hierarchical nature of the SASS design. Principals may often act as gate-keepers for the teachers in sample by not providing lists of the teachers in their schools; principals may complete the principal/administrator form but not the school form. Districts may also serve as gate-keepers for their schools.

Nonresponse

A project to investigate the characteristics of nonrespondents in the 1990-91 SASS is under development. Characteristics of SASS units - districts, schools, principals/administrators, and teachers - respondents and nonrespondents will be compared across many dimensions with a view to providing an understanding of nonresponse in the SASS.

Work has also begun on assessing the nonresponse adjustment cells chosen for the SASS school survey and the associated cell-collapsing strategy. These cells had been selected based on intuitive analytic judgement. This study, however, is intended to quantify these judgments and propose alternatives if necessary. Some preliminary results for the school component of the SASS are available in Shen, Parmer,and Tan (1992). A replication of the study on other SASS analytic units - principals/administrators, school districts, and teachers - is also desirable.

Plans are underway to increase the number of followups in the Teacher Followup Survey component of the SASS, thereby making this component a multiwave longitudinal study of teachers. Even though item nonresponse is relatively small in this survey, research on imputation methods that use previously collected data is desirable.

Variance Estimation

In SASS, the sampling unit is the school. School districts are brought into the sample because a school in the district has been selected in the SASS sample. Thus, the school district collection unit is an aggregation of schools(the sampling unit) belonging to the district. Kaufman(1992) addresses the issue of how well balanced half-sample replication methods estimate variances when the collection unit (school districts) is an aggregation of sampling units (schools).

Generalized variance models provide data users an easy way of obtaining variance estimates for complex sample surveys. A project is under way to develop generalized variance models for each component of the SASS.

Measurement Issues

Much attention has been given to resources and students as the principal measures of institutional improvement -- expenditures per pupil, for example. Statistics such as these, however, provide little information about school quality or the quality of the educational experience in schools. To remedy this situation, a number of ideas, as discussed in the report of a Panel on Education Indicators (1991), will be developed as potential measures of educational experience and institutional quality. A series of research projects, field studies, and feasibility studies will be developed during the next several years. The research will be a combination of conceptual research (appropriate measures), item and content research (the possible ways the measures can be implemented), and field and validation research (do these measures work in practice and do they work in large scale surveys). Originally conceived as a component of the SASS, the direction of the research may either lead to a new survey or significant modifications to an existing survey.

Evaluation of Survey Data

Evaluation of the quality of survey data can take several approaches. Microdata studies which evaluate the quality of the individual response, field performance statistics, experiments to test competing methodologies, and macrocomparisons with other established data sources are all used to establish the validity of a data set. SASS has several projects, ongoing and in the planning stages, which will bring information to bear on the quality of the SASS data.

SASS Reinterview Program

SASS has fielded a reinterview program in each cycle. Each reinterview was aimed at measuring simple response variance, a measure of the inconsistency between responses over repeated applications of the question. Thus, the purpose of the reinterview was to identify questions needing improvement in the next cycle of SASS. Bushery, Royce, and Kasprzyk(1992) describe results from the reinterview programs and show how these programs can be a tool for identifying problem items in a questionnaire. While the Bushery et al paper shows results from a limited number of questions in both the 1987-88 and 1990-91 SASS, an analysis of the quality of substantially more items is available in an internal memorandum(Royce, 1992).

The 1991-92 Teacher Followup Survey(TFS) component of the SASS also conducted a reinterview to ascertain the quality of individual items. This reinterview program, however, featured the use of a probing, reconciled interviews to improve the reinterview's diagnostic power. In this case, we expect to obtain information not only on questions that are unreliable, but also on the reasons for the inconsistency in responses.

Comparing Estimates across Forms

The SASS obtains the same or similar data across several survey forms. Thus, for example, it is possible to obtain rates of attrition and separation not only in the SASS but also in the TFS. National counts of teachers are available from the teacher, school, and district forms. Estimates of the number of certified teachers are available from both the teacher and district questionnaires. The relationship between these seemingly equivalent estimates is not well understood. During the next year, a project will begin to identify all estimates of the same phenomena across the different components, to quantify the differences if they exist, and to try to understand the reasons for the differences.

Evaluating Self-Reports of Urbanicity

In both the 1987-88 and 1990-91 cycles of SASS, the question, "Which best describes the community in which this school is located?" was asked of the principal(for the administrator/principal survey) and the respondent to the school survey. The response categories were given as rural, small city or town, medium-sized city, suburb of a medium city. etc. These reports are highly subjective and have exhibited moderate response variance as determined through the reinterview program(Bushery et al, 1992). Recently Johnson(1989) developed a methodology for assigning "type of locale" codes based on the school mailing address being matched to Bureau of the Census data files containing population and population density data, Standard Metropolitan Statistical Area(SMSA) codes, and a Census code defining urban and rural areas. A study is in progress to compare codes derived through the Johnson algorithm to the self-reported classifications found in the two cycles of SASS. These comparisons will give us a better understanding of this survey item. Since the self-report method is used on many NCES surveys, the results of this study have wider applicability than the SASS.

SASS Quality Profile

Work on developing a quality profile for the SASS has begun. The quality profile will summarize methodological and evaluation research related to the SASS and will provide an overview of procedures for all phases of the survey - sample selection, data collection, data processing and estimation. It is intended to provide an overview of what is known about the sources and magnitude of errors in the SASS, and thus a sourcebook of information on the quality of the SASS data.

SASS User Survey

In designing the SASS approximately six years ago, NCES anticipated a variety of users - education planners, policymakers, managers, government analysts, and academic researchers. By the end of 1992, SASS will have released several major reports, a number of E.D. Tabs, public use microdata tapes and CD ROMs, and restricted access data tapes. In February 1992, the SASS Review Board, a working group of researchers interested in the use and evaluation of the SASS, suggested the need for a SASS user survey. The purpose of the survey would be to identify uses/users of the SASS and to assess whether they are consistent with the uses and users as identified in the design phase. The survey would also attempt to assess whether the available survey products meet user needs and how dissemination efforts could be improved. In the near future after the goals of survey are more clearly defined, we anticipate developing such a survey.

Data Analysis: Quantity and Quality of Teacher Labor Supply

This year an analytic project that focusses on estimating the effects of compensation and other policy variables on the quantity and quality of the teacher labor supply will begin. This project will address: 1) the estimation of the external labor supply facing schools and the effect of compensation and other school-level variables on this supply; 2) the estimation of internal labor supply, i.e. the retention of the teaching workforce and the effect of compensation and other policy variables on temporary and permanent flows in and out of the teaching profession. While a decidedly substantive analytic project, this project is intended to also re-evaluate the vacancy matrix data on the 1987-88 Teacher Supply and Demand Questionnaire, data that were not released on the public use tapes due to response inconsistencies.

Endnote

Any large complex data collection raises numerous questions about methods and data quality. SASS is still in its infancy in terms of understanding and use. The 1990-91 SASS included a methods research panel to help assist in answering unresolved methods issues. The 1993-94 SASS will also include such a panel. This commitment as well as the research commitment described in the projects above will provide a much deeper understanding of the SASS data as well as improve the quality of the survey operations.

REFERENCES

Blank, R. (1992), "Improving Reliability and Comparability of NCES Data on Teachers and Other Education Staff," Proceedings of the Section on Survey Research Methods, American Statistical Association.

Bushery, J., Royce, D. and Kasprzyk, D. (1992), "The Schools and Staffing Survey: How Reinterview Measures Data Quality," Proceedings of the Section on Survey Research Methods, American Statistical Association.

Frazier, R. (1992), "1991 SASS Research Study Final Report," Memorandum to the Record, October 10, 1992, Demographic Statistical Methods Divison, U.S. Buerau of the Census.

Jenkins, C. (1992), "Questionnaire Research in the Schools and Staffing Survey: A Cognitive Approach," Proceedings of the Section on Survey Research Methods, American Statistical Association.

Johnson, F. (1989), Assigning Type of Locale Codes to the 1987-1988 CCD Public School Universe, National Center for Education Statistics Technical Report, Data Series:SP-CCD-87/88-7.4, CS89-194.

Kaufman, S. (1992), "Balanced Half-Sample Replication with Aggregation Units," Proceedings of the Section on Survey Research Methods, American Statistical Association.

Panel on Education Indicators (1991), Education Counts, National Center for Education Statistics, Office of Educational Research and Improvement, U.S. Department of Education, Washington, DC.

Parmer,R., Shen, P., and Tan, A. (1992), "Mail versus Telephone Response in the 1991 Schools and Staffing Survey," Proceedings of the Section on Survey Research Methods, American Statistical Association.

Royce, D. (1992), "1991 Schools and Staffing Survey (SASS) Reinterview Response Variance Report," Demographic Statistical Methods Division, U.S. Bureau of the Census.

Shen, P., Parmer, R., and Tan, A. (1992), "Characteristics of Nonrespondents in the 1991 Schools and Staffing Surveys School Sample," Proceedings of the Section on Survey Research Methods, American Statistical Association.

THE SCHOOLS AND STAFFING SURVEY: HOW REINTERVIEW MEASURES DATA QUALITY

John M. Bushery and Daniel Royce, Bureau of the Census
and Daniel Kasprzyk, National Center for Education Statistics

KEY WORDS: Data quality, reinterview

1. INTRODUCTION

The Schools and Staffing Survey (SASS) is a good example of how a reinterview program can contribute to improved data quality by identifying questions which need improvement. We believe we have improved one aspect of SASS data quality, simple response variance – in part because the SASS reinterview program identified questions needing improvement.

The 1991 SASS reinterview results also suggest that mail respondents provide more reliable data than those interviewed in a telephone follow-up operation.

1.1 The SASS Surveys

The National Center for Education Statistics (NCES) sponsors, and the U.S. Bureau of the Census conducts, the Schools and Staffing Survey (SASS) to provide data on teachers, school administrators, schools, and local education agencies.

The SASS runs on a three-year cycle, the first in 1987-88 and the second in 1990-91. The Census Bureau conducts the SASS by mail, with telephone follow-up of cases not responding by mail.

Mail response rates range from 49 percent (for private schools) to 80 percent (for public school administrators), with final response rates between 83 (private school teachers) and 97 percent (public school administrators again). We completed one-sixth to one-third of the cases using telephone follow-up.

1.2 The SASS Reinterview Program

Two major purposes of reinterview programs are quality assurance and estimating response error [1].

The SASS reinterviews estimate simple response variance, a measure of the inconsistency between responses over repeated applications of a question. Our main goal is to identify questions needing improvement for the next cycle of SASS. We identify problematic questions in the reinterview and follow up with cognitive research and other questionnaire design techniques to make the improvements.

To estimate response variance accurately, the survey error model assumptions require the reinterview to be an independent replication of the original interview. Independence is difficult to achieve because the respondent might remember his or her answer to the original interview question. To the extent a reinterview lacks independence, response variance may be underestimated. Operational constraints often make it difficult or impossible to conduct the reinterview as an exact replication of the original interview. When a reinterview does not replicate the original interview perfectly, the differences in methodology may overstate the response variance.

The SASS reinterviews fail to replicate the original interview in two respects:

- All SASS reinterviews contained fewer questions than their original counterparts.
- The original SASS surveys used self-administered mail-return questionnaires (with telephone follow-up of non-respondents). Except for the 1991 SASS School Survey, all the reinterviews were conducted by telephone.

We conducted the Census Bureau's first-ever mail reinterview in the 1991 SASS School Survey. Some of the 1988 SASS reinterview findings suggested that for some questions, the reinterview model assumptions were not adequately met [2]. Section 2.3 discusses this topic in more detail. These results prompted us to evaluate the 1991 SASS School questionnaire through a mail reinterview.

1.3 Response Variance Measures

Response error consists of response variance and bias. The Census Bureau estimates two main metrics (from unweighted data) to quantify response variance, the **gross difference rate** and the **index of inconsistency**. In a categorical variable, one-half the gross difference rate equals the simple response variance. The gross difference rate also represents the proportion of respondents who change their answers from one interview to the next. In a question with a gross difference rate of 20 percent, one fifth of the respondents changed their answers.

The index of inconsistency is a relative measure of response variance. A simplified definition of the index is the ratio of the simple response variance to the total variance of a characteristic. The L-fold **index of inconsistency** is a weighted average of the indices over all categories in a multi-category question. An index of 50 means that half the total variance of a characteristic can be attributed to response variance. Experience provides a rough rule of thumb for interpreting the index of inconsistency. If the index is:

- less than 20, response variance is low.
- between 20 and 50, response variance is moderate.
- greater than 50, response variance is high.

High response variance means the question itself causes at least as much of the variability in the data as the variability among respondents in the population. Two reasons for high response variance are:

- The question is poorly worded and confuses the respondent.
- The information requested is too difficult for the respondent to provide.

Because the index of inconsistency estimates the ratio of two variances, the index itself has high variability. If the data don't provide enough cases in each original-by-reinterview outcome cell, a reliable estimate of the index cannot be computed.

2. REINTERVIEW RESULTS

This paper compares response variance results for questions reinterviewed in both the 1988 and 1991 cycles of SASS. Table 1 shows reinterview sample sizes and completion rates for 1988 and 1991. We used unweighted data and tested all comparisons at $\alpha = 0.10$. Tables 3 through 6 display 90 percent confidence intervals in parentheses.

The Administrator and Teacher surveys ask both attitudinal and factual questions. In 1988 the attitudinal questions we reinterviewed showed high levels of inconsistency [2]. Inconsistency in attitudinal questions may result from simple response variance or from actual changes in attitudes between the original interview and reinterview. In 1991, we decided to concentrate the reinterview on factual questions -- with the aim of improving future cycles of the SASS.

In the 1988 SASS, we could estimate the index of inconsistency reliably for 35 of the 45 factual questions we reinterviewed. We estimated the index reliably for 109 of the 126 factual questions reinterviewed in 1991 [3]. Table 2 summarizes the results of both SASS reinterviews.

Keep in mind that the distributions in Table 2 are not strictly comparable. We purposively selected different sets of questions for the two reinterviews. We evaluated 15 factual questions common to both cycles of SASS. Eleven of these questions received significant revisions in 1991. Four of the revised questions displayed reduced response variance. Our question improvement efforts have paid off, at least partially.

Table 1. SASS Reinterview Sample Sizes

	1988	1991
Administrator Survey		
Eligible for Reinterview	1309	1123
Response Rate	87%	87%
Teacher Survey		
Eligible for Reinterview	1126	1101
Response Rate	75%	74%
School Survey		
Eligible for Reinterview	1309	1123
Overall Response Rate	87%	84%
Attempted Mail Reinterview		50%
Percent Completed by Mail		43%
Attempted Telephone Reinterview *		57%
Percent Completed by Telephone		41%

* Includes 80 reinterviews not returned by mail and 85 original mail interviews returned too late for mail reinterview.

Table 2. Summary of SASS Reinterview Results *

Response Variance	1988	1991
All Three Components		
Low	4 (11%)	43 (39%)
Moderate	14 (40%)	38 (35%)
High	17 (49%)	28 (26%)
Administrator and Teacher Surveys		
Low	4 (19%)	26 (36%)
Moderate	8 (38%)	26 (36%)
High	9 (43%)	21 (29%)
School Survey		
Low	0 (0%)	17 (47%)
Moderate	6 (43%)	12 (33%)
High	8 (57%)	7 (19%)

* Questions for which index could be reliably estimated.

2.1 Administrator and Teacher Survey Results

The two Administrator questions reinterviewed in both SASS cycles ask whether the respondent earned a bachelor's degree and a master's degree. These "degree earned" questions are virtually the same as the corresponding Teacher survey questions. The results for Administrators were nearly identical to the

Teacher results shown in table 3. The 1988 question provided a list of possible degrees and asked the respondent to "mark all that apply." The 1991 question asked, "Do you have a bachelor's degree?" If "Yes," the next question asked "Do you have a master's degree?" The remaining degrees (associate, doctor's, etc.) used a "mark all that apply" approach. Table 3 suggests the direct question format produces more reliable data for degree earned.

Table 3. Teacher Survey Reinterview Results
-- Degrees Earned --

	1988	1991
Bachelor's Degree		
Percent Yes[1]	97.6	98.1
GDR *	7.5	0.6
	(6.0 - 9.2)	(0.3 - 1.3)
Index	79.5	Too few cases
	(64.2 - 98.5)	did not mention
Master's Degree		
Percent Yes[1]	41.5	41.4
GDR *	4.3	1.1
	(3.2 - 5.7)	(0.6 - 1.9)
Index *	8.9	2.2
	(6.7 - 11.8)	(1.2 - 3.9)
Professional Diploma / Ed. Specialist		
Percent Yes[1]	4.4	4.7
GDR	7.0	5.2
	(5.6 - 8.7)	(4.1 - 6.8)
Index	69.8	62.7
	(56.0 - 87.1)	(48.2 - 81.6)
Associate Degree		
Percent Yes[1]	13.7	6.7
GDR	8.1	6.9
	(6.6 - 9.9)	(5.5 - 8.6)
Index	36.9	54.2
	(30.1 - 45.3)	(43.0 - 68.2)

[1] Responded "Yes" in original interview.
* Statistically significant difference between 1988 and 1991.

In the Teacher survey in both SASS cycles we also reinterviewed questions on teaching assignment, years in teaching, and plans to remain in teaching (an attitude type question). None of these questions exhibited significantly improved response variance.

The teaching assignment questions reinterviewed in 1988 and 1991 were similar but not strictly comparable. In 1991 we reinterviewed a screener question used to identify teachers, which asked about full and part-time status and included categories for itinerant teachers, long-term substitutes, other professional staff, and administrators (the last two are out of scope for the Teacher survey). The 1988 question simply asked about full-time and four levels of part-time teaching. The 1988 question includes all full-time teachers, the 1991 figure includes only regular full-time teachers. These design differences make it difficult to compare the two questions, but response variance on the number of full-time teachers showed no significant change between 1988 and 1991. The new categories seem to cause respondents some uncertainty -- about six percent (s.e. = 0.8) of the respondents described their assignment as itinerant teacher, long-term substitute, other professional staff, or administrator in the original interview. Only three percent (s.e. = 0.6) selected one of these answers in the reinterview. The data suggest the "itinerant teacher" category is the main source of this inconsistency. It may help to define "itinerant" more clearly on future questionnaires.

The 1988 "years teaching" questions asked, "... how many years have you worked as a full-time teacher in public and/or private schools ..." (repeated for part-time) and provided a cross-tabulation for the respondent to complete:

	Years full-time	Years part-time
Public		
Private		

In 1991 we changed the format to ask four separate questions:

- " ... how many years have you worked as a full-time teacher in private ...,"
- "... part-time in private ...,"
- "... full-time in public ...," and
- "... part-time in public ..."

Table 4. Teacher Survey Reinterview Results
-- Years Teaching --

	1988	1991
Full-time, Public		
GDR	7.6	7.0
	(6.1 - 9.5)	(5.5 - 8.9)
L-fold Index	10.8	9.8
	(8.7 - 13.4)	(7.7 - 12.4)
Part-time, Public		
GDR	9.0	6.6
	(6.7 - 12.0)	(5.0 - 8.6)
L-fold Index	44.4	42.5
	(33.2 - 59.3)	(32.5 - 55.7)
Full-time, Private		
GDR	5.2	5.3
	(3.6 - 7.4)	(3.3 - 8.7)
L-fold Index	12.4	8.8
	(8.7 - 17.7)	(5.4 - 14.4)
Part-time, Private		
GDR *	3.4	7.5
	(2.1 - 5.8)	(4.8 - 11.6)
Index	38.5	37.8
	(23.0 - 64.4)	(24.4 - 58.4)

* Statistically significant difference between 1988 and 1991.

We grouped the responses into the four categories of interest to the NCES:

- less than three years,
- three to nine years,
- 10 to 20 years,
- more than 20 years.

Unfortunately, no improvement resulted. The full-time estimates enjoyed low response variance in both years, and the part-time estimates exhibited moderate response variance in both cycles of SASS (Table 4.).

The final Teacher question reinterviewed in both SASS cycles was, "How long do you plan to remain in teaching?" The consistency of this attitude-type question decreased between 1988 and 1991. The gross difference rate increased from 39.5 percent (36.8% - 42.6%) to 46.8 percent (44.0% - 49.9%) and the L-fold index increased from 55.4 (51.6 - 59.6) to 66.6 (62.6 - 71.1). Since we did not change this question, we speculate that teachers' attitudes in 1991 were less stable than in 1988.

Increased response variance among public school teachers drove the overall decrease in consistency -- private school teachers showed no significant change in response variance between 1988 and 1991

2.2 School Survey Results

In the School survey, we reinterviewed four questions in both 1988 and 1991. Although these questions were virtually unchanged between the two cycles, they showed a small but statistically significant decrease in response variance.

We think a better replication of the original interview by the reinterview in 1991 caused some of this decrease. Table 5 shows the reinterview results for these questions.

The question, "Which best describes the community in which this school is located?" contained ten categories in 1988 and 1991.

1 rural or farming community
2 small city or town, not a suburb of a large city
3 medium-sized city
4 suburb of medium city
5 large city
6 suburb of large city
7 very large city
8 suburb of very large city
9 military base or station
10 Indian reservation

The index of inconsistency for these categories ranged from 21.1 to 68.8 in 1988 and from 22.2 to 62.1 in 1991. The overall response variance (L-fold index) for this question improved slightly, but remains in the moderate range. "Community" is an important variable in the NCES' analyses. Fortunately, the NCES is now able to obtain this information from geographic data files [6], instead of asking the schools. The result will be more accurate data, with reduced respondent burden.

We reinterviewed three questions about programs offered by the school, "Which of the following programs and services are available to students in this school, either during or outside of regular school hours, and regardless of funding source -

- bilingual education
- English as a second language
- extended day or before-or-after-school day-care."

Table 5. School Survey Reinterview Results

	1988	1991
Which best describes the community in which this school is located?		
GDR *	34.7	30.4
	(32.3 - 37.1)	(27.9 - 32.9)
L-fold Index *	42.4	37.6
	(39.6 - 45.4)	(34.7 - 40.9)
bilingual education		
Percent Yes¹	15.3	14.2
GDR *	16.2	12.1
	(14.5 - 18.2)	(10.5 - 14.1)
Index	53.5	45.1
	(47.7 - 60.0)	(39.0 - 52.3)
English as a second language		
Percent Yes¹	31.6	28.3
GDR *	16.1	13.7
	(14.4 - 18.1)	(12.0 - 15.8)
Index *	37.1	30.1
	(33.1 - 41.7)	(26.3 - 34.6)
extended day or before-or-after-school day-care		
Percent Yes¹	16.3	23.0
GDR	9.3	8.8
	(7.9 - 11.0)	(7.4 - 10.6)
Index	31.7	24.7
	(26.8 - 37.4)	(20.5 - 29.7)

¹ Responded "Yes" in original interview.
* Statistically significant difference between 1988 and 1991.

2.3 Mail versus Telephone Results (1991)

In 1991 we revised the School survey reinterview procedures:

- We used a mail reinterview for mail respondents and a telephone reinterview for telephone follow-up cases.
- We requested the same respondent complete the

reinterview questions as answered the original School survey.

Both procedural changes helped the reinterview replicate the original survey better. We decided to specify the original school respondent as the reinterview respondent, because in the 1988 reinterview we inadvertently changed the reinterview's respondent-selection rules by combining the Administrator and School reinterview questionnaires. We suspect many administrators had an assistant or secretary complete the original School survey. Changing respondents between the original and reinterview tends to overstate response variance in the 1988 School survey.

We did not conduct a controlled experiment, but reinterviewed by mail whenever possible and by telephone when necessary, obtaining about 465 mail-mail cases and 270 telephone-telephone cases. This analysis covers the same four School survey questions discussed in section 2.2. Under the mail-mail procedure almost all the School questions reinterviewed in 1991, including the four in Table 6, displayed lower simple response variance than under the telephone-telephone procedure.

Table 6. Mail Original/Reinterview versus Telephone Original/Reinterview

	Mail-Mail	Telephone-Telephone
Community School Located		
GDR *	19.0	39.9
	(16.3 - 22.2)	(35.5 - 45.2)
L-fold Index *	24.0	48.6
	(20.6 - 28.2)	(43.2 - 55.1)
Bilingual Education		
GDR *	6.9	18.6
	(5.2 - 9.1)	(15.2 - 23.0)
Index *	31.5	55.3
	(23.5 - 42.0)	(45.3 - 68.2)
English as 2nd Language		
GDR *	10.9	15.7
	(8.8 - 13.6)	(12.6 - 19.8)
Index	24.2	33.5
	(19.6 - 30.1)	(26.8 - 42.3)
Extended Day Care		
GDR *	6.7	11.5
	(5.1 - 8.9)	(8.8 - 15.2)
Index *	19.7	31.9
	(14.7 - 26.4)	(24.5 - 42.2)

* Statistically significant difference between mail-mail and telephone-telephone.

We observed lower response variance in both numerical data (for example, head counts of students enrolled) and non-numerical data. Royce [3] details results for all School survey questions reinterviewed in 1991. We can think of four possible reasons for this result.

- Only respondents who answered the original survey by mail were eligible for the mail reinterview. These respondents were likely to be more cooperative and answer the questions more carefully in both interviews.
- Respondents interviewed by mail may take time to look up the answers to questions from records or they may go through a more careful, but more lengthy, thought process to provide the needed facts. In contrast, those interviewed by telephone may feel the interviewer prefers a speedy response to an accurate one, so give their "best guess-timate." Research has shown some respondents employ what survey practitioners call "satisficing." [4] In satisficing, the respondent expends just enough effort to satisfy the interviewer. Also, respondents interviewed by telephone may not feel free to take the time to look up records while the interviewer is waiting on the phone [5].
- Mail respondents may leave more difficult or uncertain questions blank. The Census Bureau's interviewers work very hard to get responses to all questions. An interviewer may manage to obtain an answer to a difficult question, but an unreliable answer. Mail respondents, on the other hand, may simply leave that question blank. We have found higher item non-response among the mail returns than in the telephone follow-up cases.
- Mail respondents may photocopy the original questionnaire after completing it and refer to their original answers when completing the mail reinterview.

We think some combination of the first three explanations is the most reasonable. Mail respondents, by definition, are more cooperative and motivated than those we must follow-up by telephone. And mail interviewing probably promotes more careful responses and more use of records.

We eliminated the last possibility. Mail respondents using photocopies of their original interviews can account for only a small part of the mail-mail versus telephone-telephone differences. We concluded that only a small fraction of the mail reinterview respondents might have used photocopies, and that these cases had little effect on the response variance differences between the two procedures. We hypothesized that respondents using photocopies would give consistent answers to all questions in the reinterview. We discarded all cases where the first 11 of the 21 reinterview questions matched. These cases accounted for only 6.5 percent of the reinterview sample and had only a negligible effect on the comparisons.

These findings on the quality of mail response data have implications beyond the SASS. Perhaps mail surveys can provide as good or better data than some surveys now conducted by telephone or in person -- and at lower cost. For the SASS, we need to determine whether the more consistent data achieved through mail results from the type of respondent who answers by mail and whether increased item non-response will cancel the gains of improved consistency.

3. PLANS FOR THE FUTURE

Reinterview programs can be a valuable diagnostic tool to identify questions which need improvement, or which perhaps should be dropped. The NCES and the Census Bureau are committed to producing accurate and reliable SASS data. They have heeded the reinterview's diagnosis and have acted to make improvements -- with some success.

What about the future? Both agencies are firmly committed to developing a first-class survey. The 1992 Teacher Follow-up Survey (TFS), which surveyed a subsample of 1991 SASS teachers, used a probing, reconciled reinterview to learn the reasons for inconsistent responses. We hope not only to identify the less reliable questions, but to gather information about why inconsistencies occur.

Plans for the future include:

- Focus at least some cognitive research on the reinterview findings.
- Consider using reconciled, probing reinterviews in the SASS to learn more about why inconsistencies occur.
- Consider expanding the mail reinterview to the Teacher and Administrator surveys.
- Apply quality assurance methods to data collection.
- Reinterview small, non-random samples to solve specific data quality problems, for example unacceptably high pre-edit rejects.
- Use reinterview methods to evaluate coverage in teacher listings (the frame of the SASS teacher sample).
- Maintain a strong commitment to a continual cycle of evaluation and improvement of SASS questionnaires, methods, and procedures.

REFERENCES

[1] Forsman, G. and Schreiner, I., "The Design and Analysis of Reinterview: An Overview," in Biemer, P. P., et al., editors, Measurement Errors in Surveys, 1991, John Wiley & Sons, Inc., pp. 280-281.

[2] Newbrough, J., "Report of the SASS-2/3(R) and SASS-4(R) Reinterview," June 28, 1989, internal Census Bureau Report.

[3] Royce, D., "1991 Schools and Staffing Survey (SASS) Reinterview Response Variance Report," July 1992, internal Census Bureau Report.

[4] Krosnick, J. A., "The Impact of Satisficing on Survey Data Quality," Proceedings of the 1990 Annual Research Conference, U.S. Bureau of the Census, March, 1990, pp. 835-845.

[5] Dillman, D. and Tarnai, J., "Mode Effects of Cognitively Designed Recall Questions: A Comparison of Answers to Telephone and Mail Surveys," in Biemer, P. P., et al., editors, Measurement Errors in Surveys, 1991, John Wiley & Sons, Inc., p. 76.

[6] Johnson, F., "Assigning Type of Locale Codes to the 1987-88 CCD Public School Universe," National Center for Education Statistics Technical Report, July 1989.

[7] Parmer, R.J., Shen, P., and Tan, A.I., "Mail Versus Telephone Response in the 1991 Schools and Staffing Surveys," Joint Statistical Meetings, Boston, 1992.

MAIL VERSUS TELEPHONE RESPONSE IN THE 1991 SCHOOLS AND STAFFING SURVEYS**

Randall J. Parmer, Pao-Sheng Shen, Andre I. Tan, U.S. Bureau of the Census
Randall J. Parmer, U.S. Bureau of the Census, Washington, D.C. 20233

I. Introduction

The 1991 Schools and Staffing Surveys (SASS) were designed to be primarily mail-out/mail-back surveys. Sample units not responding by mail are contacted as part of the telephone follow-up. Due to the high cost of conducting a telephone interview as compared to an interview conducted by mail, attempts are made to maximize the mail response rate. Mail responses alone, however, are unacceptably low due to the great potential for bias nonresponse adjustment would produce. Telephone follow-up, therefore, is necessary to increase overall survey response rates. This mixed mode of data collection, however, causes some concern about response bias due to mode.

In this paper, we shall address the issue of possible response bias as well as identify particular subgroups where mail response is low so resources may best be concentrated in improving overall mail response for the surveys. Section II describes the SASS surveys in general. Section III presents the methodology we will use to identify possible mode bias. Section IV presents the results. Section V gives our conclusions and suggestions for further research.

This paper analyzes the effect upon the data caused by mode of interview for school data only. Teacher, administrator, and public school district data could also be analyzed in the same way.

II. Background

A. General Survey Description

1. Frame Construction

The 1991 Schools and Staffing Surveys consists of a school, a teacher, and for public schools a Local Education Agency or school district survey. Public schools were identified on the Common Core of Data (CCD), a file containing all public schools in the nation, created by the National Center for Education Statistics from lists provided by the states. This CCD was matched to the previous SASS public school sampling frame. Non-matches from the previous frame were included with the CCD to make up the public school sampling frame for 1991.

The private schools were selected from a list frame, constructed by matching multiple lists obtained from private school organizations, State Departments of Education, and a private vendor. This frame is thought to include 80-90% of private schools. To increase the coverage of the survey, an area frame was constructed by selecting 120 Primary Sampling Units (PSUs), consisting of counties or groups of counties. Within these sample counties, lists of schools were obtained from local sources, such as yellow pages, churches and fire marshals. These lists were unduplicated with the list frame. The remaining schools, not matching to the list frame, make up the area frame.

2. Design

Public schools were stratified by state, grade level, and Indian/non-Indian. Probabilities of selection were computed, proportional to the square root of the number of teachers in the school conditioned on the 1988 selection. The probabilities were adjusted to obtain the desired proportion of overlapping schools from 1988. Approximately 9900 sample public schools were selected systematically within each of the 165 strata.

Private schools were stratified by 18 affiliations, 3 grade levels, and census region for the list frame, and by PSU and grade levels for the area frame. Probabilities of selection were computed and adjusted similarly to the public schools. Approximately 3300 private schools were selected, systematically within each stratum.

3. Data Collection

School questionnaires were mailed to schools. They were asked to fill them out and mail them back to the Census Bureau. After four weeks, if the school hadn't responded, we sent out a second questionnaire. If after three more weeks the school hadn't responded, we called them and attempted to complete the interview by telephone. Schools still not responding by telephone were classified as noninterviews.

4. Estimation

Schools' probabilities of selection were adjusted for school merges and other situations that would affect the probability of selection. The inverse of the probability of selection became the basic weight. This basic weight was adjusted to account for noninterviews using noninterview adjustment cells. A ratio adjustment was also applied which adjusted the characteristics of the sample schools to the characteristics of the whole sample frame.

B. Issues to be Addressed

Four issues will be addressed in our discussion of mode of interview. The first issue is what types of respondents are more likely to respond by mail. We examine this issue in order to identify certain subgroups of schools where a more concentrated effort at improving mail response rates has the greatest potential benefit, thereby lowering overall survey costs.

The second issue involves comparing response categories by mode of interview so as to identify items with mode differences. At this point, we still won't know if the response differences represent inherent differences in the types of respondents or if it represents response bias. It is merely being used as a tool to narrow down the number of items we need to look at further.

The third issue involves conducting covariance analysis on the items identified with mode differences to try to filter out inherent differences in the characteristics of the respondents and measure the difference due solely to mode of interview. Since this analysis has been done for more than one item, a rank-sum test was used to make an objective probability statement that addresses the question of whether or not there is response bias due to mode.

The fourth issue involves item nonresponse and comparing item nonresponse rates between the two modes of interview.

III. Methodology

A. Comparison of Response Categories

Responses to questionnaire items were compared using a chi-square test for independence, whereby the two modes of interview (mail, telephone) were compared across response category. Continuous variables were categorized into approximately five categories.

The usual Pearson Chi-Square test produced in SAS by PROC FREQ is inappropriate for this analysis due to the complex sample design. So, Rao and Scott's (1984) correction to the standard chi-square, which requires knowledge of the cell

** This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the author(s) and do not necessarily reflect those of the Census Bureau.

design effects, was used in our analysis. Design effects were obtained based on the estimated variance using 48 pseudo-replicates.

Comparisons were made using unweighted and weighted data. Unweighted data was analyzed as a preliminary step in this analysis. Items showing significant differences were analyzed using weighted data, adjusted by the appropriate design effect.

B. Analysis of Covariance

Regression models were fit to the data within each block constructed using the stratification variables (for example, within affiliation and grade level). Questionnaire items were treated as the dependent variables and some selected variables which were believed to be related to dependent variables and "untainted" by mode of interview were used as the covariates. The square root of the number of teachers was also included in the model to take into account the effect of the probability of selection on the covariance analysis (see Nathan and Holt (1980)). Finally, mode of interview and its corresponding interaction with the covariate were also included in the model. Our goal is to filter out the effects of inherent differences in the respondents and the effects of the design upon the responses by mode. This section describes this covariance analysis.

The mode research methodology uses a combination of parametric and nonparametric approaches:

To perform a rank-sum test, it is necessary first to express the data from different questionnaire items in common units via a transformation to relative deviate within each block. This is done by subtracting the overall mean from each observation and dividing by the within-block sample standard deviation.

Assumption:

The linear model for our study can be written as

$$Y_{ijkl} = \alpha_{ijl} + \lambda_{il} X_{ijkl} + \beta_{il} Z_{ijkl} + e_{ijkl} \qquad (1)$$

where Y_{ijkl} represent the lth variable (questionnaire items) after standardization for the kth subject (school) in block i (association x grade level) receiving the jth treatment (mail, telephone). X_{ijkl} and Z_{ijkl} are the corresponding covariate and the square root of the number of teachers and e_{ijkl} is random error.

$$E(Y_{ijkl}) = \alpha_{ijl} + \lambda_{il} X_{ijkl} + \beta_{il} Z_{ijkl} = \mu_{ijkl}$$

$$\begin{aligned}(cov(Y_{ijkl}, Y_{i'j'k'l'})) &= \sigma^2_{ll'} \ \forall\, i=i',\, k=k' \\ &= 0 \ \text{otherwise}\end{aligned}$$

In matrix notation the vector $Y_{ijk} = (Y_{ijk1}, \ldots, Y_{ijkL})'$ are independently distributed with mean $\mu_{ijk} = (\mu_{ijk1}, \ldots, \mu_{ijkL})'$ and covariance matrix:

$$\sum_{L \times L} = \begin{pmatrix} \sigma^2_{11} & \cdots & \sigma^2_{1L} \\ \vdots & & \vdots \\ \sigma^2_{L1} & \cdots & \sigma^2_{LL} \end{pmatrix}$$

The null hypothesis is:

$$Ho: \alpha_{i1} = \alpha_{i2} \ \text{for } i=1,\ldots,I \quad \text{where } \alpha_{ij} = \begin{pmatrix} \alpha_{ij1} \\ \vdots \\ \alpha_{ijL} \end{pmatrix}$$

Now, perform analysis of covariance for different variables within each block using ordinary least squares (OLS).

The adjusted mean for $\alpha_{ijl} + \lambda_{il}\bar{X}_{i..l} + \beta_{il}\bar{Z}_{i..l}$ is estimated by

$$\begin{aligned}&\hat{\alpha}_{ijl} + \hat{\lambda}_{il}\bar{X}_{i..l} + \hat{\beta}_{il}\bar{Z}_{i..l} = \\ &\bar{Y}_{ij.l} - \hat{\lambda}_{il}(\bar{X}_{ij.l} - \bar{X}_{i..l}) - \hat{\beta}_{il}(\bar{Z}_{ij.l} - \bar{Z}_{i..l})\end{aligned}$$

and denoted by $\bar{Y}^a_{ij.l}$, where $\hat{\alpha}_{ijl}$, $\hat{\lambda}_{il}$ and $\hat{\beta}_{il}$ are ordinary least square estimates within block i. Note that the difference between adjusted means of treatments 1 and 2 is

$$\begin{aligned}\bar{Y}^a_{i1.l} - \bar{Y}^a_{i2.l} &= (\bar{Y}_{i1.l} - \bar{Y}_{i2.l}) - \hat{\lambda}_{il}(\bar{X}_{i1.l} - \bar{X}_{i2.l}) \\ &= \hat{\alpha}_{i1l} - \hat{\alpha}_{i2l}\end{aligned}$$

Since treatments are homogeneous with respect to X_{ijkl} and Z_{ijkl} under the model (1), the difference between the adjusted treatment means can be interpreted as $\bar{X}_{ij.l}$ and $\bar{Z}_{ij.l}$.

After the analysis of covariance for different variables within each block, we have I independent pairs of vectors $(\bar{Y}^a_{i1.}, \bar{Y}^a_{i2.})$ for $i=1,\ldots,I$ where:

$$\bar{Y}^a_{i1.} = \begin{pmatrix} \bar{Y}^a_{i1.1} \\ \vdots \\ \bar{Y}^a_{i1.L} \end{pmatrix} \text{ and } \bar{Y}^a_{i2.} = \begin{pmatrix} \bar{Y}^a_{i2.1} \\ \vdots \\ \bar{Y}^a_{i2.L} \end{pmatrix},$$

Note that $\bar{Y}^a_{ij.l}$ is correlated to $\bar{Y}^a_{ij.l'}$ for all j, where j = 1, ..., J, and l, where l = 1,...,L. Even though $\bar{Y}^a_{i1.}$ and $\bar{Y}^a_{i2.}$ are not best linear unbiased estimates (the best unbiased estimate can be obtained by the generalized least squares method (GLS) which requires estimation of Σ), covariates were considered and, $\bar{Y}^a_{i1.}$ and $\bar{Y}^a_{i2.}$ are consistent estimates of α_{i1} and α_{i2} under the model (1).

Finally perform a rank-sum-type test. Let R_{ij} represent the rank of $\bar{Y}^{*}_{ij}$ among all values of variables in the pooled set of J x L sample in block i.

Since data from different items have been standardized, define S_{ij} as the sum of the rank assigned to the i block in sample j (treatment). Perform a one-way analysis of variance on the $\{S_{ij}\}$ values, when the number of blocks is large enough (based on asymptotic normality) and perform a sign test on the $Z_i = S_{i1} - S_{i2}$ for i = 1, ..., I when i is small.

IV. Results

A. Mail Response Rates for Selected Subgroups

Tables A-1 through A-4 present mail response rates for selected subgroups. Tables A-1 and A-2 present mail response rates for private schools. Tables A-3 and A-4 present mail response rates for public schools. Note that this analysis is conditioned on the sample that was selected in 1991, so no standard errors are used.

As Table A-1 reveals, mail response rates show great difference by affiliation. Lutheran, Catholic, Military, and Christian Schools International show the highest mail response rates, tending toward 60% or more, which we would consider high for private schools. Jewish, Friends, and American Association of Christian Schools show low rates - 45% or less.

Table A-2 shows a high mail response rate for the Chicago and Kansas City Regional Offices, and a low mail response rate for New York. The affiliation differences may be the cause of the differences seen in these three tables, but that cannot be determined from this analysis.

Other results reveal a fairly low mail response rate for combined schools and a high mail response rate for nonmetropolitan schools.

Table A-3 shows a low mail response rate for large central cities. Table A-4 shows a low mail response rate for the New York Regional Office. The low response rate for large city schools may be the cause of this.

Other results show the mail response rate for public schools by state. Rates vary from 48% in the District of Columbia to 81% in Delaware. There appears, however, to be no geographic patterns, such as by size or region.

B. Comparison of Response Categories

A fairly substantial number of items show a significant effect by mode of interview. Based on chi-square analysis alone, however, it is impossible to tell if these differences are due to mode or represent inherent differences in the characteristics of the respondents for each mode of interview. If, for example, from our results presented in Section A above, we believe Jewish schools have a low mail response rate, then this analysis will show mode differences for any item correlated with Jewish schools, even if mode does not influence the actual responses given. For this reason, chi-square analysis is used only as a tool to further narrow the scope of the covariate analysis to follow, and is not being used to draw conclusions about any biases that may be caused by mode of interview.

C. Covariate Analysis

Tables B-1 and B-2 list the results of the covariate analysis. Table B-1 shows the number of significant paired comparisons (blocks) for selected public school items. Table B-2 lists the results for selected private school items. See Attachment C for an example of the output produced in SAS by PROC GLM, which was used to carry out the covariance analysis.

Table B-1 shows that for the items where a reasonable linear regression model could be fit, 3 of 27 paired comparisons were significantly different at the a = .10 level. This seems to indicate no effect upon the data due to mode. However, there appears to be some trend in the block level adjusted means (not shown) whereby the telephone respondents seem to give larger values than do the mail respondents, even when the size covariate is corrected for.

As explained in Section III. B., due to the possible correlation among the questionnaire items being analyzed and due to the possible phenomena being observed, we shall need to undertake a rank-sum type test using standardized block-level means. This analysis is presented in Section IV.D.

Table B-2 gives the results of the covariate analysis for private school data items. It shows 17 of 203 significant paired comparisons at a = .10. This would seem to indicate no differences due to mode. Again, however, this analysis suffers the same difficulties as mentioned previously for the public school data items. Thus, rank-sum type tests also need to be conducted for these items.

D. Nonparametric Testing

As described in Section III, the adjusted means within each block (stratum) were standardized across treatment (mode) and item (questionnaire item). Standardized values were ranked and one-way testing was conducted on the sums.

For the public school items and some of the private school items, there were only nine blocks, resulting in too few degrees of freedom. Thus, a sign test was conducted on the ranked sums rather than a one-way analysis of variance.

The result of sign testing for the public school items listed in Table B-1 did not reveal a significant difference at α = .10. Thus, we would fail to conclude there is evidence of an effect due to mode of interview.

For the three private school items from Table B-2 with only nine blocks, the sign test, again, did not reveal a significant difference due to mode. For the three items with 41 blocks, however, the result of the one-way analysis of variance revealed a significant effect at α = .01.

Due to this strong piece of evidence, we would generally conclude there is evidence of a difference in the data due to mode for private schools.

E. Item Nonresponse

Item nonresponse rates were computed for every item from all questionnaires from the 1991 SASS by Census Bureau staff. It is generally believed that mail responses produce a higher item nonresponse rate. Thus, a null hypothesis that there is no difference in item response rates was tested using a sign test. For both public and private schools, this hypothesis was rejected at α = .10. However, since we used all the items from the questionnaires and there is believed to be substantial correlation in response among the items, particularly between adjacent items, this result must be viewed with some skepticism. As a method of analyzing sets of items with reduced correlation, five samples were selected systematically across all the items. The sign test was conducted on all five samples and all five revealed a significant difference at α .10. Thus, our evidence is consistent with the belief that mail responses have a higher item nonresponse rate.

V. Conclusions

Based on the results of the covariate analysis presented in Section IV.C., we would conclude that there is little if any effect upon the data due to mode of interview. The results of the nonparametric testing, however, revealed some evidence of a difference at least for private schools. It is important to note that the items from the school questionnaire that we have been studying are generally "objective" in nature. They are items that could be considered descriptive of the school and not items we would consider to be greatly subject to the feelings and opinions of the respondent. Some such questions are included on the teacher questionnaire, and will be studied by the Census Bureau in the near future.

In the absence of any large bias due to mode of interview, it is in the interest of the SASS surveys for the Census Bureau and the National Center for Education Surveys to undertake methods for improving the overall mail response rate in order to reduce cost. Section IV.A. has identified some subgroups for which the mail response rate is relatively poor, specifically for large city public schools, and for specific affiliations of private schools. Dillman (1991) suggests methods for improving mail response rates, such as questionnaire design, use of reminders, and length of the questionnaire. Also, establishment of better contact with the specific school organizations mentioned should help to improve mail response rates. Mail response rates are generally good for the SASS surveys, but we believe there is room for improvement.

VI. References

Dillman, D. (1991). The Design and Administration of Mail Surveys, Annual Review Sociology, 17, 225-249.

Hendrix, L., Carter, M., Scott, D. (1982). Covariance Analyses with Heterogeneity of Slopes in Fixed Models, Biometrics, 38, 641-650.

Nathan, G., Holt, D. (1980). The Effect of Survey Design on Regression Analysis, Journal of Royal Statistical Society, Sec. B, 42, #3, pp 377-386.

Rao, J.N.K., Scott, A.J. (1984). On Chi-squared Tests for Multi-way Contingency Tables With Cell Proportions Estimated from Survey Data. Ann. Statist., 12, 46-60.

Rao, J.N.K., Scott, A.J. (1987). On Simple Adjustments Chi-squared Tests With Sample Survey Data. Ann. Statist., 15, 385-397.

Table A-1: Private School Mail Response Rate by Affiliation (List Frame Only)

Affiliation	Mail Response Rate
Association of Military Colleges and Schools - US	66.7%
Catholic	63.0%
Friends	42.3%
Episcopal	50.5%
National Society for Hebrew Day Schools	35.1%
Solomon Schecter	42.5%
Other Jewish	36.1%
Lutheran - Missouri Synod	73.6%
Ev Lutheran Ch - Wisconsin Synod	66.0%
Ev Lutheran Ch in America	71.3%
Other Lutheran	58.2%
Seventh-day Adventis	57.0%
Christian Schools International	64.0%
American Association of Christian Schools	30.7%
NA of Private Schools for Exceptional Children	58.1%
Montessori	48.5%
NA of Independent Schools	48.8%
All Other	50.3%
TOTAL	55.3%

Table A-2: Private School Mail Response Rate by Regional Office (List Frame Only)

Affiliation	Mail Response Rate
Boston	56.7%
New York	42.9%
Philadelphia	56.1%
Detroit	55.6%
Chicago	69.5%
Kansas City	65.1%
Seattle	57.2%
Charlotte	54.7%
Atlanta	53.0%
Dallas	53.8%
Denver	55.2%
Lc Angeles	52.3%
TOTAL	55.7%

Table A-3: Public School Mail Response Rate by Type of Locale

Affiliation	Mail Response Rate
Large Cen City	54.9%
Mid-size Cen City	66.4%
Urban fringe of large cen city	65.2%
Urban fringe of mid-size cen city	69.5%
Large town - nonMSA	73.7%
Small town	71.4%
Rural	67.0%
TOTAL	67.3%

Table A-4: Public School Mail Response Rate by Regional Office (List Frame Only)

Affiliation	Mail Response Rate
Boston	68.4%
New York	54.5%
Philadelphia	70.0%
Detroit	63.5%
Chicago	71.7%
Kansas City	65.4%
Seattle	67.8%
Charlotte	71.4%
Atlanta	71.9%
Dallas	65.7%
Denver	65.3%
Los Angeles	64.4%
TOTAL	67.3%

Table B-1: Results of Covariate Analysis for Public School

Item	R-square	Model Variables	Number of Significant Paired Comparison (a = .10)
Number of Students	0.93	grade mode urbanicity CCD # students	2 of 9
Number of teachers	0.86	grade mode urbanicity CCD # teachers	1 of 9
Number of teachers-education beyond bachelor's	0.66	grade mode urbanicity CC # teachers	0 of 9
Number of new teachers	0.21	*	*

* Good fit if could not be found

Table B-2: Results of Covariate Analysis for Private School Items

Item	R-square	Model Variables	N umber of Significant Paired Comparison (a = .10)
Number of Students	0.93	urbanicity grade mode PSS # students	1 of 9
Student % minority	0.18	*	*
Enrollment in chapter 1	0.53	association PSS # teachers mode	1 of 11
Tuition	0.62	grade mode urbanicity PSS # teachers	4 of 42
FTE teachers	0.83	grade mode urbanicity PSS # teachers	2 of 9
# state certified teachers	0.68	grade mode urbanicity PSS # teachers	2 of 41
Number of teachers	0.81	grade mode urbanicity PSS # teachers	2 of 9
Number of teachers-education beyond bachelor's	0.69	grade mode urbanicity PSS # teachers	2 of 41
Number of new teachers	0.58	grade mode urbanicity PSS # teachers	3 of 41
Starting Salary	0.16	*	*

* Good fit could not be found

QUESTIONNAIRE RESEARCH IN THE SCHOOLS AND STAFFING SURVEY: A COGNITIVE APPROACH

Cleo R. Jenkins, Bureau of the Census
Washington, DC 20233

KEYWORDS: Self-administered Questionnaires, Institution Surveys

The Schools and Staffing Survey (SASS) is conducted by the Census Bureau for the National Center for Education Statistics. It is a relatively new set of integrated surveys first conducted in the 1987-88 and 1990-91 school years and scheduled to be conducted every three years hence. Self-administered questionnaires, of which there are eight, are mailed to public school districts, and to both public and private school administrators, the schools themselves, and to teachers within the schools, asking questions about enrollment, teaching positions, as well as other school and teacher characteristics.

This survey has recently been the focus of questionnaire design research at the Census Bureau. One particular Schools and Staffing Survey, the Public School Questionnaire, was chosen for in-depth study. This survey had an especially high pre-edit failure rate in 1991 (Jenkins, 1992). This means that information from the 1991 school questionnaire disagreed with comparable data for the same school from a survey conducted by the National Center for Education Statistics, known as the Nonfiscal Surveys of the Common Core of Data (CCD).

This paper describes both the methods that were used to conduct this study and some results of the research.

I. METHODOLOGY

One objective of this research was to gain in-depth knowledge about questions that had high pre-edit failure rates in the 1991 surveys. Another was to test newly developed questions. A condensed version of the Public School 1991-92 Field Test Questionnaire (SASS-3A) served this purpose well.

Once the scope of the questionnaire was defined, the researchers used their intuition and experience in questionnaire design to uncover potential problems in the questionnaire and to develop questions to probe respondents' understanding of the items.

After establishing the protocol, twenty in-depth interviews were conducted, four in each of five mid-western states: Oklahoma, North Dakota, South Dakota, Nebraska, and Iowa. These particular states were chosen because they exhibited the highest pre-edit failure rates in student and teacher counts in 1991. Together the National Center for Education Statistics (NCES), the agency that sponsors the SASS, and the Demographic Survey Division (DSD), the division within the Census Bureau responsible for conducting the SASS, provided the researchers with a list of approximately ten schools within each state. The reason for supplying the researchers with more than the final four schools was to allow for scheduling conflicts and refusals.

Not wanting to burden respondents, while at the same time wanting to study the reject phenomenon, a compromise was reached in which three of the schools selected in each state were not in any other SASS. The remaining school, however, was a 1991 pre-edit failure. A final constraint on sample selection was that the schools needed to be within a few hours' drive of the major city in which the researchers were based in each state.

The Public School Questionnaire is addressed to the school principal. During the actual survey, a label is affixed to the right-hand side of the cover page. The researchers mimicked this (see Figure 1).

The researchers contacted the principals, explained the nature of the study, and established a date and time to conduct the interview. The final sample consisted of respondents who were willing to participate. The interviews tended to last an hour and a half. They were tape-recorded and summaries of the interviews were written (see Jenkins, 1992, for summaries of each interview).

II. RESULTS AND DISCUSSION

The results of the cognitive interviews are discussed in the remainder of this paper. I have decided to focus on errors that resulted from the misunderstanding of concepts, the layout of the questions themselves, and finally, the use of records. I use two examples from the cognitive interviews to describe each of these errors in detail (see Jenkins et al., 1992, for a complete review of the results of the cognitive interviews).

A. Misunderstanding of Concepts

The cognitive interviews revealed two concepts that were widely misunderstood: one was respondents' understanding of the school for which they were to report and the other was the classifying of employees in full-time or part-time positions. The many reasons for these misunderstandings are described below.

1. Respondents' Understanding of the School For Which They Are to Report

A very important concept and one which affects the way respondents answer every item on the questionnaire is their understanding of the unit for which they are supposed to report. The cover page contains a very important instruction in the paragraphs on the left-hand side of the page that tells respondents to "Please complete this questionnaire with information about the SCHOOL name on the label." (See Figure 1.) This is the unit for which respondents are supposed to report. I have categorized the result of this understanding into three types: general agreement between their understanding and the intent of the questionnaire, ambiguity between the two, and finally, disagreement between the two.

The first group is made up of nine respondents whose understanding of the school for which they were to report generally agreed with the questionnaire's intent. Basically, respondents in this first group were inclined to report in terms of the school named on the questionnaire label, whether they read the school named there or not. Six of the nine respondents actually read the label.

Respondents who were principals over schools that clearly stood apart (i.e., functioned separately and/or were not in close proximity to any other school) seemed to fall into this first group. These respondents were not confused as to their school's identity. For the most part there was a clear demarcation such that the principals did not consider reporting for any other school(s).

Group II was made up of eight respondents whose understanding of the school for which they were to report was ambiguous. Generally, this was the case in relatively small school systems in which two or three schools comprised the entire school district. Often the schools were housed in one building or they were housed in separate buildings that were clustered around one another. There was a principal for each of the two or three schools within the district, but the principal of the school named on the label saw himself as capable of reporting for the other school(s), if he thought, however begrudgingly, that's what was being asked of him. Because of the schools' close association with one another, the line of demarcation wasn't as clear for these respondents.

These respondents had alternative definitions of the unit for which they could report and they relied on the questionnaire to inform them which one to use. On one hand, they could define their school as the grades over which they had jurisdiction. Because of their organization, however, it was conceivable to them that the questionnaire might be asking about the entire school system, kindergarten (K) through 12th grades. As a result, they were quite receptive to cues from the questionnaire. Unfortunately, these cues were conflicting.

Half of them began to complete the questionnaire by reading the cover page. Generally, they read through the title information and then the first two paragraphs on the left-hand side of the page. Because the paragraphs refer to the label, they turned the questionnaire sideways to look at the label. The other half of the respondents, however, never noticed the school named on the label. Neither the instruction referring to the label nor the school's name itself is prominent. Both are buried among a lot of information on the cover page. In fact, the school's name is not only buried, but it is turned sideways (see Figure 1).

After providing their name and address in item a, which is the first question on the form, these respondents turned to item b. Item b asks if the school serves students in any of grades 1 through 12. If the school doesn't, they are instructed to return the questionnaire to the Census Bureau. If they do, they are to continue. The reference to "grades 1 through 12" in this item seemed to trigger these respondents into thinking that the questionnaire might be asking about the entire school system rather than just their school. They weren't sure, but they now had a reason to believe this was the case.

After answering item c, which asks about their School State Identification Number, they turned to item 1. The question of item 1a asks for what grade levels the school offers instruction, and the question of item 1b asks how many students were enrolled in each grade on October 1 of the school year. In both cases, prekindergarten through 12th grade answer categories are provided (see Figure 2). The answer categories seemed to provide these respondents with more evidence that the questionnaire might be asking about the whole school system. As a result, some began interpreting item 1 as asking about the entire school, but most didn't.

Most waited until they reached item 2, which asks how many students were enrolled in the school in grades K-12 on or about October 1 of this school year. For the most part, these respondents voiced their ambiguity at item 1, but still they answered item 1 in terms of their school. In some cases, they may have done this simply because it was easier, but in other cases, it seemed that these respondents needed more evidence before they could be swayed into reporting for the entire school system. And the fact that item 2 seemingly asks for the number of students in grades "K-12" became the evidence they needed.

Once made, this interpretation was continuously reinforced by the many "K-12" references in the questions that follow item 2, until eventually it became solidified in the minds of some respondents. They stopped questioning the unit for which they should report and began to report for the entire school system. This is not to say, however, that this was painless. The fact was they needed to go through a great deal of work to obtain information to answer for the entire school.

It is not surprising, therefore, that their interpretations weren't always the final determinant of how they reported. Sometimes the records they had on hand became the limiting factor. This meant that although their interpretations might be relatively consistent from item 2 onward, their answers were not necessarily consistent. Sometimes they answered in terms of the school system. This was often the case with item 2. This item requests a summary statistic they often had on hand. It asks how many students were enrolled in the school on or about October 1 of this school year. At other times, they answered in terms of their school only, as was often the case with item 3. This item requests information they couldn't conveniently obtain. It asks for a breakdown of the student population into ethnic categories.

Also, it should be noted that some respondents continued to express ambiguity. These respondents didn't settle on one definition, but instead interpreted questions in which they noticed the "K-12" reference as asking about the entire school system and questions in which they didn't notice this reference as asking about their school.

Group III was made up of three respondents whose definitions simply didn't agree with the questionnaire's. Two of these respondents had jurisdiction over both the elementary and secondary portions of a relatively small school, with both portions housed in one building. In another case, the respondent was principal of both the middle and high school portions of the school system, which again were housed in one building. In these cases, the respondent's definition of his school was clearly different from the school named on the label, and the problem was that the questionnaire tended to reinforce this wrong definition. In fact, two of these respondents never looked at the school named on thr abel.

2. Respondents' Understanding of Full-time Versus Part-time Status.

Another concept respondents had a very difficult time with was that of full-time versus part-time employment, as asked for in item 30 (see Figure 3). To understand why respondents misreported, it may be best to begin with a situation in which respondents were likely to report correctly. They were likely to correctly report an employee as part-time if that employee was exclusively part-time and the job itself could be considered full-time. For example, respondents were likely to report an instructional aide as part-time if that aide only worked for part of the day, meaning he/she didn't work the rest of the day, and there were others who did work all day as an aide. In this case, the part-time aide could be compared to a full-time aide and there wasn't any confounding information with which to be confused (i.e, any other assignment or job). As a result, the situation was clear to them, but this was also one of the less frequent situations.

The more frequent situations were less clear. For example, employees who worked at jobs that by definition could never be considered full-time jobs were difficult for respondents to categorize. This was the case with bus drivers. Respondents could agree that bus drivers always work less than a full day, but they couldn't agree if that meant they should be categorized as part-time or full-time. The reason they couldn't agree on this is that the bus driver's job is not full-time relative to other full-time jobs, such as the principal's job; however, it is full-time if the unit of comparison is limited to a bus driver's job. Looked at from this perspective, it is as "full" a job as a bus driver's job can get.

Also, problems arose when an employee worked part-time in more than one assignment, but full-time at the school. One reason respondents misunderstood this concept was that they were used to thinking in terms of an employee's employment status at the school overall and not by assignment. Take, for instance, an aide at the school, who works full-time, but whose assignment is divided between being an instructional aide and librarian aide. More often than not, the respondent would report this employee as a full-time instructional aide and full-time librarian aide. The same happened with a teaching principal. He reported himself as a full-time teacher and then again as a full-time administrator. In these instances, respondents thought of the employees as full-time and had difficulty thinking of them as part-time.

Respondents also had difficulty if an employee worked part-time at this school, but full-time for the school district, meaning the employee was shared among the schools. In the smaller schools, many of the staff were shared, including librarians, guidance counselors, clerical staff, the student support services staff, and the other support staff. Here again, respondents had a tendency to report these employees as full-time.

B. Format Considerations

Errors occur when an item is laid out such that respondents don't see, and consequently don't read, information that is necessary to correctly answer the item. Respondents commonly overlooked information that was placed beyond what they considered to be the answer space, including "none" boxes and skip instructions. As a result, they were likely to spend a great deal of time and energy trying to answer questions that didn't apply to them, as demonstrated below with item 15. There were also instances in which an entire item was laid out poorly, as demonstrated below with item 29.

1. "None" Boxes and/or Skip Instructions

Item 15 asks a series of questions about limited-English proficient students (see Figure 4). Part a of this item asks "How many students attending this school as of October 1, 1991, were identified as limited-English proficient (LEP)." In response to this question, quite a few respondents made the mistake of reporting "0" on the answer line because they didn't notice the "none" box that was placed about half an inch beneath the answer line. The cognitive interviews revealed quite a bit about how respondents interpret questions that don't apply to them from this.

Respondents who had previously had LEP students but who didn't have any now used their past experience to answer part b, which asks what methods were used to identify LEP students. They reported the methods they had previously used to identify LEP students. Another respondent whose school had never had any LEP students answered the best he could by marking the "other" answer category and writing in "never been a problem." It became evident as a result of this research that respondents commonly marked the "other" box and wrote something in when they thought they were supposed to answer a question, but they couldn't understand it. Either it was ambiguously worded or it wasn't applicable to them, as was the case here.

In general, respondents who had previously had LEP students came to realize that part c, which asks about the number of LEP students in specified programs, didn't apply to them and correctly skipped to the next item at this point. These respondents seemed to be familiar with the notion of limited-English proficiency and its acronym. This helped them realize that this question didn't apply to them.

Unfortunately, respondents who never had LEP students just plowed away, trying to answer questions they shouldn't have. This was probably due to the fact that only the acronym LEP is used in this question and although it was defined previously, they weren't really familiar with the notion of limited-English proficiency in the first place, let alone its acronym.

It became obvious as respondents tried to answer this part of the item that they didn't really know what programs (1) through (6) were, since they didn't have and never had any LEP students. Consequently, they transformed these answer categories into something that had meaning to them. All sorts of misinterpretations arose as a result. One of the more reasonable interpretations was to think it was asking for the number of regular students in the listed programs. In this case the respondent reported "none" in all but the fourth category, where he reported all of his special education students. In other words, he didn't change the meanings of the individual programs per se, just the population to which they applied.

Matters really broke down, however, when respondents not only thought the question applied to regular students, but they changed the meanings of the programs as well. This happened most for the first two programs. These were written such that respondents couldn't comprehend the entire sentence, but they could find meaning in individual words. For instance, one respondent thought the first category (subject matter in home language) was asking for the number of classes in grades 7 through 12. This respondent seemed to key in on the words "subject matter." To her these words were associated with the number of classes in grades 7 through 12. To understand this, one must realize that usually subject matter is taught in subject matter classes in grades 7 through 12, and not at the elementary level. Another respondent interpreted the second category (maintaining fluency in home language) as asking if the school offered foreign language instruction in Spanish. Obviously, this respondent noticed the word "Spanish" in the example and extrapolated from that a program that had meaning to him. The point is these respondents were not answering the questions asked of them.

2. Item Layout

Item 29 asks a series of questions about teaching vacancies in the school. There is a problem with the layout of part d in this item (see Figure 5), which asks how difficult or easy it was to fill the vacancies in the listed fields. The first problem was that not all subparts of the question applied to all schools. Elementary schools weren't sure how to answer parts (3) through (9), since these are subject matter courses not offered at the elementary level. It didn't seem quite right to mark "no vacancy in that field" when the truth was they didn't even have that field. Conversely, high schools weren't sure how to mark the first two categories.

Also, there was a special problem with the layout of the ninth category (9), vocational-technical education. Very few respondents read the follow-up question on the left-hand side beneath this category. Consequently, they didn't understand this part of the item. They were supposed to mark how difficult or easy it was to fill the vacancies they had in vocational-technical education in the boxes on the right-hand side. Then they were supposed to identify the subfield(s) of vocational-technical education to which that mark applied in the follow-up list of subfields. However, respondents were misled by the fact that the boxes fall under the "no vacancy in that field" column. Because they tended to see the list of subfields on the right-hand side as just a continuation of the fields specified on the left-hand side, they often continued right on down the column, marking these boxes (as they had the others) to indicate they didn't have these fields.

C. Use of Records

The most striking aspect of respondents' use of records was how varied their recordkeeping systems were. They ranged from slips of handwritten papers that were found in the top drawers of their desks or hanging near them to more formal systems. Some used report forms that came from files in either their office or the secretary's office; some even used state-of-the-art computer databases.

The cognitive interviews revealed that using records did not necessarily guarantee the data would be accurate. Errors arose when respondents didn't use appropriate records. Sometimes this was because they didn't have the appropriate records. Other times it was because they didn't think they had them. Still, other times it was because they didn't recognize the record was inappropriate. One very common error resulted from respondents thinking they didn't have records for the time period specified in a question when in fact they did, as demonstrated below with item 1b. Another more complicated error occurred when respondents applied misguided heuristics to the use of their records, as described below with items 2 and 3.

1. Use of Inappropriate Records

Item 1b asks how many students were enrolled in each of the listed grades on October 1 of this school year (refer to Figure 2). By law, schools are required to submit reports with student enrollment by grade for around October 1st to either the school district or the state. Respondents should have used this report to answer this question. It would have saved them from having to reproduce numbers and probably would have yielded more accurate data. However, a third of them didn't.

One reason for this may be that the question fails to tell them to use it, and in fact, it may even hinder them from considering it. According to the framers of this question, they had the official fall reporting date in mind when they used the date "October 1." They expected respondents to associate this date with the official fall reporting date. Among respondents who focused on the "October 1" date, however, this was either not enough to trigger them to think about their fall report, or if it did, it caused them to dismiss it. For instance, one respondent dismissed using the state report because it was dated September 10th rather than October 1st. He had the business office go through the trouble of producing October 1st numbers from their computer database when the report dated September 10th was already available. Although he reported for the right time period, the office spent more time than necessary answering this question.

Another reason respondents didn't use the fall official report was because they weren't aware of its existence. These respondents, who were the principals of the schools, either weren't as familiar with the school records as their secretaries or they were new to the job. In these cases, the respondents just didn't realize they could comply with the reference period, so they did what they thought best: they ignored it and reported data for the time period they had.

Relying on other records forced this group to report data for a time period different from the one requested. They reported numbers for the beginning of the year, end of the first quarter (November 3rd), end of second semester (January 13th), as well as current figures.

2. Heuristics Applied to the Use of Records

Item 3 asks for a break-down of students into ethnic categories (refer to Figure 6). The majority of respondents used a heuristic to answer this question. First, they relied either on their knowledge of the student population or on some kind of record to report the number of students in the ethnic categories in parts a through d. After this, they calculated the number of white students in part e by subtracting the total number of minority students from the total they had reported in item 2a (refer to Figure 7). As a result, the total number of students reported in item 3 was consistent with the number reported in item 2a. However, the number of white students was not always accurate.

This approach was fine, as long as the record they used to answer item 3 was for the same time period as the record they had used to answer item 2a. Then the data were not only consistent, but they accurately reflected the ethnic counts at a given point in time. However, since item 3 doesn't specify a time period, a few respondents answered item 3 using current data, whereas they had used records as of October 1st to answer item 2a. It wasn't obvious to these respondents that they might be introducing an error into the data by deriving the number of white students as they did.

Also, their method of calculating white students was flawed if the minority counts themselves were off, which was the case a number of times. For instance, one respondent reported the number of American Indians as of last year. He initially interpreted this question to be asking for last year's numbers because of the reference period given in item 2b. In addition, he reported the wrong number of black students because he made a mistake when he manually counted up these students from a student list. When he was done reporting these wrong counts, he proceeded to calculate the number of white students by the method mentioned above. As a result, the white count was off as well.

Another respondent double counted the number of minorities she reported in parts a through d because of the way she answered here. According to the secretary, the school actually reported all minorities as American Indian on a report they submit to the Office of Indian Education. Since the respondent used this report to answer part a, she inadvertently reported all minorities as American Indian. Following this, she went on to report the minorities again in parts b, c, and d. As a result, the number of white students was also erroneous.

In these cases, the numbers didn't accurately reflect the ethnic counts, but the values reported in items 2 and 3 were consistent. In some cases, these mistakes seemed to be the result of respondents not paying close attention to what they were doing. In other cases, it seemed to be because the questionnaire asks for data the respondents didn't have in the requested format. And in still other cases the questionnaire asks for data with which the respondents weren't wholly familiar.

CONCLUSION

In this paper, I have described questionnaire research with the Public School Questionnaire from a cognitive perspective, meaning how and why respondents interpreted information as they did. Examples of respondent errors from the cognitive interviews were presented, including errors that resulted from the misunderstanding of concepts, the layout of the questions themselves, and from the use of records.

The cognitive interviews revealed that errors occur because information presented on the questionnaire is not always perceived as intended. Many respondents did not understand the school for which they should report. In large part, this was due to the fact that the school's name is hidden from view on the cover page and suggestive references to the entire school system are used throughout the questionnaire. In general, this error should be relatively easy to correct. Most respondents were inclined to report their school correctly, but were just confused by the questionnaire. On the other hand, many respondents didn't understand the concept of full-time versus part-time employment as intended by the questionnaire; however, this may be more difficult to correct because asking respondents to think as the questionnaire does is asking them to think in a relatively complex and foreign way.

The "none" boxes and skip instructions present respondents with problems, and this seems to be due to the method respondents use to answer questions. Once respondents answer a question, they seem to think the response task is over. As a result, they do not take in new information until they begin what they perceive to be the next "question-answer" cycle. Also the layout of the questions themselves sometimes give respondents difficulty. However, mistakes such as these may be relatively easy to correct.

Respondents' use of records is one of the most complex areas of questionnaire research to study, since it requires in-depth knowledge about respondents' records as well as how they use those records, and very little is known about this process to date. Certainly this is an area in need of further research. As demonstrated earlier, problems can occur when respondents use records. Some of the errors that were witnessed during the cognitive interviews may be correctable, some need further research, and some seem to be intractable. Errors that arise from questionnaire miscues, such as the use of inconsistent time periods and not providing clear references to particular records may be relatively easy to correct. However, mistakes that occur for other reasons may be difficult to correct. An error needing further research is one that arises because respondents do not have information in the requested format. In-depth studies are needed to design questions that ask for information in appropriate formats. An example of a mistake that may be intractable, however, is one in which respondents do not pay close attention to what they are doing.

The next step in this process will be to redesign the questionnaire using guidelines resulting from this research. The first and probably most important guideline is that the school's name and grade levels should be prominently displayed. The final step will be to conduct a test of alternative questionnaires. Discussions are underway on how best to conduct this test.

REFERENCES

Jenkins, C. J. 1992. "A Review of the Post-edit Item Response Rates, Pre-edit Reject Rates, and Edit Change Tallies from the 1991 Schools and Staffing Surveys." Internal Census Bureau memorandum. March 23, 1992.

Jenkins, C. J. 1992. "Cognitive Interviews on the Public School Questionnaire." Internal Census Bureau memorandum. April 30, 1992.

Jenkins, C. J., S. Ciochetto, and W. Davis. 1992. "Results of Cognitive Research on the Public School 1991-92 Field Test Questionnaire for the Schools and Staffing Survey (SASS-3A). Internal Census Bureau memorandum. June 15, 1992.

Jenkins, C. J., S. Ciochetto, and W. Davis. 1992. "Recommendation for Changes to the Public School 1991-92 Field Test Questionnaire for the Schools and Staffing Survey (SASS-3A) Based on Cognitive Interviews." Internal Census Bureau memorandum. Sept. 24, 1992.

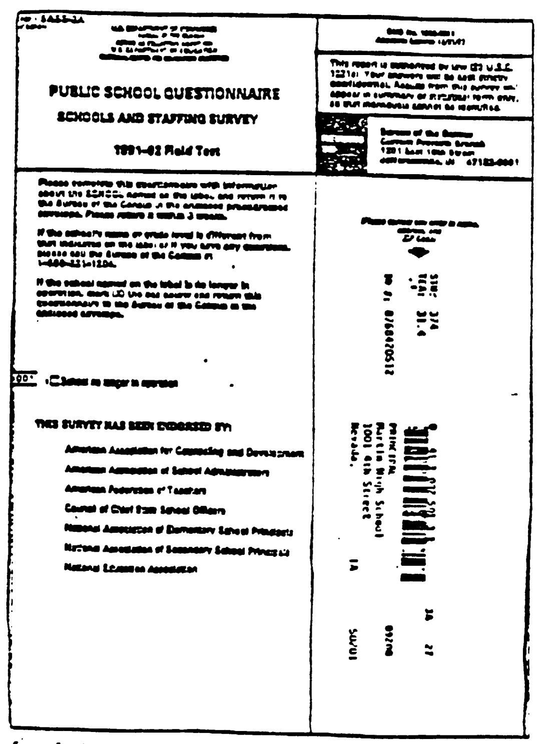


Figure 1. Cover Page of the Public School Questionnaire

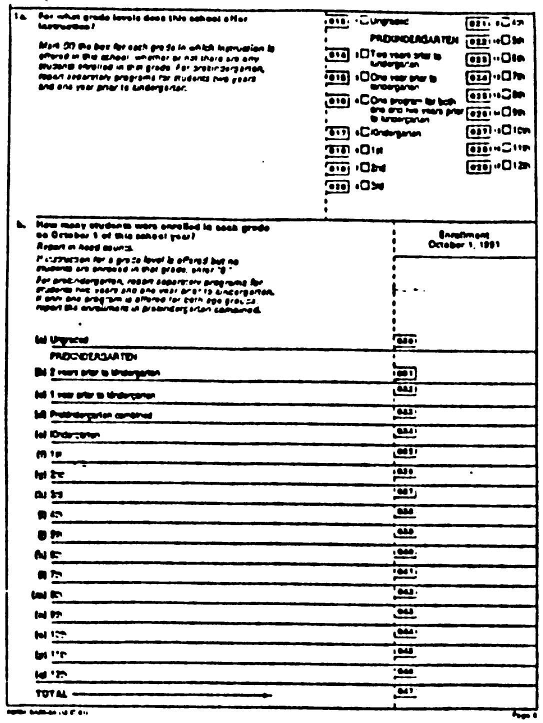


Figure 2. Student Enrollment by Grade Item

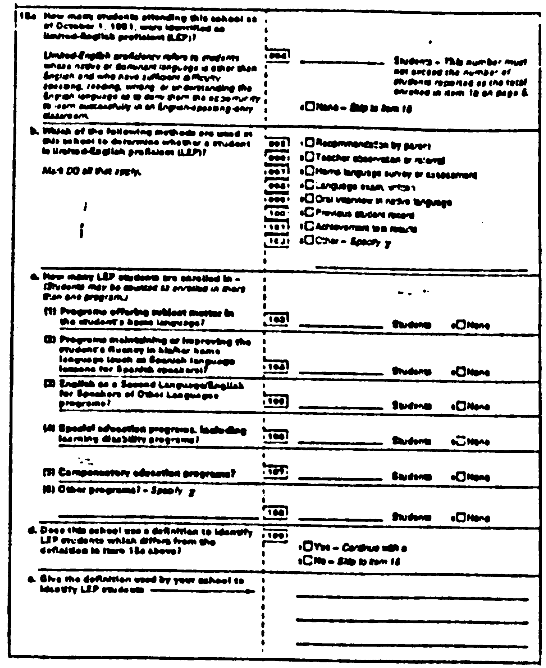


Figure 4. Number of Limited-English Proficient Students Item

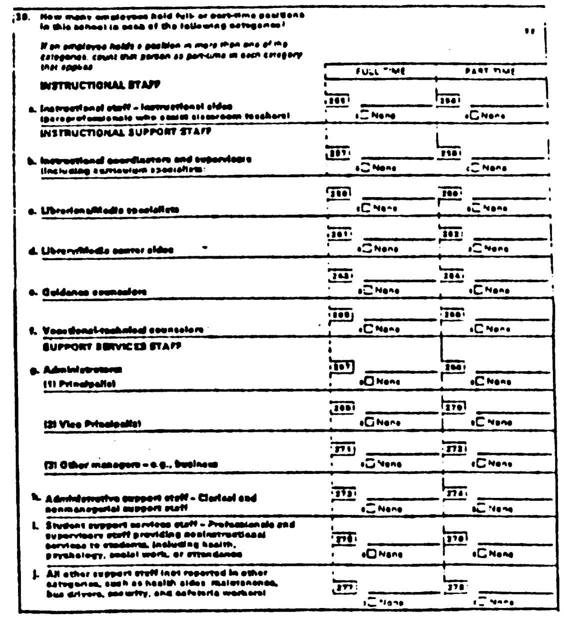


Figure 3. Classification of Employees Item

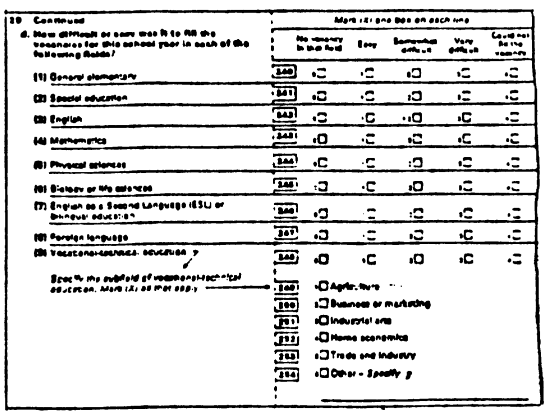


Figure 5. Teacher Vacancies Item

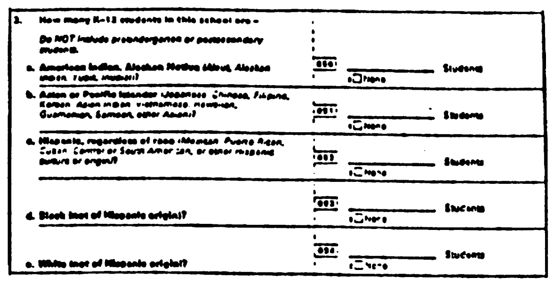


Figure 6. Student Enrollment by Ethnicity Item

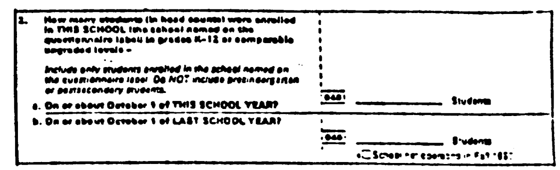


Figure 7. Total Student Enrollment Item

BALANCED HALF-SAMPLE REPLICATION WITH AGGREGATION UNITS

:

Steven Kaufman, National Center for Education Statistics
Room 422e, 555 New Jersey Ave., Washington, D.C. 20208

KEY WORDS: Half-sample replication, Bootstrapping, Variances, aggregation units

Introduction

Given a list of sampling units (frame), most sample designs select a sample proportional to a known variable and collect data for the selected units (i.e., the sampling and collection unit are the same). One example of this type of design is sampling schools proportional to the number of teachers in the school, and collecting school data from the selected schools. It is well known that balanced half-sample replication provides appropriate variance estimates for such designs. A slightly different type of design is when the collection unit is an aggregation of the sampling units. In the example above, if school district data is also collected from all districts with sampled schools then this is an example of the second type of design. In this case, the school district (collection unit) is an aggregation of the schools (sampling unit) belonging to the district.

The question that this paper addresses is whether balance half-sample replication is appropriate when the collection unit is an aggregation of the sampling units. Using the usual BHR design assumptions (i.e., two units independently selected per stratum, the replicates are fully balanced, and the collection and sampling units are the same) then the following is true concerning BHR:

$$E(V_{BHR}(X_n))=V(X_n)=V((X_{n/2}^1 + X_{n/2}^2)/2)$$

$$=1/2V(X_{n/2}^1)$$

E: is expectation with respect to all possible samples

V_{BHR}: is the BHR variance estimate

X_n: is a linear estimate based on the full sample of n units

V: is the true variance estimator

$X_{n/2}^1$: is the half-sample replicate estimate of X based on the first unit selected within each stratum

$X_{n/2}^2$: is the same thing as $X_{n/2}^1$ except the estimate is based on the second unit selected in each stratum

This says that BHR assumes the true variance is inversely proportional to the sample size. It is this property of BHR that might not be true when the collection unit is an aggregation of the sampling units. If the inverse of the selection probability is used as the weight then the possibility of a biased variance estimator can be seen by looking at the form of the selection probability.

When the selection and collection units are the same, the selection probabilities are usually linear with respect to the sample size. If this is true with the aggregation unit selection probabilities, one might expect BHR to work well too.

The selection probability for an aggregate unit, A, has the following form:

$$P_A=1-\prod_{Ah\in A}(1-P_{Ah})$$

P_{Ah}: is the selection probability for the part of aggregate A that is within stratum h.

If aggregation A is composed of one sampling unit for each of two stratum then:

$$P_A=P_1 + P_2 - P_1P_2$$

P_i: is selection probability for unit i

This selection probability is not linear with respect to the sample size, assuming the P_is are linear in the sample size. Hence, BHR may be inappropriate. However, if P_1P_2 is small relative to P_is then BHR may provide a reasonable approximation. Whether BHR is appropriate, depends on the distribution of the P_As.

One possible alternative to BHR is the bootstrap variance estimator. Bootstrap samples use the full sample to approximate the distribution of the frame. From this approximate frame, bootstrap samples are selected using the initial sample size. Since the bootstrap estimates are based on the initial sample size, the half-sample assumption that the variance is inversely proportional to the sample size is not necessary.

The goal of this paper is to investigate, how well BHR estimates the variance when the collection unit is an aggregation unit. Two weighting schemes will be tested - 1) the inverse of the selection probability and 2) a weighting scheme that is linear with respect to the number of selection units selected in the aggregation unit. In addition, a bootstrap variance estimator will be tested using the inverse of the selection probability as a weight.

I'll show, using simulations with a weight based on the inverse of the selection probability, that for the National Center of Education Statistics' Teacher Demand and Shortage survey, BHR works reasonably well for most states. For eight states, BHR does not provide reasonable variance estimates. For these few states, a bootstrap estimator provides reasonable estimates. Based on simulations, I will also show when the second weighting scheme is used, BHR appears to provide unbiased variance estimates.

Simulations

The Teacher Demand and Shortage Survey (TDS) is one of four linked surveys the Center produces to study the critical aspects of teacher supply and demand, the composition of the administrator and teacher work force, and the status of teaching and schooling generally. School districts, schools, administrators, and teachers are all surveyed through a common sample design. These surveys are called the Schools and Staffing Surveys (SASS) and are designed to provide state estimates. The focus of this paper is the initial sampling unit, schools; and the survey of school districts, the aggregation unit. The simulations will be based on data from two frames - the SASS public school frame and a frame of all matching public school districts.

The school frame will be used to select multiple school samples using a design similar to the SASS school sample. Each school sample will be matched to the district frame to produce the district sample, as is done for the TDS survey. The district frame has teacher, student, graduate, and school counts. From each district sample, estimates and their BHR variances will be produced. The true variances can be estimated and compared to the BHR estimates.

Design of SASS School and TDS Surveys[1]

The school survey uses NCES's public school file as the frame. The frame is stratified by state by school level (elementary, secondary and combined). The school sample is selected using a systematic probability proportionate to size procedure. The measure of size is the square root of the number of teachers in the school. The school districts that include a sampled school comprise the TDS district sample. In order to simplify the computation of the district selection probabilities, the schools are sorted by school district within each stratum.

This design does not satisfy all the BHR assumptions. The selection is done systematically, so units selected within a stratum are not independent and a finite population correction is required. In addition, more than two schools are selected per stratum. To satisfy the BHR assumptions, the simulation sample design is modified from the above TDS design in the following manner:

1) Each sampling stratum is further stratified by substrata. The substrata are chosen to be the set of all schools that could be selected in two consecutive selections from the systematic selection procedure described above (i.e., the schools within two sampling intervals). Two schools will be selected within each substratum.
   To simplify the district weight computation, a district spanning two substrata is place in only one substratum depending on which substratum contains most of the district's measure of size.
   If the cumulative school measures of size within a district is larger than two sampling intervals than the district is subdivided into pseudo-districts that are approximately equal to two sampling intervals. Each of these pseudo-districts comprise a substratum. Such districts are certainty districts. The purpose of this modification is to maintain the original school sample distribution. Otherwise, by selecting at most two schools for very large districts, more of the smaller districts will be selected then in the original design.
2) Two schools are selected with replacement within each substratum. The first selection will be assigned to panel 1, while the second will be assigned to panel 2. This maintains the BHR assumption of independent selections and eliminates any finite population adjustment for the variance that might be part of the original design.

Bootstrap Implications of Simulation Design

When the sample and collection unit are the same, the bootstrap variance estimator for simple random sampling is biased by a factor of $n_h/(n_h-1)$. Since the simulation sample design selects two units per stratum, this bias would be significant with the school as the collection unit. With the district as the collection unit, it isn't clear what the appropriate stratum sample sizes are when determining the bias. If all districts are defined completely within a single substrata then the bias will be large (i.e., the effective district sample will be close to 2 providing a bias close to $2/(2-1)=2$). If all districts are defined across substrata, the bias might be smaller (i.e., if districts are defined within three substratum with a sample of 6 schools then the bias might be close to $6/(6-1)=1.2$). In reality, the district definitions are somewhere in the middle, so the magnitude of the bias is unclear. However, I will assume there is an effective sample size in the "stratum" which can partition states that will or won't be significantly biased.

Weighting

Two weighting schemes will be analyzed - one based on the district selection probability (probability weight), and the other based on the school selection probability (expected hits weight). The sample estimate, BHR variance estimates and the estimate of true variance will be computed for each weighting scheme.

The probability weight for district d (PWT_d) is:

$$PWT_d = 1/(1-((1-p_{de})(1-p_{ds})(1-p_{dc}))^2)$$

p_{de}: $\sum_{i \in S_{de}} p_i$, S_{de} is the set of all elementary schools in district d

p_i: is the selection probability for school i.

p_{ds}: $\sum_{i \in S_{ds}} p_i$, S_{ds} is the set of all secondary schools in district d

p_{dc}: $\sum_{i \in S_{dc}} p_i$, S_{dc} is the set of all combined schools in district d

If p_{de}, p_{ds} or p_{dc} is greater than or equal to one then the district is selected with certainty and $PWT_d = 1$.

The expected hit weight for district d (HWT_d) is:

$$HWT_d = H_d/\sum_{i \in S_d} p_i$$

H_d: is the number of schools selected in district d

S_d: is the set of all schools within district d

The unbiasness of this weight follows from the fact that the expectation of the numerator (the expected number of schools selected within a district) is equal to the denominator (the sum of all school selection probabilities within a district).

BHR should be unbiased using the expected hits weight because any linear district estimate can be written as a normalized school estimate.

Let X_d be a district variable and suppose we want to estimate the total value of X_d within the set of all districts in some set U, say all urban districts then:

$$\sum_{d \in U} HWT_d \; X_d = \sum_{d \in U} (H_d / \sum_{i \in S_d} p_i) \; X_d$$

$$= \sum_{i \in Uk_d} (p_i / \sum_{i \in S_d} p_i) \; X_d / p_i, \quad Uk_d \text{ is the set of all sampled schools within districts in U}$$

This is now written as a school estimate weighted by $1/p_i$, normalized by $p_i / \sum_{i \in S_d} p_i$, and where X_d is assigned to every school within district d. Since this is a school estimate, the BHR variance estimate should be unbiased.

Balanced Half-sample Replicates

The selected schools are placed into half-sample replicates using the usual textbook methodology[7]. The r^{th} district half-sample replicate is defined to be the set of districts that have schools in the r^{th} school half-sample replicate. Since the SASS replicates are based on 48 replicates, the simulations will be based on 48 replicates. The district replicate weights are:

For the probability weight, the replicate weight is:

$$RPWT_d = 1/(1-((1-p_{de}/2)(1-p_{ds}/2)(1-p_{dc}/2))^2)$$

The probabilities are divided by 2 because with half the sample, each school has half the chance of being selected.

For the expected hits weight, the replicate weight is:

$$RHWT_{rd} = H_{rd} / \sum_{i \in S_d} (p_i/2)$$

H_{rd}: is the number of schools within replicate r and district d.

District Bootstrap Samples

The idea behind the bootstrap samples is to use the sample weights from the selected units to estimate the distribution of the school and district frames. From the estimated bootstrap frame, B bootstrap samples can be selected using the simulation TDS design. For each selected school i in district d the weights say, PWT_d districts should be generated for the bootstrap frame. The PWT_d districts should have a total cumulative school measure of size equal to $W_i p_i$, where W_i is the school sampling weight $(1/p_i)$. The bootstrap frame and selection are described below for a specific sample.

1) Generate a file of selected schools. If a school is selected twice, it is on the file twice.
2) Divide each school into PWT_d bootstrap-districts (indexed by bd), each with a $(W_i p_i)/PWT_d$ school bootstrap measure of size. If PWT_d is not an integer then the bootstrap-district representing the noninteger part has a $C_d W_i p_i / PWT_d$ school bootstrap measure of size, where C_d is the noninteger part of PWT_d.

   If a selected district has selected schools in the elementary and secondary strata then the bd^{th} bootstrap-district generated in the elementary stratum should match to the bd^{th} bootstrap-district in the secondary stratum. This relationship should exist for all school levels that are selected for the district. Since this relationship exists for the selected districts, it is important to reflect it in each bootstrap-district.

   The sum of the school bootstrap measures of size for school i is $W_i p_i$, which is the appropriate representation based on the school weight. The number of districts being represented by school i is PWT_d, which is the appropriate representation based on the district weight.

   Each bootstrap-district within a stratum could be divided into W_i bootstrap-schools. Since the school is only a selection unit, not a unit of analysis, it's only required to know which district is selected and not which school. To compute the bootstrap-district weights, the bootstrap-school selection probabilities would be summed within a bootstrap-district, anyway. This would yield the same results as the procedure described above. One method is computational less intensive.
3) Using the frame generated in step 2 and assuming two units are independently selected, within the substrata, proportional to the bootstrap measures of size, compute the bootstrap-district weight, $BPWT_{bd}$. Let u denote a selection unit on the frame and p_u the selection probability for u.

   If pd is a bootstrap-district representing an integer part of PWT_d then:

$$BPWT_{bd} = 1/(1-((1-p_{bde})(1-p_{bds})(1-p_{bdc}))^2)$$

p_{bde}: $\sum_{u \in S_{bde}} p_u$, S_{bde} is the set of all elementary units in bootstrap-district pd

p_{bds}: $\sum_{u \in S_{bds}} p_u$, S_{bds} is the set of all secondary units in bootstrap-district bd

p_{bdc}: $\sum_{u \in S_{bdc}} p_u$, S_{bdc} is the set of all combined units in bootstrap-district bd

If p_{bde}, p_{bds} or p_{bdc} is greater than or equal to one then the bootstrap-district is selected with certainty and $BPWT_{bd} = 1$.

If pd is a bootstrap-district representing a non-integer part of PWT_d then:

$$BPWT_{bd} = C_d/(1-((1-p_{bde})(1-p_{bds})(1-p_{bdc}))^2)$$

4) With the frame and bootstrap selection probabilities define in step 2, independently select two units per substratum proportional to bootstrap measures of size. The weights for the selected bootstrap-districts are defined in step 3.

5) Since the available data is defined by the districts selected in the original sample, a bootstrap-district weight indexed by d ($BPWT_d$) is required:

$$BPWT_d = \sum_{bd \in S_{dB}} BPWT_{bd}, \quad S_{dB} \text{ is the set of all } bd \in d \text{ selected in the } B^{th} \text{ bootstrap}$$

6) Repeat steps 4 and 5 until B bootstrap

samples are selected. Since there are 48 balanced half-sampled replicates, there will be 48 bootstrap samples.

Sample Estimate

For each of the simulation samples, totals, averages and ratios will be computed within each of the fifty states and the District of Columbia. The averages are average number teachers per district and average number of schools per district. The ratios are the ratio of the number pupils to the number teacher and ratio of the number of teachers to number of schools. The totals are number of student, teachers, graduates, schools and districts. The student, teacher and graduate totals are highly correlated with the measure of size, while the school and district totals have a lower correlation with the measure of size. For each of the 90 simulation samples, 459 sample estimates and respective sample variances are computed (51 states * 9 estimates). The average of these estimates across the 90 simulations is an estimate of the expectation of the respective sample estimate. It is these averages that are the building blocks of this analysis. An estimate of the true variance for the sample estimates can be obtained by computing the simple variance of the sample estimates across the 90 simulations. The expected values for the sample variances can now be compared with the estimate of the true variances.

A number of other analysis statistics are required. They are described below.

Analysis Statistics

Confidence Coefficient

To measure the accuracy of the variance estimates, a one sigma two tailed confidence coefficient is computed by determining what proportion of the time the population estimate is within the respective confidence interval. If the variance estimates are appropriate then the confidence coefficients should be close 0.68. One sigma confidence coefficients are used because there aren't enough simulations (90 of them) to accurately measure the 5% tail.

BHR Bias Indicator

The main task of this paper is to measure whether the BHR assumption, that the true variance is inversely proportional to the sample size, is violated. If it is violated, what is the impact on the district variance estimates. The following statistic can be used to partition states into those that will or won't be significantly biased.

BHR Bias Indicator = $V/(1/2(V_1 + V_2)/2)$

V: is the simple variance of the 90 simulation sample estimates that are computed from districts selected in both panels, using the methodology of interest (PWT_d or HWT_d weighting schemes),

V_1: is the simple variance of the 90 simulation sample estimates that are computed from the districts selected in panel 1, using the methodology of interest,

V_2: is the simple variance of the 90 simulation sample estimates that are computed from the districts selected in panel 2, using the methodology of interest.

The numerator is an estimate of the true variance and the denominator is another estimate of the true variance assuming the true variance is inversely proportional to the sample size. Hence, the ratio should be close to one when the true variance is proportional to the sample size.

Within each state, this ratio is computed for each of the nine sample estimates. An average state ratio is then computed using the weights described in the 'Weighting the Estimate' section below. When producing the tables, this state average is assigned to each of the nine state estimates and associated statistics.

In the tables, B, the bias indicator is partitioned into three sets:

Bias Indicator (B)	Expected Result
B > 1.05	BHR underestimate the variance
1.05 > B > 0.95	BHR provides appropriate variance estimates
B < 0.95	BHR overestimates the variance

Bootstrap Bias Indicator

As stated before, the concern with the bootstrap variance estimator is its biased when two units are selected within a stratum, as is the case, with the simulation design. With districts, this bias is difficult to measure. However, the bias should be larger in states that have more districts solely defined in only one substratum. To measure this, the proportion of each state's districts that are totally within a single substratum is calculated.

In the tables, the proportion of districts in 1 stratum is divided into three groups:

Bootstrap Bias Indicator (B)	Expected Results
B > 0.2	most bias
0.2 > B > .08	some bias
B < 0.08	least bias

This proportion will be used as a potential bias indicator for the bootstrap variance estimates.

S_p/σ_p, S_e/σ_e and S_b/σ_b

Besides the confidence coefficient, the ratio of the average estimated standard error (probability or expected hits weight with BHR; or probability weight with Bootstrapping), across the 90 simulation, over the estimated true standard error (probability or expected hits weight with BHR; or probability weight with Bootstrapping) is another measure of the accuracy of the variance estimates.

σ_p/σ_e

Since the accuracy of BHR variance estimation for two weighting schemes are being compared, it is important to know which has the smaller standard error, irrespective of whether the BHR techniques work. This is done by looking at the ratio of the estimated true standard error using the probability weight (σ_p) divided by estimated true standard error using the expected hits weight (σ_e).

Weighting the Estimates

Each of the statistics described above is computed for the nine estimate within each of the 51 states. These 459 estimates with their respective sample variance estimates, estimated true variances and other statistics are summarized by type of estimate - averages, ratios and totals. Since there are differential numbers of these type of estimates (five totals for every two averages and ratios), an important consideration is how these estimates should be weighted. Within a state, the estimates are equally weighted by estimate type with high and lower correlated totals being weighted equally. All summary statistics in the tables are weighted averages using the weights describe above.

Results

Probability Weight and BHR Variances

When analyzing table 1, BHR overestimates the true standard errors. Where the bias is expected to be positive, averages, ratios and total all have a large upward bias, with confidence coefficients as high as 81%, on average. Where the bias is not expected to be positive, there is still a positive bias, but at a more acceptable level. The confidence coefficients using the true standard errors are all close to 0.68, so the difference is caused from the BHR procedures and not from the distribution of the estimates. There are eight states, where the bias is expected to be positive.

Expected Hit Weight and BHR Variances

Table 2 is the same as tables 1 except the expected hits weight is used instead of the probability weight. The results for the expected hits weight are very different than the probability weight results. There doesn't appear to be any significant bias with the expected hits weight. The BHR standard errors are all close to the true standard error. The largest difference occurs where the bias is expected to be positive, in which case, BHR overestimates the standard error by 10% for averages. The BHR confidence coefficients are all close to the coefficients based on the estimated true standard error.

Probability and Expected Hits Weight

Since the BHR variance estimate are less bias using the expected hits weight, one must ask whether estimates using the expected hits weight are as reliable as the estimates based on the probability weight. If the answer is yes, then the expected hits weight should be used instead. Table 1 and 2 provide the ratio of the true standard error using the probability weight over the true standard error using expected hits weight (σ_p/σ_e).

For averages and ratios the probability weight estimates have smaller standard errors. In table 1, the gains in precision range from an average of 2% to an average of 19% over the precision of expected hits weight estimates. In table 2, the gains range from an average of 3% to an average of 12%.

For totals, the expected hit weight estimates have smaller standard errors. In table 1, the probability weight estimate's precision ranges from an average of 12% to an average of 48% larger than the expected hits weight's precision. In table 2, the probability weight's loss of precision ranges from an average of 26% to an average of 34%.

For two out of the three types of estimates the probability weight estimates are better than the expected hits weight estimates. Overall, the probability weight is better. However, if totals are the primary interest then the expected hits weight provides the best estimates. Since none of the totals in the simulation study are uncorrelated with the selection measure of size, performance of such totals is unknown. If the expected hits weight performs poorly with uncorrelated totals, it may not be advisable to use the expected hits weight.

Probability Weight and Bootstrap Variances

Table 3 uses the proportion of districts within a state that are solely in 1 stratum as a bias indicator, to compare the bootstrap standard error to the true standard error. Where the bias is expected to be smallest, the bootstrap standard error estimator using the probability weight provides good standard error estimates. The bootstrap estimates are on average 10%, 6% and 2% smaller than the estimated true standard error respectively for averages totals and ratios. The confidence coefficients are 0.66, 0.72 and 0.71 on average. Where the bias is not expected to be smallest, the bootstrap estimator doesn't do as well, and underestimates the true standard error.

As stated before the BHR variance estimator does not work in eight states. In these eight states, the bootstrap variance estimator did work well.

Overall, the bootstrap standard error estimates preform poorly. This poor performance seems to be related to the inherent n/(n-1) bias of bootstrap variance estimator. When a state has all of their district solely in one stratum, this bias will be large because the sample design only selects two units per stratum. This implies that the bootstrap variance will be 1/2 the true variance. When a state has few district solely in one stratum, the bias is small and the result show this.

Using a sample design that selects more than two units per stratum should improve the bootstrap variance estimates.

Distribution of District's Selection Probabilities

In the introduction, I suggested the selection probability distribution would determine whether BHR would provide reasonable variance estimates. For this simulation, BHR does not work well when more than 20% of the district selection probabilities are larger than 0.95. Other surveys, using an aggregation collection methodology, should review the selection probabilities of the aggregation units. If more than 20% of the probabilities are larger than 0.95 then the BHR variances may be biased.

Conclusions

This simulation study has shown that when the collection unit is an aggregation of the selection units then BHR may not provide reasonable variance estimates. If the weight is based on the aggregation unit's selection probability then the bias can be large when more than 20% of the probabilities are larger than 0.95. BHR assumes that the true variance is inversely proportional to the sample size. This assumption is not necessary true with this design and it appears that the violation of this assumption is the cause of BHR bias in eight states.

If the expected hits weight is used then the variances do not appear to be biased. However, average and ratio estimates derived using the expected hits weight are not as precise, as estimates based on the aggregation unit's selection probability. If totals are the only estimates of importance then the expected hits weight is better.

Using the simulation design and bootstrap procedure described in this paper with the probability weight, some state's variances are best using the bootstrap methodology. In these states, the effective sampling sizes in the "stratum" are large enough to introduce only a small bias. However, bootstrapping did not work for most states because the effective sample sizes in a "stratum" are to small.

Ongoing Activities

Currently, I am trying to extend the bootstrap methodology to a systematic probability proportionate to size selection procedure where n (>2) schools are selected per stratum. With a larger stratum sample size, I'm hoping the bootstrap bias will be smaller. So far, the preliminary results are encouraging when compared with BHR variance estimates.

References

1. Kaufman, Steven (1991). 1988 Schools and Staffing Survey Sample Design and Estimation. Technical Report NCES 91-127, U.S. Department of Education, National Center for Education Statistics.

2. Wolter, K. M. (1985). Introduction to Variance Estimation. New York: Springer-Verlag, p110-145

Table 1 -- Probability weight BHR standard errors by estimate type and bias indicator

Bias Indicator	Estimate Type	σ_p --- σ_e	s_p --- σ_p	One s_p Conf. Coeff	One σ_p Conf. Coeff
Neg Bias	average	0.98	1.04	0.67	0.68
Unbiased	average	0.96	1.10	0.69	0.67
Pos Bias	average	0.94	1.24	0.76	0.68
Neg Bias	total	1.48	1.06	0.69	0.69
Unbiased	total	1.35	1.15	0.73	0.68
Pos Bias	total	1.12	1.43	0.81	0.68
Neg Bias	ratio	0.94	1.12	0.71	0.69
Unbiased	ratio	0.89	1.18	0.74	0.68
Pos Bias	ratio	0.81	1.42	0.80	0.68

Table 2 -- Expected hits weight BHR standard errors by estimate type and bias indicator

Bias Indicator	Estimate Type	σ_p --- σ_e	s_e --- σ_e	One s_e Conf. Coeff	One σ_e Conf. Coeff
Neg Bias	average	0.93	0.94	0.60	0.66
Unbiased	average	0.97	1.03	0.67	0.68
Pos Bias	average	0.96	1.10	0.70	0.68
Neg Bias	total	1.31	0.98	0.65	0.67
Unbiased	total	1.34	1.03	0.68	0.68
Pos Bias	total	1.26	1.07	0.68	0.67
Neg Bias	ratio	0.88	0.95	0.64	0.68
Unbiased	ratio	0.88	1.02	0.66	0.69
Pos Bias	ratio	0.88	1.07	0.69	0.67

Table 3 -- Bootstrap standard errors by estimate type and bias indicator

Bias Indicator	Estimate Type	Bias Indic. Size	s_b --- σ_b	One s_b Conf. Coeff	One σ_b Conf. Coeff
High Bias	average	0.48	0.79	0.56	0.69
Some Bias	average	0.14	0.77	0.53	0.67
Least Bias	average	0.02	0.90	0.66	0.73
High Bias	total	0.48	0.92	0.63	0.69
Some Bias	total	0.14	0.97	0.65	0.68
Least Bias	total	0.02	1.06	0.72	0.71
High Bias	ratio	0.48	0.85	0.58	0.70
Some Bias	ratio	0.14	0.91	0.63	0.69
Least Bias	ratio	0.02	1.02	0.71	0.72

Characteristics of Nonrespondents in the Schools and Staffing Surveys' School Sample**

Pao-Sheng Shen, Randall J. Parmer and Andre I. Tan, U.S. Bureau of the Census
Pao-Sheng Shen, U.S. Bureau of the Census, Washington, D.C. 20233

1. Introduction

In the 1991 Schools and Staffing Survey (SASS), there are some nonrespondents. One strategy for adjusting for nonresponse is to estimate the variables of interest with a poststratification estimator. Each respondent observation is weighted by the inverse of the respondent proportions of the observations in its cell, which is defined on the auxiliary variables such as grade level, enrollment and urbanicity. In doing this, one is implicitly modeling the nonresponse mechanism by assuming that the probability of nonresponse may vary among cells but not within cells. Hence, it is important to choose suitable adjustment cells such that the response probabilities of individuals within cells are as homogeneous as possible. This approach is discussed in detail by Schaible (1979).

The first objective of our research is to identify the auxiliary variables correlated with nonresponse and make recommendations for nonresponse adjustment cells. The second objective is to identify subpopulation with low response rate where field resources can be concentrated to improve the overall response rate. The data used are from a sample of 8995 public schools and 2741 list frame private schools.

Section 2 of the article presents a brief description of the 1991 SASS. In Section 3, we discuss the methodology. To identify the auxiliary variables correlated with nonresponse, adjusted Chi-square tests are used for testing the correlation between the auxiliary variables and response status. For estimation of response rates in subpopulations, due to the small subpopulation sizes, procedures depending on the distribution created by the sampling plan are unstable or not available. The logit-estimates, which are simply the application of the "pseudo" maximum likelihood estimate (pseu-MLE) from Roberts, Rao and Kumer (1987), were used to estimate the response rates for subpopulations of public schools. For private schools, subpopulation sample sizes are too sparse to support the existence of a unique pseu-MLE. Hence, empirical Bayasian-logit estimates which are based on the "pseudo" maximum posterior estimate (pseu-MPE) defined in Section 3.3 were used as alternatives. Section 4 contains a summary of our results and conclusions.

2. The 1991 Schools and Staffing Survey (SASS)

2.1 Frame Construction

The 1991 Schools and Staffing Surveys consists of a school, a teacher, and for public schools a Local Education Agency or school district survey. Public schools were identified on the Common Core of Data or CCD. This CCD was matched to the previous SASS public school sampling frame. Non-matches from the previous frame were included with the CCD to make up the public school sampling frame for 1991. Public schools were stratified by state, grade level, and Indian/non-Indian.

The private schools were selected from a list frame, constructed by matching multiple lists obtained from private school organizations, State Departments of Education, and a private vendor. This frame is thought to include 80-90% of private schools. To increase the coverage of the survey, an area frame was constructed by selecting 120 PSUs, consisting of counties or groups of counties. Within these sample counties, lists of schools were obtained from local sources, such as yellow pages, churches and fire marshals. These lists were unduplicated with the list frame. The remaining schools, not matching to the list frame, make up the area frame.

2.2 Design

Public schools were stratified by state, grade level, and Indian/non-Indian. Probabilities of selection were computed, proportional to the square root of the number of teachers in the school conditioned on the 1988 selection. The probabilities were adjusted to obtain the desired proportion of overlapping schools from 1988. Approximately 9900 public sample schools were selected systematically within each stratum.

Private schools were stratified by affiliation, grade level, and census region for the list frame, and by PSU and grade level for the area frame. Probabilities of selection were computed and adjusted similarly to the public schools. Approximately 3300 private schools were selected, systematically within each stratum.

2.3 Data Collection

School questionnaires were mailed to schools. They were asked to fill them out and mail them back to the Census Bureau. After four weeks, if the school hadn't responded, we sent out a second questionnaire. If after three more weeks the school hadn't responded, we called them and attempted to complete the interview by telephone. Schools still not responding by telephone were classified as noninterviews.

2.4 Estimation

Schools' probabilities of selection were adjusted for school merges and other situations that would affect the probability of selection. The inverse of the probability of selection became the basic weight. This basic weight was adjusted to account for noninterviews using noninterview adjustment cells. A ratio adjustment was also applied which adjusted the characteristics of the sample schools to the characteristics of the whole sample frame.

3. Methodology

3.1 Testing

The response status (yes or no) is considered to be the response variable. The continuous auxiliary variables are divided into 2 -5 groups. The standard Chi-Squared tests for independence (denoted as X_I^2 when auxiliary variables are not used for stratification) or tests for homogeneity (denoted as X_H^2 when auxiliary variables are used for stratification) are not appropriate due to the complex sample design of SASS. As a result, some adjustments that take into account the design are necessary in order to make valid inferences from survey data. Rao and Scott (1984) derived a first-order correction denoted by δ to the standard Chi-Squared test which requires the knowledge of only the cell design effects (deffs) and the deffs for marginals provided the model admits a direct solution to likelihood equations under multinomial sampling. These results are applicable to the test results in our study.

However, because of a shortage of information on cell deffs, only X^2_I and some of X_H^2 were adjusted. The reason is that the empirical study by Holt, Schott and Ewings (1980) indicated that the distortion of nominal significance level is substantially smaller with X_I^2 than with X_H^2. The deffs for adjusting tests were obtained based on the estimated variance of all the individual cells using 48 pseudo-replicates originated by the U.S. Bureau of the Census (Simmons and Baird (1968)).

** This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

3.2 Subpopulation Estimation

Some variables of interest that we identified as correlated with nonresponse, were chosen to construct subpopulations. By the levels of the variables chosen, the populations of public schools and list frame private schools were divided into 20 and 48 subpopulations respectively. In certain subpopulations, sample sizes are too small to have accurate estimates by using the standard methods based on the selection probabilities. One strategy is to borrow information across subpopulations by using an unsaturated logit regression model. Due to difficulties in obtaining appropriate likelihood functions for our design, "pseudo" maximum likelihood estimates (pseu - MLE) (Roberts, Rao and Kumar (1987)) can be used to replace maximum likelihood estimates (MLE) of regression coefficients. This strategy was implemented on the estimation of response rates for subpopulations of public schools and the estimates based on the regression model are referred to as logit-estimates. However, with only 2741 samples for list frame private schools, the observed response or nonresponse frequencies are zero for some subpopulations. These conditions may make pseu-MLE not unique (Albert and Anderson (1984)). To solve the existence problem, an empirical Bayes approach was proposed and Bayesian-logit estimates were used as alternatives.

The approach is described in Section 3.3. The goodness-of-fit of the model was based on a likelihood ratio test corrected by an upper bound on δ proposed by Rao and Scott (1987). The upper bound can be obtained using information on cell deffs and marginal deffs. The test is conservative and applicable to the model not admitting a direct solution to the likelihood equation.

3.3 An Empirical Bayesian Approach for Subpopulation Estimation of Binary Data from Complex Sample Surveys

Without loss of generality, suppose that the population is partitioned into IxJ subpopulations according to factor A_i ($i=1,\dots,I$) by factor B_j ($j=1,\dots,J$). Let $P=(P_{11},\dots,P_{IJ})'$ where P_{ij} denotes the proportions that schools in the ij[th] subpopulation are respondents. Let $\hat{N}_{ij}$ denote the 1991 SASS survey estimate of the ij[th] subpopulation total $\hat{N}_{ij1}$ the corresponding estimate of response frequencies. With large $\hat{N}_{ij}$ and reasonably large frequencies, $\hat{N}_{ij1}$, the ratio estimate

$$\hat{p}_{ij} = \frac{\hat{N}_{ij1}}{\hat{N}_{ij}}$$

is often used to estimate P_{ij}. When the data are too few $\hat{p}_{ij}$ can be very unstable. In this situation, it seemed much more appropriate to borrow information across subpopulations by using an unsaturated logistic model. A logistic regression model for the response rate P_{ij} of the ij[th] subpopulation is given by $P_{ij}=f_{ij}(\beta)$, where

$$\log\left[\frac{f_{ij}(\beta)}{1-f_{ij}(\beta)}\right] = X'_{ij}\beta \tag{3.1}$$

In (3.1) X_{ij} is an S-vector of known constants derived from the factor levels and β is an S-vector of unknown parameters. The pseu-MLE of β can be obtained from solving the following "pseudo" likelihood equations through iterative calculations:

$$X'D(\Delta)f(\hat\beta) = X'D(\Delta)\hat{p} \tag{3.2}$$

where $X = (X_{11},\dots,X_{IJ})$ is an SxIJ matrix of rank S, $D(\Delta) = diag(\Delta_{11},\dots,\Delta_{IJ})$,

$\Delta_{ij} = \frac{\hat{N}_{ij}}{\sum_{ij}\hat{N}_{ij}} = \frac{\hat{N}_{ij}}{\hat{N}}$ is the estimated subpopulation relative size ω_{ij},

$\hat{p} = (\hat{p}_{11},\dots,\hat{p}_{IJ})'$ and $f(\hat\beta)=(f_{11}(\hat\beta),\dots,f_{IJ}(\hat\beta))'$.

Under the assumption that $n^{1/2}(\hat{p} - f(\hat\beta))$ converge in distribution to N(0, V), the estimated asymptotic covariance matrix of $f(\hat\beta)$ is (Robert, Rao and Kumar (1987))

$$\hat{V}_f = [D(\Delta)]^{-1}\hat{\Delta}X\hat{V}_\beta X'\hat{\Delta}[D(\Delta)]^{-1} \tag{3.3}$$

where

$\hat{\Delta} = diag(\Delta_{11}\hat{f}_{11}(1-\hat{f}_{11}),\dots,\Delta_{IJ}\hat{f}_{IJ}(1-\hat{f}_{IJ}))$ and

$\hat{V}_\beta = n^{-1}(X'\hat{\Delta}X)^{-1}[X'D(\Delta)\hat{V}D(\Delta)X](X'\hat{\Delta}X)^{-1}$

where $\hat{V}$ denote the survey estimate of the covariance matrix V.

However, when any of $\hat{N}_{ij1}$ or $\hat{N}_{ij0}$ $(=\hat{N}_{ij}-\hat{N}_{ij1})$ is zero, a unique pseu-MLE $\hat\beta$ may not exist for the regression model considered. A sufficient condition for the existence of a unique $\hat\beta$ is $0 < \hat{N}_{ij1} < \hat{N}_{ij}$ for all i, j (Albert and Anderson (1984)). The empirical Bayesian approach developed next solves the existence problem and has an intuitively appealing interpretation. First, we model the distribution of P_{ij} as a Beta distribution with parameters a_i and b_i. That is, we will assume the hierarchical prior $P_{ij} \overset{iid}{\sim} Beta(a_i,b_i)$ for $i=1,\dots,I$ and $j=1,\dots,J$, so that the P_{ij} have density function

where $B(a_i,b_i)$ is the complete beta function.

$$h(P_{ij}) = [B(a_i,b_i)]^{-1} P_{ij}^{a_i-1}(1-P_{ij})^{b_i-1}$$

Note that the hierarchial prior model is equivalent to grouping like subpopulations (with the same level of factor A) into strata (different levels of factor A) and modeling the subpopulations within a stratum to have a common distribution. For list frame private schools, based on the data 1988 and 1991 SASS, the variation of response rate within each association is smaller than the variation of response rate among the associations. Also, there were reasonably large sample sizes in each association. Hence association was used as factor A and the combination of the other variables was used as factor B to construct the hierarchical model.

Next, we estimate a_i and b_i for $i=1,\dots,I$ from the marginal distribution of data by integrating the following "pseudo" likelihood equation with respect to P_{ij}.

$$\prod_{ij}\left[(P_{ij})^{n\left(\frac{\hat{N}_{ij1}}{\hat{N}}\right)+a_i-1}(1-P_{ij})^{n\left(\frac{\hat{N}_{ij0}}{\hat{N}}\right)+b_i-1}[B(a_i,b_i)]^{-1}\right] \qquad (3.4)$$

The result of integration of (3.4) is

$$\prod_{ij}\left[B\left(n\left(\frac{\hat{N}_{ij1}}{\hat{N}}\right)+a_i,\ n\left(\frac{\hat{N}_{ij0}}{\hat{N}}\right)+b_i\right)[B(a_i,b_i)]^{-1}\right] \qquad (3.5)$$

The expression in equation (3.5) is maximized under the constraint $a_i>0$, $b_i>0$ using numerical method to obtain the pseu-MLE of a_i and b_i denoted by $\hat{a}_i$ and $\hat{b}_i$ for $i=1,\dots,I$. The value of β obtained by solving the following equation will be called the pseu-MPE of β and denoted by $\hat{\beta}$ and the estimator $f(\hat{\beta})$ will be referred as empirical Bayesian-logit estimator.

$$X'D(\hat{\omega})f(\hat{\beta}) = X'D(\hat{\omega})\hat{P} \qquad (3.6)$$

where $\hat{P} = (\hat{P}_{11},\dots,\hat{P}_{IJ})'$, $\quad \hat{P}_{ij} = \dfrac{n\left(\frac{\hat{N}_{ij1}}{\hat{N}}\right)+\hat{a}_i}{n\left(\frac{\hat{N}_{ij}}{\hat{N}}\right)+\hat{a}_i+\hat{b}_i}$

$D(\hat{\omega}) = diag(\hat{\omega}_{11},\dots,\hat{\omega}_{IJ})$,

$f(\hat{\beta}) = (f_{11}(\hat{\beta}),\dots,f_{IJ}(\hat{\beta}))' \qquad \hat{\omega}_{ij} = \dfrac{\left[n\left(\frac{\hat{N}_{ij}}{\hat{N}}\right)+\hat{a}_i+\hat{b}_i\right]}{n+J\sum_i(\hat{a}_i+\hat{b}_i)}$

The pseu-MPE always exists since $\hat{a}_i > 0$ *and* $\hat{b}_i > 0$ for all i.

REMARK.

First, note that $\hat{P}_{ij}$ can be written as

$$\left[\pi_{ij}\hat{P}_{ij} + (1-\pi_{ij})\left(\frac{\hat{a}_i}{\hat{a}_i+\hat{b}_i}\right)\right]I_{(\hat{N}_{ij}>0)} + \left(\frac{\hat{a}_i}{\hat{a}_i+\hat{b}_i}\right)I_{(\hat{N}_{ij}=0)}$$

where

$$\pi_{ij} = \frac{\tau_i^2}{\tau_i^2+\delta_{ij}^2} \qquad \tau_i^2 = \frac{\hat{a}_i\hat{b}_i}{(\hat{a}_i+\hat{b}_i)^2(\hat{a}_i+\hat{b}_i+1)}$$

$$and \quad \delta_{ij}^2 = \frac{\hat{P}_{ij}(1-\hat{P}_{ij})}{n\left(\frac{\hat{N}_{ij}}{\hat{N}}\right)}$$

Note that τ_i^2 is the estimated variance of the i^{th} stratum based on the superpopulation model (prior distribution imposed on P) and δ_{ij}^2 is an intuitively estimated sampling variance for the ij^{th} subpopulation when the subpopulation sample size is not zero. The smaller the sampling variance relative to stratum variance, the more weight $\hat{P}_{ij}$ gets. Just as intuitively reasonable, for large relative sampling variance, which can be defined as infinity when sample size is zero, little weight should be given to $\hat{P}_{ij}$, and there should be a borrowing of strength from the other observations in the same stratum. Secondly, under the model (3.1), it follows that β has a prior $\pi(\beta)$ in the form

$$\pi(\beta)=\prod_{ij}[f_{ij}(\beta)]^{a_i}[1-f_{ij}(\beta)]^{b_i}$$

Solving the Equation (3.6) is equivalent to maximizing the following "pseudo" posterior likelihood function with respect to β

$$\prod_{ij}[f_{ij}(\beta)]^{n\left(\frac{\hat{N}_{ij1}}{\hat{N}}\right)+\hat{a}_i}[1-f_{ij}(\beta)]^{n\left(\frac{\hat{N}_{ij0}}{\hat{N}}\right)+\hat{b}_i}$$

The conditional asymptotic covariance of $f(\hat{\beta})$ can be derived as follows:

LEMMA. Let β_0 denote the conditional expected value of $\hat{\beta}$ when $P = P_0$.

Suppose that

(A) The conditional distribution of $n^{1/2}(\hat{P} - f(\beta_0))$, as n tends to infinity, is normal with mean 0 and variance V_0 and,

(B) For all i,j, we have $\hat{\omega}_{ij} - \omega^*_{ij} = o_p(1)$ where ω^*_{ij} 's are some design-dependent constants.

then the conditional asymptotic variance of $f(\hat{\beta})$, denoted by V_f , is $V_f = [D(w^*)]^{-1} \Delta_0 X V_\beta X' \Delta_0 [D(w^*)]^{-1}$

where

$$V_\beta = n^{-1}(X'\Delta_0 X)^{-1}[X'D(\omega^*)V_0 D(\omega^*)X](X'\Delta_0 X)^{-1} \quad (3.7)$$

$D(\omega^*) = \mathrm{diag}(\omega^*_{11},\ldots,\omega^*_{IJ})$ and

$\Delta_0 = \mathrm{diag}\{\omega^*_{11} f_{11}(\beta_0)[1-f_{11}(\beta_0)],\ldots,\omega^*_{IJ} f_{IJ}(\beta_0)[1-f_{IJ}(\beta_0)]\}$

Proof:

Let $U(\beta, D(\hat{\omega})) = \sum_i U_i(\beta, D(\hat{\omega})) = \sum_i X'_i \Delta_i (P_i - f_i(\beta))$.

By Equation (3.6), $U(\hat{\beta}, D(\hat{\omega})) = 0$

Under the assumption (B),

$$U(\beta, D(\hat{\omega})) = U(\beta, D(\omega^*)) + o_p(1) \text{ for all } \beta \quad \text{as } n \to \infty$$

Now, treat $U(\beta, D(\omega^*))$ as a function of β only and denoted by $L(\beta)$. Regularity conditions are satisfied by $L(\beta)$ and as n is large $n^{1/2}(\hat{\beta} - \beta_0)$ can, using a Taylor expansion, be approximated by

$$\left\{\left[\frac{\partial}{\partial\beta}L(\beta)\right]_{\beta=\beta_0}\right\}^{-1} L(\beta_0) = \left\{\sum_i\left[-\frac{\partial}{\partial\beta}U_i(\beta, D(\omega^*))\right]_{\beta=\beta_0}\right\}^{-1}\left[n^{1/2}\sum_i U_i(\beta_0, D(\omega^*))\right]$$
$$= (X'\Delta_0 X)^{-1} X' D(\omega^*)\, [n^{1/2}(P - f(\beta_0))]$$

Under the assumption (A), it follows that $n^{1/2}(\hat{\beta} - \beta_0)$, as n tends to infinity, converges in distribution to $N(0, V_\beta)$.

Similarly, noting that

$$n^{1/2}(f(\hat{\beta}) - f(\beta_0)) = \left[\frac{\partial f(\beta)}{\partial\beta}\right]_{\beta=\beta_0}[n^{1/2}(\hat{\beta} - \beta_0)] = [D(\omega)]^{-1}\Delta_0 X\, [n^{1/2}(\hat{\beta} - \beta_0)]$$

it follows that $n^{1/2}(f(\hat{\beta}) - f(\beta_0))$, as n tends to infinity, converges in distribution to $N(0, V_f)$.

Let $\hat{V}$ denote the survey estimate of the covariance matrix V_0 (given the prior parameters a_i and b_i for $i = 1,\ldots,I$). Then (3.7) can be estimated by

$$\hat{V}_\beta = (X'\hat{\Delta}X)^{-1}[X'D(\hat{\omega})\hat{V}D(\hat{\omega})X](X'\hat{\Delta}X)^{-1}$$

where $\hat{\Delta} = \mathrm{diag}\ \{\hat{\omega}_{11} f_{11}(\hat{\beta})[1-f_{11}(\hat{\beta})],\ldots,\hat{\omega}_{IJ} f_{IJ}(\hat{\beta})[1-f_{IJ}(\hat{\beta})]\}$.

Similarly, the asymptotic covariance of $f(\hat{\beta})$ can be estimated by

$$\hat{V}_f = [D(\hat{\omega})]^{-1}\hat{\Delta}X\hat{V}_\beta X'\hat{\Delta}[D(\hat{\omega})]^{-1} \quad (3.8)$$

In our study, the computer programs were written in SAS® to perform the required maximization of the logarithm of equation (3.5) to obtain the estimated prior parameters $\hat{a}_i$ and $\hat{b}_i$ for all i. Then SAS/CATMOD was used to obtain the $\hat{\beta}$ and $\hat{\beta}$ for public schools and private schools respectively. Due to small sample sizes for certain subpopulations, a pseudo-replication scheme is not applicable to the estimation of V and V_β . One way around this is to aggregate, temporarily, some of the subpopulations of small sample sizes to the same group. In other words, define disjoint groups of subpopulations and implement a pseudo-replication scheme to estimate the covariance of groups. Assign the estimated group standard deviation to all subpopulations belonging to the same group. In our study, this strategy was used to obtain subpopulation design effects, $\hat{V}$ and $\hat{V}$ and then $\hat{V}_f$ and $\hat{V}_f$ were calculated.

4. Results and Conclusions

4.1 Testing

Table 1 and Table 2 illustrate the estimated deffs $\hat{\delta}$ and results of $X^2_I, X^2_W, X^2_I/\hat{\delta}$ and $X^2_W/\hat{\delta}$ for some auxiliary variables selected. From Table 1 and 2, we note that the deffs for public schools are much higher than those for list frame private schools. One explanation is that our design is not planned to reduce the variance of the estimation of response rate. However, it happened that for private schools, both grade level and association, which are strongly correlated with response status, were used for stratification, while for public schools, grade level, which is used for stratification, is weakly correlated with response status. Even though design effects for public schools are very high, it turned out that the size of the modified tests based on $X^2_I/\hat{\delta}$ was significant at $\alpha = .001$ for urbanicity and at $\alpha = .01$ for enrollment and the modified test based on $X^2_W/\hat{\delta}$ was significant for state at $\alpha = .001$. For private schools, the size of the modified tests based on $X^2_W/\hat{\delta}$ was significant at $\alpha = .001$ for grade level and association. The size of the modified tests based on $X^2_I/\hat{\delta}$ was significant at $\alpha = .001$ for affiliation and urbanicity and at about $\alpha = .10$ for enrollment.

Table 1
Test Results for Public Schools

Variables	df	δ	X^2_G	$*X^2_G/\delta$	X^2	$**X^2/\delta$
Grade Level	1	12.1	[illegible]	1.8	.	.
[illegible]	20	[illegible]	[illegible]	[illegible]	.	.
Urbanicity	1	[illegible]	.	.	[illegible]	47.9
[illegible]	4	13.6	.	.	213.6	13.8

Table 2
Test Results for Private Schools

Variables	df	δ	X^2_G	X^2_G/δ	X^2	X^2/δ
Grade Level	1	2.3	[illegible]	[illegible]	.	.
Urbanicity	1	1.7	.	.	157.4	[illegible]
Enrollment	1	2.0	.	.	4.1	2.0
Affiliation	1	1.8	.	.	172.7	[illegible]
Association	4	2.5	[illegible]	[illegible]	.	.

*For testing homogeneity:

$$\delta = \frac{1}{(I-1)(J-1)} \sum_{i=1}^{I} \sum_{j=1}^{J} (1-\hat{P}_j)\, \hat{d}_{j(i)}\, (1-n_{i+}/n)$$

where n is sample size, n_i+ is row margin, $\hat{P}_j = \sum_i (n_{i+}/n)\hat{P}_{j(i)}$, $\hat{P}_{j(i)}$ is the estimated cell proportions within the i^{th} row population and $\hat{d}_{j(i)} = n_{i.}\, \hat{V}(\hat{P}_{j(i)})/\hat{P}_j(1-\hat{P}_j)$ is the estimated deffs of $\hat{P}_{j(i)}$.

where $\hat{V}(\hat{P}_{j(i)})$ denotes the estimated variance of $\hat{P}_{j(i)}$.

**For testing independence:

$$\delta = \frac{1}{(I-1)(J-1)} \sum_{i=1}^{I} \sum_{j=1}^{J} (1-\hat{P}_{i.}\hat{P}_{.j})\hat{d}_{ij} - \sum_{i=1}^{I}(1-\hat{P}_{i.})\hat{d}_i(r) - \sum_{j=1}^{J}(1-\hat{P}_{.j})\hat{d}_j(c)$$

where $\hat{P}_{i.}$ and $\hat{P}_{.j}$ are the estimated row and column marginals proportion, $\hat{d}_i(r)$ *and* $\hat{d}_j(c)$ are the estimated deffs of $\hat{P}_{i.}$ and $\hat{P}_{.j}$ respectively, and $\hat{d}_{ij}$ is the estimated deff of $\hat{P}_{ij}$, which is the estimated proportion for the ij^{th} cell.

4.2 Subpopulation Estimation

For public schools, the population was divided into 20 subpopulations by grade level, urbanicity and enrollment. Based on the unadjusted chi-square of each term, some interaction terms appear to be nonsignificant and are excluded from the full model. The following reduced model was chosen to explain the variation in the response rate.

$$V_{ijk} = \log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} \qquad (4.1)$$

where π_{ijk} denotes the response rate of the ijk^{th} subpopulation.

α_i denotes the effect of the i^{th} enrollment, $i = 1, \ldots, 5$.

β_j denotes the effect of the j^{th} urbanicity, $j = 1, 2$.

γ_k denotes the effect of the k^{th} grade level, $k = 1$,

$(\alpha\beta)_{ij}$ denotes the interaction of the i^{th} enrollment by the j^{th} urbanicity.

$(\alpha\gamma)_{ik}$ denotes the interaction of the i^{th} enrollment by the k^{th} grade level.

Similarly, for list frame private schools, the population was divided in 48 subpopulations by association, grade level, urbanicity, and enrollment. The following model was chosen to explain the variation in the response rate.

$$V'_{ijkl} = \log\left(\frac{\pi'_{ijkl}}{1-\pi'_{ijkl}}\right) = \mu' + \alpha'_i + \beta'_j + \gamma'_k + \lambda'_l + (\alpha'\beta')_{ij} + (\alpha'\gamma')_{ik} + (\alpha'\lambda')_{il} + (\beta'\lambda')_{jl} + (\alpha'\gamma'\lambda')_{ikl} + (\beta'\gamma'\lambda')_{jkl} \qquad (4.2)$$

where π'_{ijkl} denotes the response rate of the $ijkl^{th}$ subpopulation

α'_i denotes the effect of the i^{th} association, $i = 1, \ldots, 6$.

β'_j denotes the effect of the j^{th} grade level $j = 1, 2$.

γ'_k denotes the effect of the k^{th} urbanicity, $k = 1, 2$.

λ'_l denotes the effect of the l^{th} enrollment, $l = 1, 2$.

$(\alpha'\beta')_{ij}$ denotes the interaction of the i^{th} association by the j^{th} grade level, and so on.

For testing the goodness-of-fit of the model, the adjusted likelihood ratio test proposed by Rao and Scott (1987) were used. The adjustment is based on the upper bound on δ which requires the information of cell deffs (subpopulation deffs). The deffs of the subpopulation were estimated using 48 pseudo-replicates. The estimated deff for the i^{th} subpopulation is equal to

$$\hat{Var}(\hat{P}_i) \,/\, (n\hat{W}_i)^{-1}\, \hat{P}_i\, (1-\hat{P}_i)$$

where $\hat{P}_i$ is the estimated response rate for the i^{th} subpopulation

$\hat{Var}(\hat{P}_i)$ is the estimated variance of $\hat{P}_i$ using 48 pseudo-replicates

$\hat{W}_i$ is the estimated relative size for the i^{th} subpopulation

n is the total sample size.

For public schools, the upper bound on δ is estimated by the average deffs available (= 6.4) and multiplied by $R_1/R_1 - m_1$, where R_1 is the number of subpopulations (= 20; and m_1 (= 15) is the number of parameters to be estimated for model (4.1). Hence the upper bound was estimated by (6.4)(20/5) = 25.7. The result for the adjusted likelihood ratio = (2.4)/25.7 = 0.09, which is not significant at the 5% level when compared to $X_5^2(0.05) = 11.1$ Note that due to the high deffs for public schools, the test is very conservative.

Similarly, for list frame private schools, the upper bound on δ is estimated by the average deffs available (= 2.1) and multiplied by $R_2/R_2 - m_2$, where R_2 is the number of subpopulation (= 48) and m_2 (= 31) is the number of parameters to be estimated for model (4.2). Hence the upper bound was estimated by (2.1)(48/17) = 5.9. The result for the adjusted likelihood ratio = 40.1/5.9 = 6.8, which is not significant at the 5% level when compared to $X_{17}^2(0.05) = 27.6$.

Based on model (4.1) and (4.2), the estimated response rate for subpopulations of public schools and private list frame schools are presented in Table 3 and Table 4 respectively. The corresponding estimated asymptotic standard deviations are also listed.

4.3. Conclusion

The empirical Baysian strategy used here for estimating response rate of subpopulations is a two-staged approach (one stage to estimate the prior, one stage to estimate the parameters given the estimated prior). The prior used is data-dependent. Although this strategy is not classical Baysian, it is in the spirit of an empirical-Bayesian procedure. This approach has the advantage of allowing information from all subpopulations to be used to provide estimates of response rate within each subpopulation. The disadvantage is that the computations are difficult. Under the hierarchical prior assumption, the estimated subpopulations' response rates were shrinked toward the marginal (association) response rate. The estimated asymptotic standard deviations did not include the uncertainty in the pseu-MLE of prior parameters. A possible remedy for this problem was suggested by Carlin and Gelfand (1991).

In summary, the variation of response rate for public schools is much smaller than that for private schools. For public schools, the nonresponse adjustment cells currently used by the U.S. Bureau of the Census are state by grade level by enrollment by urbanicity. Based on our results of testing, it seems to be a good choice. When further collapsing is necessary, cells can be collapsed with grade level first, enrollment second and urbanicity third. For private list frame, the nonresponse adjustment cells currently used by the U.S. Bureau of the Census are association by grade level by urbanicity. Based on the results of testing, it indicated that enrollment may also be a good candidate for creating nonresponse adjustment cells. If further collapsing is necessary, the cells can be collapsed with enrollment first, grade level second, urbanicity third and association fourth.

5. References

[1] Albert, A., and Anderson, J. A. (1984). On the Existence of Maximum Likelihood Estimates in Logistics Regression Models, Biometrika, 71, 1-10.

[2] Carlin, P. Bradley and Gelfand, E. Alan (1991). A Sample Reuse Method for Accurate Parametric Empirical Bayes Confidence Interval. Journal of Royal Statistical Society. Sec. B 53, No. 1, 189-200

[3] Holt, D., Scott, A. J., and Ewings, P. D. (1980). Chi-squared Test With Survey Data, Journal of Royal Statistical Society, Ser. A., 143, 302-320.

[4] Rao, J. N. K. and Scott, A. J. (1984). On Chi-squared Tests for Multi-way Contingency Tables With Cell Proportions Estimated from Survey Data. Ann. Statist., 12, 46-60.

[5] Rao, J. N. K. and Scott, A. J. (1987). On Simple Adjustments Chi-squared Tests With Sample Survey Data. Ann. Statist., 15, 385-397.

[6] Roberts, G., Rao, J. N. K. and Kumar, S. (1987). Logistic Regression Analysis of Sample Survey Data. Biometrika, 74, 1-12.

[7] Schaible, Wesley L. (1979), "Estimation of Finite Population Totals from Incomplete Sample Data: Prediction Approach," in Symposium on Incomplete Data, Preliminary Proceedings, Washington, D.C., U.S. Department of Health, Education, and Welfare, 170-187.

[8] Simmons, W.R. and Baird, J. (1968). Use of Pseudo-replication in the NCHS Health Examination Survey, Proceedings of the Social Statistics Section of the American Statistical Association.

Table 3

Estimated Response Rate and Asymptotic Standard Deviation for Subpopulation of Public Schools

		Urbanicity	
Enrollment	Grade Level	1	2
1	1	* [illegible]	* [illegible]
	2	[illegible]	[illegible]
2	1	[illegible]	[illegible]
	2	* [illegible]	* [illegible]
3	1	[illegible]	[illegible]
	2	* [illegible]	* [illegible]
4	1	[illegible]	[illegible]
	2	* [illegible]	* [illegible]
5	1	[illegible]	[illegible]
	2	[illegible]	[illegible]

The asymptotic standard deviation of the cells marked by "*" have been estimated based on the aggregation of subpopulations.

Table 4

Estimated Response Rate and Asymptotic Standard Deviation for Subpopulations of List Frame Private Schools

			Association					
Enrollment	Grade Level	Urbanicity	1	2	3	4	5	6
1	1	1	[illegible]	* [illegible]	* [illegible]	* [illegible]	* [illegible]	* [illegible]
1	1	2	[illegible]	* [illegible]	* [illegible]	* [illegible]	* [illegible]	* [illegible]
1	2	1	* [illegible]	* [illegible]	* [illegible]	* [illegible]	* [illegible]	* [illegible]
1	2	2	* [illegible]	* [illegible]	* [illegible]	* [illegible]	* [illegible]	* [illegible]
2	1	1	[illegible]	* [illegible]	* [illegible]	* [illegible]	* [illegible]	* [illegible]
2	1	2	[illegible]	* [illegible]	* [illegible]	* [illegible]	* [illegible]	* [illegible]
2	2	1	* [illegible]	* [illegible]	* [illegible]	* [illegible]	* [illegible]	* [illegible]
2	2	2	* [illegible]	* [illegible]	* [illegible]	* [illegible]	* [illegible]	* [illegible]

The asymptotic standard deviation of the cells marked by "*" have been estimated based on the aggregation of subpopulations.

Improving Reliability and Comparability of NCES Data on Teachers and Other Education Staff

Results of a CCSSO/NCES Workshop

Rolf K. Blank, Ph.D.

Paper presented at the 1992 Joint Statistical Meetings, August 11, 10:30 a.m.
Boston, MA

Council of Chief State School Officers
State Education Assessment Center
One Massachusetts Avenue, NW, Suite 700
Washington, DC 20001-1431
202/408-5505

This paper was prepared with assistance from colleagues at the Council: Barbara Clements, Melanie Dalkilic, and Lisa Solomon. The data workshop described in the paper was supported under a contract to the Council from the National Center for Education Statistics. Support for developing the paper was also provided by the National Science Foundation, Division of Research, Evaluation, and Dissemination.

Improving Reliability and Comparability of NCES Data on Teachers and Other Education Staff

INTRODUCTION

The National Center for Education Statistics (NCES) and the Council of Chief State School Officers (CCSSO) convened a Workshop on Improving Reliability and Comparability of Staffing Data on November 6-7, 1991. The Workshop was organized as a final summary activity of the State Data Project of the Schools and Staffing Survey (SASS), which was a field test of the use of state-collected data for a portion of SASS. Participants in the Workshop were representatives of state education agencies, local education agencies, and staff of NCES and CCSSO. The purposes of the Workshop were to analyze reasons for differences in how states and local districts aggregate and report staffing data, to understand constraints on data collection and reporting, and to recommend steps at federal, state, and local levels for improving data reliability and comparability.

This paper summarizes the proceedings and recommendations from the Workshop. The paper is based on a report of the Workshop prepared by staff of the Council of Chief State School Officers, State Education Assessment Center. There are four sections in the paper: Workshop Objectives, Analysis of 10 Issue Areas, Suggested Revisions of SASS Items, and Summary of Recommendations.

WORKSHOP OBJECTIVES AND DESIGN

The discussion of issues in collecting and reporting data on school staff was based on results from the SASS State Data Project. In 1989-90, NCES worked with eight states in the State Data Project of the Schools and Staffing Survey (SASS). The project was designed to test whether state education agencies have the capacity to provide data from their automated record systems that would otherwise be collected from individual local education agencies (LEAs) in the Schools and Staffing Survey, thereby saving the LEAs several hours of reporting burden. The participating state education agencies reported requested staffing data for a selected sample of LEAs, and the data were compared with data reported directly by the selected LEAs.

The Workshop was designed to identify reasons for differences in the data reported by states and local districts for the same items on the District Teacher Demand and Shortage survey of SASS. Participants were state education agency data managers from seven of the states that participated in the SASS State Data Project, an LEA data manager from one district in each of the seven states, and six state education agency data managers from other states.

The NCES and CCSSO staff organizing the Workshop viewed the issues to be discussed as having broader use than only an analysis of the results of the SASS

State Data Project. State and local differences in aggregating and reporting data on school staff are also of vital significance to the Common Core of Data, which includes annual NCES collection of staffing data through the state education agencies. States collect and report their own staffing data with state forms (sometimes called "administrative records"). These data are aggregated and reported to NCES in the CCD State Nonfiscal Survey. Another use of state staffing data is demonstrated in the CCSSO *Science/Math Indicators Project,* which reports state-by-state statistics on characteristics of science and mathematics teachers. The workshop was also an opportunity for NCES to receive input from local and state data managers on the SASS data collection items.

In the workshop, state and local education staff were divided into two discussion groups. The items included in the SASS State Data Project were organized in 10 issue areas, and the groups were asked to identify reasons for differences in state and local data under these topic areas. Then, the groups were asked to develop recommendations on ways to improve the reliability and comparability of data collected by SASS, CCD, states, and local districts.

Objectives Outlined by NCES

* A unique characteristic of the workshop was the inclusion of participants from local school districts. All of the participants recognized that the ultimate sources of these data are from the LEAs.

* The meeting had a fairly specific although wide ranging objective--to attempt to figure out why reports summarized at the state level differ from the reports that are received from the LEAs--which is the better set of numbers?

* NCES is concerned about the burden at the local level and how data quality can be improved.

* It is important to discover the problems of reporting teachers by field and in FTEs.

* NCES would like to identify if there is a better way to collect data on teacher demand and shortage.

* NCES and CCSSO can benefit from state and local experience and knowledge and develop an understanding of the common problems.

ANALYSIS OF 10 ISSUE AREAS

Outlined below are the two discussion groups' analyses of the problems with data quality and comparability and their recommendations. The findings of the two groups have been merged into one summary.

Issue 1. PreKindergarten and Kindergarten Enrollments and FTE Teachers

Problems:

There is some confusion as to whether districts are reporting enrollment vs. membership, and there is a clear difference in NCES definitions. Possibly a substitute phase might be used, e.g., pupil count as determined by your state.

Some Pre-Kindergarten programs do not start until after October 1 so an undercount is produced.

Special education Pre-K students are being counted here but some districts do not report data on locally funded programs in pre-kindergarten. This area needs more discussion and definition. Not all states collect enrollments for PreK.

Recommendations:

States and NCES should work out and agree on a common definition for Pre-K. In the meantime, districts should provide Pre-K enrollment and FTE teacher data to the extent that they can.

Most states can report on Kindergarten membership and FTE teachers.

Issue 2. Ungraded Enrollments and FTE Teachers

Problems:

Students moving from grade to grade sometimes get counted in an ungraded status.

SASS does not differentiate ungraded vs. non-graded students.

The number of ungraded FTE teachers is very difficult to report in districts without automated systems.

Some states require students to be reported in a grade for state reporting, yet districts reported ungraded students on SASS.

Recommendations:

Definitions need to be clarified among states and NCES, and the definitions transmitted to districts.

The ungraded category should be broken into special education vs. students in non-graded (multi-graded) classrooms.

This should be a state responsibility for data reporting. Most states have already determined how to count ungraded students, but it needs to be consistent from state to state. Non-automated districts may not be able to provide ungraded FTE teachers.

If a state or district does not have this category, the respondent should leave it blank.

Districts should check their state policies regarding the counting of ungraded students.

Issue 3. Grades 1-6 vs. 7-12 Enrollments and FTE teachers

Problems:

Those districts that do not have the grade breakouts for teachers as requested by SASS tend to cross out/over the questions to conform with own categories. Consistency within district is achieved, but totals

: don't match those reported by the states.
It is difficult to classify teachers by grade level for districts and most states.

Districts report teacher data by school. Middle school teachers are masked when arbitrary grade levels are used to categorize teachers in ways that do not exist in a particular state.

Recommendations:

States collect and report grade by grade membership (enrollment on or around October 1) and can report to SASS with greater reliability and comparability than districts. Postsecondary enrollments and FTEs should be dropped.

The 1-6, 7-12 categories for FTE teachers could stay the way they are, but states and NCES must develop some algorithm for computing from state teacher numbers collected by school so that the data will be consistent.

States should have responsibility for grade breakouts of FTE teachers, because it is better to have 50 agencies computing the grade breakouts than 5,400 districts (in SASS sample).

Issue 4. Grades 7-12 FTE teachers by subject assignment

Problems:

States now have to make assumptions to aggregate their assignments into the categories requested by SASS, for example, reducing 600 assignment codes to 10 categories.

Districts simply look at the master schedule, but by doing this they tend to over-report the number of FTEs, i.e. they produce more of a headcount.

Computing FTE teachers by subject and grade level categories adds a considerable burden and requires assumptions about grade level.

Most states have an assignment for K-8 general elementary; not too many have it for 7-8 general elementary.

Recommendations:

States should report FTE teachers by subject assignments. Combine the elementary and secondary matrices into one matrix. Do not worry about grade level by subject. Total FTEs by subject (or self-contained) is important.

Break out science category into separate science fields.

Each state should have codes for self-contained classroom teachers at K-6 and 7-8.

Issue 5. New Hires

Problems:

SASS does not differentiate teachers returning from a leave of absence.

Some states could not count transfers between districts.

The October 1 reporting date issue is

sometimes a problem. Districts often hire teachers after Oct. 1. Some teachers may be waiting for certification. There is a question of whether positions are really vacant is problematic for many districts and for states.

Long term substitutes may be in a position for which they will be hired later.

The definition is unclear: States, and some districts, count teachers who are hired from Oct 1 to Oct. 1, i.e. some service in prior school year. Districts can report on teachers hired in the current school year.

Recommendations:

NCES should analyze new hire data to see local vs. state ability to provide.

Local districts should be responsible for reporting data on items having to do with new hires in districts, substitute teachers, and lay offs.

The definition of the SASS category should be clarified: New hire = Teachers not on the district payroll last year.

Vacancy questions on the matrix should be handled by districts.
If definitions were modified, then districts could do the middle columns.

Issue 6. Special Education

Problems:

Sometimes Pre-K gets counted within special education.

Recommendations:

Pre-K counts should be separated from special education counts.

One total FTE count for special education is fine--do not split by 1-6 vs. 7-12.

Issue 7. Certified Teachers

Problems:

States have different standards and definitions for certification.

There are some differences in definitions for: probationary, provisional, temporary, and emergency certification.

Districts may not apply definitions consistently and would have more difficulty counting certified by subject assignment than states.

About 30 states can report subject assignments by certification in the assignment.

Recommendations:

"Certified" varies from state to state but everybody has to know whether a teacher meets all of the state's requirements for certification. Certification is a state responsibility, so states can be expected to provide more accurate data.

States should report FTE teachers' certification. They are better equipped because districts that are not computerized would have to manually go through their files.

Issue 8. Total FTE for PreK though 6 Total FTE for 7-12

Problems:

Totals reported in different places on SASS should match each other. Many are not equal as reported by states or districts.

Recommendations:

Editing checks should be made on all grade totals and compared with the rest of the survey.

References from one place to another on the SASS form would cue the respondent on how the totals should look.

Issue 9. Instructional Support and Non-Instructional Staff

Problems:

Principals and assistant principals are not accounted for and some LEAs may put it in somewhere anyway.

There may be a double count when "health" workers are mentioned in both Part (3) and Part (4) of this item, as it now stands.

Recommendations:

Either put principals in or literally exclude them; account for them in some way.

A better definition of administrative staff has been developed for the CCD. These should be used in SASS.

States should handle these responses because they will be providing the data through CCD survey (beginning in 1992-93).

Issue 10: Librarians

Problems:

Some states define librarians as teachers, others do not.

There is too much emphasis on librarian issue.

Recommendations:

States should report the current status of librarians, guidance counselors, and other support.

Districts should report on vacancies of librarians, guidance, and other support.

There is not enough emphasis on guidance counselors; equal emphasis is requested because of difficulty in staffing and the aging workforce of these people.

General Issue: How is FTE computed?

Problems:

All states and districts may not compute FTEs on the basis of a 40-hour week for full time teachers (As stated in SASS survey).

Recommendations:

Use whatever a full time week is and put it in the denominator.

General Issue:

When are data collected and reported?

Problems:

There is an October 1 reporting date problem. Some counts are easily available for that date, e.g., enrollments/membership. Others are harder, e.g., teacher certification.

There is confusion in the definitions of enrollment vs. membership. Oct. 1 is usually defined as the date for student membership counts.

Recommendations:

If a snapshot picture is what is desired, then October 1 is what most states use.

Other Recurring Issues/Suggestions

The form is too complex: Although definitions were those that have been used for a long time, having been through several revisions, misinterpretation of questions/definitions was still considered to be major cause of incomparability of data from states and locals.

There should be a logical sequence of questions; questions should follow a logical order, so that the person who is filling it out can see clearly what the survey is trying to collect.

Periodically, the respondent should be asked to check his/her totals with other totals/columns in survey, to verify consistency, and eliminate discrepancies early on at the local level.

A task force should be developed to work on specifics of the issues of concern discussed in this workshop. The task force could work on the possible design of an instrument tailored to states. States can therefore work with their districts, knowing which items the districts will be working on and which items they are responsible for.

A crosswalk study with all the states should be conducted while the field test is going on.

Requests were made for participants to send in marked-up versions of the current survey, so as to add other items that may not have been discussed.

SUGGESTED CHANGES TO SASS DATA COLLECTION INSTRUMENT AND ITEMS (LEA Teacher Demand and Shortage 1992 Field Test)

Question #1:

Ungraded:

Need better definition for ungraded.
Need to separate and account for multi-graded and special education students/teachers.
If state does not have such a category, they should leave it blank.
Ungraded should not be used as a catch-all category to make the matrix all add up.

Grade Breakdowns:

Need better definition for pre-Kindergarten vs. special ed. vs. Head Start.
Instructions should be clear that if state does not have the proposed grade breakdowns, they should add up students by grade and provide the categories requested.

Drop postsecondary from survey.

Question #4:

Same issue about defining ungraded and breaking down pre-Kindergarten to separate out special education students/teachers.

Drop postsecondary from matrix.

Question #7:

Somehow account for those positions that were abolished not because a suitable candidate could not be found, but because those positions were no longer needed due to lack of demand, or budgeting problems, etc. As it stands, it leaves person completing survey confused as to why they are not completing the picture.

Question #8:

Rewording of this question to make it clearer.

Question #9:

New definition for newly hired teachers is needed. Define: New teachers = those teachers who were not employed on the district's/state's files last year.

Also need clearer definition of "emergency certification", what it may include and what it shouldn't. Define: "...a teacher who has not fully completed all state requirements that define a certified teacher in that state."

Question #10:

The phrase "laid off" should not be used, it is jargon-like.

Question #11:

Suggestion to define "general secondary" as to what it is, rather than what it is not.
Drop "exclude PK" from question 11b.
May be able to combine this matrix with #12. The grade levels are not really what's important; it is more vital to get the subject areas.

Question #12:

Clearly define "general elementary" in grades 7 and 8.
Science should be broken down by

field and not lumped all together. Suggestion to put librarians in this matrix, as a separate subject. Some states consider librarians to be teachers and this is a potential shortage area. This eliminates the wasting of space in question #13. Again, new definition needed for "newly hired".

Question #17:

Need to account for principals and assistant principals as part of non-instructional support. If unaccounted for, survey respondent may include them and throw numbers off.

May need to include the word "non-professional" in question 17b(4).

Guidance counselors should be given more emphasis/importance.

Part (4) of this question should be clearer about being non-professional although it includes health support staff.

General Suggestions:

Periodically, respondents should be requested to check column totals against other column totals to verify accuracy and consistency.

Logical sequence to questions would be desirable so as to make it clearer to respondent that the question is in line with the goal and what the survey is trying to accomplish with a given question.

Define all terminology before using it in data collection instrument, and/or provide an attachment of definitions. Then require respondents to account for and explain any differences in their definitions with those of the survey.

Instructions should include the provision that if district/state does not have a particular category, the respondent should leave that question blank or put N/A.

SUMMARY OF RECOMMENDATIONS FROM WORKSHOP

o State education agencies can report district-level data on selected items in the LEA Teacher Demand and Shortage Questionnaire of the Schools and Staffing Survey.

o The SASS items from the State Data Project that were analyzed at the Workshop can be divided into four categories:

(a) items that should be reported by states because the data are more reliable and comparable and most/all states can report the data, e.g., enrollments by grade (K-12) and FTE teachers by subject (K-12);

(b) items that should be reported by states because the data are more reliable and comparable but not all states are able to report the data, e.g., FTE certified teachers and instructional support and non-instructional staff (More states will be able to report in the future.)

(c) items that should be reported by districts because the data are more reliable or most/all states do not have the data, e.g., new hires; and

(d) items that could be reported by states with work on a consensus definition and categories, e.g., pre-K enrollments and FTE teachers, ungraded enrollments and FTE teachers, FTE teachers by grade level categories.

o Seven state education agency representatives volunteered to serve on a Task Force that would try to pursue the next steps in developing a state role with the Teacher Demand and Shortage survey. At least four tasks for a Task Force on State Reporting in SASS were identified:

1. Plan a second field test, or "trial run," of state reporting in SASS. The Task Force could determine how the items in categories (a) and (b) listed above, should be tested, recommend timing and procedures for the field test, and recommend how to get states to participate and report data. It would be voluntary, but it might be possible to have all interested states participate. (For example, in our Science/Math Indicators, 30 states participated in a "Trial Run" of our reporting system prior to the official start the following year.)

2. Plan a crosswalk study with all 50 states to determine any differences in reporting and how many states could report on category (b). The data in category (a) have been cross-walked, but may need double-checking with states.

3. Select items (e.g., from categories (a) and (b)) from the 1992 SASS Field Test that could be compared with state data for 1991-92 and identify a small number of states to participate. The Task Force would work on analyzing the results and reasons for differences.

4. Work on consensus definitions and data categories for category (d). States could subsequently use the results for state data reporting with SASS.

o State participants in the November workshop agreed that they would like to continue to work with NCES to develop a state reporting role in SASS, particularly with the LEA Teacher Demand and Shortage questionnaire. A State Task Force would require some staff resources and funds, and it might be a project of the Cooperative Statistics System.

The advantages of a state role are:

- reducing burden on districts;
- expanding the district-level data collected through SASS (e.g., the items and matrices that were dropped for the 1990-91 survey);
- getting more reliable and comparable data on some items; and
- reducing double-reporting by districts (to SASS and to states).

Possible disadvantages are:

- adding up to 50 more data collection respondents and
- if all 50 states do not participate on some selected items, determining a way to match state-reported data from states that participate with district-reported data from non-participating states.

o Suggestions on the instrument and items for the 1992 Teacher Demand and Shortage Field Test:

1. Many comments and suggestions were offered on the wording of items and organization of matrices (see this report). State and local representatives agreed that the form is complex and appears long and forbidding to a local respondent. Possibly some of the recommended changes can be incorporated for 1992.

2. The kind of review and comment on the form provided through the small group sessions may be a method of gaining feedback that NCES should consider for future instruments.

National Center for Education Statistics
Council of Chief State School Officers

WORKSHOP ON IMPROVING RELIABILITY AND COMPARABILITY OF STAFFING DATA

November 6-7, 1991

State Education Agency Representatives
Bob Barker, Special Projects Manager, Reports Management Division, Texas
Curtis Cronin, Coordinator of Data Processing, Alabama
Bob Eliason, Civil Rights Specialist, District Data Unit, Minnesota
Elizabeth Ford, Coordinator, Wisconsin Center for Education Statistics, Wisconsin
James Fulton, Manager, Education Demographics Unit, California
Clyde Hatton, Division Director, Management Information Systems, Mississippi
John Jones, Director of School Data, Missouri
Jo Ann Kerrey, Chief Supervisor, Management Information, South Carolina
Dori Nielson, Education Analyst, Management Information Services, Montana
Dennis Pool, Director of Information Center, Nebraska
Len Powell, Chief, Bureau of Educational Data Systems, New York
Peter Prowda, Education Consultant, Bureau of Research and Teacher Assessment, Connecticut
M. Eugene Wallace, Director, Statistical Services Section, Georgia

Local School District Representatives
Thel Kocher, Director for Testing, Evaluation, and Information Services, St. Paul Public Schools, St. Paul, Minnesota
Ralph Madrid, Executive Director of Personnel, Lubbock Independent School District, Lubbock, Texas
Gerald Peck, Assistant Superintendent for Personnel, School District of Superior, Superior, Wisconsin
Andrea Rosen, Director of Research, Evaluation, and Special Services, Great Neck Union Free School District, Great Neck, New York
Diane Shea, Business Administrator, Farmington Public Schools Farmington, Connecticut
Donald Shipp, Director of Employee Processing, Selection, and Placement, St. Louis Public Schools, St. Louis, Missouri
Ellen Tindle, Attendance Coordinator, Choctaw County Schools, Butler, Alabama

Southern Regional Education Board
Lynn Cornett, Vice President for State Services

U.S. Bureau of the Census
Susan Dougherty, Director, Demographic Services Division
Patrick Healy, Survey Statistician, Demographic Services Division

National Center for Education Statistics
Paul Planchon, Associate Commissioner, Elementary/Secondary Education Statistics Division
Lee Hoffman, Chief, General Surveys and Analysis Branch
Dan Kasprzyk, Branch Chief, Special Surveys and Analysis Branch
Mary Rollefson, Statistician, Special Surveys and Analysis Branch
John Sietsma, Education Statistician, General Surveys and Analysis Branch

National Research Council
Dorothy Gilford, Study Director, Committee on National Statistics

National Science Foundation
Larry Suter, Program Officer, Office of Studies, Evaluation, and Dissemination, Education and Human Resources Directorate

Council of Chief State School Officers
Rolf Blank, Project Director, Science/Math Indicators
Barbara Clements, Project Director, Education Data System Implementation
Melanie Dalkilic, Project Assistant, Science/Math Indicators
Lisa Solomon, Project Associate, Education Data System Implementation

SAMPLING FRAMES AT THE UNITED STATES NATIONAL CENTER FOR EDUCATION STATISTICS

M. McMillen, D. Kasprzyk, and P. Planchon
National Center for Education Statistics

Introduction

Federal data collection of education statistics began in the 1869-70 school year when the Office of Education implemented a biennial voluntary education survey that included data on elementary and secondary school student attendance, teaching staff, and finance aggregated to the state level. It was 1954 before the Federal collection efforts moved to an annual collection. And National reporting of school district and school level data did not begin until 1975.

Throughout the 124 years of education data collection and reporting, considerable attention has focused on the coverage, quality, comparability, and timeliness of the data. A number of special studies and commissions have addressed these issues resulting in at least six reconfigurations of the data collection system and four bureaucratic relocations of the agency.

The current set of elementary-secondary data collections grew out of the 1985 Elementary-Secondary Redesign Project. The Redesign Project was charged with the task of reviewing the thrust and scope of the elementary and secondary data collection system. Ultimately a ten year plan was formulated as a result of a set of public meetings, a series of invited papers and comments, and a synthesis of the papers and comments.

The Hawkins-Stafford Education Improvement Amendments of 1988 strengthened the structure of the National Center for Education Statistics as a statistics-gathering agency and established a federal-state cooperative statistics system. The resulting changes provided the impetus for the implementation of the basic elements of the ten-year plan.

The combined outcome from the 1988 Education Improvement Amendments, the redesign project and related ten year plan is a data collection program that is considerably different and more comprehensive in scope than the one that had existed previously. One significant change has been an increased reliance on NCES data collections as sampling frames for sample surveys. This paper describes the major data systems currently used as sampling frames at the U.S. National Center for Education Statistics. As will be described below, the NCES sampling frames are of two types; universe data collection systems and sample survey data systems. While both types of systems are typically institution-based(school/school district/postsecondary institution), they are often used as the first stage sampling frame for multi-level longitudinal and cross-sectional surveys of students, teachers, or administrators.

First, we describe the three principal institution-based universe data collection systems of the NCES: 1) the Common Core of Data (CCD) and its five components is the NCES primary data base on elementary and secondary public education in the United States; 2) the Private School Survey (PSS) is the principal data base on elementary and secondary private schools in the United States. This data system is comparable to the CCD Universe Survey for public schools; and 3) the Integrated Postsecondary Education Data System (IPEDS) is the core postsecondary education data collection program, its multiple components(like CCD) encompass all identified institutions whose primary purpose is to provide postsecondary education. Second, we describe two sample survey data systems, the Schools and Staffing Survey (SASS) and the National Postsecondary Student Aid Study (NPSAS), and how these data systems are used as sampling frames for noninstitution-based studies.

Universe Data Collection Systems: Common Core of Data

The Common Core of Data (CCD) is the basic NCES database on elementary and secondary public education. The CCD is an annual national data set with statistical information for all public schools and school districts in the U.S. and its territories; data reported on the CCD are comparable across all states.

The Common Core of Data has two purposes: first, to provide basic information and descriptive statistics on public elementary and secondary schools and schooling in general; second, to provide an official list of public elementary and secondary schools and districts in the nation, thus providing NCES the universe from which to select samples for NCES surveys.

CCD Design

The CCD survey collects data about all public elementary and secondary schools, all local education agencies (LEAs), and all state education agencies (SEAs) in the United States. CCD contains basic data on schools and school districts, students and staff, in addition to fiscal data. Basic data are name, address, phone number, and type of locale; students and staff data contain demographic characteristics; and the fiscal data cover revenues and current expenditures.

The CCD is made up of a set of five surveys sent to state education departments, including the Virgin Islands and outlying areas. Most data are obtained from administrative records maintained by the state education agencies (SEAs). The SEAs compile CCD requested data into prescribed formats and transmit the information to NCES.

Components of CCD

The CCD data system has five parts:

1. The Public School Universe contains data on public elementary and secondary schools in operation during a school year, school location and type, enrollment by grade and counts of students by race/ethnicity, counts of students eligible for free lunch, and the number of classroom teachers(FTE).

2. The Local Education Agency Universe contains name, phone number, location and type of agency, current number of students, and number of high school graduates and completers in the previous year, counts of dropouts by sex for grades 7 through 12 for all LEAs in the nation.

3. The State Aggregate Non-Fiscal Report contains state level aggregates of students by grade level, full-time equivalent staff by major employment category, and high school graduates and completers in the previous year.

4. The State Aggregate Fiscal Report contains state level data on average daily attendance, school district revenues by source (local, state, federal), and expenditures by function (instruction, support services, and non-instruction).

5. The School District Fiscal Data contain data by school district, including enrollment, revenues by source and expenditures by function.

The Public School Universe Survey and the Local Education Universe Survey are the two key universe components of the CCD used for sampling schools and local education agencies, respectively. The addresses in the universe surveys provide the means for contacting a sampled school, while the basic data the surveys obtain provide information needed to design and stratify the sample.

Universe Data Collection Systems: Private School Survey

To obtain a complete picture of elementary/secondary education, activity comparable to the CCD public school universe survey is needed for the private elementary/ secondary education sector. At a 1988 meeting with private school associations, NCES introduced a proposal to develop such a private school data collection system. This data collection system, the Private School Survey (PSS), is designed to build an NCES universe of private schools in the U.S.

Private school universe data are obtained every two years by a mail-out/mail-back collection design. A lack of response by the school elicits a telephone followup. Data collected include: grade range, enrollment by grade, number of graduates(if a high school), number of teachers, student race/ethnicity distribution, and school's religious orientation.

Private School Frames

The primary sources for building the universe list of private schools are: a commercial list, state lists of private schools, and private school association lists. To identify schools overlooked in the list building component, an area frame component is included. The universe list and additional schools identified in the area search comprise schools included in the Private School Survey.

List Frames

NCES has used a dual frame approach in surveying private schools since 1983. A commercial list from Quality Education Data (QED) served as the base list for the private school universe in 1987 and 1989. NCES checked all schools on the QED file to determine their eligibility for inclusion on the list per criteria defined by NCES. Many schools on the QED base list did not meet the criteria and were eliminated, leaving approximately 23,000 private schools on the base list.

To improve coverage NCES collected membership lists from 20 private school associations and denominations. Schools on private school membership lists were compared and added to the base list when appropriate. As a result of these efforts, 1,261 schools were added in 1987, and 866 schools were added in 1989 for a total of 24,727 schools on the NCES private school universe list. Despite these efforts, the list frame undercoverage of schools was estimated to be approximately 20%.

The 1991-92 PSS made a substantial effort to increase the number of schools on the master list of private schools by not only adding schools from the sources previously mentioned (commercial lists and private school association lists), but also by adding schools obtained from lists maintained state education agencies. A significant number of additional schools were added, resulting in a school undercoverage rate of about 12% and an enrollment undercoverage rate of about 3%. A project is underway at the Census Bureau to evaluate these efforts and the impact of these new sources.

Area Frame

Additional schools are identified through an area search of randomly selected primary sampling units (PSUs). The first NCES area search for private schools was conducted in 1983, and this method has been used to improve coverage in private school surveys since that time.

The 1989-90 PSS area frame sample consisted of 123 PSUs from two sets of sample PSUs. Within each of the 123 PSUs, the Census Bureau attempted to find all eligible private schools. An area canvas was not attempted; however, regional field staff created the frame by using such sources as yellow pages, local education agencies, chambers of commerce, and local government offices. The schools found were matched with the NCES private school universe list from the list frame. Schools that did not match the list were contacted to verify eligibility. Eligible schools located and identified in the area frame and not on the master list were assigned a sample weight and an estimate of the number of private schools represented by the area frame calculated. This estimate when combined with the number of private schools on the master list yields the national estimate of the number of private schools.

During the last administration of the PSS, the area frame component accounted for a smaller contribution to the overall national estimate of the number of private schools and the number of students in private schools in the U.S. The acquisition of new lists and improved unduplication procedures has improved the private school list. Unfortunately, the fact that the 1991-92 area frame still constitutes 12% of the student estimate and 3% of the student estimate indicate that the universe list is still missing a significant number of schools. Since they are small schools (i.e, 12% of schools, but only 3% students), this suggests their exclusion would yield biased estimates; thus, the elimination of the area frame component of the PSS is not yet warranted.

Universe Data Collection Systems: Integrated Postsecondary Education Data System

The Integrated Postsecondary Education Data System (IPEDS) is the core postsecondary education data collection program. It contains all institutions whose primary purpose is to provide postsecondary education. This includes academic, vocational, and continuing professional education programs, and excludes avocational and adult basic education programs.

The approximately 11,000 IPEDS institutions include: baccalaureate or higher degree granting institutions, 2-year award institutions, and less-than-2-year institutions(i.e., institutions whose awards usually result in terminal occupational awards or are creditable toward a formal 2-year or higher award). Compatible reporting formats have been developed for the different sectors(public, private nonprofit, private for-profit) of postsecondary education providers.

IPEDS Components

The IPEDS data system contains:

1. Institutional Characteristics Survey which includes annual data on the institution's address, telephone number, tuition, types of programs, levels of

degrees, and accreditation.

2. Fall Enrollment Survey which includes full-and part-time enrollment data by sex, and racial/ethnic categories.

3. Fall Enrollment in Occupationally Specific Programs Survey which provides fall enrollment in occupationally specific programs, by sex and race/ethnicity.

4. Completions Survey which provides numbers of associate, bachelor's, master's, doctor's and first professional degrees by discipline and sex, numbers of awards by racial/ethnic composition, program area, and sex.

5. Salaries, Tenure, and Fringe Benefits of Full-Time Instructional Faculty Survey which provides annual data on the number of full-time instructional faculty by rank, sex, tenure status, length of contract, and salaries and fringe benefits.

6. Financial Statistics Survey which annually provides current revenues by source, current expenditures by function (e.g.,instruction, research), assets and indebtedness, and endowment investments.

7. College and University Libraries Survey which provides staffing, collection, transaction, and operating expenditures data.

8. Fall Staff Survey which provides the number of staff by occupational activity, full-time and part-time status, sex, and race/ethnicity.

Since the Institutional Characteristics Survey identifies and characterizes institutions offering postsecondary programs, it is used as the basis for sampling postsecondary institutions. The data the survey obtains on the institution and its programs provide the background information necessary to stratify postsecondary institution samples; however, individual components of the IPEDS data system are also used together for sampling as the need arises; for example, to build the frame for the National Postsecondary Student Aid Study (discussed below) IPEDS institutions on the Institutional Characteristics and the Fall Enrollment files were used, whereas the IPEDS Completions File was used to develop data on race/ethnicity trends in degrees conferred.

Sample Survey Data Collection Systems: Schools and Staffing Survey

SASS Overview

The Schools and Staffing Survey(SASS) uses the two elementary/secondary universe data systems, CCD and the PSS, as frames for drawing a sample of elementary/secondary schools in the public and private sectors, respectively. In addition, the SASS sample of schools is then used to draw samples for other SASS components - - principals, teachers, local education agencies, and most recently, students and libraries.

SASS was initially implemented to meet the need for information on the characteristics and experience of teachers and administrators, to describe the essential features of the school as a place to work and a place to learn, and to provide data on aspects of teacher supply and demand and attrition. The SASS design permits state and national estimates for public schools and affiliation and national estimates for private schools. The SASS was first fielded in the 1987-88 school year, was repeated in the 1990-91 school year, and will be conducted every three years.

The SASS is an integrated system of surveys of public and private schools, school districts, school principals and administrators (public and private), and teachers (public and private). The data collection consists of seven mail-out/mail-back surveys implemented during one school year.

In the year following SASS, a subsample of teachers in the SASS teacher sample are selected for the SASS Teacher Followup Survey. This mail survey, a survey of public and private school teachers, is designed to provide information on teacher attrition and retention in public and private schools. In the 1993-94 school year, SASS will also implement a sample of students (of a subsample of SASS teachers). Student data will be reported from administrative records the school maintains on the students.

SASS As A Sampling Frame

Schools are the primary sampling unit in SASS, the sample being drawn from the CCD for the public school sample and from the PSS for the private school sample. School administrators/principals are in sample if the school is in sample, and public school districts are included in sample when one or more schools in the district are selected.

To develop a sampling frame of teachers for the SASS, all schools in the SASS sample are asked to provide a list of teachers in the school. The list includes name as well as limited information about the teacher, such as years teaching experience, race, and teaching specialty by level. Schools are asked to complete and mail back a form requesting this information, provide the list of teachers over the telephone, or if neither of these alternatives are acceptable to draw the sample themselves with instructions from the Census Bureau. On average, between four and eight teachers are selected in each sample school selected for the SASS.

In each round of SASS, a subsample of teachers responding in SASS serve as the sampling frame for the Teacher Followup Survey. A sample of teachers responding to SASS is drawn. The sample is stratified by whether or not the teachers are in the teaching profession one year after the SASS is conducted.

In school year 1993-94, a subsample of public and private schools in the SASS sample will be asked to to participate in a survey of library media centers (staffing collection, expenditures, technology, and equipment) and librarians/media specialists (qualifications and working conditions).

In school year 1993-94, the sample of teachers selected in the SASS sample constitutes the sampling frame for a new student records component of the SASS. For a subsample of the teachers selected in the teacher sample, class rosters for a specific day and class period will be requested from the school in order to provide a list of students eligible for sample selection; thus, the national probability sample of schools has served as the frame for a national probability sample of teachers, and finally a national probability sample of students distributed across elementary and secondary levels.

Finally, in school year 1993-94 public schools in the 1990-91 SASS sample will serve as a sampling frame for an NCES conducted national survey on curricular options in public high schools. The use of schools in the SASS sample permits analyses using the extensive school-based data collected in the SASS, such as the school's enrollment and racial composition, the size, structure, and experience of the faculty, along with the curricular options data obtained in the survey.

Sample Survey Data Collection Systems: National Postsecondary Student Aid Study

NPSAS Overview

The National Postsecondary Student Aid Study (NPSAS), conducted every three years, is a nationwide study of students enrolled in less-than-2-year institutions, community and junior colleges, 4-year colleges, and major universities located in the U.S. and Puerto Rico. NPSAS obtains data on student demographics, family income, education expenses, employment, education aspirations, parental demographic characteristics, parental support, and how students and their families meet the costs of postsecondary education. The first NPSAS was conducted during the 1986-87 school year and repeated in 1989-90. Data were gathered from students' institutional records, from the students themselves, and parents.

NPSAS Design

The sample design for NPSAS was a multi-stage probability sample of students enrolled in postsecondary institutions. The first stage sample consisted of geographic areas of the country; institutions within the selected geographic areas were selected in the second stage of sampling; the third stage of sampling was the selection of students in sampled institutions. The 1993 NPSAS sample includes about 78,000 students at 1,200 institutions and about 25,000 parents. NPSAS data come from multiple sources, including institutional records, and student and parent interviews. Detailed data concerning participation in student financial aid programs are extracted from institutional records. Beginning with the 1990 NPSAS, student and parent data were collected using a computer-assisted interview.

The 1987 NPSAS sampled students only enrolled in the fall of 1986. Beginning with the 1990 NPSAS, students enrolled at any tim · during the year were eligible for the study. This design change provides data necessary to estimate full-year financial aid awards.

NPSAS As A Sampling Frame

NPSAS is a nationally representative sample of institutions, students, and parents. It, thus, provides an efficient way of identifying a nationally representative sample of beginning students in postsecondary education as well as an efficient way to identify a nationally representative sample of baccalaureate degree completers in postsecondary education. Thus, NPSAS serves as a sampling frame for the Beginning Postsecondary Study (BPS) a longitudinal study of students from the beginning of their postsecondary education and the Baccalaureate and Beyond (B&B) study, a longitudinal study of students from graduation on. Using NPSAS as a sampling frame for these two studies has the obvious benefit of having available data from all components of NPSAS as base year data for the samples.

BPS follows NPSAS beginning students at 2-year intervals for at least six years beginning with the 1990 administration of NPSAS. This should allow adequate time to complete postsecondary education and transit between undergraduate and graduate education and between postsecondary education and work.

B&B will follow NPSAS baccalaureate degree completers at 1,3,6,9, and 12 years after completion of their undergraduate, beginning with the 1993 NPSAS. In addition to student data, B&B will collect postsecondary transcripts and financial aid records covering the undergraduate period, providing complete information on progress and persistence at both the undergraduate and graduate levels.

Future Considerations For NCES Sampling Frames

The work of NCES programs has been highly decentralized in the past. In particular, the development and maintenance of universe data collection systems has involved both NCES staff and a variety of contractors. In recent years, the Census Bureau has become the NCES data collection agent for these universe data systems. The expected benefits of such an arrangement include a stronger approach to maintaining consistent definitions and concepts over time and where feasible across data collections, the development over the long term staff knowledgeable of NCES concepts and issues, a closer and more efficient working relationship between staff involved in universe data collection systems and survey data systems, and improvements in the use of the universe data systems for sampling.

We also expect to see over the next several years a project develop that will design and implement an integrated sampling frame, useful over an extended period of time (with updates) for all major NCES surveys. The elementary/secondary school universe will extend to cover schools with pre-kindergarten programs, thus providing a frame for sampling such programs in order to allow for the possibility of extending the scope of NCES institutional data collections. At the present time, our knowledge of early childhood education programs is limited primarily to parent reports. The addition of pre-kindergarten programs to the universe frame will allow NCES to extend the scope of NCES surveys of programs, staffs and students. Samples will be designed to minimize overlap among the various programs and take advantage of the similarities in the operations of some programs.

An integrated approach will also involve the centralized management of list frame operations and, perhaps, more importantly area frames. For example, at the present time, only private school estimates at the elementary/secondary level incorporate an area frame component for estimation. To the extent that undercoverage may exist in some sections of the postsecondary and prekindergarten universe, each collection could benefit from the shared effort of a list frame operation that spans prekindergarten and postsecondary education. For example, universe collections using an area frame for coverage improvement can share the effort of listing eligible schools.

References

Burkheimer, G., Forsyth, B., Wheeless, S., Mowbray, K., Boehnlein, L., Knight, S., Veith, K., Knepper, P. (1992). Beginning Postsecondary Student Longitudinal Study Field Test Methodology Report: BPS 90/92. National Center for Education Statistics, Office of Educational Research and Improvement, Washington, D.C., NCES 92-160

Choy, S.P., Henke R.R., Alt, M.A., Medrich, E.A., and Bobbitt, S.A.(1993). Schools and Staffing in the United States: A Statistical Profile, 1990-91. National Center for Education Statistics, Office of Educational Research and Improvement, Washington, D.C., NCES 93-146.

Davis, C. and Sonnenberg W.(1993). Programs and Plans of the National Center for Education Statistics. National Center for Education Statistics, Office of Educational Research and Improvement, Washington, D.C.

Gerald, E., McMillen, M., and Kaufman, S.(1992). Private School Universe Survey, 1989-90. National Center for Education Statistics, Office of Educational Research and Improvement, Washington. D.C., NCES 93-122.

McDowell, L. (1993). Public Elementary and Secondary Schools and Agencies in the United States and Outlying Areas: School year 1991-92. National Center for Education Statistics, Office of Educational Research and Improvement, Washington, D.C., NCES 93-328.

Morgan, F. (1992). Race/Ethnicity Trends in Degrees Conferred by Institutions of Higher Education: 1980-81 through 1989-90. National Center for Education Statistics, Office of Educational Research and Improvement, Washington, D.C., NCES 92-039.

Shepherd, J. and Malizio, A. (1992). Methodology Report for the 1990 National Postsecondary Student Aid Study. National Center for Education Statistics, Office of Educational Research and Improvement, Washington, D.C., NCES 92-080.

MONITORING DATA QUALITY IN EDUCATION SURVEYS

Samuel S. Peng and Kerry Gruber, National Center for Education Statistics
Wray Smith, Synectics for Management Decisions, Inc., and Thomas B. Jabine, Consultant
Wray Smith, Synectics Mgmt Decisions, 3030 Clarendon Blvd., Suite 305, Arlington, VA 22201

KEY WORDS: Nonsampling error; Quality assessment; Quality profiles; Sampling error; Statistical standards

Achieving high quality in education surveys is a major goal of the National Center for Education Statistics, U.S. Department of Education. Various features have been routinely built into the design of surveys and operational procedures to ensure that, for example the sample is selected according to specifications, the response rate is high and nonresponse bias is minimized, and the data are valid, accurate, and reliable. To assess the achievement of these procedures and to identify areas for improvement, NCES has developed a set of statistical standards against which project staff can determine the strengths and weaknesses of each survey system. NCES has also initiated a series of studies to develop and examine the quality profiles of survey systems, such as the Schools and Staffing Survey and the Common Core of Data, and to evaluate specific quality issues, such as the potential nonresponse biases of State NAEP (National Assessment of Educational Progress) trial assessments and the undercoverage of certain kinds of institutions in the Integrated Postsecondary Education Data System. The quality profiles are to include consideration of both sampling and nonsampling errors. This presentation describes these activities and shares our recent experience and findings.

1. Overview of NCES Data Quality Concerns

1.1 General Responsibility for Education Statistics

The National Center for Education Statistics (NCES), which is a part of the U.S. Department of Education, has major responsibilities assigned to it by the U.S. Congress for collecting, analyzing, and disseminating statistics and other data related to education in the United States and in other nations. The General Education Provisions Act and the Hawkins-Stafford Amendments of 1988 assign specific responsibilities to NCES for maintaining and improving the quality of education data and for assisting state and local educational agencies, including postsecondary education agencies, in improving and automating their statistical and data collection activities. NCES is directed to collect and report, on a state-by-state basis where feasible, full and complete statistics on the condition of education in the United States.

Recent legislation as well as recommendations from various advisory panels and organizations have led NCES to put renewed emphasis on the development of written standards for the conduct of its work. The first comprehensive written standards were adopted in 1987 and a major effort to revise and update the standards began in 1989 and was completed and adopted by NCES in January 1992. The current version of the *NCES Statistical Standards* (see Flemming, 1992) includes twenty "standards" (procedures that must be followed) and two "guidances" (procedures that are desirable but not mandatory). Many of the standards are directed toward the attainment of high quality data, both from sample surveys and from universe surveys, and other standards are directed toward documenting and evaluating survey designs and the resultant data quality.

1.2 Current Examples of Monitoring and Evaluating Data Quality

There is an ongoing program within NCES directed toward monitoring data quality in education surveys. As part of that program a number of special studies have been recently initiated and are in the process of being completed. Among these are the development of a "quality profile" for the Schools and Staffing Survey (SASS), the design and development of an ongoing assessment of the Common Core of Data (CCD), an examination of potential nonresponse biases in the state trial assessments of the National Assessment of Educational Progress (NAEP), and an investigation of the effects of undercoverage of certain types of postsecondary institutions on the estimates produced by the Integrated Postsecondary Education Data System (IPEDS).

Each of these periodic data collections are in fact "survey systems" rather than individual surveys.

- SASS is a sample survey system which has distinct components for public and private schools (elementary and secondary schools), for teachers and principals/headmasters in those schools as well as for administrators in local school districts.

- CCD is a universe survey which gathers fiscal and nonfiscal data from administrative records at the local public school and local education agency level as reported through state education agencies. Although sampling error is not involved, the components of nonsampling error are sometimes difficult to identify and to evaluate, including difficult questions which arise from the use of different (non-NCES standard) definitions in various states for some of the key data elements in the CCD.

- The NAEP state assessments of educational progress involve testing of individual students at designated grade levels and complex rules for substitution of schools when the designated schools fail to participate.

- The IPEDS Fall Enrollment Survey, conducted annually as part of a postsecondary education data system, provides an example of coverage and response problems when certain categories of institutions (particularly proprietary two-year postsecondary institutions) may come into existence or close their doors rather frequently and may often not be inclined to offer full cooperation to a government data collection. Imputation procedures for missing data present some subtle problems in this survey system.

For some of these survey systems, data collection and processing are conducted by the U.S. Bureau of the Census on behalf of NCES and for some components private contractors are involved. NCES's approach to data quality issues recognizes that in the case of public education there are distinct state and Federal roles and responsibilities and that private institutions at all levels present a special challenge of voluntary cooperation. NCES also recognizes its responsibility for balancing the tradeoffs between accuracy and timeliness and for balancing survey accuracy and survey costs (see Groves, 1989). There are also important sets of interaction effects between universe surveys (such as CCD) and sample surveys (such as SASS). Currently, CCD provides the frame information on public elementary and secondary schools for SASS.

In addition to the SASS quality profile development covered in this presentation, there is a broader SASS research program and other efforts aimed at improving SASS data, including special studies of nonresponse in SASS, a SASS reinterview program, a study of mode of data collection for SASS, intercomparisons of SASS and CCD data, an upcoming study of optimal periodicity for SASS, and development of a SASS users' manual.

2. A Quality Profile for the Schools and Staffing Survey

2.1 Background, Purposes, Scope, and Current Status

A quality profile is a document that summarizes, in convenient form, what is known about the quality of data in a particular survey. It describes the nature and sources of errors in the survey data and the findings from methodological experiments conducted to test alternative design components. A description of the survey design and procedures is included as background. A survey quality profile has two main audiences: data users, to inform them of the strengths and limitations of the data, and those responsible for the design and operation of the survey, for whom it can be an important tool for total quality management.

An early version of the quality profile was a 1978 "error profile" by Brooks and Bailar which provided this kind of information for estimates of employment from the Current Population Survey. The Census Bureau's Survey of Income and Program Participation, which began operation in 1983, was the subject of the first major quality profile for an entire survey (King, Petroni and Singh, 1987). An expanded version of that quality profile was released 3 years later (Jabine, 1990).

Work on a quality profile for the Survey of Schools and Staffing (SASS) began in 1992 and is nearing completion. Although its name suggests a single survey, SASS is actually a periodic, integrated *system of surveys* of schools, school districts (generally called local education agencies, or LEAs), school administrators, and teachers, conducted by the NCES. Users of the survey data include educators, researchers, policymakers and others interested in educational issues. The survey data are collected by mail, with telephone followups to nonrespondents. Survey data collection operations began in 1987 and two complete rounds of surveys have been conducted, with a third scheduled to start in 1993.

Development of a quality profile for a system of surveys, rather than a single survey, posed a new question about how to organize the materials. Should each chapter present information for a single major source of error, such as response error, for all five surveys or should the material be organized by survey? The question was further complicated by the sequential nature of the SASS sample selection process, which begins with the selection of samples of public and private schools, followed by selection of samples of teachers and public school districts associated with the sample schools. The decision was to organize the information by survey, starting with the School Survey, and to avoid undue repetition of design and procedural information by referring back to earlier chapters as needed. Each of the chapters covering the individual surveys has sections covering: frame development and sampling; data collection procedures and associated errors; data processing and estimation; and evaluation of estimates.

The draft of the *SASS Quality Profile* is in the final stages of review. The remainder of this section summarizes the findings that will be included with respect to major sources of error and identifies several ongoing research, development and evaluation activities that were underway but not completed in sufficient time for inclusion in the first *SASS Quality Profile*.

2.2. Principal sources of error in SASS

Coverage error There are no direct estimates of gross or net coverage errors available for any of the SASS surveys. However, there are several indications, some of them quantitative, of potential coverage error. These include:

- The use, for both the public and private school surveys, of list frames constructed two years prior to the reference school year for the survey.

- The need to use an area sample to supplement the list frame for private schools. The area sample accounted for about 22 percent of the estimated number of private schools in Round 1 and about 21 percent in Round 2, indicating no significant improvement of coverage by the list frame in Round 2.

- In Round 2, it was discovered that some multi-site special education programs of the State of California were listed on the sampling frame as single schools. Adjustments were required to eliminate duplication for those sites located at existing schools and to select a sample of the other sites.

- Discovery in both rounds, subsequent to sample selection, of some duplicate listings in the private school list frame.

- In Round 1, exclusion from the public school frame of 275 small Nebraska LEAs with about 2,800 students.

- For the teacher surveys, use of teacher listing forms that ask only for teachers working at the sample schools at the time the forms were being completed. Teachers who begin working later in the reference year have no chance of inclusion.

- In Round 1, counts of teachers on teacher listing forms were, in the average state, about 5 percent below the counts reported for the same schools on their School Survey questionnaires.

Sample estimates of the number of schools were also affected in both rounds by school survey respondents who provided data for a unit other than the one intended on the basis of the sample selection. Some respondents, especially in Round 2, reported combined data for two different schools at the same location, and some, especially in small LEAs, reported combined data for all schools in the LEA. Conversely, in the Teacher Demand and Shortage Survey, a few LEAs reported data for a single school rather than the entire LEA. Many of these erroneous reports were identified and corrected prior to data release, but some may have escaped detection.

Nonresponse error Response rates for public schools have consistently exceeded those for private schools. Response rates improved in Round 2 for each of the four basic surveys for both sectors. Response rates for the Teacher and Teacher Followup Surveys are composite rates, reflecting losses from schools that did not supply teacher lists and nonresponding teachers from schools that did supply lists. Consequently these rates were, with one exception, lower than those for the other three surveys.

There was considerable variation in response rates within each sector. For the public school sector in Round 1, in each of the four basic surveys a few states had response rates of less than 80 percent. This was due in part to a small number of LEAs, some of them fairly large, that declined to participate in any of the surveys. For the private school sector, one or more affiliation groups had response rates of less than 60 percent in each of the four basic surveys in Round 1.

The forthcoming report (Jabine, 1993) will also present detailed data on item nonresponse. The analysis of item nonresponse that occurred in Round 1 led to significant changes in the content and format of the questionnaires used in Round 2.

Measurement error Information about measurement (response) errors associated with SASS data collection comes from several sources: reinterviews, a record-check study, in-depth interviews using cognitive research techniques, methodological experiments, reviews of completed questionnaires and analyses of errors and inconsistencies detected during data processing. The main findings from these sources were:

- Reinterviews have shown that the items asking for the opinions, perceptions and future expectations of teachers and school administrators are, almost without exception, subject to high response variability.

- Evidence from several sources suggests that the quality of information obtained by mail is superior to that obtained in telephone followups to nonrespondents.

- An experiment, the State Data Project, was undertaken in connection with the Pretest for Round 2 of SASS to test the feasibility of obtaining data for the public sector Teacher Demand and Shortage Survey from state rather than local education agencies. A comparison of data collected from both sources for the same sample of LEAs showed a high frequency of substantial differences (more than 10 percent in either direction) for several variables. Based on these findings, it was decided not to try to collect the data for LEAs from state agencies in Round 2.

- Some of the concepts adopted for SASS data collection appear to be unfamiliar to respondents and to cause them considerable difficulty in formulating appropriate responses. One such concept is that of full-time equivalent (FTE) teachers used in the School and the Teacher Demand and Shortage Surveys. A school that has part-time teachers should report numbers of FTE teachers that are lower than their teacher counts. Nevertheless, many such schools reported the same numbers for teacher counts and FTE teachers.

- A record-check study, the Teacher Transcript Study, compared teachers' self-reports of their educational backgrounds with data from college transcripts. The main conclusion was that self-reports of types and years of degrees earned and major fields were reasonably accurate, but that self-reported information on courses and credit hours in specific fields was less accurate.

- For all surveys and in both rounds of SASS, it was common for respondents to ignore skip instructions and consequently to try to answer questions that did not apply to them. Such errors have little or no direct effect on the quality of data because most inapplicable responses can readily be deleted in clerical and computer edits.

The foregoing and other findings relating to measurement error led to numerous changes in survey instruments and procedures between Rounds 1 and 2, and additional changes are planned for Round 3.

<u>Data processing and estimation error</u> In contrast to the preceding sources of error, there is not much quantitative information available for data processing and estimation errors in the SASS surveys. A recent study of the correlates of nonresponse in the School Survey led to a recommendation for some changes in the definitions of the nonresponse adjustment cells and the order of collapsing small cells in the weighting process.

<u>Sampling error</u> At present, there are two ways for users of SASS data to determine the sampling errors of estimates that are of interest to them. Publications of SASS data include standard errors for many of the published estimates. Users of microdata files can compute standard errors for any estimate by employing readily available software for variance estimation by the balanced half-sample replication method. Half-sample replication weights for this purpose are included in the microdata files.

A recent study has confirmed the feasibility of including generalized variance functions in SASS publications. These functions, which relate the sampling error of an estimate to its size, can be used by those who do not work with microdata files, or lack the software for the replication method, to produce approximations to the sampling errors associated with their estimates of interest.

<u>Comparisons with data from external sources</u> Results of comparisons of SASS data with data available from sources other than NCES include the following:

- The Census Bureau collects data on school enrollment annually in the October Supplement to the Current Population Survey (CPS). SASS estimates of *private* elementary and secondary school enrollment from Round 1 exceeded the CPS estimates for the same school year by 15 percent. NCES surveys of private schools prior to SASS had shown similar differences with CPS enrollment estimates during the 1980s.

- The National Catholic Education Association conducts an annual census of Catholic schools. SASS Round 1 estimates of the number of Catholic schools and their enrollment exceeded the Association's census counts by 6.1 and 7.8 percent, respectively.

- Public school administrators' salaries reported in the Round 1 School Administrator Survey were compared with data obtained directly from state education agencies in selected states. The values were similar and there were no obvious inconsistencies.

- Round 1 estimates of teachers' salaries were compared with data from private organizations. The Teacher Survey estimate of average base salary, $26.231, was 6.6 percent below a $28,071 estimate of average salary for the same school year from a 1989 survey by the American Federation of Teachers and 6.4 percent below an estimated average salary of $28.029 reported by the National Education Association.

2.3 Research in progress

Several SASS-related research, development and evaluation activities are in various stages of completion. Some are just getting under way. For others, data have been collected or compiled and the results are being analyzed.

Two projects are related to plans to expand the coverage and content of SASS. As part of a pretest for Round 3 of SASS, questionnaires for collecting data about public and private school library media centers and library media staff specialists were tested. Item nonresponse and other features of the pretest responses are being analyzed and the questionnaires are being redesigned for use in Round 3 of SASS.

Collection of data about students is another possible area of expansion for SASS. Procedures for selecting samples of students and obtaining information about them from school records were tested in 1991. The completeness and quality of the data provided by the schools for the sample students are being evaluated.

Possible changes in the modes of data collection for SASS are being evaluated. Development and testing of computer-assisted methods of response for schools and LEAs has begun. Interactive diskettes with the survey questions will be mailed to respondents, who will complete them using their own computers. This method of data collection has already been used successfully by the NCES for completion, by state offices, of questionnaires relating to public libraries and completion of questionnaires for academic libraries.

A first attempt to evaluate the feasibility of collecting data for LEAs from state education agencies was inconclusive. There were substantial differences between items reported directly by LEAs and the corresponding values reported by the state agencies. However, further testing of the collection of at least some of the LEA information from the states is planned.

Efforts to improve response rates are continuing. When telephone followups are necessary for teachers who do not mail in their questionnaires, it has proved difficult to reach them at their schools and complete the interviews by telephone while they are there. In the pretest for Round 3 of SASS, conducted during school year 1991-92, postcards were sent to teachers during the mail followup phase asking them to supply their home telephone numbers if they were willing to be contacted at home. The results of this test are being analyzed. A study is underway, using data from all of the SASS surveys in Round 2, to compare the characteristics of nonrespondents and respondents, based on the sampling frame information that is available for both groups. It is hoped that the results of the study will suggest methods of improving response rates for problem groups and also possible improvements in the nonresponse adjustments used in developing estimates from the data for responding units.

The quality of SASS data is affected in many ways by the quality of the sampling frames for schools, LEAs and teachers. Several current evaluation and research projects are aimed at the improvement of the sampling frames and other features of the SASS surveys that relate to coverage. For public schools and LEAs, the Common Core of Data was adopted, starting in Round 2, as the frame of choice. As discussed in the next section of this paper, a plan has been developed for a detailed assessment of the quality of data collected in the CCD surveys, including the data that are used to create and maintain the LEA and public school sampling frames. For private schools, NCES has requested the Census Bureau to undertake a detailed analysis of the private school list and area frames and the procedures for updating them. As part of this study, the two frames will be matched for the sample of areas that are covered by the area frame.

Two other activities are also relevant to coverage improvement. Work is continuing on efforts to redesign the instructions and initial items on the school and LEA questionnaires to make it clearer to respondents which schools or LEAs they are being asked to report for. The forms and procedures for the teacher listing operations that provide the sampling frames for the Teacher and Teacher Followup Surveys are being evaluated, with emphasis on completeness and on the accuracy of information about teacher characteristics used in the sample selection processes.

For several variables, SASS obtains information from more than one survey. Estimates of the number of teachers, for example, can be obtained from the School, Teacher and Teacher Demand and Shortage Surveys. When aggregate estimates for school districts, states and other domains are compared, the differences are sometimes larger than could be accounted for by sampling variability. A Cross-Questionnaire Estimates Comparison Study is being undertaken to document comparable estimates that can be produced from more than one SASS survey, compare them at several levels of aggregation, and identify possible reasons for differences.

Results of all of these ongoing research, development and evaluation activities will be documented in internal memoranda, contractor reports and, where appropriate, in NCES working papers, technical reports and papers presented at professional association meetings or in journals. NCES also expects to provide updates to the *SASS Quality Profile* at appropriate intervals, possibly after each round of the survey. References to documentation for all of the findings mentioned in this presentation will be included in the *Quality Profile*.

3. Development of a Design for an Ongoing Assessment of Data Quality in the Common Core of Data (CCD) Survey System

3.1 Background Information on the CCD Surveys

Survey descriptions The CCD surveys provide basic statistical information about public elementary and secondary students, staffs, schools, and agencies. The CCD survey system is managed and directed by NCES, with major operational responsibilities delegated to the U.S. Bureau of the Census under an interagency agreement. The CCD system collects annual universe data reflecting three levels of aggregation (state, agency and school) from state education agency (SEA) administrative records. In summary, the information collected includes:

- State Aggregate Fiscal. Detailed information (for 56 states and outlying areas) about revenues and expenditures for public elementary and secondary education, reported in accord with the NCES Fiscal Handbook.

- State Aggregate Nonfiscal. Counts of public education staff, students, school completers (56 states and outlying areas).

- Agency Universe. Public education agency name, mailing address, telephone, agency type, county code, locale code; counts of education staff, students, school completers and dropouts (approximately 17,000 agencies).

- School Universe. School name, mailing address, telephone, school type, operating status and locale codes; counts of students and teachers (approximately 83,000 schools).

Counts of dropouts by sex within racial/ethnic status for each of grades 7 through 12 were added to the Agency Universe Survey for the 1992-1993 school year. Education staff counts were added to the Agency Universe Survey in that year also.

The nonfiscal surveys are distributed to SEAs in December of the reported school year (that is, December 1992 for the 1992-1993 school year and 1992 fiscal year reports). Completed reports are due on March 15; late and revised data are accepted until approximately September 1. The fiscal surveys follow a different schedule. Data edits are conducted by screening for missing or unacceptable responses, incorrect totals for summed variables, and values that diverge widely from the previous year's reports.

Work to improve data Over the past eight years, NCES has engaged in considerable redesign of the CCD surveys. The State Aggregate Fiscal Survey increased in detail from approximately 30 to 130 items, and NCES contracted for the development of individualized state "crosswalk" software programs that reconcile the state's fiscal reports with the requirements of the NCES survey. Through a contract with the Council of Chief State School Officers (CCSSO), NCES and the majority of the states agreed upon definitions for the data elements collected on the CCD surveys and negotiated Data Plans and Technical Assistance Plans that outline each state's existing and projected capability to comply with CCD reporting requirements. CCSSO technical reports, based on analysis of state reporting forms and conversation with state personnel, document all of these activities. By the end of 1991, both NCES and the majority of SEAs had subscribed to common definitions and reporting procedures for the CCD data, with documentation of those items each state did not report.

The CCD redesign effort to date has concentrated on establishing standard definitions and reporting periods for data items, and identifying which items

SEAs cannot report. Most of this work has been conducted through discussions between NCES contractors and CCD Coordinators at the individual SEAs. The fiscal crosswalk project has examined SEA fiscal reports and state procedures for converting these into CCD fiscal reporting requirements, with some on-site examination of SEA records and consultation with SEA staff. There has been virtually no on-site examination of SEA nonfiscal record systems.

3.2 A Perspective on Assessing CCD Data Quality

For the most part, the CCD gathers from the states information that the states already gather at their own initiative, following data requirements and definitions designed to accord with state education law and policy and to meet state needs. The questions cover all public schools and districts in the state. The state is under no legal compulsion to respond. This arrangement has great strengths and significant limitations. The strengths are:

- **The data are objective**, although not without error. Because the data are drawn from records, responses are not subject to the errors of recall, perception of meaning, and sensitivity to question wording and question sequence that create response problems in many surveys. The key respondent, the State CCD Coordinator, is an experienced professional who has worked in the state education department for some time, has secured from the schools and school districts the administrative information required by the state, and in many cases has responded to CCD surveys in prior years.

- **Coverage is likely to be generally good,** and

- **Response rates of schools and school districts are typically high**. The states have direct administrative relationships with the school districts and schools that help to ensure that they have current information on active and inactive districts and schools. The authority that the states exercise over the districts and schools helps to ensure that these subordinate units answer state inquiries promptly and accurately and respond to followup questions prompted by review of their information. As we have seen in our review, not all schools, districts, and states have the information that CCD requests. This means that item nonresponse rates may be high, especially for new topics, such as dropouts, and topics that are perceived to be outside the core interests of school administrators, such as school support staff.

The limitations of the arrangement are:

- **NCES has only a "cajoling authority"** over the coverage, content, and quality of the survey. Public education is the province of the states and localities, which provide 93% of the funding for public education. Accordingly, **control of the CCD data** rests mainly with the states and their representatives such as the Council of Chief State School Officers (CCSSO), not with the federal government. The states, not NCES, control survey coverage, the availability of information, and the definitions and classifications according to which the data are collected. NCES operates separate data collections for private schools because they are inadequately covered by state governments. NCES consults intensively with the states and their representatives and conducts technical assistance projects that seek to improve state capabilities to produce data that meet national standards. But NCES does not exercise control. NCES does have more than a cajoling authority in regard to the reporting of State Per Pupil Expenditure (SPPE). Because the figure is used to distribute $6 billion in federal aid each year, the incentive to report is high. Even so, the states employ a variety of definitions in calculating the number of pupils used in the denominator of the SPPE (see Morgan, 1991).

- **There is great variety in the availability, content, and quality of data**, including nonstandard definitions of measures, some of which are governed by state law, nonstandard names and identification codes. The variety in state definitions seems to be part of the price the nation pays for a federal system of education.

- **Complexity.** The CCD enumerates a variety of populations simultaneously (school districts, schools, students, staff), and data users expect to find sensible relationships among the figures for these populations. Moreover, CCD is conducted at several levels of aggregation simultaneously (state, school district, school). Data for each of these levels have to make sense in relation to one another. However, the encompassing of multiple levels of administration and multiple populations in the CCD produces many complexities and anomalies. There are students and staff that do not nest within schools; students, staff, and schools that do not nest within districts; and districts that do not nest within states. At the state level these anomalies may reflect, for example, educational programs run by correctional institutions and by state health and welfare agencies. At the local level, anomalies may reflect, for example, students for whose educational expenses the district is responsible but who are not schooled within the district. Such students may be "assigned" to existing schools even though they never attend there. It may be that figures for higher levels of aggregation (school district, state) cannot be arrived at by a simple summing of counts at lower levels because complex counting rules are necessary to avoid double counting; hence there are schools without pupils, schools without teachers, teachers without pupils, and pupils without teachers.

- **Errors are "lumpy"**. While it is rare for an entire state to fail to respond to one of the CCD surveys, it can be serious when it happens because the reporting units are so large. For example, CCD data for the school year 1991-92 were never submitted by the SEA in Virginia because of difficulties experienced in the changeover to a new computer system. That omission alone means that data on one in every 40 American schoolchildren were unavailable. There are about 16,000 school districts in the United States. Together, the largest 16 of them enroll one in every 10 American schoolchildren. Data problems in any one of the large districts are likely to mean data problems for their states and the national totals.

- In some states, the **CCD may be perceived as peripheral** and its data requests may be accorded less attention than in other states.

The strengths of the CCD and its shortcomings derive from the same source: its nature as a voluntary, universe, administrative record survey. On balance, these characteristics make the CCD a difficult data source from a statistical administrator's point of view. It is difficult in the lack of a central decision process, the lack of uniformity in definitions, the variety of state units responsible for data collection and initial processing, and the varying level of statistical capability among states. It is difficult from a data user's point of view for many of the same reasons, which raise doubts in the user's mind about the level of trust to be placed in the data. That is why it is essential to evaluate the data quality of the CCD, and to find a way to assess it that respects the nature of the CCD. NCES is now considering a new design framework within which it can specify and implement an ongoing assessment of data quality in the CCD survey system.

4. Nonresponse in the NAEP Trial State Assessments

4.1 Background of the Study

In 1990 the National Center for Education Statistics (NCES) launched a National Assessment of Educational Progress (NAEP) Trial State Assessment (TSA) of eighth-grade public school students in mathematics. This State Assessment Program assessed mathematics skills among over 2,000 eighth grade public school students in each of thirty-seven participating states and the District of Columbia.

Approximately 100 schools were selected in each state. The sample of schools in each state was selected with probability proportionate to size, where the measure of size was equal to the number of students enrolled in the eighth grade per school. The schools within each state were stratified by the following variables: urbanicity, percentage of black and hispanic students enrolled and median household income. All states, except for those with 100 schools or fewer, were stratified by urbanicity and income variables. Only states with significant minority populations were stratified based on minority enrollment. A sample of about 30 students were selected within each sample school. The student sample size of 30 for each school was chosen to ensure at least 2,000 students participating from each state, accounting for school nonresponse, exclusion of students, inaccuracies in the measures of enrollment, and student absenteeism from the assessment. Some students were excluded from the sample for various reasons and the number and reason for each excluded student was accounted for in each state. Each sample student completed a 55 minute assessment including 10 minutes of background information and 45 minutes of mathematics items.

4.2 Preliminary Results

There was nonresponse at the school district level, the school level, and the student level. When the state coordinator reported the nonparticipation of a school, a substitute school was selected. The process of selecting a substitute for a school involved identifying the most similar school in terms of the following characteristics: urbanicity, percent of black enrollment, percent of hispanic enrollment, eighth grade enrollment, and median income. Schools that substituted for a refusing school were assigned the base weight of the refusing school, if they agreed to participate. The base weight assigned to a school was the reciprocal of the probability of selection of that school.

In cases where there was nonresponse of the substituted schools there were also separate weight adjustments. Further there was adjustment for student nonresponse and poststratification adjustments as a result of a raking process. The base weight for a participating school was adjusted for nonparticipating schools for which no substitute participated. This procedure involved creating nonresponse classes based on urbanicity and minority strata. In states where no minority stratification was used, nonresponse classes were created based on median household income. The objectives in forming the nonresponse classes was to create as many classes as possible, as homogeneous as possible, but such that the resulting nonresponse adjustment factors were not subject to large random variations resulting from sampling error.

Nonresponse adjustments had to be recalculated according to the initial nonresponse. The schools were sorted into nonresponse classes and the following counts and ratios were listed for each initial nonresponse class:

- Total in-scope schools from the original sample
- Participating in-scope schools from the sample (both original and substitutes)
- Total in-scope schools from the original sample divided by participating in-scope scope schools from the sample.

The following procedures were adopted for reviewing these counts and ratios and determining what collapsing should be done. Within an initial nonresponse class, if the ratio of inscope schools to participating schools was less than 1.35, with at least six participating schools in the class, there was no need to collapse the particular cell. If any nonresponse class had fewer than 6 schools or a ratio greater than or equal to 1.35, it was collapsed with another class such that the new class met these conditions.

The table below gives those states that have a significantly lower weighted school participation rate in 1992 compared with 1990.

NAEP STATE TRIAL ASSESSMENTS

States with School Participation Rates (%) that Decreased from 1990 to 1992

State	Rate (1992)	Rate (1990)
Alabama	92	97
Arkansas	97	100
Idaho	91	97
Maryland	91	100
Michigan	94	97
Nebraska	85	94
New Hampshire	92	97
New Jersey	78	98
New Mexico	94	100
New York	83	86
North Dakota	97	100
Ohio	90	98

5. References

Bateman, D. and Cowan, C. (1984). "Plans for the 1980 Census Coverage Evaluation." Pp. 347-348 in Beth Kilss and Wendy Alvey, editors, *Statistical Uses of Administrative Records: Recent Research and Present Prospects, Volume I.* Statistics of Income Division, Internal Revenue Service, U.S. Department of the Treasury (March 1984).

Brooks, C. and Bailar, B. (1978), *An Error Profile: Employment as Measured by the Current Population Survey.* Statistical Policy Working Paper 3. Federal Committee on Statistical Methodology, Office of Federal Statistical Policy and Standards, U.S. Department of Commerce.

Flemming, E. (1992). *NCES Statistical Standards.* NCES 92-021. National Center for Education Statistics, U.S. Department of Education.

Groves, R. (1989). *Survey Errors and Survey Costs.* New York: Wiley.

Hogan, H. (1992). "The 1990 Post-Enumeration Survey: an Overview." *The American Statistician*, 46:4, November 1992, pp.261-269).

Jabine, T., assisted by King, K. and Petroni, R. (1990). *Survey of Income and Program Participation: SIPP Quality Profile.* Bureau of the Census, U.S. Department of Commerce.

Jabine, T. (1993). *Schools and Staffing Survey: SASS Quality Profile.* U.S. Department of Education, National Center for Education Statistics (forthcoming).

King, K., Petroni, R. and Singh, R. (1987). *Quality Profile for the Survey of Income and Program Participation, SIPP Working Paper No. 8708,* Bureau of the Census, U.S. Department of Commerce.

Kish, L. (1987). *Statistical Design for Research.* New York: Wiley.

Morgan, R. (1991). "Comparison of State Methods for Collecting, Aggregating, and Reporting State Average Daily Attendance (ADA) totals to the National Center for Education Statistics," NCES 91-049, April 1991.

Plewes, T. et al. (1988). *Quality in Establishment Surveys.* Statistical Policy Working Paper 15. Subcommittee on Measurement of Quality in Establishment Surveys, Federal Committee on Statistical Methodology, Statistical Policy Office, Office of Management and Budget.

[1.08 8-29-93]

GENERALIZED VARIANCE FUNCTIONS FOR THE SCHOOLS AND STAFFING SURVEYS

Sameena Salvucci, Synectics for Management Decisions, Inc., Glenn Galfond, Price Waterhouse, and Steven Kaufman, National Center for Education Statistics
Sameena Salvucci, Synectics for Mgmt Dec., 3030 Clarendon Blvd, Ste 305, Arlington, VA 22201

KEY WORDS: Variance estimation, Complex survey designs, Relative variance, iteratively re-weighted least squares, Repeated surveys

I. Introduction

This paper presents the results of an empirical examination of relative variances of selected statistics estimated from a complex sample survey. This study looked at the data gathered during the 1987-88 Schools and Staffing Survey (SASS) which was a national survey of elementary and secondary schools conducted by the National Center for Education Statistics (NCES). The target populations for the SASS were school administrators (principals and heads), and classroom teachers in public and private elementary/secondary schools. The survey design consisted of two parallel but essentially separate schemes, one for the public schools and one for private (nonpublic) schools. The components of SASS were (1) Survey of Teacher Demand and Shortage (TDS), (2) Survey of Schools (3) Survey of School Administrators, and (4) Teacher Survey. Approximately 13,000 schools and administrators, 65,000 teachers, and 5,600 Local Education Agencies (LEA's) composed the SASS sample.

NCES prepared eight SASS data files corresponding to the four types of surveys of both public and private schools, each of which contains a set of 48 replicate weights. These weights were designed to produce variances using balanced half-sample variance estimation. However, these replicate weights can be utilized only by users who have half-sample replication software available. The purpose of this task is to develop and test a new procedure using generalized variance functions for approximating the sampling error associated with an estimate of interest.

There were a large number of estimates of interest for the SASS. Estimates of proportions, totals and averages at the national level for various subdomains (i.e., region, school level, minority status, school size, community status and combinations of these) were made. Examples include (1) the total number of administrators who earned a bachelors degree, (2) the proportion of Hispanic students (regardless of race) (3) the number of FTE teachers, and (4) the average hours of teaching basic subjects in private schools.

The school sample was a single stage sample stratified by state by school level in public schools, and state by affiliation by school level in private school. Schools were systematically selected using a probability proportionate to size (pps) algorithm.

Within the first stage school sample, a second stage teacher sample was selected stratified by teacher experience level (teachers with three or fewer years of experience were classified into the new teacher stratum, and all other teachers were classified into the experienced teacher stratum). Within a school, teachers were selected systematically with equal probability.

The goal of this effort was to produce generalized variance functions for each of the Schools and Staffing Surveys. The generalized variances are designed for the user who does not have half-sample replication software available, but requires an approximation to the sampling error associated with his/her estimates of interest.

II. Method of Generalizing Variances

A generalized variance function (GVF) is a mathematical model describing the relationship between the variance or relative variance (relvariance) of a survey estimator and its expectation. If the parameters of the model can be estimated from past data or from a small subset of the survey items, then variance estimates can be produced for all survey items by evaluating the model at the survey estimates, rather than by direct computations.

Denote the estimator of a certain attribute of interest as Xhat and let $X = E\{Xhat\}$ denote its expectation. Then the relvariance can be expressed as follows:

$$V^2 = Var(Xhat)/X^2$$

Most of the GVFs to be considered are based on the premise that the relative variance is a decreasing function of the magnitude of the expectation X.

A simple model which exhibits this property is:

$$V^2 = A + B/X, \text{ with } B > 0. \qquad \text{(Model 1)}$$

The parameters A and B are unknown and to be estimated. Experience has shown that Model 1 often provides an adequate description of the relationship between V^2 and X. In fact, the Census Bureau has used this model for its Current Population Survey since 1947.

However, in an attempt to achieve an even better fit to the data than is possible with Model 1, the following are alternative forms of relvariance models which may be considered

$$V^2 = A + B/X + C/X^2 \qquad \text{(Model 2)}$$
$$\log(V^2) = A + B\log(X) \qquad \text{(Model 3)}$$
$$V^2 = (A + BX)^{-1} \qquad \text{(Model 4)}$$
$$V^2 = (A + BX + CX^2)^{-1} \qquad \text{(Model 5)}$$

where

V^2 = Relative variance

X = Expectation of the selected survey estimate

A,B,C = Unknown parameters to be estimated

Unfortunately, there is very little theoretical justification for any of the models discussed above. There is some limited justification for Model 1 (Wolter (1985).

III. Technical Approach

As a first step, a pilot test was conducted and based on the pilot test conclusions an exploratory analysis procedure was determined. The findings from the exploratory analysis determined which fitted model was to be used as the GVF.

a. Pilot Test

Step 1: Direct estimates of totals for selected student and teacher headcount variables from the School and the Teacher Demand and Shortage surveys at the national level (by sector and community type) were calculated. These estimates were chosen as a provisional group of similar items to be used for model estimation. A direct calculation of the variance of each of the totals using a balanced half-sample replication technique was used to derive the relvariance and the coefficient of variation (CV). Scatter plots of the log of the estimate versus the log of the CV were used to form "final" groups of statistics that followed a common model. These final groups were formed by simply removing from the provisional group those statistics that appeared to follow a different model than the majority of statistics in the group, and added other statistics, originally outside the provisional group, that appeared consonant with the group model.

As noted in Section II, there is no rigorous theoretical justification for any of the models that relate V^2 to X. Because we were unable to be quite specific about any of the models and their attending assumptions, it was not possible to construct, or even to contemplate, optimum estimators of the model parameters. Discussions of optimality would require an exact model and an exact statement of the error structure of the estimator $Vhat^2$ and Xhat. In the absence of a completely specified model, we attempted to achieve a good empirical fit to the data (Xhat, $Vhat^2$) as we considered alternative fitting methodologies.

Step 2: Using the calculated estimates and their CV's, un-weighted nonlinear models using SAS NLIN procedure were fit in order to produce least-squares estimates of the parameters of all five of the relvariance models described in section II above for each of the six subdomains groups (made up of combinations of public/private and urban/suburban/rural). The iterative method specified for the NLIN procedure was the modified Gauss-Newton method which regresses the residuals onto the partial derivatives of the model with respect to the parameters until the estimates converge.

Step 3: The results of the NLIN runs were summarized in terms of the RMSE and bias by quartile.

Step 4: An overlay of the scatterplot of the CV's versus the log of the estimate onto the fitted regression curve was plotted for each of the fitted models described in step 2.

Step 5: Finally, the results of steps 3 and step 4 were examined to help determine a viable subset of models to be used for the overall analysis. This determination was made by looking at both how well the data fit the model and how well the shape of the curve was in accord with reality.

Preliminary Results:

Both models 2 and 5 produced inappropriate shapes for the regression curve fit to the data in terms of a danger that extrapolation could lead to a result that was far from in accord with reality. Of the remaining models (1, 3 and 4), model 1 was the worst because the shape of the regression curve often dropped off too fast and leveled off too quickly. The shape of the curve for Model 3 seemed reasonable and appeared to fit fairly well overall, but had a higher RMSE than model 4. Also, model 3 resulted in a conservative (but possibly very large) predicted CV for small estimates. Model 4 had the best overall RMSE, largely due to a downward curvature on the left side of the regression curve. Model 4 also resulted in a possible bias (understatement) of CV's for large estimates. (See Figures 1 through 5 for examples representative of the regression curve plots produced during the pilot test. See Figures 6 and 7 as examples where model 4 had lower RMSE than model 3 and Figures 7 and 8 as examples where model 4 had lower RMSE than model 3.)

Preliminary Conclusions

Models 2 and 5 were to be excluded from any further analysis based on the inappropriate shape of the regression curve fit to the data. More data would be needed for small estimates to choose between models 3 and 4. Model 1 would be included for further analysis because it is the only model with limited theoretical justification. It was therefore decided to fit all three viable models (models 1, 3 and 4) using three alternative fitting methodologies: unweighted, weighted, and iteratively reweighted non-linear regression approach.

b. Exploratory Analysis

Step 1: Percentages, totals and averages for selected variables from each of the four SASS data sets (School, School Administrator, Teacher, Teacher Demand & Shortage (TDS)) for various subdomains (i.e., region, state, school level, minority status, school size, community status and combinations of these) were calculated.

Step 2: CV's for the estimates in step 1 were calculated using balanced half-sample replication techniques. Plots of the log of the estimate versus the log of the CV were used to finalize groups to be used for model estimation.

Step 3: Using the calculated estimates in each of the subdomain groups from step 1 and their respective CV's from step 2, nonlinear models using SAS NLIN procedure were fit in order to produce ordinary least-squares (OLS), weighted least squares (WLS), and iteratively re-weighted least squares (IRLS) estimates of the parameters and respective R-squared values for each of the relvariance models 1, 3 and 4 described in section II. The WLS procedure was specified to work with the sum of squares which weighted inversely to the square of the observed CV and the IRLS method was specified to work with the sum of squares which weighted inversely to the square of the predicted CV. The minimizing values from the OLS technique were used as starting values in the WLS and IRLS runs. A plot of the regression curve fit for each of the three methods (OLS, WLS, IRLS) of fitting a model was used to determine which method for fitting the model worked best. Based on these plots, the IRLS technique of model fitting proved to be best. The OLS technique gave too much weight to the small estimates whose corresponding relvariance was usually large and unstable and the WLS technique was a better procedure because it gave the least reliable terms in the sum of the squares a reduced weight, but the IRLS technique fit most of the data better than either of the other two techniques. A plot showing the R^2 values of one model versus another model was used to determine which GVF model fit best. (See separate volumes for the above mentioned plots).

Step 4: An out of sample test was performed to validate conclusions made from step 3.

Findings: The following are the selected IRLS models within each survey based on the exploratory analysis:

-- **The School Survey**

Student Totals - GVF Model 3 was selected
Teacher Totals - GVF Model 3 was selected
Averages - GVF Model 1 was selected

-- **The TDS Survey**

Student Totals - GVF Model 1 was selected
Teacher Totals - GVF Model 1 was selected
Averages - GVF Model 3 was selected

-- **The School Administrator Survey**

Admin Percents - GVF Model 1 was selected
Admin Totals - GVF Model 1 was selected
Averages - GVF Model 3 was selected

-- **The Teacher Survey**

Teacher Percents - GVF Model 1 was selected
Teacher Totals - GVF Model 1 was selected

-- **Salary Averages**

- GVF Model 3 was selected

Standard Error of a Ratio

To estimate the relative variance of an estimated ratio, R = X/Y, where Y is an estimator of the total number of individuals in a certain subpopulation and X is an estimator of the number of individuals in another subpopulation, use

$$V^2_R = V^2_X - V^2_Y$$

where the relvariances of X and Y are read from the appropriate GVF table. This formula has been shown to produce useful approximations. The approximation is appropriate when the correlation between the ratio X/Y and the denominator Y is close to 0; the approximation is an overestimate if the correlation is positive.

IV. References

Kish, L. (1967). *Survey Sampling*. New York: John Wiley and Sons.

Wolter, K. M. (1985). *Introduction to Variance Estimation*. New York: Springer Verlag.

Figure 1

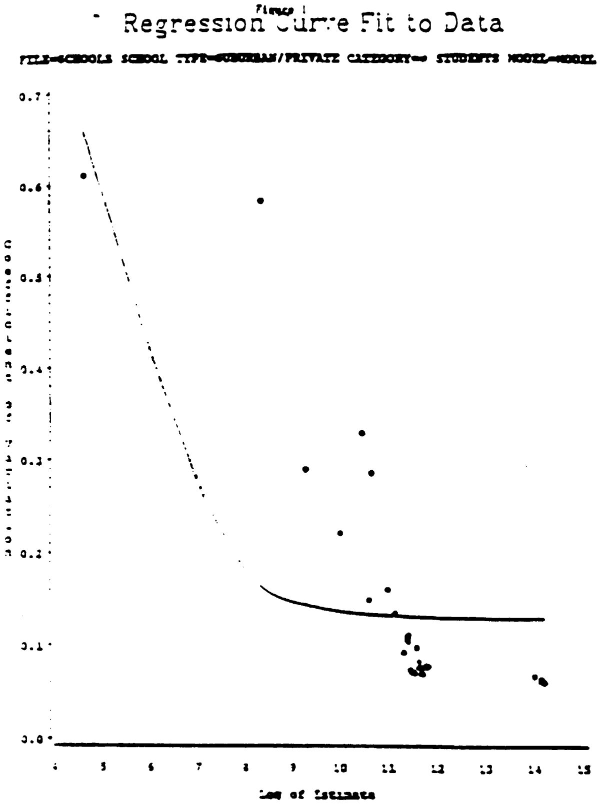


Figure 2

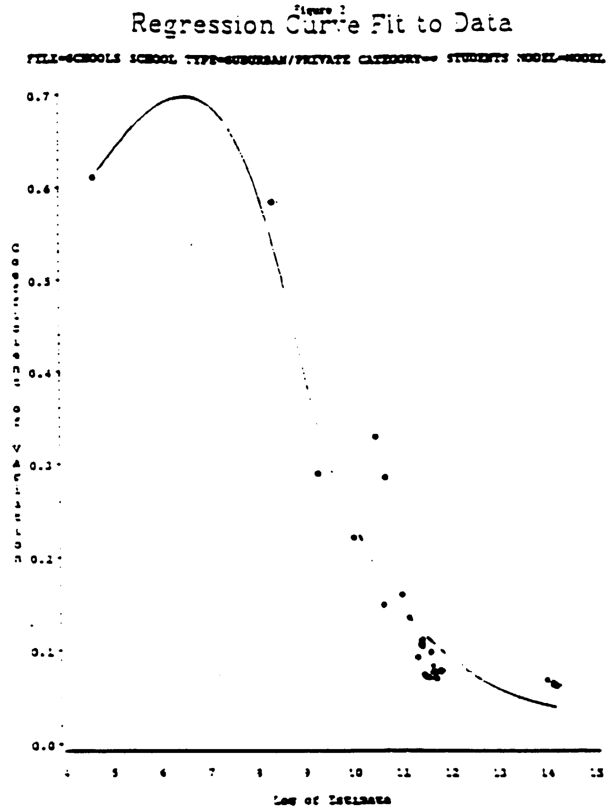


Figure 3

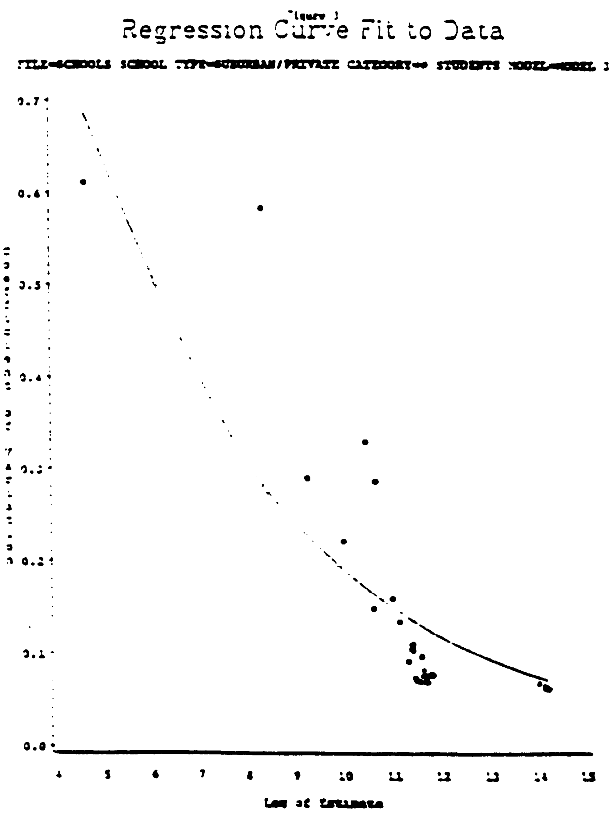


Figure 4

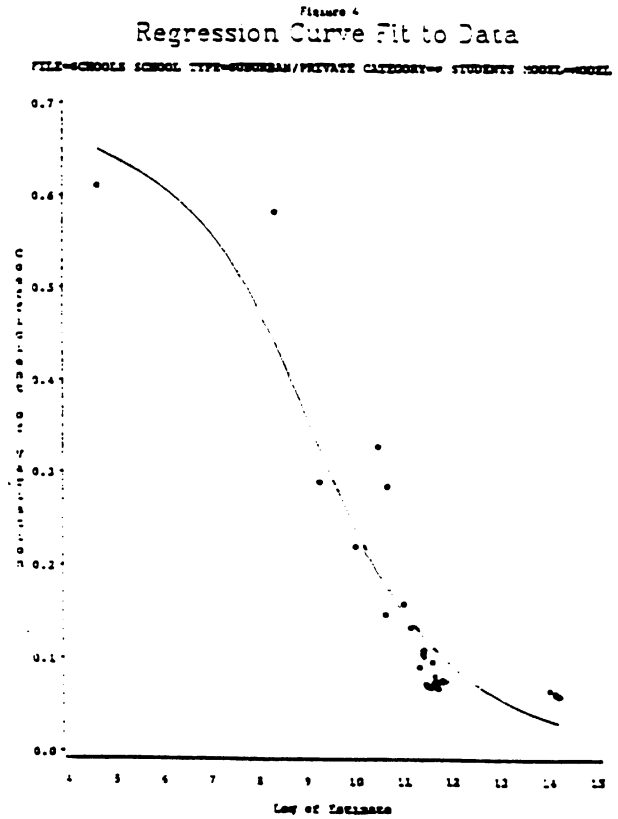


Figure 3

Regression Curve Fit to Data

FILE=SCHOOLS SCHOOL TYPE=SUBURBAN/PUBLIC CATEGORY=> STUDENTS MODEL=MODEL 3

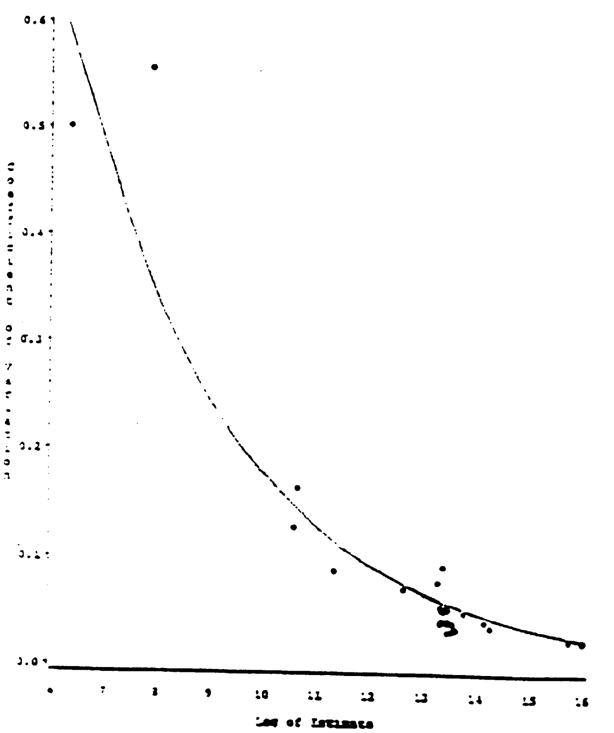


Figure 4

Regression Curve Fit to Data

FILE=SCHOOLS SCHOOL TYPE=SUBURBAN/PUBLIC CATEGORY=> STUDENTS MODEL=MODEL 4

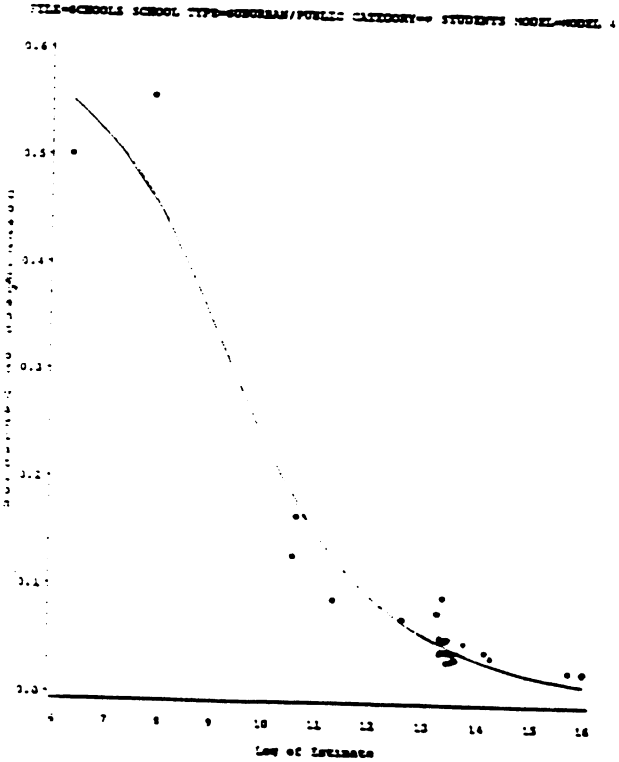


Figure 5

Regression Curve Fit to Data

FILE=SCHOOLS SCHOOL TYPE=SUBURBAN/PRIVATE CATEGORY=> STUDENTS MODEL=MODEL 5

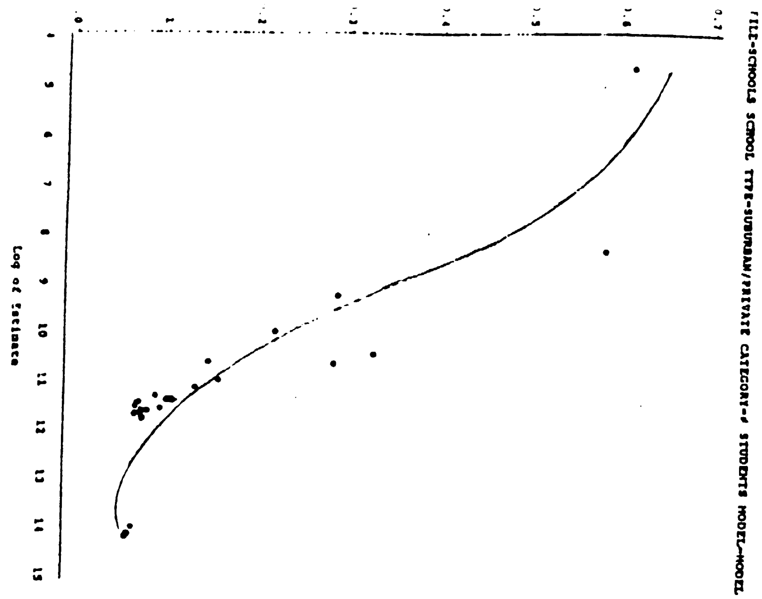


Figure 6

Regression Curve Fit to Data

FILE=SCHOOLS SCHOOL TYPE=RURAL/PUBLIC CATEGORY=> STUDENTS MODEL=MODEL 3

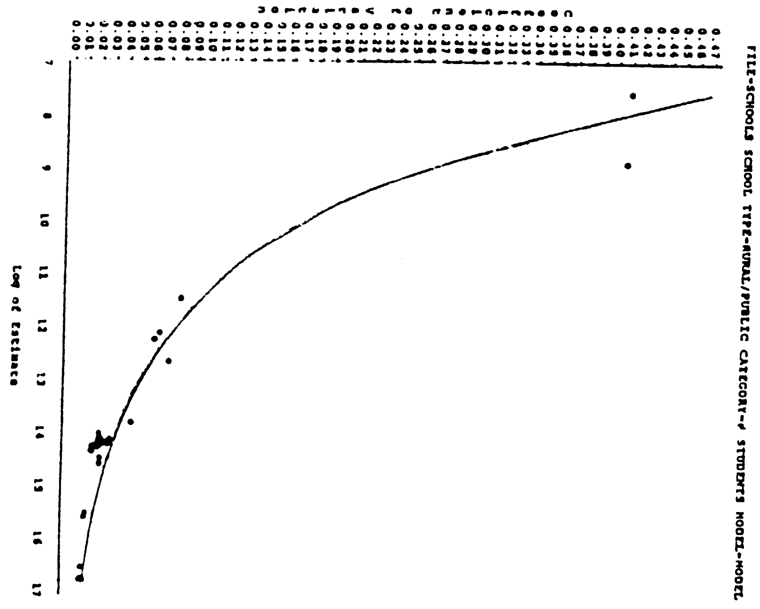
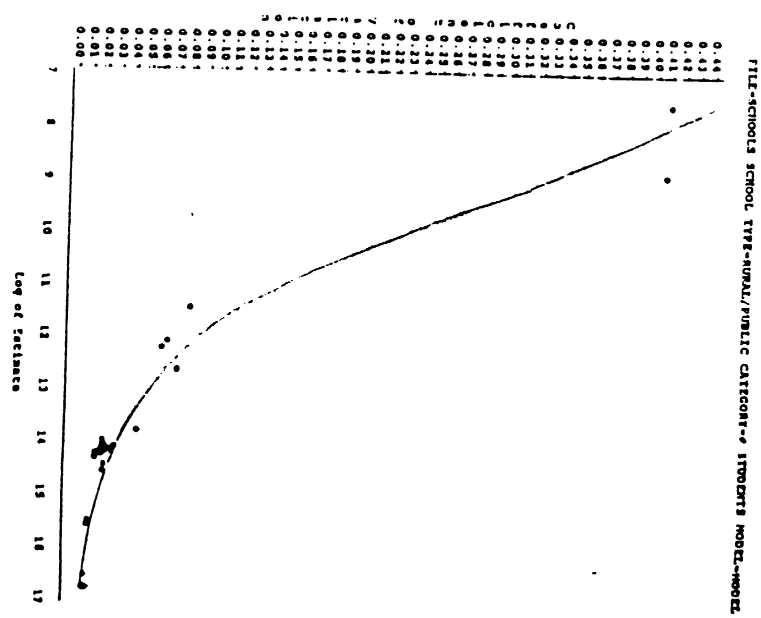


Figure 7

Regression Curve Fit to Data

FILE=SCHOOLS SCHOOL TYPE=RURAL/PUBLIC CATEGORY=> STUDENTS MODEL=MODEL 4



A BOOTSTRAP VARIANCE ESTIMATOR FOR THE SCHOOLS AND STAFFING SURVEY

Steven Kaufman, National Center for Education Statistics
Room 422e, 555 New Jersey Ave., Washington, D.C. 20208

Keywords: Half-Sample Replication, Bootstrapping

Introduction

The National Center for Education Statistics' (NCES) School and Staffing Survey (SASS) conducted by the Census Bureau has a complex sample design. Schools are selected using a stratified systematic PPS (unequal selection probabilities) sample design. From this design, data are collected at the school and school district level. The school district is an aggregation unit (i.e., the district selection probability is computed by aggregating school selection probabilities containing the district across the school strata). The probability is nonlinear with respect to the school sample sizes. It has been demonstrated (Kaufman,1992) under the usual Balanced Half-Sample (BHR) sample design that the BHR variance estimator for these district estimates can overestimate the variance. The apparent reason for the bias in the BHR estimator is that the district variances decrease faster than the inverse of the sample size, which BHR assumes. Since the bootstrap variance estimator doesn't necessarily make this assumption, this simulation study investigates whether a bootstrap variance estimator can perform better than the BHR variance estimator.

Another aspect of this paper is to investigate whether the bootstrap variance estimator reflects the finite population correction generated from the SASS sample design without using the joint inclusion probabilities. If independent systematic samples are selected, using the original sample design, then the simple variance of the estimates produced for each of the samples will reflect the appropriate variance. In this situation, units with selection probabilities close to one will appear in each sample more often then units with smaller selection probabilities. Since the bootstrap variance estimator mimics this process better than the BHR variance estimator, it might provide a better variance estimate, when the sampling rates are large.

The goal of this paper is to investigate, using a simulation study, whether a bootstrap variance estimator: 1) provides better variance estimates than the BHR estimator when estimates are based on aggregation units (school districts); and 2) reflects the impact of large sampling rates better than BHR. **The proposed bootstrap variances can be computed using any BHR program without any modifications.**

The SASS sample design for schools and school districts will be used for the simulation. The SASS district sample design will be used to study goal 1. Since the SASS is designed to produce State estimates, the sampling rates in small States are high; therefore, the SASS is a good design to demonstrate goal 2. Since the SASS sample design sorts the frame in a specific nonrandom order, four methods of sorting the bootstrap frame will be tested in the simulations.

Design of SASS School and District Surveys

The school survey uses NCES's public school Common Core of Data file as the frame. The frame is stratified by State, and within State by school level (elementary, secondary and combined). The school sample is selected using a systematic probability proportionate to size sampling procedure. The measure of size is the square root of the number of teachers in the school. Before sample selection, the school frame is sorted by a specific nonrandom order. The school districts that include a sampled school comprise the school district sample. In order to simplify the computation of the district selection probabilities, it is important, within each stratum, to keep schools belonging to the same district together.

Weighting

The school weight for school i (W_i) is:

$$W_i = 1/p_i$$

p_i: is the selection probability for school i.

The district weight for district d (W_d) is:

$$W_d = 1/(1-(1-p_{de})(1-p_{ds})(1-p_{dc}))$$

p_{de}: $\sum_{i \in S_{de}} p_i$

S_{de}: the set of all elementary schools in district d

p_{ds}: $\sum_{i \in S_{ds}} p_i$

S_{ds}: the set of all secondary schools in district d

p_{dc}: is $\sum_{i \in S_{dc}} p_i$

S_{dc}: is the set of all combined schools in district d

- If p_{de}, p_{ds} or p_{dc} is greater than or equal to one then the district is selected with certainty and $W_d = 1$.

Balanced Half-sample Replicates

The r^{th} school half-sample replicate is formed using the usual textbook methodology (Wolter, 1985) for establishment surveys with more than 2 units per stratum. The r^{th} district half-sample replicate is defined to be the set of districts that have schools in the r^{th} school half-sample replicate. Since the SASS half-sample variances are based on 48 replicates, the simulations will be based on 48 half-sample replicates.

The school replicate weight is:

$$RW_i = 2/p_i \, .$$

The district replicate weight is:

$$RW_d = 1/(1-(1-p_{de}/2)(1-p_{ds}/2)(1-p_{dc}/2))$$

The probabilities are divided by 2 because with half the sample, each school has half the chance of being selected.

Three BHR variance estimates will be presented based on the methodology described above. The first (**BHR no FPC**) is the variance estimates described above. This estimate does not make any type of Finite Population Correction (FPC) adjustments.

The other two make simple FPC adjustments. The second BHR variance estimate (**BHR Prob FPC**) adjusts the first variance estimator by $1-P_h$, where P_h is the average of the selection probabilities for the selected units within stratum h. For the district P_h's h represents a State.

The third BHR variance estimate (**BHR SRS FPC**) adjusts the first variance estimator by $1-n_h/N_h$, where n_h is the number of sample units in stratum h and N_h is the number of units on the frame in stratum h. For the district adjustments h represents a State.

School-bootstrap Frame

The idea behind the bootstrap samples is to use the sample weights from the selected units to estimate the distribution of the school and district frames. From the estimated school-bootstrap frame, B bootstrap samples can be selected. The school-bootstrap frame is generated in the following manner:

For each selected school i and associated district d, W_d bootstrap-districts (bd) are generated, as well as, W_i/W_d bootstrap-schools (bi) within each bootstrap-district. If W_d or W_i/W_d have a noninteger component then a full school is generated with a reduced selection probability and weight. As shown below, the bootstrap expectation of the bootstrap weights (W_{bi} or W_{bd}) equals the full-sample weight (W_i or W_d). The bi^{th} bootstrap-school has the following measure of size (m_{bi}):

$$m_{bi} = I_{bd} * I_{bi} * 1/W_i,$$

$$I_{bd} = \begin{cases} 1 & \text{if bd is an integer component of } W_d \\ C_{bd} & \text{if bd is a noninteger component of } W_d, \\ & \quad C_{bd} \text{ being the noninteger component} \end{cases}$$

$$I_{bi} = \begin{cases} 1 & \text{if bi is an integer component of } W_i/W_d \\ C_{bi} & \text{if bi is a noninteger component of } W_i/W_d, \\ & \quad C_{bi} \text{ being the noninteger component} \end{cases}$$

The sum of the m_{bi}s, generated from a selected school, equals one; so one bootstrap-school would be selected to represent school i, provided the bootstrap stratum sample size and sort order are the same as in the original design.

Each bootstrap-school, bi, generated within a bootstrap-district, bd, has the bd^{th} bootstrap-district's id. If the d^{th} district has selected schools in the elementary and secondary strata then the bd^{th} bootstrap-district id generated in the elementary stratum should match to the bd^{th} bootstrap-district id in the secondary stratum. This relationship should exist for all school levels that are selected for the district. This is important to compute the appropriate bootstrap-district weights.

Bootstrap Sample Size

The bootstrap sample size is usually chosen to provide unbiased variance estimates. When the original sample is a simple random sample of size n then Efon (1982) shows a bootstrap sample size should be n-1. Sitter (1990) has computed the bootstrap sample size for the Rao-Hartley-Cochran method for PPS sampling. A variation of this result is used in this simulation. The Sitter's bootstrap sample size (n^*) is the sample size which make the following quantity closest to 1:

$$\left(\sum_{g=1}^{n^*} (N_g^{*2}-N^*)\right) / \left(\sum_{g=1}^{n} (N_g^{2}-N)\right) * \left(N^2-\sum_{g=1}^{n} N_g^2\right) / \left(N^* * (N^*-1)\right)$$

n^*: is the bootstrap stratum sample size

g: represents a sampling interval in the stratum

N_g^*: is the number of bootstrap-schools in the g^{th} sampling interval, where the bootstrap-schools are in a random order

n: is the sample size in the stratum

N^*: is the number of bootstrap-schools in the stratum

N : is the number of schools in the stratum

N_g: is the number of schools in the g^{th} sampling interval, where the schools are in their original order; either a random order for the Rao-Hartley -Cochran method or the specific nonrandom order for the SASS method

n^* can not be calculated directly. The quantity above is computed for each n^* from n-10 to n. The n^* that is closest to one is used in the bootstrap selection.

The variation to Sitter's formulation is in the computation of N_g^* and N_g. Two modifications are made. The first occurs when either I_{bd} or I_{bi} are not equal to 1. Instead, of using 1, as Sitter does when counting units; $I_{bd} * I_{bi}$ is used to calculate N_g^*. To reduce the incidence of $I_{bd} * I_{bi}$ being not equal to 1, the districts are ignored when determining n^*. This is accomplished by generating a bootstrap frame as described above, assuming $W_d = 1$ (i.e., W_d never has a noninteger component). The second modification is due to the fact that a school or bootstrap-school can be in two sampling intervals. When this happens, N_g and N_g^* are not increased by one. Instead, they are increased by the proportion of the unit that actually goes into the sampling interval. If either I_{bd} or I_{bi} are not equal to 1, and the bootstrap-school is in two sampling intervals then N_g^* is increased by the product of the two modifications described above. If n is large, n^* should not be affected much by these modifications.

Sorting the School-Bootstrap Frame

If the bootstrap variance estimate is to work correctly, it is important that the school-bootstrap frame be randomized in an appropriate manner. In one extreme, when the bootstrap frame is sorted by the order of selection from the original sample and $n^*=n$, the variance estimate will be zero. In the other extreme, when the bootstrap frame is sorted randomly, the variance estimate ignores the original ordering and may overestimate the variance. Four orderings will be tested in this simulation study.

Sort Method 1

Schools within a stratum are sorted by order of selection. Next, schools are consecutively paired within each stratum. Each pair is assigned a random number. The bootstrap-districts and bootstrap-schools generated within each pair of schools are assigned bootstrap-district and bootstrap-school random numbers, respectively. Bootstrap-schools are sorted by the school pair random number; within each pair, bootstrap-schools are sorted by the bootstrap-district random number; and within the bootstrap-district, the bootstrap-schools are sorted by the bootstrap-school random number.

Sort Method 2

If the weights are relatively uniform within the set of paired schools, method 1 may underestimate the true variance. Sort method 2, tries to adjust for this. Sitter (1990) shows when the sample weights are uniform that his n^* will equal n-1. Hence, for this simulation, when n^* is between n and n-2, it will be assumed the stratum weights are relatively uniform and sort method 1 may underestimate the true variance. Instead, the bootstrap-schools are sorted by the bootstrap-district random number; and within the bootstrap-district, the bootstrap-schools are sorted by the bootstrap-school random number. If $n^* < n-2$, for a stratum, then the bootstrap-schools are randomized as described in sort method 1.

Sort Method 3

Sort method 3 is the same as sort method 2, except that the weights are assumed to be uniform when n^* is between n and n-3, instead of sort method 2's n and n-2. In this case, the bootstrap-schools are sorted by the bootstrap-district random number; and within the bootstrap-district, the bootstrap-schools are sorted by the bootstrap-school random number. If $n^* < n-3$, for a stratum, then the bootstrap-schools are randomized as described in sort method 1.

Sort Method 4

Sort method 4 does not use the school pairings; instead, bootstrap-schools are placed in a district and school random order. With this sort, the bootstrap-schools are sorted by the bootstrap-district random number; and within the bootstrap-district, the bootstrap-schools are sorted by the bootstrap-school random number.

Bootstrap Sample Selection

Given the bootstrap frame, m_{bi} as the measures of size, stratum bootstrap sample sizes and bootstrap-school ordering, select the bootstrap sample using the same sampling scheme as in the original sample. The bootstrap frame must be randomize with each sample selection. Bootstrap-schools, generated from noncertainty schools, with measures of size larger than the sampling interval are not removed from the sampling process. If a bootstrap-school is selected more than once, the bootstrap-school weight is multiplied by the number of times it is selected.

Number of Replicates and Bootstraps

Since the SASS BHR variances are based on 48 replicates, 48 bootstrap samples are computed for each simulation sample. Given the time it take to select a set of bootstrap samples, only 60 simulation samples are used.

Bootstrap Weights

The bootstrap-school weight, W_{bi}, is:

$$W_{bi} = I_{bd} * I_{bi} * M_{bi}/p_{bi}$$

M_{bi}: is the number of times the bi[th] bootstrap-school is selected

p_{bi}: is the bootstrap selection probability for the bi[th] bootstrap-school

$$E_{\bullet}(\sum_{bi} W_{bi})=\sum_{bi} I_{bd} * I_{bi} =\sum_{i} W_{i}, \text{ as desired.}$$

$E_{\bullet}$: is expectation over the bootstrap samples

Since the available data is defined by the districts selected in the original sample, a bootstrap-school weight indexed by i (BW_i) is required:

$$BW_i = \sum_{bi \in S_{iB}} W_{bi}$$

S_{iB}: is the set of all $bi \in i$ selected in the B[th] bootstrap sample.

The bootstrap-district weights, W_{bd} is:

$$W_{bd} = I_{bd}/(1-(1-p_{bde})(1-p_{bds})(1-p_{bdc}))$$

p_{bde}: is $\sum_{bi \in S_{bde}} p_{bi}$

S_{bde}: is the set of all elementary bootstrap-schools in bootstrap-district pd

p_{bds}: is $\sum_{bi \in S_{bds}} p_{bi}$

S_{bds}: is the set of all secondary bootstrap-schools in bootstrap-district bd

p_{bdc}: is $\sum_{bi \in S_{bdc}} p_{bi}$

S_{bde}: is the set of all combined bootstrap-schools in bootstrap-district bd

If p_{bde}, p_{bds} or p_{bdc} is greater than or equal to one then the bootstrap-district is selected with certainty and $W_{bd} = 1$.

$$E_{\bullet}(\sum_{bd} W_{bd})=\sum_{bd} I_{bd} =\sum_{d} W_{d}, \text{ as desired.}$$

Since the available data is defined by the districts selected in the original sample, a bootstrap district weight indexed by d (BW_d) is required:

$$BW_d = \sum_{bd \in S_{dB}} W_{bd}$$

S_{dB}: the set of all $bd \in d$ selected in the B[th] bootstrap sample.

Sample Estimate

For each of the simulation samples, totals, averages and ratios are computed within a number of the States and the District of Columbia, using variables available on the sample frame. For district samples, two averages are computed using teachers and schools; two ratios are computed using students, teachers and schools; and five totals are computed using students, teachers, graduates, schools and districts. For the school samples, two averages are computed using teachers and students; one ratio is computed using students and teachers; three totals are computed using students, teachers and schools. For each of the 60 simulation samples, the sample estimates and respective sample variances are computed for both district and school samples. An estimate of the true variance for the sample estimates can be obtained by computing the simple variance of the sample estimates across the 60 simulations. The bootstrap and BHR sample variance can now be compared with the estimate of the true variance.

A number of other analysis statistics are used. They are described below.

Analysis Statistics

Coverage Rates

To measure the accuracy of the variance estimates, a one sigma two tailed coverage rate is computed by determining what proportion of the time the population estimate is within the respective confidence interval. If the variance estimates are appropriate then the coverage rates should be close .68.

Coverage Rate Bias (**Bias**)

$$\text{Bias} = R_e - R_t$$

R_e: is the coverage rate based or either a bootstrap or BHR variance estimate

R_t: is an estimate of the true coverage rate. For a given estimator, it is based on the simple variance of the simulation estimates for that estimator

Tables 1-6 presents the coverage rate Bias's.

CV of Variance Estimate (**CV**)

To measure the variability of the variance estimate, the coefficient of variation (CV) of the variance estimate is calculated.

$$CV = [(1/59) \sum_{t=1}^{60} (V_t - \bar{V})^2]^{1/2}/\bar{V}$$

V_t: is the variance estimate for the t^{th} simulation estimate,

$\bar{V}$: is the average variance estimate across the 60 simulation samples.

Table 7 presents the CV of the variance estimates averaged across the States included in the study.

Results

Due to the time to complete the simulations, simulations for 4 large States (more than 2,000 schools) did not include bootstrap sort 1 or sort 2. First, tables 1-6 are discussed which are based on the 25 States in the simulations. The worst variance estimator is **BHR no FPC**. A large percent of the time the one σ coverage rates are better 2σ coverage rates than one σ coverage rates (i.e., Bias GE 0.14). The worst case is in table 5 with 68% of the estimates being better 2σ coverage rates than one σ coverage rates. One reason for this is because the sampling rates are very high in some States. The other two BHR variance estimate are better; but in 4 out of the 6 tables, there are still a reasonable number of estimates that are better 2σ coverage rates. In table 2, 24% of the estimates are better 2σ coverage rates. In general, the BHR variances tend to be overestimates.

An additional problem with the two FPC adjusted BHR variance estimates is that a number of the coverage rates are better $.5\sigma$ coverage rates than one σ coverage rates (i.e., Bias LT -0.14). The worst cases are found in table 4, where the Prob and SRS adjusted estimates have 60% and 36% of the coverage rates being better $.5\sigma$ coverage rates, respectively.

The best bootstrap variance estimator is the **bootstrap sort 4** estimator, with the **bootstrap sort 3** estimator a close second. There are still some coverage rates that are better 2σ coverage rates, but now the worst case is table 2 with 16%. The bootstrap variances for school estimates do tend to be underestimates, while district estimates tend to be overestimates. However, except for school ratios, the **bootstrap sort 4** estimator appears to be better than any of the BHR estimators. For school ratios, **BHR prob FPC** or **BHR SRS FPC** appear to be best. Some of the sort 4 estimates are better $.5\sigma$ coverage rates, but except for school ratios, the BHR FPC adjusted estimates are still worst overall with respect to this point. The worst **bootstrap sort 4** coverage rates are in table 3 (school ratios) with 20% being better $.5\sigma$ coverage rates. However, the absolute bias of the standard errors for these 20%, averages less than -0.04. Since the -0.04 bias is so small, even for school ratios **bootstrap sort 4** performs well.

If there is a desire to make an FPC adjustment for large sampling rates, the **bootstrap sort 4** appears to be the best variance estimator from those tested. However, if the desire is to always provide a conservative variance estimate then the **BHR no FPC** is the most conservative.

The major drawback with Bootstrap variances is that the calculation of the bootstrap replicate weights is far more complicated and computer intensive than the calculation of BHR replicate weights. However, this work only needs to be done once. Given the bootstrap weights, any BHR variance program can compute the bootstrap variance estimates, without any special adjustments. The bootstrap weights use most of the sample cases in each replicate, so when computing variances for ratios, there is not as much need to worry about zero denominators, as is the case with BHR variances.

When the sampling rates are lower one expects the **BHR No FPC** to provide good results. Although not presented here, this is true for the States in this study with low sampling rates. For these States, the **bootstrap sort 4** also provides good results, especially for school estimates.

Table 7 presents the CV of the variance estimates. For the most part, the BHR CV's are smaller than the Bootstrap CV's. However, the differences are small. For practical purposes, BHR and Bootstrap CV's are the same. One reason for this result is that the BHR replicates are only partially balanced.

References

Efron, Bradley(1982). The Jackknife, the Bootstrap and Other Resampling Plans. SIAM No. 38, p62.

Kaufman, Steven(1992). Balanced Half-Sample Replication with Aggregation Units. ASA 1992 Survey Research Methods Proceedings.

Sitter, R.R.(1990). Comparing Three Bootstrap Methods for Survey Data. Technical Report Series of the Laboratory for Research in Statistics and Probability, No. 152, p9-10.

Wolter, K. M.(1985). Introduction to Variance Estimation. New York: Springer-Verlag, p110-145.

Table 1 -- Frequency Distribution of 1 σ Coverage Rate Bias (**Bias**) for School Averages by Type of Variance Estimator

Bias Col Pct	Type of Variance Estimator: Bootstrap Sort 1	Sort 2	Sort 3	Sort 4	BHR Prob FPC	SRS FPC	No FPC
LT -0.14[1]	38.10	9.52	0.00	0.00	8.00	8.00	8.00
[-0.14., 0.0)	52.38	61.90	56.00	56.00	16.00	16.00	4.00
[0.0, 0.14)	9.52	23.81	32.00	32.00	64.00	56.00	44.00
GE 0.14[2]	0.00	4.76	12.00	12.00	12.00	20.00	44.00

Table 2 -- Frequency Distribution of 1 υ Coverage Rate Bias (**Bias**) for School Totals by Type of Variance Estimator

Bias Col Pct	Type of Variance Estimator: Bootstrap Sort 1	Sort 2	Sort 3	Sort 4	BHR Prob FPC	SRS FPC	No FPC
LT -0.14[1]	42.86	9.52	8.00	4.00	8.00	8.00	0.00
[-0.14., 0.0)	33.33	57.14	44.00	52.17	12.00	8.00	12.00
[0.0, 0.14)	23.81	23.81	36.00	28.00	56.00	60.00	48.00
GE 0.14[2]	0.00	9.52	12.00	16.00	24.00	24.00	40.00

Table 3 -- Frequency Distribution of 1 σ Coverage Rate Bias (**Bias**) for School Ratios by Type of Variance Estimator

Bias Col Pct	Type of Variance Estimator: Bootstrap Sort 1	Sort 2	Sort 3	Sort 4	BHR Prob FPC	SRS FPC	No FPC
LT -0.14[1]	66.67	52.38	20.00[3]	20.00[3]	0.00	0.00	0.00
[-0.14., 0.0)	33.33	38.10	64.00	64.00	48.00	48.00	16.00
[0.0, 0.14)	0.00	9.52	16.00	16.00	44.00	44.00	60.00
GE 0.14[2]	0.00	0.00	0.00	0.00	8.00	8.00	24.00

Table 4 -- Frequency Distribution of 1 σ Coverage Rate Bias (**Bias**) for District Averages by Type of Variance Estimator

Bias Col Pct	Type of Variance Estimator: Bootstrap Sort 1	Sort 2	Sort 3	Sort 4	BHR Prob FPC	SRS FPC	No FPC
LT -0.14[1]	19.05	19.05	16.00	16.00	60.00	36.00	0.00
[-0.14., 0.0)	38.10	38.10	32.00	32.00	24.00	32.00	20.00
[0.0, 0.14)	33.33	33.33	44.00	44.00	16.00	28.00	44.00
GE 0.14[2]	9.52	9.52	8.00	8.00	0.00	4.00	36.00

Table 5 -- Frequency Distribution of 1 σ Coverage Rate Bias (**Bias**) for District Totals by Type of Variance Estimator

Bias Col Pct	Type of Variance Estimator: Bootstrap Sort 1	Sort 2	Sort 3	Sort 4	BHR Prob FPC	SRS FPC	No FPC
LT -0.14[1]	14.29	19.05	16.00	20.00	32.00	16.00	0.00
[-0.14., 0.0)	52.38	52.38	44.00	40.00	60.00	48.00	4.00
[0.0, 0.14)	28.57	23.81	36.00	36.00	8.00	36.00	28.00
GE 0.14[2]	4.76	4.76	4.00	4.00	0.00	0.00	68.00

Table 6 -- Frequency Distribution of 1 σ Coverage Rate Bias (**Bias**) for District Ratios by Type of Variance Estimator

Bias Col Pct	Type of Variance Estimator: Bootstrap Sort 1	Sort 2	Sort 3	Sort 4	BHR Prob FPC	SRS FPC	No FPC
LT -0.14[1]	0.00	0.00	0.00	0.00	28.00	4.00	0.00
[-0.14., 0.0)	42.86	42.86	32.00	36.00	56.00	56.00	0.00
[0.0, 0.14)	57.14	52.38	64.00	60.00	16.00	40.00	40.00
GE 0.14[2]	0.00	4.76	4.00	4.00	0.00	0.00	60.00

Table 7 -- CV of the Variance (**CV**) by Type of Estimate and Type of Variance Estimator

	CV Type of Variance	Type of Estimate: School AVE	School RATIO	School TOTAL	District AVE	District RATIO	District TOTAL
B	No FPC	0.31	0.32	0.31	0.40	0.28	0.37
H	SRS FPC	0.31	0.32	0.32	0.43	0.35	0.38
R	PROB FPC	0.31	0.32	0.32	0.43	0.35	0.42
B	Sort 4	0.30	0.35	0.32	0.50	0.36	0.35
O	Sort 3	0.31	0.35	0.33	0.52	0.36	0.36
O	Sort 2	0.34	0.36	0.36	0.52	0.34	0.36
T	Sort 1	0.33	0.39	0.36	0.53	0.34	0.35

[1] Coverage rates in this category are better .5σ coverage rates than σ coverage rates

[2] Coverage rates in this category are better 2σ coverage rates than σ coverage rates

[3] The absolute bias of the standard errors in these States, averages less than -0.04

ADJUSTING FOR NONRESPONSE BIAS OF CORRELATED ITEMS USING LOGISTIC REGRESSION[1]

Pao-Sheng and Robin Fisher
Robin Fisher, U.S. Bureau of the Census, Washington, D.C. 20233

1. Introduction

Item nonresponse in sample surveys is the failure to obtain a specific question that should have been answered. In particular, item nonresponse rate on the survey of the economic is usually high. This can result for many reasons, the most frequent being "refusals to answer", which can relate to the underlying data value (non-ignorable nonresponse) and human behavior. Most designs of the survey questionnaire incorporate procedures for following up on missing responses to items identified as either especially important to the overall quality of the survey data or with previously noted high nonresponse rate. For example, the design of the Survey of Income and Program Participation (SIPP) questionnaire incorporated procedures for following up on missing responses to the items of wage and salary income, income received from self-employment and interest and dividend income. The response status on these items by the same individual are most likely correlated. The problem of missing items for categorical variable has been examined from the perspective of modeling the mechanisms of nonresponse by Fay (1986), Chambers and Welsh (1993), Alho (1990), and Särndal (1981).

This paper proposes a method of adjusting item nonresponse in presence of callback based on a generalized logistic regression model that can account for the correlation among responses on items. The probability of response for any item is represented by a logistic regression model, in which the value of that item, the response status of the rest of the items and the available covariates, which may include the observed item variables for all the individuals by the last callback, are explanatory variables. The respondents are assumed to answer some or all of items after one or more call-backs. The parameters of our model can be estimated by taking a conditional maximum likelihood approach based on the respondents. This approach has the advantage of the simple expression of conditional logistic model. The estimated individual probabilities of responding are used in a Horvitz-Thompson type estimator to reduce bias in the estimation of sample means for every single item.

2. The Logistic Regression for Correlated Responses

2.1 A Class of Conditional Logistic Models

Let I = {1, ..., n} be a set of indices for n individuals selected in a simple random sample. Let $X_i = (X_{i1}, ..., X_{iL})$ be the set of item outcomes from individual i and they suffer the nonresponse, i=1, ..., n, where X_{il} expresses the outcome of the l^{th} item from individual i, the value of which becomes known when individual i responds for the item l. The vector of covariates of individual i is denoted by Z_i. Suppose up to $J \geq 2$ attempts are made to capture the data for an individual. Define the nonresponse indicator vector $U_{ij} = (U_{ij1}, ..., U_{ijL})^T$, where, for l=1, ..., L with $U_{ijl} = 1$ if individual i was captured at the j^{th} attempt for the l^{th} item, and $U_{ijl} = 0$ otherwise.

Define $y_{ijl} = \sum_{k=1}^{j} U_{ikl}$ (i=1, ..., n; l=1, ..., L) for short. Then $y_{ijl} = 1$, if and only if individual i was captured by the j^{th} attempt for the l^{th} item. If U_{ijl} are correlated, the probability for $U_{ijl}=1$ not only depends upon X_{il} and Z_i but also depends on the responses for the rest of the items. First, consider the class of conditional logistic models when j=1

$$logit\ Pr\ (U_{i1l} = 1\ |U_{i1}^{-l}, X_{il}, Z_i) = F_{1l}\ (U_{i1}^{-l}) + G_{1l}(X_{il}, Z_i) \tag{1}$$

when j > 1

$$logit\ Pr\ (U_{ijl} = 1\ |U_{ij}^{-1}, S_{ij-1}, X_{il}, A_i) = F_{jl}\ (U_{ij}^{-1}) + G_{jl}(X_{il}, Z_i)\ if\ y_{ijl} = 0 \tag{2}$$

$$Pr\ (U_{ijl} = 1\ |\ U_{ij}^{-l}, S_{i,j-1}, X_{il}, Z_i) = 0 \quad if\ y_{i,j-1,l}=1$$

where U_{ij}^{-l} is U_{ij} with the exclusion of U_{ijl}, for j=1, ..., J, $S_{ij-1} = (U_{i1}^T, ..., U_{ij-1}^T)^T$.

---

[1] This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

F_{1l} is an arbitrary function of U_{i1}^{-1} such that $\sum_{l=1}^{L} U_{i1l} F_{1l} (I_{i1}^{-1})$ is invariant under permutation of U_{i1l}'s, where I_{i1}^{-1} is U_{i1}^{-1} with $U_{i1k} = 0$ for $k > 1$, F_{jl} is an arbitrary function of U_{ij}^{-1} such that $\sum_{l=1}^{L} U_{ijl} F_{jl} (I_{ij}^{-1}, S_{i,j-1})$ is invariant under permutation of U_{ijl}'s, and where I_{ij}^{-1} is U_{ij}^{-1} with $U_{ijk} = 0$ for $k > 1$

Thus F_{jl} is a function describing the dependence of item l on the response status of the other items in and before call-back attempt j. The function G describes the dependence on the outcomes X and the covariates Z.

For given F_{1l} and F_{jl}, from (1) and (2) we have the joint probability of U_{ijl}'s uniquely defined as when j = 1

$$Pr(U_{i1} \mid X_i, Z_i) = \exp\left[\sum_{l=1}^{L} U_{i1l}[F_{1l}(I_{i1}^{-1}) + G_{1l}(X_{il}, Z_i)]\right] \Big/ d_{1i} \qquad (3)$$

where

$$d_{1i} = \sum_{u_{i11},\ldots,u_{i1L}} \exp\left[\sum_{l=1}^{L} u_{i1l}[F_{1l}(I_{i1}^{-1}) + G_{1l}(X_{il}, Z_i)]\right],$$

$$u_{i1l} = 0,1,\ l=1,\ldots,L$$

when j > 1

$$Pr(U_{ij} \mid S_{i,j-1}, X_i, Z_i) = \exp\left[\sum_{l=1}^{L} U_{ijl}[F_{jl}(I_{ij}^{-1}, S_{i,j-1}) + G_{jl}(X_{il}, Z_i)]\right] \Big/ d_{ji} \qquad (4)$$

where

$$d_{ji} = \sum_{u_{ij1},\ldots,u_{ijL}} \exp\left[\sum_{l=1}^{L} u_{ijl}[F_{jl}(I_{ij}^{-1}, S_{i,j-1}) + G_{jl}(X_{il}, Z_i)]\right]$$

$$u_{ijl} = \begin{cases} 0,\ 1 & if\ y_{i,j-1,l} = 0 \\ 0 & if\ y_{i,j-1,l} = 1 \end{cases}$$

Equations (3) and (4) follow from an argument similar to that given Liang and Zeger (1989) in the appendix. Note that d_{1i} and d_{ji} are the normalizing constants which involve a sum of 2^L exponential terms.

2.2 An example

1. A special case is that where the response probability of item l in attempt j depends on the responses of the other item only through their number in attempt j (denoted by $r_{ijl} = \sum_{k \neq l} U_{ijl}$ and their number by attempt j-1 (denoted by $t_{i,j-1,l} = \sum_{k \neq l} y_{i,j-1,k}$). That is, when j = 1

$$logit\ Pr\ (U_{i1l} = 1 \mid U_{i1}^{-l}, X_{il}, Z_i) = F_1\ (r_{i1l}) + G_{1l}(X_{1l}, Z_i)$$

when j > 1

$$logit\ Pr\ (U_{ijl} = 1 \mid U_{ij}^{-l}, S_{i,j-1}, X_{il}, Z_i) = F_j(r_{ijl}, t_{i,j-1,l}) + G_{jl}(X_{il}, Z_i) \quad if\ y_{i,j-1,l} = 0$$

$$Pr\ (U_{ijl} = 1 \mid U_{ij}^{-l}, S_{i,j-1}, X_{il}, Z_i) = 0 \quad if\ y_{i,j-1,l} = 1$$

When j=1, we have the joint probability of U_{i1} uniquely defined as

$$Pr(U_{i1} \mid X_i, Z_i) = \exp\left[\sum_{l=1}^{L} U_{i1l}\left[F_1\ (B_{i1l}) + G_{1l}(X_{il}, Z_i)\right]\right] \Big/ d_{1i} \qquad (5)$$

where $B_{i1l} = \sum_{k=1}^{l-1} U_{i1k}$ and we assume that $B_{i11} = 0$

Similarly, when j > 1, the joint probability of U_{ij} conditional on $S_{i,j-1}$ is uniquely defined as

$$PR(U_{ij} \mid S_{i,j,l}, X_i, Z_i = \exp\left[\sum_{l=1}^{L} U_{ijl}\left[F_j(B_{ijl}, t_{i,j-1,l}) + G_{jl}(X_{il}, Z_i)\right]\right] \Big/ d_{ji}$$

$$\text{where } U_{ijl} = 0 \text{ when } y_{i,j-1,l} = 1$$
$$= 0 \text{ otherwise} \tag{6}$$

$$,\quad B_{ijl} = \sum_{k=1}^{l-1} U_{ijk} \quad \text{and we assume that } B_{ij1} = 0$$

We might choose the model to describe a situation in which some respondents are more willing to respond than others. Consider item l in attempt j. If the subject has already responded to a large number of items, we might expect him or her to be more likely to respond to item l. We index the function F with the call-back attempt to allow the dependence to change with the call-back number.

This type of model has been considered by Qu, William, Beck and Goormastic (1987) for the case J=1.

If we let $\sum_{t=0}^{T_i} F_1(t) = 0$ when $T_{i1} = 0$, and

$\sum_{t=t_{ij}-1}^{T_{ij-1}} F_j(t) = 0$ when $T_{ij} = t_{i,j-1}$ and $F_j\,(B_{ijl}\,,\, t_{i,j-1}) = F_j\,(B_{ijl} + t_{i,j-1})$, l=1,...,L,

We get the following class of conditional logistic models.

When j=1

$$Pr(U_{i1l}=1 \mid U_{il}^{-l}, X_{il}, Z_i) = \frac{e^{F_1(r_{i1l}) + G_{1l}(X_{il}, Z_i)}}{1 + e^{F_1(r_{i1l}) + G_{1l}(X_{il}, Z_i)}}$$

When j > 1

$$Pr(U_{ijl} = 1 \mid U_{ij}^{-l}, S_{i,j-1}, X_{il}, Z_i) = \frac{e^{F_j(r_{ijl}^{*}) + G_{jl}(X_{il}, Z_i)}}{1 + e^{F_j(r_{ijl}^{*}) + G_{jl}(X_{il}, Z_i)}} \text{ if } y_{i,j-1,l} = 0$$
$$= 0 \text{ if } y_{i,j-1,l} = 1$$

Where $r_{ijl}^{*} = r_{ijl} + t_{i,j-1,l}$
Here is one simple model.

$$G_{jl}(X_{il}, Z_i) = \alpha_{jl} + X_{il}\beta_{1l} + Z_i^T\beta_{2l} \tag{7}$$

$$F_1(r_{i1l}) = r_{i1l}\delta, \text{ and } F_j(r_{ijl}^{*}) = r_{ijl}^{*}\delta \quad \text{for } j>1 \tag{8}$$

Therefore the characteristics X_i and covariates Z_i affect the capture probabilities for the same way for each attempt. The effect of X_i is felt on P_{ijl} only through the l[th] characteristics X_{ij}. Different attempts may have different capture probabilities depending on the α_{ij}'s. This may reflect varying methods at callback or the possibility that the respondent's probability of response changes after a number of calls. We can imagine, for example, that a respondent may develop some resistance after even a small number of attempts have been made in which case α_{jl} decreases in j. Also notice that the number of responses in other items affect each item the same way.

3. **Estimation Procedure (Conditional Maximum Likelihood Approach)**

Without loss of generality we can order the data so that by observation 1 through n_i are the responded items for the i[th] individual and n_i+1 through L are the nonresponded items, we can estimate the probabilities based on the following 'working' conditional likelihood.

$$L^{\circ} = \prod_{i \in (I_A \cup I_S)} \prod_{l=1}^{n_i} \prod_{j=1}^{J} v_{ijl}^{U_{ijl}}$$

where I_A denotes the set of individuals who answer all the items, I_S denotes the set of individuals who only answer some of the items,and

$$v_{ijl} = \frac{\mu_{ijl}}{1-\mu_{i,J-1,l}}$$

$$\text{where } \mu_{i,J+1,l} = 1 - \sum_{k=1}^{J} \mu_{ikl}\text{, and where}$$

$$\mu_{i1l} = P_{i1l} = \frac{e^{F_{1l}(U_{il}^{-l}) + G_{1l}(X_{il}, Z_i)}}{1 + e^{F_{1l}(U_{il}^{-l}) + G_{1l}(X_{il}, Z_i)}},$$

$$\mu_{ijl} = P_{ijl} \prod_{k=1}^{j-1} (1-P_{ikl}), \quad j=2,\ldots,J, \quad \text{and where}$$

$$P_{ijl} = \frac{F_{jl}(U_{ij}^{-l}, S_{i,j-1}) + G_{jl}(X_{il}, Z_i)}{1 + F_{jl}(U_{ij}^{-l}, S_{i,j-1}) + G_{jl}(X_{il}, Z_i)}.$$

Maximum 'working' conditional likelihood estimates of the parameters can be found by numerically maximizing the log of this function with respect to the parameters involved. Consider the assumptions of (9) and (10). Does not have a unique maximum. One way to solve this problem is to use the available additional information in conjunction with the likelihood equation. For computational advantage, we use the approach proposed by Alho (1991).

Let $I_{1l} \subset I$ be the set of individuals captured at the first attempt for the l^{th} item ($l=1,\ldots, L$), I_{2l} the set of individuals captured at the second attempt for the l^{th} item, etc. Let $I_{J+1,l}$ be the set of individuals that are not captured for the l^{th} item at all. Define n_{jl} = card (I_{jl}), for $l=1,\ldots,L$ and $j=1,\ldots,J+1$; thus $n = n_{1l} + \ldots + n_{J+1,l}$ for $l = 1,\ldots,L$.

Note that for the l^{th} item, given $S_{l,J-1}$ and U_{ij}^{-1}, we have, for $j=1,\ldots,J$

$$E\left(\sum_{i\in \bigcup I_{k}} U_{ijl}\frac{1-P_{ijl}}{P_{ijl}}\middle| S_{i,j-1}, U_{ij}^{-1}\right) = (n_{jl} + \ldots + n_{J+1,l}) - \sum_{i \in \bigcup I_k} P_{ijl}$$

Notice $E(n_{jl} \mid S_{i,j-1}, U_{ij}^{-1}) = \left(\sum_{i\in\bigcup I_k} P_{ijl}\right)$; suppose we use n_{jl} to estimate this. We estimate the expectation on the left-hand side with the observed value also. This yields the likelihood equations

$$\sum_{i\in I_l} \exp(-\alpha_{jl} - W_{il}^T\beta_l - r_{ijl}\delta) = n - (n_{1l} + \ldots + n_{jl}),\ l=1,\ldots,L$$

Given β_l and δ we can thus solve for α_{jl}, by taking

$$\alpha_{jl} = -\log\,(n - n_{1l} - \ldots - n_{jl}) \Big/ \sum_{i\in I_l} \exp(-W_{il}^T\beta_l - r_{ijl}\delta)$$
$$(j=1,\ldots,J, l=1,\ldots,L)$$

To solve for

$$\alpha = (\alpha_1^T, \ldots, \alpha_L^T)^T \text{ where } \alpha_l = (\alpha_{1l},\ldots,\alpha_{Jl})^T, \text{ for } l = 1,\ldots,L$$

$$\beta = (\beta_1^T, \ldots, \beta_L^T)^T \text{ and } \delta$$

we use an iteration based on Newton's method.

Differentiating the log likelihood L with respect to β_l we get

$$\frac{\partial L}{\partial \beta_l} = -\sum_{y_{il}=1}\left[\sum_{j=1}^{J}(U_{ijl} - \nu_{ijl})\sum_{k=1}^{j} P_{ikl}\right] W_{il} = 0$$

We can solve numerically for α, β, and δ.

Having calculated the estimate $\hat\alpha$, $\hat\beta$, $\hat\delta$, the Horvitz-Thompson type estimator was considered based on the requirement of unbiasedness. Define $X_l = (X_{1l}, \ldots, X_{nl})^T$, $l=1, \ldots, L$.

The true sample mean for the item l is $\bar{X}_l = X_l^T 1_n / n$ where 1_n is a vector of n ones. By translating $\hat\alpha$, $\hat\beta$, $\hat\delta$, into (7) and (8), we can calculate the estimates $\hat\mu_{l,J+1,l}$ to get the conditional unbiased Horvitz-Thompson type estimator of $\bar{X}_l$ ($l=1,\ldots L$) as

$$\overline{X}_l = \frac{1}{n}\sum_{i\in I_l \cup I_l} X_{il}\, Y_{iJL} / \hat\mu_{l,J+1,l}$$

Let $\hat\gamma = (\hat\alpha, \hat\beta, \hat\delta)$ be the estimates of $\gamma = (\alpha, \beta, \delta)$ subject to regularity conditions on the X_{il}'s and Z_i's, the constraints used to ensure identifiability, coverage to continuous relation between the α_{jl}'s and (β_l, δ) which is satisfied by the true parameter vector γ. The proof for the consistency of the parameter estimator $\hat\gamma$ can be given by emulating the standard argument from Fahrmeir & Kaufmann (1985).

4. References

Alho, J. (1990). "Adjusting for Nonresponse Bias Using Logistic Regression." Biometrika 77, 617-624.

Besag, J. (1974). "Spatial Interaction and the Statistical Analyses of Lattice Systems." Journal of the Royal Statistical Society, Ser. B, 36, 192-236.

Chambers, R. L. and Welsh A. H. (1993). "Log-linear Models for Survey Data with Non-ignorable non-response." J. R. Statis. Soc. B, 55, 157-170.

Fay, R. E. (1986). "Causal Model for Patterns of Nonresponse." J. Am. Statist. Ass., 81, 354-365.

Fahrmeir, L and Kaufman, H. (1985). "Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models." Ann. Statist. 13, 342-368.

Liang, K. and Zeger, S.L. (1989). "A Class of Logistic Regression Models for Multivariate Binary Time Series." Journal of the Am. Statist. Assoc. 84, 447-451.

Qu, Y., Williams, G. W., Beck G. J., & Goormastic, M. (1987). "A Generalized Model of Logistic Regression for Clustered Data." Commun. Statist. - Theory Meth. 16, 3447-3476.

Särndal, C.E. (1981). "Frameworks for Inference in Survey Sampling with Applications to Small Area Estimation and Adjustment for Nonresponse." Bull. Int. Statist. Inst. 49, 494-513

Zhao, L.P. & Prentice, R.L. (1990). "Correlated Binary Regression Using a Quadratic Exponential Model." Biometrika 77, 642-648.

COMPARISONS OF SCHOOL LOCALE SETTING: SELF-REPORTED VERSUS ASSIGNED

Frank Johnson, National Center for Education Statistics
U.S. Department of Education, Washington, DC 20208-5651

KEYWORDS: SASS, CCD, Urban, Rural

This paper focuses on the geographic locale settings reported in two surveys conducted by The National Center for Education Statistics (NCES), part of the U.S. Department of Education. The two surveys are the School Universe component of the Common Core of Data (CCD) survey for school year 1988-89, and the Public School component of the Schools and Staffing Survey (SASS) for school year 1990-91. Instances where the self-reported locale setting code from SASS disagree with the assigned locale setting code from CCD are analyzed.

CCD Locale Code

The Common Core of Data (CCD) School Universe Survey is an annual collection, containing a record for every public elementary and secondary school in the United States and territories. NCES assigns each school a locale code by matching each school address to Census Bureau files. Census data used in assigning locale codes are 1) population and population density, 2) Standard Metropolitan Statistical Area (SMSA) codes, and 3) a Census code characterizing places as rural or urbanized areas. All Census data used in this project are based on the 1980 Census of Population and Housing. (For more information on the locale code assignment see Johnson, 1989.) The seven CCD locale codes are:

1. <u>Large City</u>: Central city of an SMSA, with the city having a population greater than or equal to 400,000 or a population density greater than or equal to 6,000 people per square mile.

2. <u>Mid-Size City</u>: Central city of an SMSA, with the city having a population less than 400,000 and a population density less than 6,000 people per square mile.

3. <u>Urban Fringe of Large City</u>: Place within an SMSA of a Large City and defined as urban by Census.

4. <u>Urban Fringe of Mid-size City</u>: Place within an SMSA of a Mid-size City and defined as urban by Census.

5. <u>Large Town</u>: Town not within an SMSA, with a population greater than or equal to 25,000.

6. <u>Small Town</u>: Town not within an SMSA and with a population less than 25,000 and greater than or equal to 2,500 people.

7. <u>Rural</u>: A place with less than 2,500 people or a place having a ZIP Code designated rural by Census.

Definitions of SMSAs and urban and rural areas are given below.

<u>Standard Metropolitan Statistical Areas (SMSA)</u>

SMSAs are defined by the Office of Management and Budget (OMB). Each SMSA comprises a central city or urbanized area and one or more neighboring counties. In order to be classified as an SMSA, two conditions must be met; 1) the central city or urbanized area must have a population of at least 50,000, and 2) the entire metropolitan area (including the central city or urbanized area) must have a total population of 100,000 or more inhabitants (75,000 in New England). Contiguous counties are included if they have close social and economic links with the area's population nucleus. Census assigns each of these SMSAs a unique code. At the time of the 1980 census there were 318 SMSAs in the United States.

The SMSAs that are used in this typology are those defined in 1983 by the Office of

Management and Budget (OMB). Since that time, they have been updated and expanded, and are now called Metropolitan Statistical Areas (MSA).

Urban and Rural Areas

The Bureau of the Census defines urbanized areas as consisting of a central city and surrounding densely settled territory with a combined population of 50,000 or more inhabitants. Places designated as urban by Census are within these urbanized areas or in places of 2,500 or more inhabitants outside these areas. All other areas are classified as rural. The urban and rural classifications cut across the SMSA classifications. There can be both urban and rural territory within an SMSA as well as in non-SMSA areas.

SASS Community Descriptor Codes

The School and Staffing Survey (SASS), Public School component, surveys a sample of schools using the CCD file as a sampling frame. This survey received responses from 8,969 schools. In this survey, respondents to the questionnaire select a locale setting which "best describes the community in which the school is located". There are ten community descriptors ranging from "a rural or farming community" to "a very large city (over 500,000 people)". Two of these community designations are beyond the scope of this analysis. They are "military base or station" and "Indian reservation." Ninety-nine of the 8,969 schools sampled chose these descriptors as best representing their school's setting. These schools have been dropped from this analysis. The remaining community description choices are listed below.

SASS community descriptor codes
1. A rural or farming community.
2. A small city or town of fewer than 50,000 people that is not a suburb of a larger city.
3. A medium-sized city (50,000 to 100,000 people)
4. A suburb of a medium-sized city
5. A large city (100,000 to 500,000 people)
6. A suburb of a large city
7. A very large city (over 500,000 people)
8. A suburb of a very large city

Overall findings

A breakdown of the locale settings assigned and reported for the schools responding to the SASS survey is provided below.

CCD assigned locale codes	Schools	Percent
1. Large central city	633	7.14
2. Mid-size central city	1,318	14.86
3. Fringe of large city	894	10.08
4. Fringe of mid-size city	871	9.82
5. Large town	242	2.73
6. Small town	2,220	25.03
7. Rural	2,692	30.35

SASS self-reported community descriptors	Schools	Percent
1. A rural/farming community	3,336	37.62
2. A small city or town	2,231	25.15
3. A medium-sized city	737	8.31
4. A suburb of medium-sized city	403	4.54
5. A large city	797	8.99
6. A suburb of large city	589	6.64
7. A very large city	408	4.60
8. A suburb of very large city	369	4.16

A crosstabulation is presented in Table 1. The two distributions are remarkably similar, especially if one takes into consideration the differences in the definitions of the two location typologies.

Reconciling CCD and SASS Locale Codes

There are several important differences between these two coding schemes. First of all is the distinction between assigning codes based on measurable demographic data versus an individual's perception of a community setting. The choice of a locale setting is likely to differ from individual to individual. Some individuals may change their response over a brief period of time (Bushery et al, 1992). Many people do not know the population of the town they live in, and one person's suburban is another one's rural.

Though there are inherent problems in an individual's choice of locale setting, there are problems with the CCD computer assigned locale codes as well. CCD locale codes are assigned based on mailing addresses. Several of these addresses are not the street address, but are Post Office boxes in nearby towns, and some schools report the school district mailing address instead of

Table 1.--Distribution of locale setting codes from 1990-91 SASS public school survey, by CCD locale code and SASS community descriptor

SASS community descriptor codes Frequency Percent Row Pct Col Pct	CCD locale codes 1 Large City	2 Mid-Size City	3 Large City Fringe	4 Midsize City Fringe	5 Large Town	6 Small Town	7 Rural	Total
1 Rural or Farm	* 2	* 50	* 15	* 94	* 21	936	2,218	3,336
	* 0.02	* 0.56	* 0.17	* 1.06	* 0.24	10.55	25.01	37.61
	* 0.06	* 1.50	* 0.45	* 2.82	* 0.63	28.06	66.49	
	* 0.32	* 3.79	* 1.68	* 10.79	* 8.68	42.16	82.39	
2 Small City	* 6	162	* 150	* 263	174	1,141	335	2,231
	* 0.07	1.83	* 1.69	* 2.97	1.96	12.86	3.78	25.15
	* 0.27	7.26	* 6.72	* 11.79	7.80	51.14	15.02	
	* 0.95	12.29	* 16.78	* 30.20	71.90	51.40	12.44	
3 Medium City	* 23	414	* 94	* 116	34	* 31	* 25	737
	* 0.26	4.67	* 1.06	* 1.31	0.38	* 0.35	* 0.28	8.31
	* 3.12	56.17	* 12.75	* 15.74	4.61	* 4.21	* 3.39	
	* 3.63	31.41	* 10.51	* 13.32	14.05	* 1.40	* 0.93	
4 Suburb of Medium City	* 2	* 84	* 68	139	* 9	* 48	* 53	403
	* 0.02	* 0.95	* 0.77	1.57	* 0.10	* 0.54	* 0.60	4.54
	* 0.50	* 20.84	* 16.87	34.49	* 2.23	* 11.91	* 13.15	
	* 0.32	* 6.37	* 7.61	15.96	* 3.72	* 2.16	* 1.97	
5 Large City	201	446	* 67	* 71	0	* 2	* 10	797
	2.27	5.03	* 0.76	* 0.80	0.00	* 0.02	* 0.11	8.99
	25.22	55.96	* 8.41	* 8.91	0.00	* 0.25	* 1.25	
	31.75	33.84	* 7.49	* 8.15	0.00	* 0.09	* 0.37	
6 Suburb of Large City	* 50	* 75	240	146	* 3	* 37	* 38	589
	* 0.56	* 0.85	2.71	1.65	* 0.03	* 0.42	* 0.43	6.64
	* 8.49	* 12.73	40.75	24.79	* 0.51	* 6.28	* 6.45	
	* 7.90	* 5.69	26.85	16.76	* 1.24	* 1.67	* 1.41	
7 Very Large City	309	* 61	* 27	* 9	0	* 1	* 1	408
	3.48	* 0.69	* 0.30	* 0.10	0.00	* 0.01	* 0.01	4.60
	75.74	* 14.95	* 6.62	* 2.21	0.00	* 0.25	* 0.25	
	48.82	* 4.63	* 3.02	* 1.03	0.00	* 0.05	* 0.04	
8 Suburb of Very Large City	* 40	* 26	233	* 33	* 1	* 24	* 12	369
	* 0.45	* 0.29	2.63	* 0.37	* 0.01	* 0.27	* 0.14	4.16
	* 10.84	* 7.05	63.14	* 8.94	* 0.27	* 6.50	* 3.25	
	* 6.32	* 1.97	26.06	* 3.79	* 0.41	* 1.08	* 0.45	
Total	633	1,318	894	871	242	2,220	2,692	8,870
	7.14	14.86	10.08	9.82	2.73	25.03	30.35	100.00

NOTE: Asterisks indicate conflicting locale setting assignments.

their own. There are also the technical problems of matching city names to files. Spellings, abbreviations and even the entire name can differ greatly through custom and keying errors. And there are towns recognized by the Post Office which are not recognized by the Census Bureau. Whereas steps have been taken in the CCD locale code assignment process to reduce these types of errors, they have not been totally effective.

Another difference lies in the terms suburb and urban fringe. "Suburb" is a common term denoting the settled areas surrounding a city. An effort to capture this setting was made in the CCD survey by the "Fringe" designations. CCD employed the use of SMSA definitions in order to make the locale assignments more scientific and to agree with definitions used elsewhere by the federal government. SASS was seeking a concise definition understandable by their respondents. Because the SMSA boundaries are defined to include whole counties, there are areas over a hundred miles from a city which are defined as Fringe of a large or mid-size city. Without a map of SMSA boundaries it would appear logical for respondents to code such areas as rural or small city.

A final difference occurs in the breakdown of cities and their corresponding fringe/suburban areas. CCD uses the central city of an SMSA definition as its cut-off for being a city, and then arbitrarily makes a distinction between large central cities and mid-size central cities based on population and population density. SASS arbitrarily set up its three tier classification scheme based on population.

Because of these differences, it is impossible to establish a perfect one-to-one relationship between the two coding schemes. However, the following crosswalk was developed in order to make a comparison. In nearly every case, one item in one coding scheme is matched to two items on the other coding scheme. This crosswalk is presented twice below, once in order of the CCD assigned locale code and again in order of the SASS self reported community descriptor on SASS.

CCD definition	SASS definition
1. Large central city	= 7. Very large city
1. Large central city	= 5. Large city
2. Mid-size city	= 5. Large city
2. Mid-size city	= 3. Medium-sized city
2. Mid-size city	= 2. Small city or town
3. Fringe of large city	= 8. Suburb of very large city
3. Fringe of large city	= 6. Suburb of large city
4. Fringe of a mid-size city	= 6. Suburb of large city
4. Fringe of a mid-size city	= 4. Suburb of a medium-sized city
5. Large town	= 3. Medium-sized city
5. Large Town	= 2. Small city or town
6. Small town	= 2. Small city or town
6. Small town	= 1. Rural or farming community
7. Rural	= 2. Small city or town
7. Rural	= 1. Rural or farming community

SASS definition	CCD definition
1. Rural or farming community	= 7. Rural
1. Rural or farming community	= 6. Small town
2. Small city or town	= 7. Rural
2. Small city or town	= 6. Small town
2. Small city or town	= 5. Large town
2. Small city or town	= 2. Mid-size city
3. Medium-sized city	= 5. Large town
3. Medium-sized city	= 2. Mid-size city
4. Suburb of a medium-sized city	= 4. Fringe of a mid-size city
5. Large city	= 2. Mid-size city
5. Large city	= 1. Large central city
6. Suburb of a large city	= 4. Fringe of a mid-size city
6. Suburb of a large city	= 3. Fringe of a large city
7. Very large city	= 1. Large central city
8. Suburb of a very large city	= 3. Fringe of a large city

Schools with conflicting locale settings

After removing schools in which the locale settings from the two coding schemes agree based on this crosswalk, there remain 1,742 schools where the codes do not agree. These occurrences are highlighted in Table 1 by asterisks to the left of the data inside the box. This represents 20 percent of the entire SASS public school sample. More than half of these schools with conflicting locale codes were coded as urban fringe on CCD (1,007 schools or 57 percent of the 1,742). The distribution of self-reported SASS locale codes in these 1,742 schools was more even, with the greatest number being coded small city or town (419 schools or 24 percent of the 1,742). Of these SASS reported small city or town schools, all but 6 schools were coded urban fringe on the CCD file.

Another finding is the small percentage of rural schools with conflicting codes. Of the 2,692 schools coded rural on CCD only 139 (5 percent) were not coded rural or small town by SASS respondents. Of the 3,336 schools reported as rural on SASS, only 182 (5 percent) were not assigned rural or small town codes by CCD. This would agree with the findings of Huang's study (1993) of rural codes in CCD and SASS.

Reexamining locale code decisions

The above discussion has dealt primarily with the differences in the two locale coding schemes and the difficulty in comparing them. Since neither of the code assignments can be characterized as perfect, the two locale codes were checked for every school in the SASS public school survey in five states: Iowa, Maryland, Massachusetts, Oregon and Utah. These states had a total of 815 schools. Maryland was chosen because of the author's familiarity with the state, and the other four were chosen to get a sampling across the nation.

The CCD and SASS locale codes were checked against 1980 Census data. Each locale code was identified as being correct or wrong. The location and population of the towns of seven schools could not be determined, and these schools were dropped from the analysis, leaving 808 schools. Schools located in places within ten miles of a city of greater than 50,000 people were determined to be in a suburban area in the SASS coding scheme. Schools more than 10 miles away from these cities but still in their SMSAs were counted as correct on the SASS survey if they were coded suburban or any of the appropriate city, town or rural codes depending on the place's population. Schools located in towns of greater than 10,000 people and less than 50,000 people were determined to be in a small town or city in the SASS coding scheme. The results of this study are presented in Table 2.

These results indicate that the Locale Desciptor on the SASS survey was correct for

Table 2--Verifying Locale Codes

State	Both correct	CCD correct SASS wrong	SASS correct CCD wrong	Both wrong	Total
IA	149 (84.2%)	21 (11.9%)	3 (1.7%)	4 (2.3%)	177 (100%)
MD	61 (43.3%)	58 (41.1%)	17 (12.1%)	7 (3.5%)	141 (100%)
MA	55 (35.3%)	51 (32.7%)	22 (14.1%)	28 (17.9%)	156 (100%)
OR	104 (64.2%)	34 (21.0%)	9 (5.6%)	15 (9.3%)	162 (100%)
UT	112 (65.1%)	51 (29.7%)	5 (2.9%)	4 (2.3%)	172 (100%)
Total	481 (59.5%)	215 (26.6%)	56 (6.9%)	56 (6.9%)	808 (100%)

Table 3--Counts of schools with incorrect locale codes by corrected CCD locale codes and corrected SASS community descriptor codes

Corrected SASS community descriptor codes (rows) by Corrected CCD locale codes (columns)

Frequency Percent Row Pct Col Pct	1 Large City	2 Mid-Size City	3 Large City Fringe	4 Midsize City Fringe	5 Large Town	6 Small Town	7 Rural	Total
1 Rural or Farm	0 0.00 0.00 0.00	0 0.00 0.00 0.00	8 2.45 36.36 7.14	9 2.75 40.91 7.20	0 0.00 0.00 0.00	1 0.31 4.55 4.17	4 1.22 18.18 50.00	22 6.73
2 Small City	0 0.00 0.00 0.00	1 0.31 2.22 2.86	3 0.92 6.67 2.68	15 4.59 33.33 12.00	2 0.61 4.44 100.00	23 7.03 51.11 95.83	1 0.31 2.22 12.50	45 13.76
3 Medium City	0 0.00 0.00 0.00	12 3.67 80.00 34.29	2 0.61 13.33 1.79	1 0.31 6.67 0.80	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	15 4.59
4 Suburb of Medium City	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	29 8.87 93.55 23.20	0 0.00 0.00 0.00	0 0.00 0.00 0.00	2 0.61 6.45 25.00	31 9.48
5 Large City	0 0.00 0.00 0.00	19 5.81 79.17 54.29	0 0.00 0.00 0.00	5 1.53 20.83 4.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	24 7.34
6 Suburb of Large City	0 0.00 0.00 0.00	2 0.61 3.03 5.71	2 0.61 3.03 1.79	61 18.65 92.42 48.80	0 0.00 0.00 0.00	0 0.00 0.00 0.00	1 0.31 1.52 12.50	66 20.18
7 Very Large City	18 5.50 100.00 85.71	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	18 5.50
8 Suburb of Very Large City	3 0.92 2.83 14.29	1 0.31 0.94 2.86	97 29.66 91.51 86.61	5 1.53 4.72 4.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	106 32.42
Total	21 6.42	35 10.70	112 34.25	125 38.23	2 0.61	24 7.34	8 2.45	327 100.00

66.4 percent of the schools investigated, whereas the locale code on the CCD file was correct for 86.1 percent of the schools. Or put another way, the SASS Locale descriptor was wrong in twice as many instances as the CCD assigned locale code.

The schools which initially had incorrect locale codes assigned to them by NCES or whose respondent chose the wrong community descriptor codes were subsetted and a cross tabulation performed by the corrected locale codes. These data are presented in Table 3.

Table 3 indicates that schools located in suburban or fringe areas are more likely to be coded incorrectly. Of the 327 schools with incorrect locale codes, 237 (72.5 percent) were found to be in an SMSA outside the central city, and 203 (62.1 percent) were found to be within 10 miles of a city.

These problems are due to the difficulty in defining suburban areas. This difficulty occurs on the SASS survey when respondents do not have a common understanding of what "suburban" means. Even when there are clear operational definitions, problems exist in the CCD locale code assignment process. These problems appear to be in matching mailing addresses (i.e., suburban post offices) with census place names and identifying their central city.

References:

Bushery, J., Royce, D., and Kaspryzk, D., "The Schools and Staffing Survey: How Reinterview Measures Data Quality", Proceedings of the Section on Survey Research Methods, American Statistical Association, Alexandria, Virginia, 1992

Huang, G., Defining the Rural: Johnson Code and Self-Report by School Personnel, Unpublished manuscript, 1993

Johnson, F., Assigning Type of Locale Codes to the 1987-88 CCD Public School Universe, NCES Technical Report, CS89-194, July 1989

THE ACCURACY OF TEACHERS' SELF-REPORTS ON THEIR POSTSECONDARY EDUCATION

Bradford Chaney, Westat, Inc.
1650 Research Blvd., Rockville, MD 20850

Key words: transcripts, recall, reporting error

This study compared teachers' self-reports of their academic qualifications, as provided on survey questionnaires, with the use of data from teachers' college transcripts. Teachers' self-reports are subject to problems with bias and recall, but the collection and analysis of transcripts, though more accurate, is also more complex. The U.S. Bureau of the Census conducted the initial mail and telephone survey of the teachers, and Westat, Inc. conducted the transcript portion of the study.

Data Collection

The 1990-91 Schools and Staffing Survey (SASS) was sent to 835 eligible teachers at 174 eligible schools, divided roughly equally between public and private schools, and between elementary and secondary schools. Of the 637 responding teachers, 45 either refused participation in the transcript portion of the study, or failed to supply information on which colleges they attended. Teachers who refused were left out of the transcript study.[1] This left a total of 592 teachers.

The SASS survey collected teachers' self-reports of their degrees earned, their majors and minors, the number of courses or credits taken in teacher education and in the teachers' two main teaching areas, and the number of courses taken in science and mathematics (among teachers who taught at least one course in science or mathematics). Teachers were also asked to list all colleges (both undergraduate and graduate) that they attended, whether or not they graduated from those colleges, and transcripts were sought from all colleges listed.

The total number of school responses was 1,658 out of 2,003 identifiable transcript requests, or 83 percent. At least one transcript was received for 92 percent of the teachers, and complete data were obtained for 51 percent of the teachers.

Teachers' Reports of the Schools they Attended

For 7 percent of the teachers, the teacher reported attendance at a college but the college stated the teacher never attended.[2] Some of these discrepancies may be due to differences in definitions of attendance. For example, one teacher said she listed a college where she had taken *noncredit* courses, because she felt the courses enhanced her perspective as an educator. Additionally, for 11 percent of teachers, colleges were unable to locate the teacher's transcripts; some of these may also represent teacher errors in listing the colleges attended.

Another type of error – a failure of the teacher to list all colleges attended – could sometimes be identified if other college transcripts included transfer credits from the missing college(s). For 9 percent of teachers, additional colleges were identified besides those listed on the questionnaire. These errors might be attributed either to poor memory on a teacher's part, or to the relative unimportance of the teacher's attendance at the college (*e.g.*, a single course during the summer). These estimates provide lower bounds on the number of omissions, since other omissions may not have been detected.

Overall, 23 percent of the teachers had at least one of the problems listed above, indicating that teachers' lists contain a significant amount of error.

Teacher Item Response Levels

Item nonresponse may lower data quality sufficiently to make transcript data preferable to questionnaire data. Also, if item nonresponse is due to teachers' inability to provide a correct response, an attempt to produce higher item response rates through increased followup may result in increased numbers of other errors.

Item response rates varied depending on the level of detail requested. They were highest for

---

[1] There is some evidence that teachers who refused had weaker backgrounds: 30 percent of teachers not in the transcript study said they had master's degrees, compared with 37 percent of those in the study, and 20 percent reported one undergraduate course or less in teacher education, compared with 14 percent of those in the study. (Neither relationship was statistically significant.) Bias might occur if teachers with lower qualifications were more reluctant to have their records reviewed. However, the fact that refusing teachers did report slightly lower academic qualifications might indicate they were willing to report their backgrounds accurately.

[2] Cases where colleges were found to have improperly reported that teachers never attended are excluded from these statistics.

general information about degrees earned (97 percent or higher). For other items, teachers generally were more likely to respond about *whether* they took courses in a subject area and whether a semester or quarter system was used than about the *number* of courses they took (Table 1). Teachers more often provided data on the number of graduate courses taken than the number of credits earned, although little difference appeared at the undergraduate level. Teachers were more likely to provide course data at the undergraduate level than the graduate level for courses in teaching methods and in the main teaching assignment, but there was little difference for courses in mathematics and science.[3] Possibly, the undergraduate level generally was more salient because of how graduation requirements are defined for a major, and because undergraduate courses are more likely to be taken over a compact time period, while responses still could be easier for the graduate level if teachers were highly likely to have taken *no* courses in a subject area.

Table 1. Item response rates

Type of information collected	Number of teachers eligible	Response rate	
		Under-graduate	Graduate
Courses in teaching methods			
Took courses	637	98	–
Semester or quarter system	594	97	–
Number of undergraduate courses	323	92	85
Number of undergraduate credits	271	90	78
Courses in main teaching assignment			
Took courses	317	94	–
Semester or quarter system	304	91	–
Number of undergraduate courses	168	92	79
Number of undergraduate credits	136	89	71
Courses in second teaching assignment			
Took courses	160	59	–
Semester or quarter system	154	56	–
Number of undergraduate courses	78	54	42
Number of graduate credits	76	53	38
Number of mathematics and science courses			
Mathematics	179	86	88
Computer science	120	73	80
Biology or life sciences	150	77	79
Chemistry	132	74	74
Physics	120	73	63
Earth or space science	103	64	66
Other natural science	93	52	53

NOTE: The questionnaire did not differentiate between undergraduate and graduate levels for the term type and for whether teachers took courses in an area.

[3]Typically, for each of these types of comparisons, the majority of comparisons were statistically significant. However, the exception was in comparing responses on the number of courses to the number of credits: the results were significant only at the graduate level, and only for courses in teaching methods or education.

Teachers' Reports on Their Degrees

Generally, teachers' self-reports on what degrees they earned showed a high correspondence with the transcripts, though up to 13 percent of the cases showed discrepancies for any particular degree.

Teachers' self-reports were most accurate on their bachelor's degrees. Essentially all teachers (528 of 538) reported they earned a bachelor's degree, and for all but 22 respondents, that report could be confirmed. Further, only partial data were received for 19 of the 22 cases, so it is possible that the receipt of additional transcripts would have confirmed those degrees. The remaining three teachers, plus three who failed to respond to the SASS question on bachelor's degrees, are the only teachers who can be clearly identified as providing incorrect or incomplete data on their bachelor's degrees.

A greater number of errors could be found for master's degrees. Nine teachers (2 percent) failed to report a master's degree, despite such an indication on their transcripts. Also, 57 teachers (11 percent) did not have their degrees confirmed: for 8 (1 percent) all transcripts were received, while for 49 (9 percent) the discrepancies may be caused by partial transcript data. Thus, the total proportion of teacher errors falls within the range of 3 to 12 percent.

Teachers displayed the same two errors for associate degrees as for master's degrees: 22 teachers (4 percent) failed to report an earned associate degree, and 16 teachers (3 percent) failed to have a self-reported degree confirmed. For 11 of the 16 teachers in the second group, only a partial set of transcripts was available.

Five teachers reported receiving doctoral degrees; of those, four degrees were confirmed, while only partial transcript data were available on the fifth. No other potential errors were detected concerning doctoral degrees.

The year the degree was earned. Teachers have more reason to err on the year of a degree. Sometimes a degree award is delayed until the next scheduled graduation ceremony, or a student may participate in a graduation ceremony with his/her peers before all requirements are met. Not surprisingly, then, the proportion of errors was higher than for listing degrees, ranging from 12 to 32 percent (Table 2). Most typically, teachers who made errors were off by 1 year.

Table 2. The year a degree was earned

Comparisons of self-reports and transcript data	Bachelor's degree	Master's degree	Associate degree	Doctoral degree
Number of teachers with complete data	427	137	19	4
Percentage of teachers with error	12	28	32	0
Too recent	5	14	11	0
Too early	7	15	21	0
Off by 1 year	7	20	26	0
More than 1 year	6	9	5	0

NOTE: Percentages may not sum to totals because of rounding.

Among the 427 teachers for whom the year of receiving a bachelor's degree was available, 12 percent made an error. The discrepancies were split between teachers who made an error of 1 year and those who off by more years, and between those who stated too early of a year and those who stated too late of a year.

A greater proportion of errors occurred for master's and associate degrees (28 and 32 percent, respectively), though most were off by only 1 year. Again, the errors were roughly evenly split between those who reported years that were too recent and those that reported years that were too early. Possibly, associate degrees were less salient. Graduate degrees are often earned part-time over many years, so the year may be less easily remembered; thus, though the relationship was not statistically significant, there were proportionally more errors if the master's degree was earned 6 to 10 years after the bachelor's degree than if it was earned earlier (38 percent versus 23 percent). However, the error rate was also lower (25 percent) if the master's degree was earned more than 10 years after the bachelor's degree, possibly because a recent degree was easier to remember.

Majors and minors. Teachers were asked to provide a code for their major for each degree. Such coding may have increased the potential for error. For example, a teacher might not notice that separate codes were provided for mathematics and mathematics education. Also, teachers might base their coding on their planned use of the major (*e.g.*, to become a mathematics teacher) rather than on the major alone.

For 65 percent of teachers earning bachelor's degrees, the major was correctly coded. Another 10 percent made errors only in whether the subject was listed as a separate discipline (*e.g.*, music) or as an area within education (*e.g.*, music education). For 12 percent, the teachers correctly reported majors within education, but gave the wrong specialty; typically, these involved differences in interpretation (*e.g.*, high school mathematics teachers listing secondary education rather than mathematics education), rather than radically different fields.

The remaining 13 percent made the greatest errors. Some errors were quite large (*e.g.*, reporting a major in biology/life science, while the transcript showed a major in art), but often they were a matter of judgment (*e.g.*, reporting an education major in counseling and guidance, while the transcript showed a major in psychology). Some errors were from poor coding (*e.g.*, classifying geography as geology/earth science rather than as a social science).

Teachers' Reports on the Courses they Took

Teachers were asked about courses they had taken in teacher education and their two main teaching assignments. Half were asked for the number of *courses* while the others were asked the number of *credits*. For the two main teaching assignments, the top two categories (*e.g.*, *5-9 courses* and *10 or more courses*) were designed to match common requirements for majors and minors. For teacher education courses, the top category was *4 or more courses*. Teachers were also asked whether the courses were taken using a semester system, a quarter system, or both.

When coding the transcripts, ambiguous cases were assumed to have been counted by a teacher. If a course might be classified within two separate disciplines, only one of which was covered in the questionnaire, teachers were assumed to have included it in their response. This coding procedure was chosen as the method of best approximating how teachers might answer the SASS questionnaire, but may result in overestimates of the courses taken.

Teacher education. Because of the questionnaire design, one might expect high accuracy in teachers' self-reports on teacher education: teachers were likely to be able to choose *4 or more courses* without having to count the exact number of courses, and if they had taken fewer courses, only a small number of courses needed to be remembered and counted.

Overall, 68 percent of the teachers gave responses that matched their transcripts at the undergraduate level (Table 3).[4] The greatest accuracy was in the category *4 or more courses*, with 81 percent giving responses that could be directly confirmed (not shown). The next highest accuracy was among teachers who reported taking no courses in teacher education; this answer may have been easier than counting the exact number of courses. Among the other two categories, a majority understated the number of teacher education courses they had taken. These categories were probably the most difficult: an error of a single course could make a teacher's response incorrect, and teachers may have difficulty remembering those areas where they took only a small number of courses.

Table 3. Courses and credits earned

Comparisons of self-reports and transcript data	Number of courses		Number of credits	
	Under-graduate	Graduate	Under-graduate	Graduate
Teacher education				
Number of teachers	280	250	228	196
Percent of teachers				
Report confirmed (total)	68	59	58	64
Partial tran. data	24	26	18	27
Teacher underestimate ..	14	18	22	12
Teacher overestimate	18	23	19	24
Partial tran. data	14	14	17	15
Main teaching assignment				
Number of teachers	141	118	111	91
Percent of teachers				
Report confirmed (total)	53	42	42	46
Partial tran. data	22	17	14	19
Teacher underestimate ..	11	3	17	7
Teacher overestimate	35	55	41	47
Partial tran. data	18	31	25	25
Second assignment				
Number of teachers	30	22	29	19
Percent of teachers				
Report confirmed (total)	37	45	14	37
Partial tran. data	10	9	3	21
Teacher underestimate ..	10	5	21	0
Teacher overestimate	53	50	66	63
Partial tran. data	27	32	21	21

NOTE: Percentages may not sum to 100 because of rounding.

Roughly the same patterns of accurate versus inaccurate responses were found in teachers' reports of the number of credit hours taken, and in their reports on graduate courses in education. However, one difference is that teachers were less likely to underestimate the number of graduate teacher credit hours taken, and more likely to overestimate the amount.

Teachers' main teaching assignment. The accuracy was lower than for teaching education, with 53 percent giving responses that could be confirmed at the undergraduate course level.[5] Generally, the balance between underestimates and overestimates changed, possibly due to the change in categories used. The percentage who overestimated the number of courses was much higher than for teacher education (35 percent); though half of these (18 percent) may be due to incomplete transcript data, the remainder was higher than for teacher education. While overestimates were more likely, underestimates were less likely. Excluding the highest category (where underestimates were impossible), teachers were less likely to give underestimates of the number of courses (from 30 to 50 percent) than in teacher education (from 44 to 63 percent).

Again, the results for teachers' self-reports on credit hours were not substantially different from those on the number of undergraduate courses taken. However, teachers were somewhat less accurate in their counts of graduate courses, with more overestimates and fewer underestimates.

Teachers' second teaching assignment. Few teachers reported a second teaching assignment, so the differences were not statistically significant.

Proportionally more errors were found for undergraduate courses than for the main assignment, with only 37 percent of the responses being confirmed. The largest group of errors was of overestimates; even excluding teachers whose transcript data were incomplete, 26 percent overestimated the number of courses. An even higher error rate was found among for the number of undergraduate *credits* earned, with 45 percent providing confirmed overestimates.

Semester and quarter systems. Teachers who reported that all courses were within the semester system were almost always correct (93 percent within teacher education), but teachers who reported all courses were within the quarter system were about equally likely to be

[4] For 24 percent of the teachers, it is possible that some teachers took more courses than were identified from the transcripts, because at least one transcript was never received. However, their reports are consistent with the data that are available. Also, two-thirds of these teachers fell within the category of those who reported taking 4 or more courses in teacher education; in their case, it is not possible for an additional transcript to conflict with their response, because there is no upper limit for this category.

[5] All comparisons in this paragraph are statistically significant.

correct (53 percent) or incorrect (47 percent; Table 4). Finally, 44 percent of teachers who reported *both* semester and quarter systems had their responses fully confirmed, and 32 percent might have had their responses confirmed if all transcripts were available.[6]

Table 4. Semester or quarter systems

Subject area	Total	Semester	Quarter	Both
Teacher education				
Number of teachers (total).	479	305	64	110
Percentage of teachers				
Report fully confirmed..	50	56	31	44
Partially confirmed	34	37	22	32
Main teaching assignment				
Number of teachers (total).	200	128	30	42
Percentage of teachers				
Report fully confirmed..	50	58	33	36
Partially confirmed	36	38	23	38
Second assignment				
Number of teachers (total).	41	26	9	6
Percentage of teachers				
Report fully confirmed..	46	58	22	33
Partially confirmed	29	27	22	50

NOTE: Percentages may not sum to totals because of rounding. *Partially confirmed* refers to transcripts that are consistent with the teachers' report, but the receipt of additional transcripts might confirm or deny the report.

The high error rate might be explained by teachers failing to consider all schools attended, and only considering where they took the most courses in the subject. Because semester systems were so common (75 percent of all courses for this study), teachers could often report using only the semester system and be correct. However, if they reported only using the quarter system, there is a good chance that at least one course was taken using the semester system.

Mathematics and science courses. Teachers who taught at least one course in science or mathematics (whether or not it was one of their main assignments) were asked to state the total number of courses taken in mathematics, computer science, biology or life science, chemistry, physics, earth or space science, and other natural science. For these questions, teachers were asked the *exact number* of courses.

The proportion who correctly stated the exact number ranged from 30 percent in mathematics to 71 percent in physics (Table 5).[7] However, teachers who had taken no courses within the discipline may have found it easy to respond. If these zeroes are excluded, the proportion giving correct answers was much lower, and ranged from 8 percent to 44 percent.[8]

Table 5. Mathematics and science courses

Subject area	Percent confirmed exactly		Mean number of courses		Mean total difference
	All teachers	Excluding teachers with zero courses	Teachers	Transcripts	
Undergraduate					
Mathematics	30	25	6.5	5.7	2.1
Comp. sci...	61	26	1.6	1.3	1.0
Life science.	49	34	3.5	3.2	1.3
Chemistry ...	66	44	1.8	2.0	0.5
Physics	71	35	1.8	1.6	0.6
Earth/space science.....	62	20	1.1	0.8	0.6
Oth. nat. sci	66	8	0.8	0.8	1.3
Graduate					
Mathematics	70	11	1.2	0.8	0.8
Comp. sci...	82	12	0.4	0.2	0.3
Life science.	85	8	0.5	0.4	0.4
Chemistry ...	88	25	0.8	0.4	0.5
Physics	89	11	0.8	0.9	0.5
Earth/space science.....	93	17	0.6	0.4	0.3
Oth. nat. sci	91	14	0.2	0.1	0.1

NOTE: Mean total difference is based on the absolute value of the difference between the teachers' self-reports and the transcripts.

The general tendency was to overstate the number of courses they had taken in a discipline. The difference was largest in mathematics, with teachers' reporting a mean of 6.5 undergraduate courses, while the transcripts showed a mean of 5.7. However, this understates the degree of teacher errors. Because some teachers gave overestimates and others gave underestimates, the errors partially balance out. If only the size of the difference between the teachers' self-reports and the transcripts is considered, the difference tends to be much larger: for example, for

[6] However, partial confirmations have less meaning in this case. When a teacher reports that all courses were taken within a single term type, then partial transcript data can confirm that, *at least as far as we know*, the teacher's report is correct. However, when a teacher reports that both term types were used, however, and partial data shows only one term type, it would be at least as accurate to say that *as far as we know* the teacher's report is incorrect as to say an additional transcript could confirm the teacher's report.

[7] Because of the emphasis on exact responses in this section, only teachers for whom all transcripts were available (or for whom full records were available if transfer courses were included) were included in the analysis.

[8] However, a side effect of excluding the zeroes is also to exclude measurement of another type of error: 25 teachers stated that they had taken no courses within one of the listed disciplines, but a transcript showed they had taken such a course. Still, these errors were less common than errors in reporting the exact non-zero number of courses taken.

undergraduate courses in mathematics, the average difference is then 2.1 (rather than 0.8).

Teachers' reports on graduate courses showed a similar pattern. However, teachers often had taken no graduate courses in the discipline. If the zeroes are excluded, teachers were actually less accurate in reporting on graduate courses than in reporting on undergraduate courses. (For example, 12 percent or less gave correct responses for mathematics, computer science, biology or life science, and physics, compared with 25 percent or more at the undergraduate level.) Also, perhaps because teachers tended to take fewer graduate courses in mathematics and science, the total distance between their self-report and their transcripts was sometimes smaller (especially for mathematics, computer science, and biology).

Table 6. Percentage giving accurate responses

Teaching area and teacher characteristic	Undergraduate		Graduate	
	Number of courses	Number of credits	Number of courses	Number of credits
Teacher education				
Gender				
Male	64	61	59	76
Female	80	67	68	69
Recency of bachelor's degree				
In last 10 years	79	60	68	75
11 - 20 years	75	69	65	59
Over 20 years ago	73	66	65	72
Institutional control				
Public	73	65	71	72
Private	78	66	56	64
Main teaching assignment				
Gender				
Male	66	49	68	63
Female	63	60	59	65
Recency of bachelor's degree				
In last 10 years	62	55	57	71
11 - 20 years	49	48	61	53
Over 20 years ago	78	62	67	62
Institutional control				
Public	61	55	60	56
Private	63	50	60	69

NOTE: Only teachers for whom complete transcript data were available were included.

Teacher Characteristics and Accurate Reports

No teacher characteristic showed a consistent pattern with relation to teacher accuracy for every statistic (Table 6).[9] For example, while some of the strongest differences were related to teachers' gender (80 percent of female teachers gave accurate responses on the number of teacher education courses, compared with only 64 percent among male teachers), for three of the eight statistics male teachers showed a higher accuracy rate. One might expect that teachers who received their bachelor's degree relatively recently could provide more accurate answers than those who had to recall their course backgrounds over longer periods of time, but again no consistent trend was found. Finally, the results were mixed based on the institutional control (public/private) of the schools where teachers taught.

Summary

For some types of data, such as general information on what degrees were earned, teachers showed high response rates and gave highly accurate data. There is little need to collect transcripts to verify these types of data, and the administration of a questionnaire is likely to be both simpler and less expensive. For more detailed data, the questionnaires were less useful. Non-response presented greater difficulties, and teachers were less likely to be accurate. One possible research strategy would be to redesign questionnaires to allow for these difficulties, while another would be to use some other source of data (such as institutional records).

In general, it appears better to request information on the number of courses than on the number of credits, given the lower item response rates and accuracy rates. Teachers were also most accurate when relatively large categories were used (*e.g.*, four courses or more in teacher education) or when identifying that they had not taken any courses in a field; they were not as effective in counting the exact number of courses. Finally, for some areas questionnaires may not provide reliable data. For areas that were not highly salient (such as the second teaching assignment), the levels of inaccuracy and nonresponse were sufficiently high that the administration of survey questionnaires seems inappropriate. For areas that might be complicated (*e.g.*, specifying term types), teachers' responses were also less reliable: given the predominance of semester systems, it would be roughly as accurate to assume all courses were semester courses as to use the teachers' responses.

[9]With only three exceptions, the relationships were also statistically insignificant. Given the lack of a consistent pattern, no teacher characteristic can be clearly related to teacher accuracy. For this section, cases where teachers' accuracy could not be clearly determined because of incomplete transcript data were excluded.

CHARACTERISTICS OF NONRESPONDENTS TO THE 1990-91 SCHOOLS AND STAFFING SURVEY

Wray Smith, Ramal Moonesinghe, Synectics for Management Decisions, Inc.,
and Kerry Gruber, National Center for Education Statistics
Wray Smith, Synectics Mgmt Decisions, 3030 Clarendon Blvd., Suite 305, Arlington, VA 22201

KEY WORDS: Hierarchical response patterns; Design effects; Unit nonresponse

The 1990-91 Schools and Staffing Survey (SASS), conducted by the Bureau of the Census for the National Center for Education Statistics, has nine interrelated components which collected data from public, private, and Indian schools, from public and private school administrators, from public and private school teachers, from public school districts (the teacher demand and shortage questionnaire), and also included a teacher followup survey one year after the main survey. This paper presents an analysis of the patterns of nonresponse (by school districts, schools, principals, and teachers) exhibited in the 1990-91 SASS, utilizing available prior information to characterize the nonresponding schools and districts and, within schools, the nonresponding teachers or administrators, and comparing the characteristics of nonrespondents with those of the respondents.

* * *

1. Overview

In the mid 1980s, the National Center for Education Statistics (NCES) conducted a number of surveys concerning schools and school personnel. In 1985, NCES undertook a critical review and redesign of its elementary and secondary data system, identifying gaps in content and inadequacies in design. As a result of this review, NCES, working with the Rand Corporation, designed a unified set of surveys that facilitates comparison between public and private schools and allows linkages of teacher, school, school district, and administrator data. This integrated set of surveys is called the Schools and Staffing Survey (SASS).

The SASS consists of four separate surveys administered simultaneously to linked samples of respondents. These surveys are the Teacher Demand and Shortage Survey, the School Administrator Survey, the School Survey, and the Teacher Survey. The SASS was designed to collect information on teacher supply and demand, the composition of the administrator and teacher workforce, and the status of teaching and schooling generally. The Teacher Survey was designed to obtain data on education and training, current assignment, job mobility, workplace conditions, and career choices of teachers. The Teacher Followup survey (TFS), conducted a year after the SASS, follows up on information obtained through the Teacher Survey and provides additional information about job mobility within the teaching profession, as well as between teaching and other careers.

In this paper, we address the issue of possible characteristics of nonrespondents in the nine components of SASS. Section 2 describes the SASS surveys in general. Section 3 presents the methodology of statistical tests we will use to test the differences. Section 4 presents the results and Section 5 gives our conclusions.

2. General Survey Description

The 1991 Schools and Staffing Survey consists of a school, a teacher, and for public schools a Local Education Agency or school district survey. The public school sampling frame was based on the 1988-89 school year Common Core of data or CCD. This CCD was matched to the previous SASS public school sampling frame. Non-matches from the previous frame were included with the CCD to make up the public school sampling frame for 1991. Public schools were first stratified by three types of schools: (A) Native American schools, (B) schools in Delaware, Nevada and West Virginia, and (C) all other schools. For the second level of stratification, the type A schools were stratified by Arizona, North Dakota, Oklahoma and all other states, the type B schools were stratified first by state and then by LEA and, the type C schools were stratified by 47 states and the District of Columbia. Within each second level there were three grade level strata: elementary, secondary, and combined schools.

The private schools were selected from a list frame, constructed by matching multiple lists obtained from private school organizations, State Departments of Education, and a private vendor. This frame is thought to include 80-90% of private schools. To increase the coverage of the survey, an area frame was constructed by selecting 120 PSUs, consisting of counties or groups of counties. Within these sample counties, lists of schools were obtained from local sources, such as yellow pages, churches and fire marshals. These lists were unduplicated with the list frame. The remaining schools, not matching to the list frame, make up the area frame. For list frame private schools, the frame was partitioned into an initial set of 216 cells. The first level of stratification was school association membership

(18). Within each association membership, schools were stratified by grade level (elementary, secondary, and combined). In some cases, when the grade level is unknown, it was imputed.

Once schools were selected, districts associated with these schools were in sample as well. Hence the district sample consisted of the set of districts that were associated with the SASS public school sample. This provided the linkage between the district and the school. This portion of the district sample represented the set of districts associated with schools. Sample size for districts with schools was 5,380. Some districts were not associated with schools. Such districts may hire teachers who teach in schools of other districts. Sample size for districts without schools was 135 units.

The selected schools (public and private) were asked to provide teacher lists for their schools. From the lists, 56,051 public and 9,166 private teachers are selected. Ten percent of the in-scope private schools and five percent of the in-scope public schools did not provide teacher lists. Within each selected school, teachers were stratified into one of five teacher types in the following hierarchical order: (A) Asian or Pacific Islander; (B) American Indian or Aleutian or Eskimo; (C) Bilingual; (D) New; and (E) Experienced. Within each teacher stratum, teachers were sorted by primary field of teaching.

3. Response Rates

The greater the nonresponse, the more one has reason to worry about its harmful effects on the survey estimates. The bias often increases with the rate of nonresponse. It is hard to get objective measures of the bias, but its relatively simple to quantify the extent of nonresponse.

A simple measure of the unit response is

$$p_r = \frac{n_r}{n_s}$$

where n_r is the number of respondents and n_s is the sample size. The unit nonresponse is consequently measured by $1-p_r$. Here p_r measures how well the survey has succeeded in obtaining at least partial response from the elements in the selected sample. Alternative measures are obtained by sample-weighted quantities.

The sample-weighted measure of unit response is where r and s denote the set of respondents and the sample respectively and Q_k is the probability of

$$pw_r = \frac{\sum_r \left(\frac{1}{Q_k}\right)}{\sum_s \left(\frac{1}{Q_k}\right)}$$

selection of the k^{th} unit.

pw_r can be interpreted as an estimated average response probability in the population. Unweighted- and weighted-measures may differ considerably. The basic weight is simply the inverse of the probability of selection, $(1/Q_k)$, as of the time of sampling.

We derived the unweighted response rates by dividing the number of interviews by the number of eligible cases (the number of sample cases minus out-of-scope cases; for example, school closed, no elementary or secondary teachers, teacher retired). The weighted response rates were derived by dividing the sum of the basic weights for the interview cases by the sum of the basic weights for the eligible cases. Since prior information on quantitative variables are not available, our response rates are based only on counts. When quantitative variables are available, the calculation of value-weighted rates may be an attractive alternative (pp562, Sarndel et al.,1992).

Characteristics of nonrespondents are compared with those of respondents, to help answer the question, "What is known about the nonrespondents to SASS?". For each component, we quantify the response rates across a number of dimensions-sampling stratum, state or private school association, school level, and other stratification variables-both weighted and unweighted. The hierarchical response patterns are also studied.

We also conducted significance tests of independence between response and the stratification variables of interest. The usual Pearson Chi-Square test produced in SAS by PROC FREQ may not be appropriate for this analysis due to the complex sample design. WESVAR is used to test for the independence of the two classification variables. The modified chi-square statistics all rely on modifying the Pearson chi-square statistic using an estimated "design effect". The Fellegi method is based on Felligi (1980), while the other three were suggested by Rao and Scott(1981 and 1984). Design effects were obtained based on the estimated variance using 48 pseudo-replicates.

4. Results

Tables A-1 through A-10 present response rates for selected subgroups. Note that this analysis is conditioned on the sample that was selected in 1991, so no standard errors are used.

As Table A-1 reveals, public schools have a overall higher response rate compared with private schools. These are the national response rates by questionnaire type. Response rates for public schools by primary cluster grouping are given in Table A-2. If the school has at least 25% American Indian students, it is classified in the "High percentage Native American" group. Otherwise, if it is located in Delaware, Nevada, or West Virginia, the school falls into the second group. The remaining cases are combined into the "all other" group.

Response rates for public schools by state are given in Table A-3. There does not seem to be a difference between weighted and unweighted response rates for public schools by state. Maryland has the lowest response rate (81%) for public schools and Indiana has the highest response rate (more than 99%).

Table A-4 shows response rates for public schools by school level. Elementary level is any school with no grade higher than the eighth grade; secondary level is a school with at least some grades higher than the ninth grade, while a combined level is any school with grade ranges below the sixth grade and above the eighth grade. For example, a kindergarten through eighth grade range is an elementary, while a school with fourth grade through twelveth grade is a combined level. There is no significant difference in response rates among the school levels. Table A-5 shows the response rates for public schools by percent minority. This reveals a slight decline in response rates for schools that have higher percentage of minority students. Table A-6 presents response rates for public schools by school enrollment. This also shows a slight decline in response rates for schools that have higher school enrollment.

Table A-7 presents the response rates for private schools by type of frame. The frame type is the source for sampling the private schools; The list frame is developed from an association membership list, such as the National Association of Independent Schools, or the prior response to the Private School Survey; while the area frame is developed from a search of selected areas' schools that do not appear on any list. The area frame was conducted in selected areas to supplement the known undercoverage of the list frames. An area frame supplements the list frame in the Private School Survey as well as in SASS. Area frame cases may be more difficult to followup - as in the case of Amish schools without telephones. The area frame cases have a much lower response rates.

Table A-8 lists the response rates for private schools by list frame association membership. The area frame cases presented in this table are for comparative purposes only. The list frame associations in this table are the ones used in stratifying the file. The response rates for some associations are below publishable standards, and the NCES is taking steps in the next SASS to ameliorate the reluctance of schools in those groups.

Table A-9 presents the response rates for private schools in the association list frame by school level. Combined schools have a lower response rate than either elementary or secondary level schools. Table A-10 presents the response rates for private schools in the area frame by school level. All area frame schools have lower response rates overall than the list frame schools, but area frame schools follow similar response rate patterns by level as the list frame schools. Combined schools have the lowest response rates, and there are more of them proportionately in the area frame than are found in the list frame.

Table A-11(1) presents the hierarchical response rates for the school survey and the district survey. There are 8397 schools that have responded to the school survey with the corresponding districts that have responded to the district survey. There are 487 schools that have responded even though their corresponding districts have not responded. Table A-11(2) shows the percentile distribution of school nonresponse at the district level. Of those districts that have responded 89.53% have 100% school response and of those districts that have not responded 78.68% have 100% school response. This shows that there is no correlation between school nonresponse and district nonresponse. A forthcoming technical report prepared for NCES will also present detailed analysis of response rates for all the surveys.

5. Conclusion

Response rates are generally good for the SASS surveys, but we believe there is room for improvement. We have identified some subgroups for which the response rate is relatively poor, specifically for large city public schools, and for specific affiliations of private schools. Combined schools have a lower response rate than either elementary or secondary level schools. Through cognitive interview research, NCES has found that small combined schools have more difficulty knowing which grades to report data for on the SASS public or private school questionnaire. The difficulty of this task may persuade some schools not to participate. Dillman (1991) suggests methods for improving mail response rates, such as questionnaire design, use of reminders, and length of the questionnaire. Also, establishment of better contact with the specific school organizations mentioned should help to improve response rates. NCES has taken number of steps to improve the response rates: the questionnaire will be designed more clearly (with the school name and expected grade range clearly marked); NCES has asked all sampled schools with questions about school identifications to call them on an 800 number if confused; Also NCES intend to meet with public representatives about the combined school problem (which should elicit strategies that work with private schools as well).

6. References

Dillman, D. (1991). The Design and Administration of Mail Surveys. Annual Review Sociology, 17, 225-249.

Fellegi, I.P. (1980). Approximate tests of Independence and Goodness of Fit Based on Stratified Multistage Samples. Journal of the American Statistical Association 71, 665-670.

Rao, J.N.K. and Scott, A.J. (1981). The Analysis of Categorical Data from Complex Sample Surveys: Chi-Squared Tests for Goodness of Fit and Independence in Two-Way Tables. Journal of the American Statistical Association 76, 221-230.

Rao, J.N.K. and Scott, A.J. (1984). On Chi-Squared Tests for Multiway Contingency Tables with Cell Proportions Estimated from Survey Data. The Annals of Statistics 12, 46-60.

Samdal, C., Swensson, B. and Wretman, J. (1991). Model Assisted Survey Sampling. Springer-Verlag.

TABLE A-1
TOTAL SURVEY RESPONSE RATES FOR SASS, 1990-91

SECTOR	WEIGHTED RESPONSE RATE	UNWEIGHTED RESPONSE RATE
PUBLIC SCH	95.30	95.10
PRIVATE SCH	83.95	85.10

TABLE A-2
RESPONSE RATES FOR PUBLIC SCHOOLS BY CLUSTER GROUP

	WEIGHTED RESP RATE	UNWEIGHTED RESPONSE RATE
HIGH % NAT AMR	97.46	97.47
DE, NV, OR WV	97.22	96.05
OTHER SCHOOLS	95.24	94.97

TABLE A-4
RESPONSE RATES FOR PUBLIC SCHOOLS BY SCHOOL LEVEL

LEVEL	WEIGHTED RESPONSE RATE	UNWEIGHTED RESPONSE RATE
ELEMENTARY	95.31	95.48
SECONDARY	95.51	95.20
COMBINED	94.12	93.03

TABLE A-5
RESPONSE RATES FOR PUBLIC SCHOOLS BY PERCENT MINORITY

PERCENT MIN	WEIGHTED RESPONSE RATE	UNWEIGHTED RESPONSE RATE
00 TO 05%	97.09	96.63
06 TO 20%	94.58	94.19
21 TO 50%	94.39	93.52
51 OR MORE	92.41	93.05

TABLE A-3
RESPONSE RATES FOR PUBLIC SCHOOLS BY STATE

STATE	WEIGHTED RESPONSE RATE	UNWEIGHTED RESPONSE RATE
ALABAMA	95.91	95.58
ALASKA	91.99	91.08
ARIZONA	94.81	96.50
ARKANSAS	97.74	97.56
CALIFORNIA	94.61	93.08
COLORADO	95.87	96.24
CONNECTICUT	93.10	92.02
DELAWARE	93.31	93.06
D.C.	86.26	86.96
FLORIDA	93.94	93.05
GEORGIA	96.65	96.81
HAWAII	98.67	98.91
IDAHO	98.62	98.11
ILLINOIS	98.71	97.17
INDIANA	99.61	99.47
IOWA	96.47	97.79
KANSAS	97.99	97.52
KENTUCKY	98.06	97.80
LOUISIANA	93.87	93.06
MAINE	94.65	96.53
MARYLAND	80.99	80.68
MASSACHUSETTS	91.13	92.98
MICHIGAN	97.11	94.44
MINNESOTA	97.39	96.50
MISSISSIPPI	97.17	96.05
MISSOURI	98.01	97.60
MONTANA	97.81	98.15
NEBRASKA	98.69	97.55
NEVADA	96.14	95.76
NEW HAMPSHIRE	96.33	93.97
NEW JERSEY	88.31	85.79
NEW MEXICO	96.01	95.14
NEW YORK	87.62	88.04
N. CAROLINA	92.63	94.12
NORTH DAKOTA	98.37	97.47
OHIO	97.00	97.03
OKLAHOMA	96.27	95.97
OREGON	95.27	95.40
PENNSYLVANIA	96.06	94.09
RHODE ISLAND	96.49	96.23
S. CAROLINA	96.55	96.15
SOUTH DAKOTA	98.52	97.89
TENNESSEE	98.06	96.45
TEXAS	97.40	96.88
UTAH	98.40	97.75
VERMONT	98.48	98.10
VIRGINIA	92.21	92.31
WASHINGTON	92.58	93.44
WEST VIRGINIA	98.20	97.59
WISCONSIN	94.57	94.08
WYOMING	97.69	97.62

TABLE A-6
RESPONSE RATES FOR PUBLIC SCHOOLS BY SCHOOL ENROLLMENT

ENROLLMENT	WEIGHTED RESPONSE RATE	UNWEIGHTED RESPONSE RATE
1 TO 299 STUDENTS	96.55	96.42
300 TO 599 STUDENTS	95.31	95.71
> 600 STUDENTS	93.92	93.47

TABLE A-7
RESPONSE RATES FOR PRIVATE SCHOOLS BY TYPE OF FRAME

FRAME	WEIGHTED RESPONSE RATE	UNWEIGHTED RESPONSE RATE
LIST FRAME	86.58	86.77
AREA FRAME	74.03	76.92

TABLE A-8
RESPONSE RATES FOR PRIVATE SCHOOLS BY ASSOCIATION MEMBERSHIP LIST

ASSOCIATION	WEIGHTED RESPONSE RATE	UNWEIGHTED RESPONSE RATE
AREA FRAME	74.03	76.92
MILITARY	90.91	90.91
CATHOLIC	90.92	90.24
FRIENDS	90.63	90.62
EPISCOPAL	89.39	84.95
HEBREW DAY	70.76	73.03
SOLOMAN SCHECHTER	85.11	85.11
OTHER JEWISH	70.36	63.14
LUTH.-MISSOURI SYN.	96.07	95.70
LUTH.-WISCONSI.SYN	97.89	97.87
EVAN. LUTH. CH. AM.	95.51	95.51
OTHER LUTHERAN	94.17	93.41
7TH-DAY ADVENTIST	93.91	94.90
CHRISTIAN SCHOOLS I	93.68	91.00
ASSOC. OF CHRISTIAN SCHOOLS INTL	59.03	70.00
EXCEPTIONAL CHILD.	86.49	88.00
MONTESSORI	85.46	85.56
NATIONAL ASSOCIAT. INDEPENDENT SCHOOLS	84.60	84.51
ALL ELSE	81.12	82.71

TABLE A-9
RESPONSE RATES FOR PRIVATE SCHOOLS
BY SCHOOL LEVEL ASSOCIATION LIST FRAME

LEVEL	WEIGHTED RESPONSE RATE	UNWEIGHTED RESPONSE RATE
ELEMENTARY	89.94	89.25
SECONDARY	90.29	90.05
COMBINED	78.22	81.43

TABLE A-10
RESPONSE RATES FOR PRIVATE SCHOOLS BY
SCHOOL LEVEL ASSOCIATION AREA FRAME

LEVEL	WEIGHTED RESPONSE RATE	UNWEIGHTED RESPONSE RATE
ELEMENTARY	77.19	78.31
SECONDARY	85.50	86.11
COMBINED	69.16	73.78

TABLE A-11(1)
HIERARCHICAL RESPONSE RATES

School Survey and District Survey

Response at the School level	Response at the district level	
	Respondent	Nonrespondent
Out of Scope	272	28
Respondent	8397	487
Nonrespondent	380	77

TABLE A-11(2)

Percentile distribution of nonresponse

Response at the District level		Percentage	
	no response	some response	100% response
Respondent	4.06	6.41	89.53
Nonrespondent	10.98	10.34	78.68

IMPLICIT MARKETS FOR TEACHER QUALITY AND SCHOOL ATTRIBUTES

Dale Ballou and Michael Podgursky, University of Massachusetts
Dale Ballou, 904 Thompson Hall, UMASS, Amherst MA 01003

KEY WORDS: Hedonic prices, teacher quality

1. Introduction

A substantial body of research has failed to find a positive relationship between per-pupil expenditures and student achievement (Hanushek, 1986; Pauly, 1991). Since teacher salaries are the largest component of public school expenditures, the absence of positive findings suggests that variation in teacher compensation is not associated with factors affecting productivity. This is surprising. Other things being equal, districts which offer higher salaries ought to enjoy advantages in recruiting personnel. Teachers with superior qualifications should be able to market themselves more successfully to districts offering the best combinations of pay and working environment. Indeed, it has been claimed that such worker-job matches are the way the market rewards qualified teachers for their skills (Murnane, 1983).

We estimate a hedonic wage equation for teachers to determine whether individuals with superior qualifications are duly rewarded by the market. Although such functions have appeared in several earlier studies of teacher labor markets, in all but one teacher quality was peripheral to the central focus (Antos and Rosen, 1975; Chambers, 1985; Levinson, 1988; Smith and Lee, 1990). In general, these studies have found low or non-existent returns to teacher attributes other than those explicitly rewarded in salary schedules. How.ver, this finding may reflect the quality of the data and the specification of the models rather than the behavior of the market. Proxies of questionable value have been used to measure teacher qualifications; in particular, only one study (Chambers, 1985) contained information about the quality of the college awarding the teacher's BA.

2. The Model

The process by which job applicants are matched with vacant positions yields an equilibrium relationship between wages (w), teacher attributes (T), and school characteristics (S): $w = W(S,T)$. "School characteristics" are understood to include community demographics and student characteristics. If tastes on both sides of the market are homogeneous, the market clears in a particularly simple fashion, with the best jobs matched to the most attractive applicants, the next best jobs to those applicants not quite so well qualified, and so on. More complicated outcomes when tastes are heterogeneous are easily conceived. A general discussion of the derivation of such hedonic functions from utility maximizing behavior and the distribution of tastes and resources in the market appears in Rosen (1974).

Coefficients on teacher qualifications represent the market's valuation of that attribute. In this research, we use the 1987-88 Schools and Staffing Survey (SASS) -- a large, national data set with information about teachers' work histories and academic backgrounds -- to examine the relationship between wages and qualifications. SASS makes it possible to include several measures of teacher quality in the model. (i) Level of formal education. There are four levels: less than a BA, BA, MA, and more than an MA. (ii) Experience. Tenure and overall teaching experience both enter the model, as districts may credit the two differently in determining a teacher's step on the salary schedule. These variables are coded so that incremental returns cease after twenty years' service, approximating the schedule cap. Non-linear returns to experence are represented by a quadratic term. (iii) Quality of the college granting the teacher's BA. As a general measure of academic qualifications, we use the selectiveness of the institution awarding the teacher's undergraduate degree. The ranking is from Barron's Profiles of American Colleges and is based on SAT and ACT scores and high school GPA and class rank. (An "unclassified" category contains colleges not listed in Barron's or recognized in the Integrated Postsecondary Education Data System (IPEDS). The data suggest that among public school teachers, this is a random assortment of institutions; however, among private school instructors, small colleges with a religious affiliation predominate.) (iv) Subject-specific qualifications. Three dummy variables deal with subject-specific qualifications. The first identifies teachers whose principal assignment is the subject they are best qualified to teach, whatever it may be. A second identifies qualified math or science instructors, and a third those teachers who majored in education rather than an academic subject. (v) Type of certification. We include two variables related to certification: a dummy variable for teachers holding temporary or emergency certification, and (among private school teachers) a second dummy variable indicating whether teachers are certified in their primary assignment area. (vi) Dedication. We use the self-assigned probability that an individual would still decide to become a teacher were the choice to be made over again.

The remaining variables in the model include teachers' demographic characteristics, a set of student behaviors and school/community characteristics likely to affect job satisfaction, and binary indicators for type of community and region. The last serve as proxies for environmental amenities and as controls for market conditions, including differences in the cost of living.

Imperfect information about job opportunities and other barriers to mobility cause salaries to deviate from those which would be predicted on the basis of observed S and T. These deviations are absorbed into an i.i.d. mean-zero error term. Economic theory suggests no specific functional form for the hedonic locus. We employ a log-linear specification. Because private schools operate in a special environment, offering unique job amenities and attracting some teachers who lack public school certification, we estimate separate hedonic functions for the two sectors.

3. Estimates

Our sample contained 30,381 full-time public school teachers and 3,790 full-time private school teachers. Catholic school teachers who had never been married were dropped from the sample in order to screen members of religious orders whose compensation is not market-based.

Selected results from least-squares estimation are presented in Table 1.[1] Teachers from better colleges are more highly paid, but the difference is modest: a graduate of the most selective undergraduate colleges earns only about 7% more than someone who attended one of the country's least selective institutions. Being qualified in one's primary assignment has a statistically insignificant effect on pay (and among math and science teachers, the sign is actually negative). Dedicated teachers earn about 7% more than teachers who would choose a different career if they had it to do over again; this is also approximately the magnitude of the penalty the market assesses for having majored in education rather than a subject area. On the whole, while most measures of teacher quality are statistically significant and of the expected sign, only the salary schedule variables (experience and level of formal education) affect compensation by more than a few percentage points.

Results for private schools are rather different (column 2). The college attended has a greater influence, with salary varying as much as 28% depending on where the teacher obtained a BA. Qualified mathematics and science teachers are also rewarded by the market, earning 5.7% more than other instructors. Private schools pay certified teachers slightly more than uncertified teachers, but a teacher with temporary or emergency certification is penalized. No premium is attached to being a committed teacher, rather the reverse: the private sector pays higher wages to teachers who think they might have done something else with their lives. This may reflect the fact that more dedicated teachers also have lower reservation wages, and are consequently over-represented in schools paying low salaries.

The coefficients in column 1 indicate that the labor market for public school teachers offers only a modest return to teachers' academic qualifications. The relative return in the private sector is substantially larger. We consider three explanations for these results: (1) misspecification of the hedonic function due to omitting the interaction of teacher attributes and school characteristics; (2) barriers to mobility; (3) low variability in salary levels within the public sector, implying low returns to mobility.

(i) Omitted interactions

The model we have estimated constrains the coefficients on school characteristics to be the same for all teachers. Yet teachers with good academic backgrounds may have a stronger than average preference for working with capable and motivated students. Schools which can offer this bundle of working conditions need not pay the same price for academic qualifications as less favored schools. Omission of such interactions from the model biases the estimates in column 1.

When the model was re-estimated with such interactions, most of the new terms were individually insignificant, though a Wald test of their joint significance just exceeded the critical F value at 5%. We report the effect on the measures of college selectivity in the third column of Table 1. The effect of the additional interaction terms has been to shrink returns to academic qualifications in the public sector toward zero. Thus the data provide no support for the misspecification hypothesis.

(ii) Barriers to mobility

Without the opportunity to choose from multiple job offers, better qualified teachers will be unable to obtain a return to their skills in the market. One possible explanation for the findings in column 1 is simply that teachers do not have enough mobility. To test this hypothesis, we run two additional regressions, one on teachers hired within five years of the survey (between 1983 and 1987) and one on teachers living in urban or suburban areas. If mobility is responsible for our earlier findings, we should find evidence to that effect upon comparing these regressions to those in column 1, given two additional, supporting hypotheses: that mobility is greater among new teachers, smaller among residents of rural areas and small towns.

The results of these regressions are given in columns 4 and 5. There is some loss of precision due to the smaller sample sizes. However, there is no evidence from these specifications that lack of mobility explains the low returns to teacher qualifications.

(iii) Low variation in public sector salaries

Since public school teachers are rewarded for better academic qualifications via job matches, the potential return to this attribute is limited by the overall variation in salaries among districts. This variation may be too small to support large returns to teacher qualifications. To examine this hypothesis we use a measure of salary variation independent of the qualifications of particular job-holders, the annual wage offered beginning teachers in each district (teachers with a bachelor's degree and no prior experience).

To facilitate comparison with our earlier results, we regress the log of starting salary on the same school, community and locational variables that entered the original hedonic equation, omitting variables specific to the job-holder. Ranking jobs by the value of the residual from this regression, we calculate a counterfactual estimate of the maximum possible return to academic qualifications by assuming that the top quartile of jobs are entirely filled by graduates of the colleges rated "Most Competitive," "Highly Competitive," and "Very Competitive". Graduates of less selective colleges are assumed to take the remaining jobs. The mean difference between the two sets of residuals is an estimate of the maximum possible gap between salaries as a function of qualifications (after controlling for location and working conditions). To the extent there is less than perfect sorting on the basis of college selectivity, the actual differences falls below this maximum.

The results for public schools are reported in column 6. The maximum possible difference between salaries in the two groups is found to be 17.9%. The actual difference (a weighted average of the coefficients in column 1) is 2.5%. Clearly, the return to academic qualifications is far below the feasible limit. Interestingly, the same is true of private schools (column 7). The maximum possible gap between the top and bottom three quartiles is 33.1%, the actual gap only 5%. Indeed, the estimated return to academic qualifications, as a fraction of the possible return, is the same in both sectors. Thus, while neither sector comes close to exhausting its potential for rewarding academic qualifications, it may be that no further explanation is required for the higher returns in the private sector.

4. Conclusion

This paper has estimated the return to personal attributes in the market for public school teachers. We find the market provides only a modest return to characteristics not officially recognized in salary schedules. Although evidence from the private sector suggests that a strong academic background adds substantially to a teacher's value, the public school market provides nothing like a commensurate return to these qualifications: a Harvard graduate who decides to teach in the public schools can anticipate earning only 7% more than a graduate of a non-selective and relatively undemanding four-year college. By contrast, the differential in the private sector is two to three times as large.

We have examined three possible explanations for these findings. The evidence suggests that our results are not the consequence of misspecifying the demand for better working conditions. Neither are they explained by insufficient mobility. There is sufficient variation within public school salaries to permit a larger return to academic qualifications, if schools were to weigh this factor more heavily in teacher recruitment.

The implications of this research are disturbing. First, rigid salary schedules that reward teachers solely on the basis of highest degree and experience are difficult to circumvent via the market. This may not be all bad, if it means that schools in poor communities attract and retain better teachers than they otherwise could. The negative side, of course, is that capable persons are discouraged from entering the profession.

If well-qualified teachers are unable to obtain satisfactory rewards via inter-district mobility, and if it is impossible to replace salary schedules with pay structures that give more weight to individual merit, then it would appear that the only way to draw more capable persons into teaching is to raise salaries across the board. The success of this policy depends on the criteria used by school officials to select prospective teachers from an enlarged applicant pool. Our results are not encouraging on this point. Indeed, if the undervaluation of academic qualifications is explained neither by lack of mobility nor by the absence of sufficient salary variation in the marketplace, it is reasonable to ask whether those responsible for hiring decisions give proper weight to these factors. The evidence in this paper suggests they do not. To that extent, prospects for renewing the profession by raising teacher salaries are not as good as they could be.

[1] Regression results for all variables in the model, along with means and standard deviations for all variables, are available from the authors.

References

Antos, Joseph R. and Sherwin Rosen. 1975. "Discrimination in the Market for Public School Teachers." Journal of Econometrics, May, 3(2), 123-150.

Barron's Profiles of American Colleges. 1991. 18th Ed. New York: Barron's Educational Services, Inc.

Chambers, Jay G. 1985. "Patterns of Compensation of Public and Private School Teachers." Economics of Education Review, 4(4), 291-310.

Hanushek, Eric. 1986. "The Economics of Schooling: Production and Efficiency in Public Schools." Journal of Economic Literature 24, 1141-1177.

Levinson, Arik M. 1988. "Reexamining Teacher Preferences and Compensating Wages." Economics of Education Review. 7(3), 357-364.

Murnane, Richard J. 1983. "The Uncertain Consequences of Tuition Tax Credits: An Analysis of Student Achievement and Economic Incentives," in T. James and H.M. Levin, eds., Public Dollars for Private Schools: the Case for Tuition Tax Credits. Philadelphia: Temple University.

Pauly, Edward. 1991. The Classroom Crucible. New York: Basic Books.

Rosen, Sherwin. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." Journal of Political Economy, January/February 82, 34-55.

Smith, Valerie E. and Julia B. Smith. 1990. "Gender Equity in Teachers' Salaries: A Multilevel Approach." Educational Evaluation and Policy Analysis, Spring, 12(1), 57-81.

Table 1:
Returns to Teacher Quality:
Selected Regressors

Selected Independent Variables	Public (1)	Private (2)	Public w/ T & S Interact (3)	Public Recent Hires (4)	Public Urban & Subrb (5)	Public Max. Return (6)	Private Max. Return (7)
Selectivity of College:							
Most Competitive	.033*** (.017)	.246*** (.040)	-.011 (.042)	.046 (.036)	.016 (.022)		
Highly Competitive	.032*** (.010)	.174*** (.031)	-.008 (.039)	.050** (.020)	.025 (.013)	.134	.248
Very Competitive	.017** (.008)	.121*** (.024)	-.021 (.038)	.012 (.016)	.007 (.011)		
Competitive	.005 (.007)	.129*** (.014)	.006 (.007)	.006 (.015)	-.003 (.010)		
Less Competitive	-.021*** (.008)	.144*** (.025)	-.021*** (.008)	-.011 (.016)	-.025** (.011)	-.045	-.083
Non-Competitive	-.043*** (.008)	.098*** (.030)	-.043*** (.009)	-.020 (.017)	-.053*** (.013)		
Unclassified	—	—	—	—	—		
Qualified in Principal Teaching Assignment (PTA)	.006 (.004)	.021 (.014)	.006 (.004)	.010 (.007)	-.005 (.006)	—	—
Qualified in PTA x Math or Science PTA	-.008 (.005)	.057** (.023)	-.008 (.005)	-.017 (.011)	-.013* (.007)	—	—
BA in Education	-.021*** (.004)	.023 (.014)	-.021*** (.004)	-.024*** (.007)	-.018*** (.005)	—	—
Temporary Certification	.001 (.008)	-.050* (.026)	.001 (.008)	.014 (.011)	-.002 (.012)	—	—
Certified in PTA	—	.023 (.015)	—	—	—	—	—
Probability Would Choose Teaching Again	.018*** (.005)	-.050** (.022)	.018*** (.005)	.026** (.010)	.018** (.007)	—	—
Sample Size	30,381	3,791	30,381	8,749	13,143	30,381	3,791

WHO DECIDES? PRINCIPALS' AND TEACHERS' VIEWS ON DECISION-MAKING

Judith I. Anderson, U.S. Department of Education
OERI/Office of Research, 555 New Jersey Ave. NW Room 610, Washington, DC 20208

KEY WORDS: Education policy; Use of survey data

School-based management (SBM) is often in the news. In practice, SBM varies from school to school, but generally it gives increased budgeting, curriculum, and staffing responsibilities to principals and teachers or to parents and community members in conjunction with school staff. The influence each group has varies from school to school, but the goal remains the same: to improve children's schooling.

Implicit in this call for greater school-level influence is the belief that those closest to the children—the principals, teachers, parents, and community members—know best what is needed to improve their schools. The purpose of this report is to examine where decision-making now occurs.

The 1987–88 Schools and Staffing Survey (SASS), a national survey conducted by the National Center for Education Statistics, asked principals and teachers a variety of questions related to SBM. For principals, SASS provides information on where decisions are made for three SBM areas: establishing curriculum, hiring new full-time teachers, and setting discipline policy. All data in this report are for public school principals and teachers.

Principals were asked how much they thought the school district, principal, and teachers actually influenced decisions on establishing curriculum, hiring new full-time teachers, and setting discipline policy. The principals' answers are categorized according to which of the three groups—district, principal, or teachers—they believed most influenced these decisions.

Teachers were asked how much actual influence they thought they had over school policy decisions and how much control they had in their classrooms over selected areas of planning and teaching.

The data provided in this report are principals' and teachers' accounts of conditions and are not based on independent observations of actual decision-making. Also, many of the differences observed between different community types may be due to district and school sizes, both of which tend to be larger in large cities.

Who Decides? The Principals' View

Public school principals painted a picture in which they and the school district considerably influence curriculum, hiring teachers, and setting discipline policy. They believed

- School district personnel were most likely to establish curriculum;
- Principals had the greatest responsibility for hiring new full-time teachers; and
- The school district or the principal was most likely to set discipline policy.

Table 1: Principals' views on who has the most influence over selected school decisions

Staff most responsible	Type of school decision: Establishing curriculum	Hiring new full-time teachers	Setting discipline policy
	(Percentage of principals responding)		
Total	100	100	100
School District	33	28	24
Teachers	12	1	1
Principals	11	49	23
Principals & teachers	19	2	18
District & teachers	3	0	0
District & principal	6	18	15
All three	15	2	17

NOTE: Figures may not total to 100 because of rounding.
Source: 1987-88 SASS.

Teachers, on the other hand, were not seen by the principals as having primary reponsibility over any of these areas. Only 12 percent of principals thought teachers were primarily responsible for establishing curriculum, and only 1 percent said teachers had primary responsibility for hiring new teachers or setting discipline policy.

Establishing curriculum. Establishing curriculum was most often viewed as a school district reponsibility:

- 33 percent of principals said that the school district was most likely to be responsible for establishing curriculum;
- 19 percent believed that teachers and principals were equally responsible for this area; and
- 15 percent said that all three—district, principal, and teacher—were equally responsible.

Hiring new full-time teachers. School principals were most likely to report that they have primary responsibility for hiring new full-time teachers:

- 49 percent said school principals were most likely to have primary responsibility for this area;
- 28 percent said that school district personnel were primarily responsible; and
- 18 percent said that the principals and the school district were equally responsible.

Teachers were seen as having little primary responsibility for hiring new colleagues.

Setting discipline policy. The school principals were equally likely to report that they and the school district personnel were most responsible for decisions on discipline policy:

- 24 percent reported that the school district was most responsible, and 23 percent reported that the school principal was most responsible.

Again, teachers were not seen as having primary responsibility for setting policy, but as working with the principals (18 percent) or with the principals and district (17 percent).

Who Decides? The Teachers' View

Teachers agree that they do not have much control over establishing curriculum and setting discipline policy, and, in addition, do not believe that they have much influence over determining the content of in-service programs and setting the policy on grouping students in classes by ability. No more than 35 percent of teachers believed that they had a great deal of control over decisions in these areas:

- Only 35 percent believed they had considerable influence over determining discipline policy or establishing curriculum;
- 31 percent believed they had much influence over determining the content of in-service programs; and
- 28 percent believed they were influential in setting the policy on grouping students in class-es by ability.

Control over classroom activities, however, is a different matter. Most teachers believed they had considerable influence over classroom decisions.

- More than half believed they had considerable control over selecting textbooks and other instructional materials (54 percent) and selecting the content, topics, and skills to be taught (59 percent);
- More than two-thirds (69 percent) believed they had a great deal of control over disciplining students; and
- Most believed they were firmly in control of selecting teaching techniques (85 percent) and determining the amount of homework to be assigned (87 percent).

Community Type

The type of community in which their schools were located influenced the control school principals believed that they and their teachers had over school decisions. Big city schools are more likely to be part of large school districts that exercise central control over decisions than are schools in small towns or rural areas. Principals in big cities thought school districts exercised greater

Table 2.—Percentage of teachers believing they had considerable influence over selected areas of school policy and classroom planning and teaching

School Policy	
(Percentage of teachers answering "5" or "6" on a scale that ran from 1 (none) to 6 (a great deal))	
Determining discipline policy	35
Determining the content of in-service programs	31
Setting policy on grouping students in classes by ability	28
Establishing curriculum	35
Classroom Activities	
(Percentage of teachers answering "5" or "6" on a scale that ran from 1 (none) to 6 (complete control))	
Selecting textbooks and other instructional materials	54
Selecting content, topics, and skills to be taught	59
Selecting teaching techniques	85
Disciplining students	69
Determining the amount of homework to be assigned	87

Source: 1987-88 SASS.

control than did their colleagues in suburbs or small towns.

School principals viewed school district staff in very large cities as being firmly in control:

- 62 percent of principals in very large cities said districts had the most influence over establishing curriculum;
- 61 percent said districts had the most influence on hiring new full-time teachers; and
- 40 percent said districts had the most influence on setting discipline policy.

Only a small minority of school principals in very large cities believed that they alone (4 percent), their teachers (5 percent), or they and their teachers together (9 percent) were primarily responsible for making decisions on curriculum. School principals reported somewhat more control over hiring teachers and setting discipline policy, but the district still was the primary influence.

The situation in small cities and towns—where schools are less likely to be part of a school district with a large centralized bureaucracy—is much different:

- Only 27 percent of principals in small cities and towns said districts had the most influence over establishing curriculum;
- 22 percent said districts had the most influence on hiring new full-time teachers; and
- 22 percent said the districts had the most influence on setting discipline policy.

The school principals in small towns and cities are much more likely to report that they or their teachers are responsible for decisions about curriculum, hiring, and discipline:

- Nearly half report they and their teachers have the most influence over establishing curriculum (13 percent say they have, while 14 percent report their teachers have, and 21 percent believe they and their teachers are equally responsible);

- Over half (53 percent) report they have the most influence over hiring new full-time teachers; and

- Nearly half report they and their teachers have the most influence over setting discipline policy (26 percent say they have, 20 percent believe they and their teachers are equally responsible, and a tiny minority—2 percent—report teachers are primarily responsible).

In general, the larger the city in which the school is located, the less the amount of control reported by school principals.

A similar pattern was found for teachers. Teachers in very large cities were less likely than their peers in smaller communities to control decisions on school policy and on classroom activities. While a minority of teachers in any type of community believed that they had considerable control over setting school policy, far fewer teachers in very large cities believed that they were in control:

Table 3.—Principals' views on who has the most influence over establishing curriculum, hiring new full-time teachers, and setting discipline policy, by type of decision and community

	School District	Teachers	Principal	Teachers & Principal	District & Teachers	District & Principal	All Three	Total
				(Percent)				
Establishing Curriculum								
Total	33	12	11	19	3	6	15	100
Very Large City	62	5	4	9	2	7	12	100
Large City	57	6	4	10	4	6	12	100
Medium City	45	11	7	11	4	5	17	100
Suburb of Very Large City	39	14	10	16	4	5	13	100
Suburb of Large City	37	12	7	18	4	5	17	100
Suburb of Medium City	30	11	10	21	3	6	19	100
Small City or Town	27	14	13	21	3	7	16	100
Rural/Farming	22	13	16	25	2	7	15	100
Hiring New Full-Time Teachers								
Total	28	1	49	2	0	18	2	100
Very Large City	61	—	20	1	—	10	7	100
Large City	45	1	32	2	1	16	3	100
Medium City	33	1	46	3	—	15	2	100
Suburb of Very Large City	24	—	56	2	—	16	1	100
Suburb of Large City	21	0	57	3	—	18	2	100
Suburb of Medium City	23	1	54	2	—	17	3	100
Small City or Town	22	1	53	2	—	20	2	100
Rural/Farming	24	0	52	2	0	20	2	100
Setting discipline policy								
Total	25	1	23	18	0	15	17	100
Very Large City	40	2	15	13	1	14	15	100
Large City	42	1	13	15	1	13	15	100
Medium City	29	1	18	19	1	14	18	100
Suburb of Very Large City	26	2	20	19	—	17	16	100
Suburb of Large City	22	1	26	19	—	14	18	100
Suburb of Medium City	24	—	23	20	—	15	17	100
Small City or Town	22	2	26	20	0	14	17	100
Rural/Farming	18	1	27	18	0	18	17	100

NOTES: (1) Figure may not total 100 because of rounding. (2) The total contains a small number of schools on military bases or Indian reservations. There were too few of these schools to include them as separate categories. (3) Community types were identified by the respondents.Very large cities are those with over 500,000 people, large cities have 100,000 to 500,000 people, and small cities and towns have fewer than 50,000 people and in addition are not suburbs of larger cities. (4) — indicates that there were too few cases for analysis.
Source: 1987-88 SASS.

- Only 27 percent of the teachers in the very large cities believed they had considerable influence over determining discipline policy; and

- Only 23 percent believed they had much influence over determining the content of in-service programs, setting the policy on grouping students in classes by ability, or establishing curriculum.

In contrast,

- 37 percent in rural and farming communities believed they had considerable influence over determining discipline policy;

- 32 percent believed they had considerable influence over determining the content of in-service programs;

- 28 percent felt they helped set the policy on grouping students in classes by ability; and

- 41 percent believed they were influential in establishing curriculum.

Teachers in big cities were also less likely to feel they had control over classroom practices than were their peers in rural areas:

- Only 41 percent in very large cities believed they had considerable control over selecting textbooks and other instructional materials, compared to 65 percent of the teachers in rural areas;

- Only 47 percent of the big city teachers reported they controlled selecting content, topics, and skills to be taught, compared to 67 percent of their rural counterparts; and

Table 4.—Percentage of teachers who said that they had considerable influence over selected areas of school policy and classroom planning and teaching, by community type and policy area

Policy area	Community type: Rural or farming	Small city or town	Suburb	Medium-sized city	Large city	Very large city
School Policy (Percentage of teachers answering "5" or "6" on a scale of from 1 (none) to 6 (a great deal))						
Determining discipline policy	37	37	34	35	33	27
Determining the content of in-service programs	32	32	32	30	29	23
Setting policy on grouping students in classes by ability	28	28	30	30	27	23
Establishing curriculum	41	38	36	31	23	23
Classroom Activities Percentage of teachers answering "5" or "6" on a scale that ran from 1 (none) to 6 (complete control))						
Selecting textbooks and other instructional materials	65	58	53	44	38	41
Selecting content, topics, and skills to be taught	67	61	58	52	47	47
Selecting teaching techniques	88	86	86	84	81	78
Disciplining students	74	71	70	68	64	60
Determining the amount of homework to be assigned	90	87	85	84	85	84

NOTES: (1) See Table 3 for an explanation of community types. (2) All three categories of suburbs are combined in the table.
Source: 1987-88 SASS.

- Only 60 percent of the teachers in the largest cities—compared to 74 percent of the rural teachers—believed they had considerable control over disciplining the students in their classes.

In general, the larger the community, the less control teachers report over their classroom practices.

Conclusions

These analyses show that much decision-making takes place at the school district level, although school principals have a considerable influence over hiring teachers. Nevertheless, studies of school-based management need to take into account the types of communities in which schools are located: school personnel in smaller communties may already have a great deal of control over curriculum, hiring, and discipline.

NOTES:

Sample sizes. These analyses were based on data from 8,580 public school principals and 40,593 teachers.

Standard errors. Standard errors were calculated using a balanced repeated replicates procedure. Tables of standard errors are available on request. For Table 1, standard errors ranged from 0.07 to 0.48; for Table 2, from 0.2 to 0.4; for Table 3, from 0.09 to 2.65; and for Table 4, from 0.4 to 1.4.

Explanation of coding for the principals' analyses. The principals were asked to indicate on a scale of from 1 (low) to 6 (high) how much actual control the district, teachers, and principals had over each of the areas. Their responses were re-coded according to which of the three (districts, teachers, or principals) they reported had the most control over the area. For example, a principal who reported a "3" for "district", a "5" for "teachers" and a "5" for "principals" would have been recoded into the "principal and teachers" category.

References

U.S. Department of Education, National Center for Education Statistics, *Schools and Staffing in the United States: A Statistical Profile, 1987–88* (Washington, D.C.: U.S. Department of Education, July 1992).

Copies of the survey questionnaires are available from the Special Surveys and Analysis Branch of the National Center for Education Statistics, U.S. Department of Education, 555 New Jersey Avenue NW, Washington, DC 20208.

The primary purpose of the presentation was to present an alternative format for the presentation of educational survey data, the *Research Report. Research Reports* are a new series begun by the Office of Research (OR) of the U.S. Department of Education. The primary purpose of the *Research Reports* is to increase the use of survey data by presenting topical information in a short, non-technical format. This paper closely approximates the *Research Report* format, although it does not contain the graphics. The principal concern we have with developing reports for non-technical audiences is how much statistical information should be presented. Should we provide: Background information on the topic under discussion? Detailed information about the survey—how it was administered, the sample design, response rates, etc.? Details about how the data were analyzed? Standard errors, sample sizes, and other statistics, or just the estimates of interest to a "typical" non-technical reader? Comments and suggestion would be appreciated, and should be addressed to the author at one of the following addresses:

MAIL: U.S. Department of Education, OERI/Office of Research, 555 New Jersey Avenue NW, Room 610, Washington, DC 20208

INTERNET: janderso@inet.ed.gov

DETERMINANTS OF PUPIL/TEACHER RATIOS AT SCHOOL SITES: EVIDENCE FROM THE SCHOOLS AND STAFFING SURVEY

Lawrence O. Picus and Minaz Bhimani, University of Southern California
EDPA WPH-901, USC, Los Angeles, CA 90089-0031

KEY WORDS: Education, SASS

One of the highest priority policy goals of the education community has consistently been the reduction of class size as a way to improve student achievement. Even though the research evidence to support the positive effect of class size reductions on student outcomes suggests certain limitations on the impact of class size reductions on performance, there is almost universal agreement that reductions in class size are important if student learning is to improve in our nation's schools (see Odden, 1990; Slavin, 1989; and Smith & Glass, 1980). What the research seems to indicate is that substantial reductions in class size are needed to have an impact on student performance. The work of Slavin (1989), Smith & Glass (1980) and Odden (1990) suggests that to be truly effective, a class size of approximately 15 students per teacher or less is needed.

A cursory review of the most recent edition of the *Digest of Education Statistics* (NCES, 1992) shows that the average pupil/teacher ratio for K-12 public schools in the United States was only 17.2:1 in 1991, very close to the 15 students per teacher emphasized in most research. Moreover, the data provided in the *Digest* suggest that this ratio has declined consistently since 1955 when is stood at 26.9 pupils per teacher (NCES, 1992, p. 73). In fact, except for an increase of 0.1 pupils per teacher between 1961 and 1962, the average pupil/teacher ratio across the United States has declined in every year since 1955. There is considerable variation in pupil/teacher ratios by state. In fact, the *Digest* shows that the pupil/teacher ratio in the fall of 1990 ranged from a low of 13.2 pupils per teacher in Vermont, to a high of 25 in Utah (NCES, 1992, p. 75).

The typical policy maker views the pupil/teacher ratio as a proxy for class size. Despite what would therefore appear to be small class sizes, teachers across the nation complain their classes are much too large. They argue that if they are to succeed in making dramatic improvements in student achievement, class sizes must be reduced. They often complain of classes with 30 or more students, and of the impossibility of meeting the needs of individual students under such conditions. The explanation for this difference between what teachers say and what the national averages seem to indicate is often that the national averages include special education classes which generally have many fewer students, and the fact that there are a number of itinerant teachers in many districts who provide special pull-out services for children through a variety of programs including Chapter 1, gifted and talented education, or for such programs as art and music instruction. Also, these national averages often include certificated personnel who have non-teaching assignments such as counselors or curriculum development specialists.

Why is there such a substantial difference between self-reported class size and nationally reported pupil/teacher ratios? More importantly, is this difference related to specific school or school district characteristics? Until recently there has been no readily available data on the size of individual teacher's classes, nor a comprehensive data base to analyze the characteristics of schools and districts that have small and/or large pupil/teacher ratios. The recently released Schools and Staffing Survey (SASS) provides an opportunity to analyze differences between district and even school reported pupil/teacher ratios, and teacher self reported class size. The survey, which was sent to over 5,000 school districts, more than 9,000 schools and over 56,000 individual teachers in those schools, provides a wealth of information on individual teacher characteristics as well as information on the schools and school districts in which they teach.

This study is part of the Finance Center for Policy Research in Education's (CPRE) in-depth study of resource allocation in elementary and secondary schools. Known as the Integrated, Multi-level Resource Allocation Study, the Center is conducting a multi-year multi-faceted study of "what dollars buy" in education. Specifically, Center researchers are conducting analyses of spending and resource allocation patterns at the national, state, district and school levels. This paper uses the SASS data on individual school teachers to consider the relationship between self reported class size and the pupil/teacher ratios generated from our earlier research at the school and district levels.

SUMMARY OF EARLIER STUDIES

Our earlier work focused on state, district and school level variables from the SASS and other data bases. At the state level we found that the single largest expenditure item for school districts is teacher salaries. On average, teacher salaries account for 45 to 50 percent of a school district's budget. Teacher compensation (salaries and benefits) generally amount to between 55 and 60 percent of expenditures (NEA, 1992). Average teacher salary in 1991-92 ranged from a low of $23,300 in South Dakota to a high of $47,300 in Connecticut (Barro, 1992). We also found that on average the pupil/teacher ratio decreases by about six percent for each ten percent increase in per pupil expenditures (Barro, 1992).

In our district level analysis, Picus, (1993a) found that there is substantially less equity in educational expenditures per pupil across school districts than is apparent when analyzing state level fiscal data bases (See Barro, 1992 for a summary of state level findings). District per pupil expenditures for education ranged from under $1,000 per pupil to over $50,000 in 1987-88, the most recent year for which SASS data are currently available (Picus, 1993a). The coefficient of variation for per pupil expenditures was 0.476, considerably larger than the coefficient of variation found in any individual state. However, most districts spent approximately 60% of their resources on direct instruction (as defined by the Census Bureau). There was considerably less variation in the share of expenditures devoted to instruction, with the coefficient of variation only 0.106.

We also analyzed the pupil/teacher ratio, concluding that as expenditures increase, the pupil/teacher ratio declines. The pupil/teacher ratio in districts spending over $5,500 per pupil averaged 12.5 while in districts spending less than $2,000 per pupil the figure was 19.0. Moreover, secondary schools had consistently lower pupil/teacher ratios than did elementary schools. Rural schools, which constituted over 40% of our sample, had the lowest pupil/teacher ratio. Similarly, suburbs had a lower average pupil/teacher ratio than the cities they surround. Interestingly, as the central city gets larger, the pupil/teacher ratio decreases both in the city schools and in the surrounding suburbs, although the pupil/teacher ratio is always lower in the suburbs than in the cities they surround.

Finally, in analyzing teacher salaries, it appeared that district location (i.e. rural, suburban, or urban) has a strong influence on teacher salaries, with higher salaries found in urban and suburban areas surrounding very large cities. Moreover, not surprisingly, districts that have higher per pupil expenditures also offer their teachers higher salaries. Finally, districts with lower pupil/teacher ratios tend to pay their teachers less, implying that many districts make a direct trade-off between class size and teacher salary.

Our analysis of teacher characteristics at the school level found that there is little difference in the mix of teacher experience across schools regardless of how those schools are categorized (i.e. by location, district spending level, pupil/teacher ratio, etc.) (Picus, 1993b). This analysis confirmed our finding that class sizes tend to be the smallest in the Northeast and largest in the West, but we found few differences by community type in the school level sample. The pattern of lower pupil/teacher ratios in secondary schools observed in our district level analyses was also found at the school level.

As with our district findings, it seems that schools with the lowest percent of students qualifying for free and reduced price lunches and the schools with the highest percent of such children seemed to have the lowest pupil/teacher ratio, with mixed results for those in-between. One new finding when data were analyzed at the school site level was that schools in districts with more money to spend tend to hire teachers with more experience and training.

Perhaps the strongest finding from our earlier studies is the consistency across schools and school districts. This pattern first emerged when we looked at expenditures for instruction as a percentage of total expenditures. This figure averaged approximately 60 percent, and varied little as other district characteristics, particularly the spending level itself, changed. This finding implies that regardless of how much money is available to school districts, they tend to allocate the same portion of their total to instruction. It is important to note that this does not imply that there are no significant differences in the way those dollars are spent. After all, 60 percent of $5,000 represents a great deal more money per pupil for instruction than does 60 percent of $2,500. Therefore, our findings indicate that districts spend available resources in similar and predictable patterns, but it is clear that districts that have more money continue to spend vastly more on direct instruction for their students. This is most clearly shown by the decrease in pupil/teacher ratios observed as spending levels increased, and the similar but weaker pattern of more experienced teachers working in higher spending districts.

VARIATION IN PUPIL/TEACHER RATIOS BY SCHOOL AND DISTRICT CHARACTERISTICS

Our teacher sample included 30,362 teachers who responded to the SASS teacher questionnaire. Eliminated from the total sample of over 56,000 teachers were those who indicated that they taught less than full-time, and those for whom a school and district match could not be made. Because it is impossible to ascertain how this reduction in the sample affects the representatives of the sample, we have elected to use the data as one large national sample rather than attempt to conduct analyses at the individual state level. The problems of assuming a representative sample on a state by state basis are considerable given this fall-off in the sample.

Our sample of 30,362 was further divided into two sub-samples. One sub-sample was established for the 12,177 teachers who indicated that they taught in a self-contained setting, while the second sub-sample included the 18,185 teachers who indicated that they were in schools that used departmentalized instruction. The self-contained setting is like that found in most elementary schools across the country while the departmentalized setting is most often found in secondary schools. At the middle or junior high school level, both models can be observed, but generally teachers reported using departmentalized instruction in the 6th, 7th and 8th grades.

District and School Pupil/Teacher Ratios vs. Individually Reported Class Size Estimates

Perhaps the most important finding from our analysis of the SASS teacher questionnaire data is the confirmation of teachers' arguments that they have much larger classes than most national and state specific pupil/teacher ratio data indicate. Table 1 provides a summary of our district, school and teacher level findings as to the pupil/teacher ratios, or self reported class sizes for various levels and types of schools. Table 1 shows the difference between aggregate data from the district and school levels, and self reported teacher data. At the district and school level, the pupil/teacher ratio for elementary grades (K-6) is between 17.68 and 18.77 pupils per teacher. However, the mean teacher reported class size for self-contained classrooms is 24.21, some 29 to 36 percent larger than estimates based on district and school data.

Similarly, the average secondary school pupil/teacher ratio as reported on the district level SASS questionnaires was 14.41. At the school level, the mean pupil/teacher ratio was 16.38 for intermediate schools and 16.55 for secondary schools. On the other hand, the self reported average class size for departmentalized classes amounted to 22.65.

Table 1
Weighted Summary of Pupil/Teacher Ratio Statistics
At the District, School Level, and Teacher Reported Class Size
1987-88 SASS

	District Pupil/Teacher Ratio			School Pupil/Teacher Ratio			Teacher Self Reported	
Statistic	Grades K-12	Grades K-6	Grades 7-12	Elementary	Intermediate	Secondary	Self-Contained	Departmentalized
Mean	15.80	17.68	14.41	18.77	16.38	16.55	24.21	23.66
Standard Deviation	4.45	8.24	5.39	4.65	3.89	5.60	11.92	11.46
Maximum	40.50	117.02	59.50	47.50	34.37	60.36	120.00	120.00
Minimum	2.00	2.00	2.00	1.55	1.92	2.4	0.5	2.00
Range	38.50	115.02	57.50	45.95	32.45	57.96	119.5	118.00
Median	16.40	17.84	14.97	18.46	16.18	16.52	23.00	22.00
Inter-quartile range	4.70	5.20	5.69	5.65	4.98	5.50	8.00	9.20
Range (99-1)	19.50	22.18	24.57	21.85	19.35	28.01	95.00	68.17
Range (95-5)	13.00	14.60	15.66	14.54	12.13	15.50	27.00	27.14
Range (90-10)	9.49	11.07	11.55	10.93	9.23	11.54	19.00	18.92
Coeff. of Variation	0.28	0.31	0.37	0.25	0.24	0.34	0.49	0.48
Number of Obs.	4,386	4,239	4,270	3,414	1,204	1,876	12,177	18,176

The difference between self-reported class size and the pupil/teacher ratios computed through district and school averages, while disconcerting, was not unexpected given that teachers have been making similar claims for a number of years. Because our earlier analyses found a number of significant factors that impact the pupil teacher ratio, it seemed fruitful to determine if those same factors have any impact on the self-reported class size. To conduct this analysis, we have shifted our focus moderately. Rather than describe differences in pupil/teacher ratios and self reported class size, we have converted these data into estimates of the number of teachers per 1,000 students.[1] The advantage of doing so is that it is possible to get a measure of how many teachers there are, on average, with assignments outside of the regular classroom.

For example, the district level pupil/teacher ratio for grades K-6 reported in Table 1 is 17.68. This is the equivalent of 56.56 teachers per 1,000 students. Similarly, the self-reported class size for self-contained classrooms was 24.05 pupils per teacher. This translates to 41.58 teachers per 1,000 students, implying that there are almost 15 teachers per 1,000 students at the elementary level who have assignments outside of the regular classroom. This would include special education teachers, who typically have smaller classes, itinerant teachers, and teachers of special subjects such as music and art. Because the SASS collected enough data to allow us to distinguish between individuals with teaching assignments and those who have non-teaching assignments such as counseling or curriculum development, the 15 teachers per 1,000 students at the elementary level are all assigned to some

[1]This is calculated by inverting the pupil/teacher ratio and multiplying by 1,000.

form of instruction, they just do not have a full-time regular education class responsibility. The balance of this paper describes how the number of teachers without regular classroom assignments varies by district and school characteristics. Except as noted, all of the differences reported below are statistically significant at the .01 level.

Variation by Enrollment

Our earlier analysis at the district level found that the pupil/teacher ratio increased along with district size. Our modeling indicated that these effects were relatively small, amounting to approximately 0.2 pupils per class if a district's enrollment increased by 1,000 students (Picus, 1993a). While this seems to be a very small effect, it is statistically significant.

Our analysis at the school level showed a similar pattern, with the average pupil/teacher ratio increasing as the enrollment in a school increased. Our modeling showed a much stronger effect, with an increased school enrollment of 100 students leading to an estimated increase in class size of approximately one half a student.

Table 2 shows how the average number of teachers per 1,000 students who have assignments outside the regular classroom varies by district and school enrollment. The table shows that as the enrollment of a district or school increases, there tend to be fewer of these teaching positions per 1,000 pupils. In fact, when district enrollment exceeds 25,000, the difference is no longer statistically significant for the self contained classes. This finding would seem to indicate that there are economies of scale to be found in the delivery of the services provided by these teachers given the lower number of teachers per 1,000 students with such assignments in districts and/or schools with higher enrollments. The lack of statistical significance when enrollments are very high indicates that there may be very few of these individuals employed by the district per 1,000 students.

Some of the numbers in the table should be viewed with caution. For example, the first row of the table implies that there are over 24 non-classroom teaching positions on average in school districts with less than 500 students. Since the district size is considerably less than 1,000, this means that on average, a district with 500 students would have approximately 12 such individuals on staff, still a rather large number.

Variation by District Expenditure Per Pupil

Picus (1993a) found that district level pupil/teacher ratios declined as expenditures per pupil and expenditures per pupil for instruction increased. However, as the percent of expenditures devoted to instruction increased, a similar pattern did not emerge. Since expenditure data are not available at the school level, Picus (1993b) compared school level pupil/teacher ratios with district per pupil expenditures. He found that at the elementary, intermediate and secondary school level, there is a trend toward smaller classes as expenditures increase.

Table 2
Number of Teachers Per 1000 Students With Assignments Outside of the Regular Classroom by District and School Enrollment

Enrollment	Avg. Number of Teachers per 1000 Students Outside the Regular Classroom: Self-Contained	Department-alized
District		
1-500	5.61	24.55
500-999	6.91	16.60
1,000-2,499	7.33	13.87
2,500-4,999	7.19	12.75
5,000-9,999	6.31	11.10
10,000-24,999	2.35	9.17
25,000-49,999	1.34	10.57
50,000 +	1.79	8.36
School		
1-99	1.82	43.38
100-199	12.98	20.30
200-299	9.09	15.16
300-399	9.14	14.89
400-499	7.07	12.81
500-599	6.86	12.29
600-799	6.32	11.01
800-999	6.74	8.77
1,000-1,499	7.87	8.46
1,500-1,999	10.57	6.08
2,000-2,499	1.02	7.07
2,500-2,999	n/a	6.76
3,000 +	n/a	10.07

Column values were calculated with the following formula: ((teacher/pupil ratio) - (self-reported teacher/pupil ratio)) * 1,000

Tables 3, 4 and 5 show how the number of non regular classroom teaching positions per 1,000 students vary with district expenditures per pupil, per pupil expenditures for instruction and the percent of total expenditures devoted to instruction respectively. Table 3 shows how the number of teachers outside the regular classroom varies with per pupil expenditures. Overall, there seems to be an increase in the number of such teachers per 1,000 students as expenditures increase. When considered with our earlier findings that the share of total expenditures devoted to instruction is fairly constant regardless of spending level (Picus, 1993 a and Picus, 1993b), Table 3 suggests that districts with more money both reduce class size, and employ more individuals with assignments outside of the regular classroom. The differences reported for districts with expenditures below $2,000 pupil were not statistically significant.

Table 3
Number of Teachers Per 1000 Students With Assignments Outside of the Regular Classroom by District Expenditure Per Pupil

Expenditures Per Pupil	Avg. Number of Teachers per 1000 Students Outside the Regular Classroom: Self-Contained	Department-alized
under $1,500	3.36	12.04
$1,500-$1,999	5.66	2.99
$2,000-$2,499	4.55	7.85
$2,500-$2,999	2.94	10.20
$3,000-$3,499	2.98	9.65
$3,500-$3,999	4.95	13.76
$4,000-$4,499	7.42	13.57
$4,500-$4,999	10.09	21.54
$5,000-$5,499	11.87	20.90
$5,500-$5,999	5.10	27.79
$6,000 +	10.79	23.08

Column values were calculated with the following formula: ((teacher/pupil ratio) - (self-reported teacher/pupil ratio)) * 1,000

This pattern is not as clear when per pupil expenditures for instruction are considered in Table 4. As expenditures for instruction increase, the number of teachers per 1,000 students with other assignments varies considerably. There is a slight tendency for the number of such positions to increase as expenditures per pupil increase in the low to middle portions of the expenditure range, but the pattern is less consistent at the high spending levels. For the 30 districts represented with expenditures between $4,500 and $5,000 per pupil, the differences reported in Table 4 are not statistically significant. Departmentalized class differences were not statistically significant for districts with expenditures between $5,000 and $6,000 per pupil.

Table 5 reports the average number of teachers outside of the regular classroom by percent of total expenditures devoted to instruction. The variation in both the self contained and departmentalized schools is relatively small. When combined with the fact that the vast majority of the districts are clustered in the center of the range presented in Table 5, it is difficult to draw any substantial conclusions about the impact of the share of expenditures devoted to instruction on the way teachers are assigned in schools. What these findings seem to indicate is that in high spending districts there are both smaller classes and more support positions than can be found in low spending districts. These findings are statistically significant at the 0.01 level for self-contained classes in districts where expenditures for instruction below 70% of total expenditures and for departmentalized classes where the percent spent for instruction is below 75%.

Table 4
Number of Teachers Per 1000 Students With Assignments Outside of the Regular Classroom by District Expenditure Per Pupil for Instruction

Exp. Per Pupil For Inst.	Avg. Number of Teachers per 1000 Students Outside the Regular Classroom: Self-Contained	Department-alized
under $1,500	4.59	7.14
$1,500-$1,999	2.99	9.86
$2,000-$2,499	5.28	12.86
$2,500-$2,999	8.19	18.03
$3,000-$3,499	10.29	27.52
$3,500-$3,999	6.08	25.03
$4,000-$4,499	14.66	42.91
$4,500-$4,999	7.11	24.62
$5,000-$5,499	-12.24	-0.22
$5,500-$5,999	40.09	4.75
$6,000 +	20.32	23.63

Column values were calculated with the following formula: ((teacher/pupil ratio) - (self-reported teacher/pupil ratio)) * 1,000

Table 5
Number of Teachers Per 1000 Students With Assignments Outside of the Regular Classroom by Percent of Total Expenditures for Instruction

Percent of Exp. For Inst.	Avg. Number of Teachers per 1000 Students Outside the Regular Classroom: Self-Contained	Department-alized
Less than 50%	5.42	9.83
50% - 54.99%	4.97	11.31
55% - 59.99%	4.45	11.68
60% - 64.99%	4.76	14.53
65% - 69.99%	7.53	14.43
70% - 74.99%	5.51	13.36
70% - 79.99%	4.66	11.77
80% +	-23.48	18.93

Column values were calculated with the following formula: ((teacher/pupil ratio) - (self-reported teacher/pupil ratio)) * 1,000

These findings indicate that the amount of money available to a school district does matter in terms of the size of classes it is able to provide for its students. To the extent that smaller classes improve student opportunities for learning, higher expenditures increase the probability that a student will attend class with fewer classmates and teachers will have smaller classes.

Variation by District and School Teacher/Pupil Ratios

Tables 6 and 7 compare the number of teachers per 1,000 students who do not have regular classroom assignments with the teacher/pupil ratios computed in our earlier research (Picus, 1993a and 1993b). Table 6 shows how the average number of such teachers varies with the number of teachers per 1,000 students at the district level, while Table 7 displays the same comparison based on the number of teachers per 1,000 students at the school level. Both tables show a very strong pattern of fewer teachers with assignments outside of the regular classroom as the number of teachers per 1,000 students declines. This implies that as the average district or school pupil/teacher ratio increases (the teacher/pupil ratio decreases), there are fewer other certificated personnel to provide additional opportunities for students.

Table 6
Number of Teachers Per 1000 Students With Assignments Outside of the Regular Classroom Compared to District Teacher/Pupil Ratios

	Avg. Number of Teachers per 1000 Students Outside the Regular Classroom	
Teachers Per 1,000 Pupils	Self-Contained	Department-alized
more than 100	72.71	49.21
90.01 - 100.00	28.26	29.52
80.01 - 90.00	17.46	22.24
70.01 - 80.00	16.84	16.38
60.01 - 70.00	10.08	10.78
50.01 - 60.00	4.97	7.02
40.01 - 50.00	-2.63	0.15
30.01 - 40.00	-7.35	-5.52
20.01 - 30.00	-23.64	-31.91
less than 20.00	-42.30	-44.84

Column values were calculated with the following formula: ((teacher/pupil ratio) - (self-reported teacher/pupil ratio)) * 1,000

The negative numbers at the lowest teacher/pupil ratios in both Tables 6 and 7 are somewhat disconcerting as they imply that the district or school estimate of the number of teachers per 1,000 students is lower than the self reported number of teachers per 1,000 students. This appears to be the result of relatively small number of schools and/or districts in the sample at these very low values of teachers per 1,000 students, resulting in a heavy influence of a small number of very large classes most likely band and physical education.

These findings makes sense if one considers the expenditure findings from the district level analysis. Remember that districts spend an average of 60 percent of their resources on direct instruction. Given that higher spending districts have smaller classes, but still spend 60 percent of their funds on instruction, the remaining 40 percent also represents a larger per pupil figure than is available in low spending districts. Thus it is not surprising that districts with more to spend tend to have higher numbers of other certificated staff available in their schools. Given this finding, the next section compares self-reported class size to the relative number of certificated, non-teaching staff in each school.

Table 7
Number of Teachers Per 1000 Students With Assignments Outside of the Regular Classroom Compared to School Pupil/Teacher Ratios

	Avg. Number of Teachers per 1000 Students Outside the Regular Classroom	
Teachers per 1,000 students	Self-Contained	Department-alized
more than 100	55.28	49.21
90.01 - 100.00	31.54	29.52
80.01 - 90.00	23.10	22.24
70.01 - 80.00	18.28	16.38
60.01 - 70.00	12.58	10.78
50.01 - 60.00	7.10	7.02
40.01 - 50.00	1.84	0.15
30.01 - 40.00	-3.28	-5.52
20.01 - 30.00	-16.06	-31.91
less than 20.00	n/a	-44.84

Column values were calculated with the following formula: ((teacher/pupil ratio) - (self-reported teacher/pupil ratio)) * 1,000

Variation by Other Variables

In our earlier work, we analyzed pupil/teacher ratios in comparison to a number of other district and school characteristics. These included the percent of minority pupils in the school or district, the number of pupils qualifying for a free or reduced price lunch (as a proxy for poverty), and the type of community in which the district and/or school is located. Similar analyses were undertaken for this project. However, analyses of the average self-reported class size by each of the variables identified above showed no identifiable patterns. This was true for both the departmentalized and self-contained teacher samples.

CONCLUSION

In our earlier work, we attempted to develop analytic models to predict the pupil/teacher ratio. These models provided additional insight into the factors that are related to pupil/teacher ratio at the district and school level. Consequently, we attempted to develop analytic models of the self-reported class size as well. Because we had data on individual teachers, individual schools, and school districts, we attempted to model self-reported class size using three equations for both self-contained and for departmentalized teachers. Each equation used either the self-contained or departmentalized class size as the dependent variable, and a series of independent variables describing either individual, school or district characteristics. While the analysis resulted in a number of significant coefficients, we were never able to explain more than 6 percent of the variance in the self-reported class size, regardless of functional form.

There are considerable difficulties with including variables from different levels in one regression equation. A solution to this problem is often to use Hierarchical Linear Modeling (HLM)techniques which control for different levels in the equations. On the surface, the SASS data appear to be ideally suited for such treatment. However, the SASS design only has four to eight teachers in each school, and between one and three schools in each district. To get significant results from an HLM model, the nested data sets require a minimum of approximately 30 observations, making HLM inappropriate for this analysis.

As a result, at this time, we are unable to explain most of the variation in self-reported class size in K-12 public schools. Although disappointing, these results may lend credence to our earlier suspicions that schools are basically consistent in what they do. The clearest example of this is the share of expenditures devoted to instruction, which averages 60 percent regardless of how much money is available to a school district. Since our initial analyses indicate that class size declines with increases in expenditures, and that as district and school pupil/teacher ratios increase, the difference between those ratios and the self reported class size declines it may be that school administrators are inclined to spend what ever resources are available to them according to the same "rules of thumb" regardless of the level of those resources. This means that districts with substantially more money, will be able to offer considerably more of everything to their students. The increased number of dollars available for instruction will translate into smaller class size and higher paid teachers, but at the same time, additional resources will be spent on non-classroom certificated staff to provide a range of support to the teachers as well.

This leads to an interesting question, which is whether or not expenditures for other functional areas such as administration and instructional support remain proportionally the same as school district expenditures increase, or if changes in those proportions can be observed with changes in spending levels. Unfortunately, the expenditure data provided by the 1986-87 Census of Governments does not allow fine enough distinctions across expenditure functions to conduct such an analysis. Hopefully, the 1990-91 Census of Governments, combined with the 1990-91 SASS, and the Census project to link Census and School District data more closely, will enable us to conduct such analyses in the future.

REFERENCES

Barro, Stephen M. (1992). *What Does the Education Dollar Buy?: Relationships of Staffing, Staff Characteristics, and Staff Salaries to State Per-Pupil Spending*. Washington D.C.: SMB Economic Research.

NCES. (1992). *Digest of Education Statistics*. National Center for Education Statistics, Washington, D.C., U.S. Government Printing Office.

Odden, A.O. (1990). Class size and student achievement: Research-based policy alternatives. *Educational Evaluation and Policy Analysis*. 12(2), 213-227

Picus, Lawrence O. (1993a). *The Allocation and Use of Educational Resources: District Level Evidence from the Schools and Staffing Survey*. Los Angeles, CA: USC Center for Research in Education Finance, Working Paper No. 34.

Picus, Lawrence O. (1993b). *The Allocation and Use of Educational Resources: School Level Evidence from the Schools and Staffing Survey*. Los Angeles, CA: USC Center for Research in Education Finance, Working Paper No. 37.

Picus, Lawrence O. and Yasser Nakib. (1993). *The Allocation and Use of Educational Resources in Florida School Districts*. Los Angeles, CA: USC Center for Research in Education Finance, Working Paper No. 39.

Slavin, R. (1989). Achievement effects of substantial reductions in class size. in R. Slavin (ed.), *School and classroom organization* (pp. 247-257). Hillsdale, NJ: Erlbaum.

Smith, M.L. & Glass, G.V. (1980). Meta-analysis of research on class size and its relationship to attitudes and instruction. *American Educational Research Journal*, 17(4) , 419-433.

NEA (1992). Estimates of School Statistics, 1991-92. Washington D.C.: National Education Association.