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: 1994
Papers Presented at the 1994 Meeting of the American Statistical Association

Working Paper No. 95-01 January 1995

Contact: Dan Kasprzyk
Special Surveys and Analysis Branch
(202) 219-1325

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

Paul D. Planchon
Associate 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).

January 1995

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 Ahmed
Acting Associate Commissioner
Statistical Standards and
Methodology Division

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

Table of Contents

SECTION ON SURVEY RESEARCH METHODS

Estimation Issues in School Surveys

Chair: Paul D. Planchon
National Center for Education Statistics

Response and Coverage Issues in School Surveys

Chair: Gary Shapiro
Abt Associates

SOCIAL STATISTICS SECTION

Education Research Using the Schools and Staffing Surveys and the National Education Longitudinal Survey

Chair: Daniel Kasprzyk
National Center for Education Statistics

Preface

The fourteen papers contained in this volume were presented at the 1994 meeting of the American Statistical Association as sessions entitled "Education Research Using the Schools and Staffing Surveys and National Education Longitudinal Survey," "Estimation Issues in School Surveys," and "Response and Coverage Issues in School Surveys." The first session was organized by Michael Podgursky of the Department of Economics of the University of Massachusetts and chaired by Daniel Kasprzyk, Special Surveys and Analysis Branch, National Center for Education Statistics (NCES). The second session was organized by Mr. Kasprzyk and chaired by Paul D. Planchon, Associate Commissioner for Elementary/Secondary Education Statistics of NCES. Mr. Kasprzyk organized the third session, and Gary Shapiro of Abt Associates served as chairman.

INTERSURVEY CONSISTENCY IN SCHOOL SURVEYS

Albert Holt, Synectics for Management Decisions; Steven Kaufman, NCES; Fritz Scheuren, George Washington University; and Wray Smith, Synectics for Management Decisions

Fritz Scheuren, 1402 Ruffner Rd., Alex. Va. 22302 U.S.A. or Email at scheuren@aol.com

KEY WORDS: Generalized least squares; Raking ratio estimation; Weight attenuation and weight bounding

1. Introduction and Background

For the first time, in 1993-1994, the private school components of the Schools and Staffing Survey(SASS) and the Private School Survey(PSS) are being fielded in the same school year. Even though these two NCES surveys measure some of the same variables, the results between the surveys will not agree.

As the PSS is used for the SASS sampling frame, the PSS results are likely to be the more accurate. Under these circumstances, it makes sense to explore whether the introduction of PSS totals into SASS might lead to improvements. Traditional post-stratification methods exist to employ auxiliary information at the estimation stage in surveys. These, however, cannot be applied to SASS without modification.

In particular, PSS and SASS both measure numbers of schools, numbers of teachers, and numbers of students. Conventional simple or raking ratio adjustment procedures could be used to adjust sample weights so that the SASS estimates agreed with PSS for each of the three totals separately. Such approaches do not work, though, if the weights are to be adjusted so that all three SASS estimates agree simultaneously.

Alternatives are possible, though, that permit simultaneous estimation. For example, the Generalized Least Squares(GLS) techniques advocated by Deville and Särndal(1992) can be used, as in Imbens and Hellerstein (1993). While the asymptotic properties of GLS and GLS-like estimators are attractive, their finite sampling properties are not necessarily desirable. Possible operational concerns with GLS procedures include: (1)Some of the resulting weights may be less than one or even may be negative.(2)The procedure may be difficult to carry out, especially when excessively small weights arise.(3)The effect on estimates not directly adjusted is unknown and could be harmful.

Modified GLS.--To discuss the basic algorithm employed in Generalized Least Squares, it is necessary to define some notation; in particular --

w_i is the original SASS weight for the ith SASS observation, i=1,...,n.

t_i is the SASS total of teachers for ith SASS observation, i=1,...,n.

s_i is the SASS total of the students for the ith SASS observation, i=1,...,n.

N is the total estimated number of schools, as given by PSS.

T is the total estimated number of teachers, as given by PSS.

S is the estimated total number of students, as given by PSS.

In reweighting SASS three constraints are imposed on the new weights u_i,

$$\sum u_i = N$$

$$\sum u_i t_i = T$$

$$\sum u_i s_i = S$$

For our application the new weights u_i, subject to these constraints, are to be chosen to minimize a loss function which can be written as the sum of squares

$$\sum (u_i - w_i)^2$$

This is perhaps the simplest and most straightforward loss function that might be chosen. Motivating it here is outside our present scope, except to say that the sensitivity of the final results to the loss function chosen seems not to be too great(but this is an application issue and will be among the areas for future study, as set forth at the end of this paper). As the literature on GLS methods also makes clear(Deville, Särndal, and Sautory, 1993), the loss function chosen determines the form of the estimators eventually developed and those obtained using squared error loss are particularly convenient in a SASS setting.

Now the usual Lagrange multiplier formulation of this problem yields, after some algebra, that the new weights are of the form

$$u_i = w_i + \lambda_1 + \lambda_2 t_i + \lambda_3 s_i ,$$

where the λ's are obtained from the matrix expression

$$\underline{d} = M\lambda$$

with the vector $\underline{d}$ consisting of three elements, each a difference between the corresponding PSS and SASS totals for schools(first component), teachers(second component), and students(third component); in particular

$$N - \sum w_i$$

$$T - \sum w_i t_i$$

$$S - \sum w_i s_i$$

The matrix M is given by

$$\begin{matrix} n & \sum t_i & \sum s_i \\ \sum t_i & \sum t_i & \sum t_i s_i \\ \sum s_i & \sum t_i s_i & \sum s_i \end{matrix}$$

and $\underline{\lambda}$ is the vector of unknown GLS adjustment factors obtained from

$$\underline{\lambda} = M^{-1}\underline{d}.$$

The M matrix is based solely on the unweighted sample relationships among schools, teachers and students. This is not an essential feature of our approach; and, indeed a weighted version of the M matrix has been tried, as discussed later.

Illustrative Example.--To fix ideas, consider the following "toy" example that may help illustrate the method being employed. In particular, suppose a SASS subgroup has ten observations; written below as column vectors where the components

$$\begin{matrix} x \\ y \\ z \end{matrix}$$

correspond to SASS schools, teachers, and students respectively:

$$\begin{matrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 1 & 6 & 2 & 7 & 3 & 8 & 4 & 9 & 5 & 10 \end{matrix}$$

Aggregating the three SASS components yields

$$\begin{matrix} 10 \\ 55 \\ 55 \end{matrix}$$

Suppose further that the PSS totals for this subgroup are

$$\begin{matrix} 10 \\ 50 \\ 50 \end{matrix}$$

Notice, the SASS school total has already been set equal to that in the PSS. This has been done so that the example starts where a standard SASS estimation procedure might end.

For the "modified GLS" the elements of the matrix M and the vector $\underline{d}$ need to be obtained. It is immediate that $\underline{d}$ is

$$\begin{matrix} 0 \\ -5 \\ -5 \end{matrix}$$

For the matrix M, after some calculation, the values are

$$\begin{matrix} 10 & 55 & 55 \\ 55 & 385 & 355 \\ 55 & 355 & 385 \end{matrix}$$

For the inverse of M^{-1}, the values turn out to be

$$\begin{matrix} .5481 & -.0407 & -.0407 \\ -.0407 & .0204 & -.0130 \\ -.0407 & -.0130 & .0204 \end{matrix}$$

Thus, solving

$$\underline{\lambda} = M^{-1}\underline{d}$$

the vector is $\underline{\lambda}' = (.4074, -.0370, -.0370)$ and the modified GLS weights are of the form

$$u_i = w_i + .4074 - .0370t_i - .0370s_i.$$

Additional General Considerations.--So far the GLS algorithms have been discussed as if the issues are simply computational. In point of fact, the real challenges arising in any SASS implementation require statistical judgments. Among these are:

- Deciding on the level of SASS at which the constraints are to be imposed. For example, from a subject-matter perspective, it seems appropriate to do GLS estimation

separately within the nine private school types. For some of the larger typologies, maybe even finer groupings might be attempted (say, school level or urbanicity). At what point will the potential benefits of a GLS adjustment outweigh the harm?

- Avoiding weights that are negative or too small (i.e., given that each SASS observation always represents at least itself, a natural requirement to impose is that $u_i \geq 1$ for all i). This concern is particularly troublesome because of the seemingly ad hoc flavor of what may be needed to get acceptable weights.

While the guidance of earlier GLS practice elsewhere is available(e.g., Bankier, 1992; Fuller et al, 1994)), neither of these challenges can be resolved for SASS, except "in the doing." Among the factors to consider are obvious ones such as --

- How difficult(expensive) is the method to implement, including to explain?
- How statistically sensitive are the constrained estimates to seemingly small but arbitrary decisions in the way the method is applied?

2. An Initial SASS Application

The basic approach taken in this Section is to analyze a small but real data set, so as to develop an understanding of the operating characteristics of the modified GLS approach being looked at here for potential use in the 1993-1994 NCES school surveys. To this end, consider, as a test, data on Catholic schools taken from the 1991-1992 PSS and the 1990-1991 SASS. These schools for SASS and PSS are divided into three subgroups: parochial, diocesan, and private. The weighted data on the last of these groups, Private Catholic Schools, are displayed below.

Item	PSS	SASS
Schools	901	894
Teachers	22340	22340
Students	354040	365367

The modified GLS application might be started by first scaling up the school total from SASS to that for PSS or simply leaving the total as is (the course taken here). In any event, after suitable calculations, familiar from Section 1, the GLS weights are obtained from the expression

$$u_i = w_i + .0415 + .0767t_i - .0046s_i .$$

One of the λ is negative; hence the u_i could be too small or even negative for a particular combination of original weight, teacher and student total. However, this did not occur.

The Private Catholic typology has the smallest sample size(at 112) and was chosen for that reason. Now three constraints are being imposed and sample size "rules of thumb" suggest that the average sample size per constraint be on the order of 25 or more. Here the average is 112/3 = 37, so reasonably good results might be expected at least on this score, provided SASS and PSS are consistent(i.e., that SASS can be treated as a representative sample of the larger PSS). Since the surveys are for different years this last condition is not guaranteed(see Section 3). Figures 1 and 2 below suggest, though, that SASS and PSS are roughly consistent, at least in this case. The SASS scatterplot lies well within that for PSS and is oriented along the same axis. Indeed, the average student/teacher ratios from the two surveys(both at about 16-to-1) are almost identical

3. A Second SASS Application

In this Section, a second GLS application is taken from the 1990-91 SASS and 1991-92 PSS. Here Nonsectarian Special Emphasis Schools are examined. That group was chosen because the weighted SASS and PSS counts are quite far apart(see below). If a problem with the GLS approach were to show up, it might well be in this group.

Item	PSS	SASS
Schools	1810	1700
Teachers	13724	18717
Students	202178	212433

First GLS Attempt.--The Nonsectarian Special Emphasis Typology has a somewhat larger sample size(at 205) than for Private Catholic Schools. Hence, standard concerns about overconstraining small numbers of cases do not bind here; indeed, it would even be possible to attempt to introduce still more PSS data into the SASS estimation --a point we will come back to later.

The modified GLS was solvable, leading to weights of the form

$$u_i = w_i - .0254 + .0101t_i - .0008s_i.$$

If sample size were our only consideration, the GLS weights should work well; however, they do not. As a matter of fact, nearly one third of these weights were less than one and many (22 in all) were negative. The SASS data are just not consistent with those from PSS. For example, the student teacher ratio in PSS is about 15 to 1; for SASS, on the other hand, it is closer to 11 to 1.

In the PSS and particularly in SASS, outliers exist which are well outside the point clouds of either source(see figures 3 and 4). One of these, circled in the SASS data is quite damaging since it has a weight of about 14 and a teacher count of 208 combined with a student count of 78--probably a data error of some sort.

Subsequent Attempts.--Removing the outlier yields the totals below.

Item	PSS	SASS
Schools	1809	1686
Teachers	13516	15836
Students	202100	211353

It would be great if we could now say that negative GLS weights or weights less than one had, with this single change, been eliminated. This did not turn out to be true; nonetheless, the results were encouraging. The number of "small" or negative weights was cut way down(from over eighty to under two dozen -- still quite sizable, however).

An examination of the SASS cases that had GLS weights that were too small revealed two patterns that might be mentioned: (1)Most of the cases were ones where the original SASS weight was close to one to begin with. (2)Some of the cases with negative weights had student/teacher ratios, that put them near the edge of the SASS and PSS point clouds -- making them possible candidates for outlier treatment too.

A series of alternatives were tried, including the use of different GLS loss functions(See Scheuren, 1994). Eventually, we settled on an alternative that fit a GLS estimator to the smaller two-thirds of the schools. The larger schools were simply too inconsistent to be fit with a GLS estimator; instead, an imputation approach was considered that might have future promise in the sample regions where the 1993-1994 SASS cases have weights of nearly one to begin with. More is said about this in the concluding section.

4. Future Plans

At this still early stage it is hard to do more than just conjecture about next steps in terms of the 1993-1994 SASS. Even so, there are some "lessons learned" and a few observations that may be of general interest. This short section makes a beginning summary of these.

First, our test plans call for more of the nine SASS typologies to be GLS-adjusted. It is plausible to speculate that still other methods may occur to us as we tackle these remaining typologies. Preliminary work, though, on some of these other typologies suggests that it is unlikely, for the 1993-1994 SASS, that we will uncover better approaches than those discussed. On the other hand, our sense of how and when to apply these techniques may grow considerably.

Second, we need to display evidence, convincing in the test SASS applications, that a GLS adjustment of the type contemplated will lead to an improvement in the estimates; or, at least, to no(or minimal) harm. On this latter point figures 5 and 6 are encouraging(because these figures show that the GLS weights are only minimally altered from their original values).

Third, methods for variance estimation need exploration . While the general GLS approach is well covered in the literature, an efficient method has to be programmed and tested in the SASS environment. Particular concerns exist, too, about the impact on variance and variance estimation of the various ad hoc adaptations needed to keep the weights reasonable.

Fourth, a general strategy for applying GLS to SASS may emerge from our work; but it appears highly unlikely that GLS procedures for SASS will become automatic any time soon. There is simply not going to be enough of an experience base to make this safe.

Fifth, some improvements in SASS and PSS processing may be a consequence of the study of GLS applications. One of those that has arisen so far is the clear possibility that edit checking could be enhanced if GLS estimation is attempted. A subtler concern is the treatment in SASS of the very largest schools, when these become nonrespondents. Here perhaps an imputation rather than a weighting approach may be preferred -- using , say, the PSS data as a starting point. Among schools above a given size this could have more benefit in reducing SASS mean square error than GLS.

Obviously, still other concerns need to be considered, even if the present modified GLS method were judged desirable; and could be made routine. Among these, of course, are the cost in time and money of its application. So stay tuned.

Selected Bibliography

Bankier, M. (1992), "Two-step Generalized Least Squares Estimation in the 1991 Canadian Census," in *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Burton, R. (1989), Unpublished Memorandum, National Center for Education Statistics.

Casady, R. and Valliant, R.(1993), "Conditional Properties of Post-Stratified Estimators Under Normal Theory," *Survey Methodology*, 19, 183-192.

Deming, W.E. and Stephan, F.F. (1940), "On a Least Squares Adjustment of a Sampled Frequency When the Expected Marginal Tables are Known," *Annals of Mathematical Statistics*, 11, 427-444.

Deville, J.C., and Särndal, C.E. (1992), "Calibration Estimators in Survey Sampling," *Journal of the American Statistical Association*, 87, 376-382..

Deville, J.C., Särndal, C.E. and Sautory,O. (1993), "Generalized Raking Procedures in Survey Sampling,"

Journal of the American Statistical Association, 88, 1013-1020.

Fuller, W., Loughin, M., and Baker, H. (1994), "Regression Weighting in the Presence of Nonresponse with Application to the 1987-1988 Nationwide Food Comsumption Survey," *Survey Methodology*, 20, 75-86.

Hansen, M., Hurwitz, W.,and Madow, W.(1953), *Sample Survey Methods and Theory*, New York: Wiley.

Holt, D. and Smith, T.M.F. (1979), "Post-stratification," *Journal of the Royal Statistical Society, Series A*, 142, 33-46.

Imbens, G.W. and Hellerstein, J.K. (1993), "Raking and Regression," *Discussion Paper Number 1658*, Cambridge, MA, Harvard Institute of Economic Research, Harvard University.

Ireland, C.T. and Kullback, S. (1968), "Contingency Tables with Known Marginals," *Biometrika*, 55, 179-188.

Kaufman, S. and Huang, H. (1993), "1990-91 Schools and Staffing Survey: Sample Design and Estimation," *Technical Report NCES 91-127*, Washington, DC: U.S. Department of Education.

Little, R.J.A. (1991), "Post-Stratification: A Modeler's Perspective," in *Proceedings of the Section on Survey Research Methods*, American Statistical Association..

Little, R.J.A. and Wu, M. (1991), "Models for Contingency Tables with Known Marginals When Target and Sample Populations Differ," *Journal of the American Statistical Association.*

Oh, H.L. and Scheuren, F. (1978a), "Multivariate Raking Ratio Estimation in the 1973 Exact Match Study," in *Proceedings of the Section on Survey Research Methods*, American Statistical Association..

Oh, H.L. and Scheuren, F. (1978b), "Some Unresolved Application Issues in Raking Ratio Estimation," in *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Olkin,I.(1958), "Multivariate Ratio Estimation for Finite Populations," *Biometrika*, 45,154-165.

Scheuren, F. (1994), "Median Raking and Other Garden Variety Sampling Tools," *Unpublished Draft NCES Working Paper.*

Särndal, C.E., Swensson, B. and Wretman,J. (1992), *Model Assisted Survey Sampling*, New York: Springer-Verlag.

Figure 1: PSS Teacher Versus Student Total for Private Catholic Schools

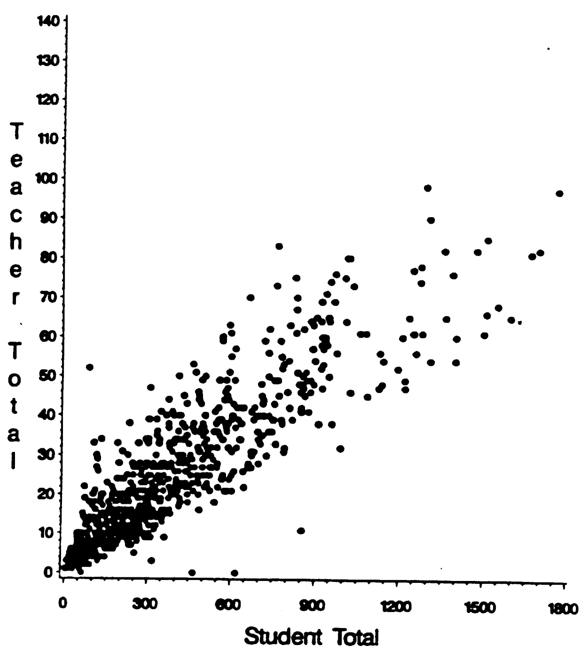


Figure 2: SASS Teacher Versus Student Total for Private Catholic Schools

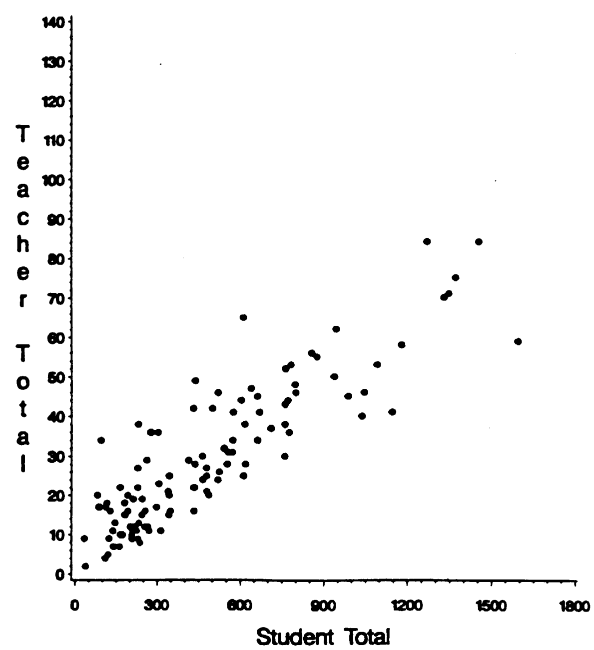


Figure 3: PSS Teacher Versus Student Total for Nonsectarian Special Emphasis Schools

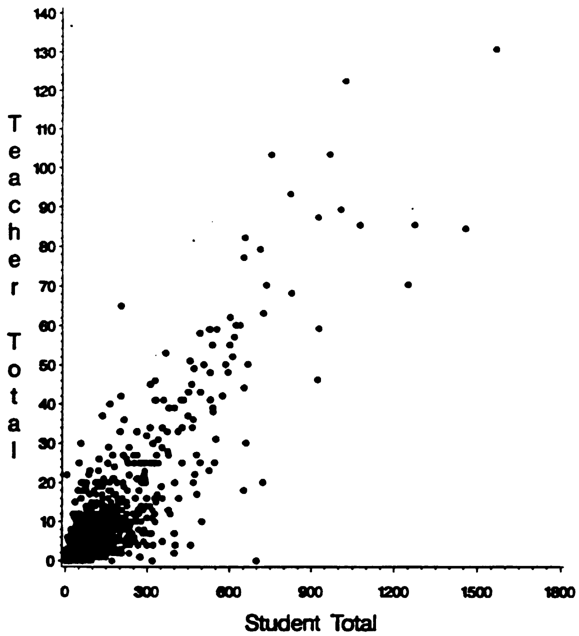


Figure 4: SASS Teacher Versus Student Total for Nonsectarian Special Emphasis Schools

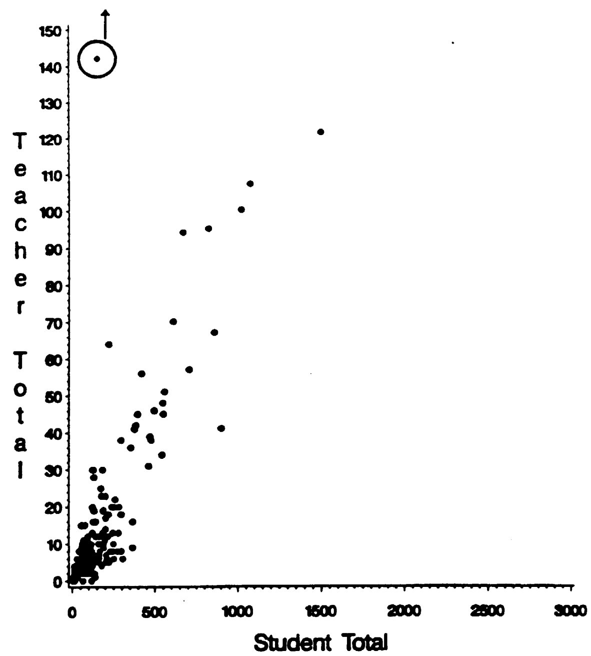


Figure 5: SASS Original Weight vs. GLS Weight for Private Catholic Schools

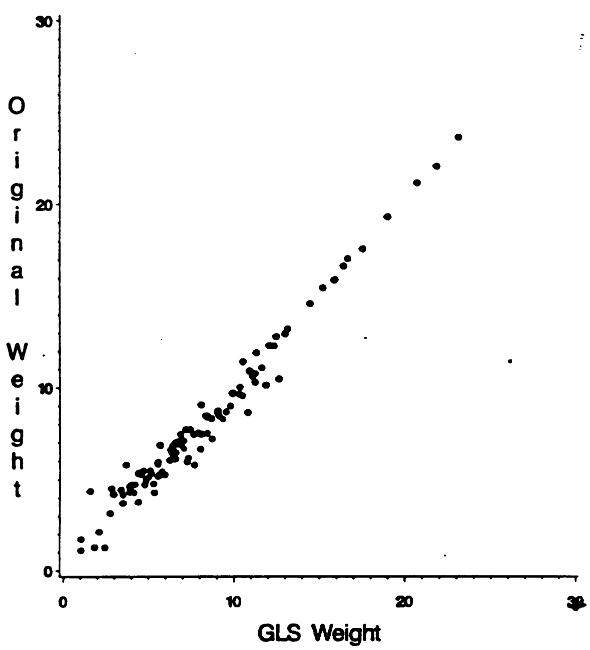
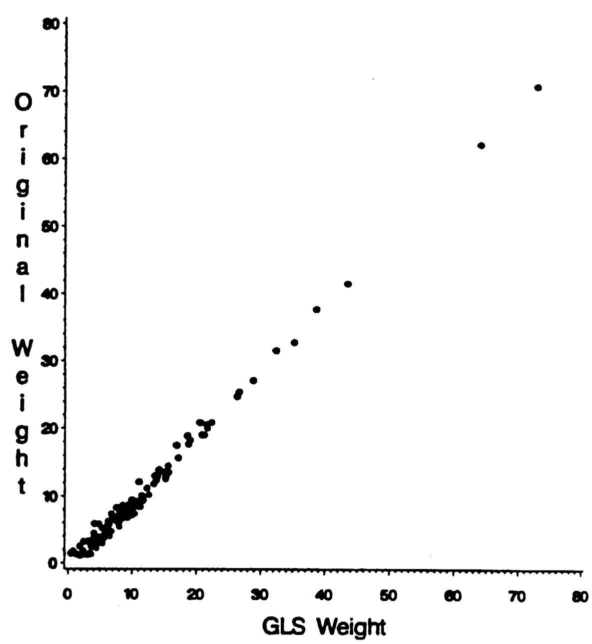


Figure 6: SASS Original Weight vs. GLS Weight for Nonsectarian Special Emphasis Schools
(smallest two-thirds)



ESTIMATION ISSUES RELATED TO THE STUDENT COMPONENT OF THE SASS

Karen Ellen King, Bureau of the Census and Steve Kaufman, National Center of Education Statistics[1]
Karen Ellen King, DSMD, U.S. Bureau of the Census, Washington, DC 20233

Key Words: Sampling, Probability of Selection, Hypergeometric Distribution

I. INTRODUCTION

The Schools and Staffing Survey (SASS) is a periodic integrated system of surveys of schools, school districts, school administrators, and teachers. For the 1993-94 SASS, a student component was added.

SASS is sponsored by the National Center for Education Statistics (NCES) of the U.S. Department of Education. Users of the survey data are educators, researchers, policy makers, and others interested in educational issues.

The survey data is collected by mail, with telephone followup of nonrespondents. All levels of the SASS are interrelated. Selection of sample schools, both public and private, is the starting point. For each sample school, a sample of its teachers is selected and data is also collected from its principal. The school district of each selected public school is also in the sample. For the current SASS, a sample of students was selected from sample teachers; continuing the relationship of one component with the other components of the survey.

The NCES planned to add a student component to the SASS for several years. The goals of this component is to examine the quality of teachers through their students and analyze student characteristics. This is accomplished by selecting a few sample students from a class taught by each sample teacher.

A student component in SASS was tested initially as part of a 1991 SASS Research Study. In this study, student sampling and the collection of administrative data on selected student was attempted for the first time. Several problems were encountered during the sampling and the collection phases which discouraged any attempt at estimation.

A second feasibility study was conducted during the 1992-93 school year to solve the operational problems encountered in the first study. It is also where we began to deal with the issue of estimation, in particular, to develop an estimator for the student's probability of selection using only the amount of information that an already over burdened school could easily provide.

This paper gives an overview of the second feasibility study and a summary of the components that make up our estimator of the probability of selection of students.

II. OVERVIEW OF SAMPLING

A. School Selection

As with all SASS surveys, the selection of samples of public and private schools was the starting point for the feasibility study. Three hundred public and 200 private schools were selected and mailed forms for listing teachers. A teacher listing form asks schools to provide the names and some demographic information for every eligible teacher at that school. Eligible teachers consist of regular full-time and part-time teachers whose main assignment was teaching in kindergarten or any of grades 1 to 12 during the school year.

Completed listing forms were returned to the Census Processing Center in Jeffersonville Indiana. Two hundred thirteen public and 133 private schools returned completed teacher listing forms.

Interviewers specially trained for this operation did the teacher selection, class period selection, and the student selection through a series of telephone conversations with participating schools.

B. Teacher Selection

Three teachers (if available) were systematically selected from each of the returned teacher listing form.

Each school was called to confirmed that each sampled teacher was eligible, i.e., did they teach at least one regularly scheduled class of K-12 grade students in a week. Once the ineligible teachers were screened out, the call continued by asking questions to classify the eligible teachers as either self-contained or

---

[1] This paper reports the general results of research undertaken by Census Bureau staff in collaboration with staff from the National Center for Education Statistics. The views expressed are attributed to the authors and do not necessarily reflect those of the Census Bureau or National Center for Education Statistics.

departmental. Sampling instructions for class period selection varied by this teacher classification.

* Self-contained is defined as teaches several different subjects to the same group of students all day.

* Departmental is defined as teaches only a limited number of subjects to more than one group of students per day.

C. Class Period Selection

For departmental teachers, a double sampling procedure was used to select the sample class period. We started by asking the school how many periods they had per week, and then, using this value, selected a set of five class periods as the initial sample. If all the teachers were departmental at the school then all three teachers had the same set of class periods.

For example, suppose the school told us that there were 25 class periods in a week (not counting homeroom). For this number of periods per week, the selected set of class periods were the fifth on Monday, the fourth on Tuesday, the third on Wednesday, the second on Thursday, and the first on Friday.

Then the interviewer probed the school about each class period in the initial set of five to determine if the teacher actually taught a class of eligible students. Eligible students are those in kindergarten through the twelfth grade, that are receiving instruction and are not in study hall, recess, lunch or homeroom. If a teacher did not teach a class in one of the class periods, the period was considered ineligible to go to the next step of sampling. Once the eligibility of each class period was determined, one out of the remaining set of eligible class periods was randomly selected.

For example, suppose teacher Jane Doe taught four out of five class periods given in the above example. (She supervises study hall the third period on Wednesdays.) To select the class period we ordered the four remaining periods by days in the week (Monday through Friday) and picked one.

The third class period in our ordered set was selected. Thus, we wanted three student names from the second period on Thursday.

For the self-contained teachers, no class period sampling procedure was needed since they only taught one class of students.

Schools were asked to get selected class period rosters. Generally the first call was terminated so that the school could look up the roster. Another time was set for a call back to do the next phase of selection.

There are two reasons to justify this elaborate scheme to select a class period. The first is the double sampling guaranteed that we selected a class period where the teacher was actually teaching. During the initial study, we selected one class period randomly in the school week for each departmental teacher. Many times the school simply said that the teacher was not teaching during the selected period. Subsequently, no students were selected for these teachers and the student sample size was much smaller than expected. The second reason was to reduce the chances of bias being introduced into the student sample. If we pick only one class period, there is the possibility that a subset of the student body would be in ineligible classes (study hall, homeroom, lunch, or recess) and have no chance of selection. When we increase the number of class periods selected to five, the chances of a student being in an ineligible class for all five class periods becomes small.

D. Student Selection

When the class period roster was available, over the phone we gave the school instructions to select three sample students from the roster. A random number table was used to indicate the line numbers of the students selected.

For example: Suppose Jane Doe's second period class on Thursdays had 26 students. Using a table, interviewers would have asked for the 3rd, the 14th, and 24th name from the top of the roster.

Student names or some other unique student identifier was requested so that we could uniquely label each student's forthcoming questionnaire. Eleven schools refused to provide student names for our survey fearing parental displeasure.

Two months after telephone sampling, student questionnaires were mailed to the schools of over 1600 public students and over 1000 private students.

III. ESTIMATOR DEVELOPMENT

If we selected our sample of students from a list of students enrolled in a school, the probability of selection within the school would be straight forward since a student would only be listed once, i.e., (1/enrollment). However, the main goal is to provide data on sample students that are taught by sample teachers in an eligible class in sample schools. This involves several level of sampling to obtain our sample student.

Due to the many levels of sampling, the probability of selection of each student for a sample teacher within a sample school is actually made up of several component probabilities and some random variables.

$$N \cdot P(\textit{class period}) \cdot P(\textit{student within class}) \cdot P(\textit{teacher}) \cdot C \cdot P(\textit{school})$$

Several of the components are straight forward and easy to define. However, several components (those dealing with sampling within the school) turned out to be quite a challenge. The first subsection defines the easier components of the estimator and the following three subsections show the more challenging components.

A. Probability of Selecting the Teacher and the Student Within the Class Period

The probability of selecting the teacher within the school is three out of the total head count of teachers (H) or

$$P(\textit{teacher}) = \left(\frac{3}{H}\right)$$

The probability of selecting the student from the selected class period (l) of teacher (j) is three out of the class size S_{lj} or

$$P(\textit{student within class}) = \left(\frac{3}{S_{lj}}\right)$$

B. Multiplicity of Teachers and Class Periods (N)

The student universe within school is a combination of every list of every class period roster of every eligible class period taught by each eligible teacher in the school during a school week. In schools containing mostly self-contained teachers, such as lower elementary schools, each student's name only appears on one teacher's class period roster. However, in schools containing mostly departmental teachers, such as high schools, each student's name can appear on many class period rosters.

The word multiplicity has come to represent the total number of ways a student can end up in the student component considering all teachers that teach the student and all class periods each teacher has the student. This is equivalent to the number of time the student's name appears on the list if we combined every class roster.

Suppose Student A has four subjects with four teachers and each subject is taught once a day or five times a week. Let us assume that the second period on Thursday was the period used to select the student.

To get the true probability of selection, we would have to obtain all this information to count all the possible ways this student could have been selected.

In the first study, we tried to get an idea of the multiplicity using the following question:

"How many class periods does the student have each week that are taught by .....only 1 teacher? two or more teachers?"

This question did not work well and went unanswered by many of the school administrators. Of course, for our example, the correct answers are twenty for only 1 teacher and zero for two or more teachers.

This particular example of all possible ways of getting Student A is very simple. When we add more teachers, more periods per day, classes that don't meet everyday, and some sort of period rotation, it gets very confusing.

When planning the second study, we debated whether to ask for all the information about a sample student's school week or reduce respondent burden by collecting for each sample student only information about the three sample teachers. It was decided to reduce respondent burden, ask for less information, and concentrate on the sample teachers only. The multiplicity question was reorganized and reworded to ask specifically for the association of the student to each of the sample teachers in the school. Basically, it was broken down into three smaller questions.

1. Does this teacher have this student?
2. Is the student with the teacher all day?
3. If not all day, what subjects does the student have with the teacher and how often does the class meet?

The same set of questions is repeated for each sample student and each sample teacher in the sample school.

A term adopted for use during this study was the "certainty" teacher. The certainty teacher is defined to be the teacher we initially went through to get the sample student. At the very least, we expected to see information for the certainty teacher filled out in the multiplicity question. Any information appearing under the other two teacher names was an added bonus.

You might wonder why we are interested in the other two teachers. We had to determine if the student had a chance of being selected through the other two teacher. If the student has more than one sample teacher then the student's probability of selection is the sum of the probability of selection through the each sample teacher (j).

$$\sum_{j=1}^{3} \left[N_j \cdot P(\textit{class period}_j) \cdot P(\textit{student within class}_{lj}) \cdot P(\textit{teacher})_{lj}\right] \cdot C \cdot P(\textit{school})$$

Most of the time in the feasibility study, the probability of selection through the other two teachers

was zero because they didn't have the student. Occasionally, a student did have more than one sample teacher and twice, the same student was selected for sample through two different sample teachers.

Let us look at the multiplicity for student A again. Suppose by chance, two of this student's teachers were selected for sample. The new question would have given us the following information. Ms. Jane Doe teaches this student English and the class meets five times per week. Mr. John Smith teaches this student Social Studies and the class meets five times per week. Jane Doe became the certainty teacher when we selected student A in her Thursday second period class and as expected, we picked up all five second period classes. The information about John Smith teaching of student A in the five class periods was a welcomed surprise. So the multiplicity or total number of ways student A could be selected through Ms. Doe is five and for Mr. Smith is also five. We also know that we probability of selection for student A will be the sum of the probability of selection through each sample teacher.

Using the multiplicity information as seen in the example, we could estimate a student probability of selection conditioned on selecting the three sample teachers in the school.

C. Probability of Selecting the Sample Class Period

Another component that we had to estimate was the probability of selecting the class period. For self-contained teachers, this probability is one since their one class is in with certainty. For departmental teachers, the double sampling procedure for selecting class period (described in section II) guaranteed an eligible class, but it added some complication to calculating this component. Recall that the procedure involved selecting a set of five class periods for the departmental teachers in a school. For each sample teacher, we determined which class periods contain an eligible class and select one of the eligible classes.

To do this, we had to calculate the probability of selecting at least one eligible class from a set of five class periods and then selecting one of them. From the start we knew that we had to consider all possible combinations of five class periods where T define the total number of class periods in the school week. Initially we came up with:

$$\textit{Initial } P(\textit{class period}) = \binom{T}{5} * \left(\frac{1}{5}\right)$$

Unfortunately, the resulting weights were large implying that the probability was too small. After several more dead ends, it occurred to us that we needed to consider the eligibility of the class period as a success in a series of trails, i.e., the probability of having at least one eligible class out of a possible set of five was a hypergeometric random variable. Actually it is a sum of hypergeometrics since we have to estimate the probability of all possible combinations of sets of five class periods that contained at least one eligible class.

Again, let T be the total periods in the school week. Let L_j define the total number of class periods that teacher (j) taught an eligible class in the school week. Finally let l be the number of eligible periods in the set of five.

The probability of selecting at least one eligible class and choosing one is:

$$\textit{Final } P(\textit{class period}) = \sum_{l=1}^{\min(5,L_j)} \frac{\binom{L_j}{l}\binom{T-L_j}{5-l}}{\binom{T}{5}} * \frac{1}{l}$$

In words this is saying the probability of selection of the class period is equal to the sum of

(the probability of getting one eligible class out of five)

PLUS (the probability of getting two eligible classes out of five and selecting one)

PLUS (the probability of selecting three eligible classes out of five and selecting one)

PLUS (the probability of selecting four eligible classes out of five and selecting one)

PLUS (the probability of selecting five eligible classes out of five and selecting one).

D. Multiplicity of Students (C)

How often can a student's name appear in the set of distinct students taught by a set of three sample teachers over all possible sets of three sample teachers? It depends on how many distinct teachers the student has during the week. This was a second multiplicity problem that we encountered and our final obstacle in a pursuit of an estimator. We didn't have any way calculating this because we didn't ask for the number of teachers the student had in the school week during student sampling. Again, due to the decision to lighten the respondent burden on school administrators, we would have to approximate this component. We felt we could estimate it as an

average across all students by using the following adjustment:

$$C = \left(\frac{X_s}{S_s}\right).$$

where S_s is the number of students in scope for the survey in the school and X_s is one over the sum of all student probability of selection within the school

$$X_s = \frac{1}{\sum_{i=1}^{s}\left[\sum_{j=1}^{3} N_{ij} \cdot \left(\sum_{l=1}^{\min(5,L_j)} \frac{\binom{L_j}{l}\binom{T-L_j}{5-l}}{\binom{T}{5}} \cdot \frac{1}{l}\right) \cdot \frac{3}{S_{lj}} \cdot \frac{3}{H}\right]}$$

One benefit of this ratio adjustment was the joint probability of selection of the three sample teachers cancels out and does not appear in the final weight.

IV. SUMMARY OF RESULTS

We have an approximation of probability of selection for each student which provides an unconditional estimator of student basic weight. This estimator depends heavily on collected data which is open to item nonresponse or response error. The basic weight for sample student i is given by:

$$BW_i = \left[\frac{1}{\sum_{j=1}^{3} N_{ij}\left(\sum_{l=1}^{\min(5,L_j)} \frac{\binom{L_j}{l}\binom{T-L_j}{5-l}}{\binom{T}{5}} \cdot \frac{1}{l}\right)\frac{3}{S_{lj}} \cdot \frac{3}{H}}\right] \cdot \frac{S_s}{X_s} \cdot B$$

Where j is a teacher.

L_j is the total number of class periods taught by teacher j.

l is a class period.

i is the student.

N_{ij} is the number of class periods student i has with teacher j.

T is the total number of class periods in the school.

S_{lj} is the number of students in teacher j's selected class period l.

S_s is the school enrollment.

X_s is one over the sum of student probabilities within school before adjustment.

H is head count of eligible teachers at the school.

V. FUTURE PLANS

Sampling and data collection has been completed for the 1993-94 SASS student component. We used the sampling methodology developed in the research studies to implement the student sampling successfully. The weighing methodology includes the estimator given earlier to generate the basic weights with one additional component as of the publishing of this paper. The component probability of $\left(\frac{l}{L_j}\right)$ has been added to the probability of selecting a class period. This probability covers the chances of selecting the particular set of eligible periods in the initial set of five sample class periods.

Tinkering with the estimator will probably continue until the weighting is run. After the estimation checks currently planned have been completed, more research may be desired.

PROPERTIES OF THE SCHOOLS AND STAFFING SURVEY'S BOOTSTRAP VARIANCE ESTIMATOR

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

Key Words: Simulation, Half-Sample Replication, Bootstrapping, Variances

Introduction

The National Center for Education Statistics' (NCES) Schools and Staffing Survey (SASS) conducted by the Census Bureau has a complex sample design. Public 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. A bootstrap variance estimator (Kaufman,93; sort method 4) has been developed that provides better variance estimates than the balanced half-sample replication (BHR) variance estimator for the public SASS estimates. The bootstrap variance estimator reflects the finite population correction associated with the SASS high sampling rates, without using the joint inclusion probabilities. **A set of bootstrap replicate weights are generated that work like BHR replicate weights, so that the bootstrap variances can be generated from any BHR variance software package.**

The goal of this paper is to provide results from simulation studies, concerning the SASS bootstrap variance estimator ('93 bootstrap variance estimator) described above. The '93 bootstrap variances estimator works well for the public SASS sample design, which uses square root teachers/school as the measure of size. With minor changes in the sample design (using school teacher counts as the measure of size), the school variance estimator can greatly underestimate the variance. However, with some changes, a new bootstrap variance estimator ('94 bootstrap variance estimator) performs better than BHR using the public SASS sample design, when the measure of size is either teacher or square root teacher counts. The '94 bootstrap procedure also performs better than BHR using the private SASS sample design.

First, the public and private sample designs are described, as well as the '94 bootstrap variance estimator. Then, simulation results are presented showing that the '93 bootstrap methodology can underestimate the variance under different PPS sample designs. Simulations also demonstrate that the '94 bootstrap estimator does perform better than BHR with a number of PPS sample designs.

Differences between the Bootstrap Methodologies

The '93 methodology computes school and district bootstraps together. To do this, the bootstrap frame represented both schools and districts. In order to compute the bootstrap weights, all bootstrap-schools within a bootstrap-district must be kept together (see Kaufman,93; weighting section). This restricts the sorting of the bootstrap-schools before the bootstrap sample is selected. It is this restriction that causes the '93 bootstrap estimator to underestimate the school based variance estimates, when a different measure of size is used (see table 1). However, the district variance estimates work well with the '93 methodology for each of the designs in this simulation study, and will not be discussed.

To improve the bootstrap school based variance estimates, the '94 methodology was developed, which ignores the district component of the design. Now, bootstrap-schools can be sorted without regard to the bootstrap-district associated with them. To compute district variance, the '93 methodology is still used.

Design of the Public School and District Samples

The public 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.

Private School Sample Design

The private school survey uses NCES's Private School Survey (PSS) file as the frame. PSS uses a list and area frame design to represent all private schools. The reason for investigating a bootstrap estimator is to find a variance estimator that reflects the finite population correction due to the large sampling rates. Since the sampling rates in the area frame are low,

they will be excluded from this study. Standard methodologies can compute the area frame variances. The list frame is stratified by School Association (19 detailed groups), within Association by Census Region (4 levels), and within Region 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.

Private schools are not associated with school districts, so the private school SASS does not have a district sample.

Weighting

The school weight for school i (W_i) is:

$$W_i = 1/p_i$$

p_i: is the selection probability for school i.

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. Since the SASS half-sample variances are based on 48 replicates, the simulations will be based on 48 half-sample replicates.

The noncertainty school replicate weight is:

$$RW_i = 2/p_i \ .$$

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.

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.

Public and Private 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 frame. From the estimated bootstrap-school frame, B bootstrap samples can be selected. The bootstrap-school frame is generated in the following manner:

For each selected school i, W_i bootstrap-schools (bi) are generated. If W_i has a noninteger component then a full school is generated with a reduced selection probability and weight. As shown in the bootstrap weighting section, the bootstrap expectation of the bootstrap weights (W_{bi}) equals the full-sample weight (W_i). The bi^{th} bootstrap-school has the following measure of size (m_{bi}):

$$m_{bi} = I_{bi} * 1/W_i,$$

$$I_{bi} = \begin{cases} 1 & \text{if bi is an integer component of } W_i \\ C_i & \text{if bi is a noninteger component of } W_i, \\ & \quad C_i \text{ being the noninteger component} \end{cases}$$

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.

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 Efron (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. Sitter's bootstrap sample size (n^*) is the sample size which makes 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-20 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 I_{bi} is not equal to 1. Instead, of using 1, as Sitter does when counting units; I_{bi} is used to calculate N_g^*. 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 I_{bi} does 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.

Determining the Sort Order for the '94 Bootstrap Methodology

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. Bootstrap variances will be computed using a number of sort orderings for each of the simulation samples. A coverage rate is computed for each ordering. The coverage rates are compared with an estimate of the true coverage rate. The ordering associated with the coverage rate closest to the true coverage rate is the ordering that is used for the bootstrap estimator. These comparisons are made at the State level for public estimates and School Association level for private estimates. The bootstrap sort orders are described below.

School Sort Method j

Selected 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-schools generated within each pair of schools are assigned bootstrap-school random numbers. If $n-n^* \leq j$, for a stratum, the bootstrap-schools are sorted by bootstrap-school random number. If $n-n^* > j$, for a stratum, the bootstrap-schools are first sorted by the school pair random number; within each school pair the bootstrap-schools are sorted by the bootstrap-school random number. In other words, if the difference between the original and bootstrap sample sizes is small, as defined by j, then ignore the original sort ordering when randomizing the bootstrap-schools. Otherwise, randomize within pairs that reflect the original sort ordering.

For the public school design with square root teachers as the measure of size, two primary sorts are used (j=1 and 2). The best ordering is then chosen between these sorts. For states that either overestimate or underestimate the coverage rate too much, new sorts are tried. For overestimates sort method j=-1 is used. For underestimate sort method j=3 is used. One state required using sort method j=n. If the coverage rate improves, the new ordering is used in the final variances.

For private schools, sort method j=1 is the primary sort used. If any of the coverage rates are large underestimates then sort method j=n is used.

For the public school design with teachers as the measure of size, sort method j=n is used most often; Sort method j=2 is the next most frequent ordering used. When these two sorts didn't work, sort method 1 or 3 turned out to be the best.

Rationale for School Sort Method j

Sitter shows that if the number of schools in a sampling interval is constant across the intervals, then n^* will be close to n-1. If schools are sorted randomly, then the expected number of schools in the intervals is constant and n^* should be close to n-1. Therefore, if n^*=n-1, the assumption is that the sort ordering is effectively random, so that the school pairing should be ignored. Sort method j=1, sorts bootstrap schools randomly if n^*=n-1. The smaller n^* is relative to n-1, the more effective the ordering is (i.e., the ordering acts less like a random ordering) and the more important the school pairings are to the sort method. Again, this is the affect of sort method j, when j is small.

When the pairings are ignored, a bootstrap-school generated for a particular school is in more sampling intervals and therefore can be selected more often. All other things kept equal, this should increase the bootstrap variance estimate. One then expects the variance from sort method j to be ≥ the variance from sort method k, when j ≥ k. This rule can be used to determine which sort to use to improve the variance estimate. The rule, however, does not always work. This might be due to random error or to the implicit bootstrap-school joint inclusion probabilities that are generated. The coverage rate from a particular sort that matches the true coverage rate is implicitly: 1) matching the effective randomness of the original sort (sort method j=1), adding variability as necessary (sort method j > 1), as well as, 3) matching the bootstrap-school joint inclusion probabilities to the true school joint inclusion probabilities.

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 is randomized 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:

$$W_{bi} = I_{bi} * M_{bi}/p_{bi}$$

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

$$E_*(\sum_{bi} W_{bi}) = \sum_{bi} I_{bi} = \sum_{i} W_i, \text{ as desired.}$$

E_*: is expectation over the bootstrap samples

Since the available data are defined by the schools 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.

Sample Estimate

For each of the simulation samples, totals, averages and ratios are computed within a number of the States for the public designs, and Private school associations for the private design. The variables used are all on the sample frame. 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. 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.

Since 4 of private school association are certainty strata (i.e., all schools classified into these associations are selected into the sample with certainty), only 15 associations will be included in the analysis tables below.

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 estimates are approximately normal then the coverage rates should be close to 0.68.

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 1-10 and 14 presents the coverage rate Bias's.

CV of Variance Estimate (**CV**)

To measure the total error in the variance estimate under the assumption that the variance estimators are almost unbiased (i.e., the sampling rates are low), the coefficient of variation (**CV**) of the variance estimate is calculated.

$$CV = [(1/59) \sum_{t=1}^{60} (V_t - \bar{V})^2]^{1/2}/VT$$

V_t: is the variance estimate for the t^{th} simulation estimate,

$\bar{V}$: is the average variance estimate across the 60 simulation samples.

VT: $[(1/59) \sum_{t=1}^{60} (X_t - \bar{X})^2]$ is an estimate of the true variance,

X_t: is an estimate from the t^{th} simulation sample,

$\bar{X}_t$: is the average of the estimates across the 60 simulation samples,

Table 13 presents the CVs.

Relative MSE of the Variance Estimate (MSE)

To measure the total error in the variance estimates, the relative mean square error (MSE) of the variance estimates is calculated.

$$MSE = [(1/59) \sum_{t=1}^{60} (V_t - \bar{V})^2 + (\bar{V}\text{-}VT)^2]^{1/2}/VT$$

For the public designs, table 11 presents the MSE of the variance estimates averaged across the States included in the study. For the private design, table 12 presents the MSE of the variance estimates averaged across the Associations.

Results based on Bias in the Coverage Rates

Table 1 shows how the '93 bootstrap methodology underestimates the variance when teachers/school is used as the measure size. 28 percent of the time, the variance for averages (**AVE**) has a very large negative bias (**BIAS** LT -.14). The variance for totals (**TOTAL**) has a very large negative bias 32 percent of the time. These are unacceptable rates; and even though the '93 bootstrap estimator works, when the measure of size is the square root of teachers/school, it does not work in a more general setting.

The '94 bootstrap variance estimator (**94 BOOT**) works much better than the '93 bootstrap estimator for a number of sample designs (public SASS design, private SASS design and public SASS design using teachers/school as the measure of size). It also works better than BHR, even when simple finite populations correction adjustments (FPC) are applied to the BHR variance estimates. The results are discussed below for each design.

SASS Public School Design (Tables 2-4)

For school averages, 52 percent of the '94 bootstrap variance estimates have a small bias (**BIAS** between -.07 to .07). BHR without any FPC adjustments (**BHR No FPC**) only has 20 percent of the variance estimates in this category. If simple FPC adjustments are applied to **BHR No FPC** the percentage increases to 48 and 44 percent for **BHR Prop FPC** and **BHR SRS FPC**, respectively. The bootstrap estimator has only one state (4 percent) which has a very large overestimate (**BIAS GE .14**), while **BHR No FPC** has 44 percent in this category. Applying simple FPC adjustment helps, but there are still a reasonable number of states with large overestimates. For the bootstrap estimator, no states have very large underestimates (**BIAS LT -.14**), while each BHR estimator has 8 percent in the very large underestimate category.

The results for school totals are similar to school averages discussed above. 56 percent of the '94 bootstrap variances are in the small bias category, while **BHR No FPC** has only 32 percent in this category. Applying an FPC helps, but the FPC adjusted BHR estimators, have only 36 percent in this category. The bootstrap estimator has 12 percent of the estimates in the very large bias category, while **BHR No FPC** has 40 percent in this category. An FPC adjustment reduces the cases to 24 percent. The bootstrap estimator has no states with very large underestimate. **BHR No FPC** likewise has no cases in this category, but the FPC adjusted variances each have 8 percent of the states in this category.

For ratio estimates, the 94 bootstrap and FPC adjusted BHR variances work well. The only problem with the **BHR No FPC** variances is that 24 percent of the states are in the very large overestimate category.

SASS Private School Design (Tables 5-7)

For school averages, 63 percent of the '94 bootstrap variance estimates have a small bias (**BIAS** between -.07 to .07). BHR without any FPC adjustments (**BHR No FPC**) only has 47 percent of the variance estimates in this category. If simple FPC adjustments are applied to **BHR No FPC** the percentage increases to 53 percent for both **BHR Prop FPC** and **BHR SRS FPC**. 11 percent of the bootstrap estimates are very large overestimates (**BIAS GE .14**), while **BHR No FPC** has 26 percent in this category. Applying simple FPC adjustment helps, but there are still a reasonable number of associations with large overestimates. For the bootstrap estimator, one association (5 percent) has a very large underestimate (**BIAS LT -.14**), while none of the BHR estimators have any associations in this category.

The results for school totals are similar to school averages discussed above. 74 percent of the 94 bootstrap variances are in the small bias category, while **BHR No FPC** has only 32 percent in this category. Applying an FPC helps with 63 and 58 percent being in this category for **BHR Prob FPC** and **BHR SRS FPC**, respectively. The bootstrap estimator has 11 percent of the estimates in the very large bias category, while **BHR No FPC** has 26 percent in this category. An FPC adjustment reduces the cases to 16 and 21 percent for **BHR Prob FPC** and **BHR SRS FPC**, respectively. Neither the bootstrap nor BHR estimators have any variances in the very large

underestimate category.

For ratio estimates, the 94 bootstrap and FPC adjusted BHR variances work well. The only problem with the **BHR No FPC** variances is that 21 percent of the variances are in the very large overestimate category.

SASS Public School Design - Measure of Size, Teachers (Tables 8-10)

Overall, the 94 Bootstrap variances are better than the BHR variances. However, the differences are not as great with this design. For averages, 76 percent of the bootstrap variances are in the small bias category; **BHR no FPC**, **BHR prob FPC** and **BHR SRS FPC** have 68, 64 and 68 percent in this category, respectively. None of the methodologies have very large overestimates, while only the FPC adjusted BHR estimates have a few very large underestimates (8 and 12 percent).

For totals, 76 percent of the 94 bootstrap variances are in the small bias category. **BHR no FPC, BHR prob FPC** and **BHR SRS FPC** have 64, 80 and 76 percent in this category, respectively. None of the methodologies have very large overestimates, while all the methodologies have a few very large underestimates (4 to 8 percent).

For ratios, all the methodologies, except **BHR no FPC**, work equally well. They all have between 52 and 56 percent in the small bias category; except **BHR no FPC**, which has only 44 percent in the small bias category. All methods have some, but minimal cases in the very large underestimate category (4 to 8 percent); and they all have substantial cases in the very large overestimate category (16 to 28 percent).

Results based on Coverage Rates of National Estimate (Table 14)

Instead of analyzing the coverage rate bias distributions by state or association, another perspective is analyzing coverage rate biases for national estimates. Since the simulations are done by a series of different sets of states, the only national estimates that can be computed are totals. The national coverage rate biases are provided in table 14. The table shows that the bootstrap biases are all less than 1 percent. The **BHR no FPC** biases vary, but are all much larger then the bootstrap bias. They range for 5.6 to 11.7 percent, depending on the type of design . The FPC adjusted BHR biases are slightly smaller than the **BHR no FPC** biases. They range from 3.3 to 7.3 percent, depending on the design.

Results Based on Relative MSE and CV of the Variance

The MSE and CV of the variance require measuring the variance of the variance, as well as the squared bias of the variance. Because they are based on only 60 simulations, these estimate may not be very stable. The coverage rate analysis should be more stable. The MSE and CVs are presented because they provide a slightly different perspective, and provide some insight into the performance of the bootstrap variance estimator when the sampling rate are not high and the finite population correction can be assumed negligible.

Results Based on the Relative Mean Square Error (Tables 11-12)

Tables 11 and 12 analyze the variance estimators with respect to their relative mean square error (MSE). MSEs are computed for each state estimate. In the tables, they are averaged two ways. One method (State or Association MSE), averages the MSEs within type of estimate (averages, ratios and totals). This method gives an overall measure of the error where each state's error is equally weighted. Another method (**All MSE**) of obtaining an overall measure of error is to sum the state variances within each total estimate, obtaining the variance of the total for all states in the analysis. The MSE of this variance can be computed using the 60 simulation estimates. These MSEs can then be averaged within the total estimate type. This error measurement gives states contributing more to the total estimate larger weight in the error measurement. A similar **All MSE** measurement can be made for the private design, using the association estimates.

The 94 bootstrap MSEs are always smaller than the **BHR no FPC** MSEs using either MSE measurement method. With respect to the **All MSE method**, the bootstrap MSEs are always the smallest. For public schools, the bootstrap MSEs are roughly comparable to the **BHR SRS FPC** and **BHR Prob FPC** MSEs when using the State error measurement method. For private schools, the Bootstrap MSEs are always smaller than any of the BHR MSEs using the Association error measurement method.

Part of the reason for these results is that the BHR replicates are not fully balanced. If they were, the BHR MSE of the variance would be smaller because the BHR variance of the variance would be smaller.

Results Based on CVs (Table 13)

The results stated above show that for the sample designs in this study, where state or associations are heavily sampled, the 94 bootstrap variance estimator does better than the BHR methods. Another question

that can be asked is whether the Bootstrap variances are better than BHR variances, when the sampling rates are small. The results presented here cannot answer this question. However, with some assumptions, one can see if its worth doing another simulation to address this question. There are two assumptions required: 1) If the sampling rates are small then all the variance estimators are unbiased; and 2) the smaller sampling rates are obtained by reducing each stratum's sample size by the same constant. If these assumptions are true then the CVs from this analysis should provide some insight into this question. The CVs are provided in table 13.

For public schools using square root teachers as the measure of size, the bootstrap and BHR CVs are about the same. For the other two designs, the bootstrap CVs are smaller than the BHR CVs. This seems to indicate that the bootstrap estimator may perform well even if the sampling rates are low. Part of the reason why this might be true is that the BHR replicate are only partially balanced.

References

Efron, Bradley(1982). The Jackknife, the Bootstrap and Other Resampling Plans. SIAM No. 38, p62.

Kaufman, Steven(1993). A Bootstrap Variance Estimator for the Schools and Staffing Survey. ASA 1993 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 - State Distribution of the Coverage Rate Bias in 93 Bootstrap Standard Errors for School Averages using Number of Teachers/School as the Measure of Size

Bias Col Pct	AVE	RATIO	TOTAL
LT -.14	28.00	8.00	32.00
[-.14, -.07)	28.00	16.00	28.00
[-.07, 0.0)	24.00	40.00	28.00
[0.0, .07)	8.00	16.00	4.00
[.07, .14)	8.00	12.00	8.00
GE .14	4.00	8.00	0.00

Table 2 - State Distribution of the Coverage Rate Bias in the 94 Bootstrap Estimator for the SASS Public Design Estimating School Averages

Bias		BHR Estimators		
Col Pct	94 BOOT	Prob FPC	SRS FPC	No FPC
LT -.14	0.00	8.00	8.00	8.00
[-.14, -.07)	4.00	0.00	0.00	0.00
[-.07, 0.0)	20.00	16.00	16.00	4.00
[0.0, .07)	32.00	32.00	28.00	16.00
[.07, .14)	40.00	32.00	28.00	28.00
GE .14	4.00	12.00	20.00	44.00

Table 3 - State Distribution of the Coverage Rate Bias in the 94 Bootstrap Estimator for the SASS Public Design Estimating School Totals

Bias		BHR Estimators		
Col Pct	94 BOOT	Prob FPC	SRS FPC	No FPC
LT -.14	0.00	8.00	8.00	0.00
[-.14, -.07)	4.00	0.00	0.00	4.00
[-.07, 0.0)	20.00	12.00	8.00	8.00
[0.0, .07)	36.00	24.00	28.00	24.00
[.07, .14)	28.00	32.00	32.00	24.00
GE .14	12.00	24.00	24.00	40.00

Table 4 - State Distribution of the Coverage Rate Bias in the 94 Bootstrap Estimator for the SASS Public Design Estimating School Ratios

Bias Col Pct	94 BOOT	BHR Estimators Prob FPC	SRS FPC	No FPC
LT -.14	4.00	0.00	0.00	0.00
[-.14, -.07)	28.00	16.00	16.00	4.00
[-.07, 0.0)	16.00	32.00	32.00	12.00
[0.0, .07)	44.00	20.00	20.00	40.00
[.07, .14)	4.00	24.00	24.00	20.00
GE .14	4.00	8.00	8.00	24.00

Table 5 - Assoc. Distribution of the Coverage Rate Bias in 94 Bootstrap Estimator for the SASS Private Design Estimating School Averages

Bias Col Pct	94 BOOT	BHR Estimators Prob FPC	SRS FPC	No FPC
LT -0.14	5.26	0.00	0.00	0.00
[-.14, -.07)	15.79	15.79	15.79	0.00
[-0.07, 0.0)	15.79	10.53	5.26	5.26
[0.0, .07)	47.37	42.11	47.37	42.11
[.07, .14)	5.26	15.79	15.79	26.32
GE .14	10.53	15.79	15.79	26.32

Table 6 - Assoc. Distribution of the Coverage Rate Bias in 94 Bootstrap Estimator for the SASS Private Design Estimating School Totals

Bias Col Pct	94 BOOT	BHR Estimators Prob FPC	SRS FPC	No FPC
[-.14, -.07)	10.53	0.00	0.00	0.00
[-0.07, 0.0)	26.32	10.53	5.26	0.00
[0.0, .07)	47.37	52.63	52.63	31.58
[.07, .14)	5.26	21.05	21.05	42.11
GE .14	10.53	15.79	21.05	26.32

Table 7 - Assoc. Distribution of the Coverage Rate Bias in 94 Bootstrap Estimator for the SASS Private Design Estimating School Ratios

Bias Col Pct	94 BOOT	BHR Estimators Prob FPC	SRS FPC	No FPC
LT -0.14	10.53	10.53	5.26	5.26
[-.14, -.07)	5.26	5.26	10.53	5.26
[-0.07, 0.0)	26.32	15.79	15.79	10.53
[0.0, .07)	42.11	42.11	36.84	42.11
[.07, .14)	5.26	10.53	15.79	15.79
GE .14	10.53	15.79	15.79	21.05

Table 8 - State Distribution of the Coverage Rate Bias in 94 Bootstrap Estimator for the SASS Public Design using Teachers/School as the Measure of Size Estimating School Averages

Bias Col Pct	94 BOOT	BHR Estimators Prob FPC	SRS FPC	No FPC
LT -.14	0.00	12.00	8.00	0.00
[-.14, -.07)	20.00	24.00	16.00	12.00
[-.07, 0.0)	48.00	44.00	44.00	28.00
[0.0, .07)	28.00	20.00	24.00	40.00
[.07, .14)	4.00	0.00	8.00	20.00

Table 9 - State Distribution of the Coverage Rate Bias in 94 Bootstrap Estimator for the SASS Public Design using Teachers/School as the Measure of Size Estimating School Totals

Bias Col Pct	94 BOOT	BHR Estimators Prob FPC	SRS FPC	No FPC
LT -.14	4.00	8.00	4.00	4.00
[-.14, -.07)	16.00	8.00	8.00	0.00
[-.07, 0.0)	44.00	56.00	36.00	20.00
[0.0, .07)	32.00	24.00	40.00	44.00
[.07, .14)	4.00	4.00	12.00	32.00

Table 10 - State Distribution of the Coverage Rate Bias in 94 Bootstrap Estimator for the SASS Public Design using Teachers/School as the Measure of Size Estimating School Ratios

Bias Col Pct	94 BOOT	BHR Estimators Prob FPC	SRS FPC	No FPC
LT -.14	4.00	8.00	4.00	4.00
[-.14, -.07)	4.00	8.00	4.00	0.00
[-.07, 0.0)	28.00	16.00	24.00	0.00
[0.0, .07)	28.00	36.00	28.00	44.00
[.07, .14)	16.00	16.00	20.00	24.00
GE .14	20.00	16.00	20.00	28.00

Table 11 -- Relative MSE of the Variance (**MSE**) by Type of Public Sample Design and Type of Variance Estimator

	Type of SASS Public School Sample Design							
	Measure of Size Square Root Teachers				Measure of Size Teachers			
MSE Type of Estimator	State[1] AVE	RATIO	TOTAL	All[2] TOTAL	State[1] AVE	RATIO	TOTAL	All[2] TOTAL
B No FPC	0.91	0.75	1.04	0.46	1.07	1.15	1.51	0.97
H SRS FPC	0.65	0.56	0.75	0.36	0.86	0.86	1.17	0.84
R PROB FPC	0.63	0.53	0.72	0.32	0.81	0.80	1.09	0.78
94 Bootstrap	0.66	0.56	0.81	0.24	0.85	0.96	1.07	0.47

Table 12 -- Relative MSE of the Variance (**MSE**) by Private Sample Design and Type of Variance Estimator

MSE Type of Estimator	Association[1] AVE	RATIO	TOTAL	All[2] TOTAL
B No FPC	1.48	1.86	2.46	0.99
H SRS FPC	0.78	0.99	1.25	0.52
R PROB FPC	0.73	0.91	1.17	0.50
94 Bootstrap	0.71	0.83	0.74	0.19

Table 13 -- CV of the Variance (CV) by Type of Sample Design and Type of Variance Estimator

Type of Design	Type of Estimator	State[1] AVE	State[1] RATIO	State[1] TOTAL	All[2] TOTAL
Public School SASS Measure of Size Square Root Teachers	BHR No FPC	0.47	0.48	0.50	0.20
	94 Bootstrap	0.48	0.43	0.54	0.19
Public School SASS Measure of Size Teachers	BHR No FPC	0.99	0.73	1.24	0.88
	94 Bootstrap	0.78	0.73	0.95	0.45
Private School SASS Measure of Size Square Root Teachers	BHR No FPC	0.73	1.02	1.08	0.27
	94 Bootstrap	0.47	0.65	0.52	0.18

Table 14 -- Coverage Rate Bias for National Estimates[2] of Totals by Type of Design and Type of Variance Estimator

Percent Type of Design	94 BOOT	BHR Prob FPC	BHR SRS FPC	BHR No FPC
Public School SASS Measure of Size Square Root Teachers	0.6	3.3	3.9	5.6
Public School SASS Measure of Size Teachers	-0.7	5.0	5.6	8.3
Private School SASS Measure of Size Square Root Teachers	0.6	7.3	7.3	11.7

1. These are the average of the state or association estimates

2. These are based on summing the state or association variances to obtain a total variance for all states or associations

OPTIMAL PERIODICITY OF A SURVEY: SAMPLING ERROR, DATA DETERIORATION, AND COST

Dhiren Ghosh, Synectics for Management Decisions, Steven F. Kaufman, National Center for Education Statistics, Wray Smith and Michael Chang, Synectics for Management Decisions
Dhiren Ghosh, Synectics Mgmt Decisions, 3030 Clarendon Blvd., Suite 305, Arlington, VA 22201

Key Words: Probable error, Loss function, ARIMA models, Repeated Surveys

Government agencies collect many different kinds of statistical data through sample surveys conducted on a periodic basis (monthly, annually, or at multi-year intervals). When the periodicity is not mandated by law, data deterioration, cost, and sampling error in the data may be considered jointly to determine optimum intersurvey time intervals. In a decision-making process, any loss due to using the survey estimate instead of the true value may be thought of as arising in part from sampling error; also, with the passage of time, the true value evolves and the survey dataset becomes obsolete. In this paper several statistical models of data deterioration are considered jointly with standard cost functions for a survey; that is, "cost-and-error models."

The concept of "probable error" is utilized in three related models in which the additivity of errors over time is assumed. A loss function is minimized in a fourth model along with a procedure for estimating the loss parameter. A fifth model assumes that there is an underlying stochastic process that is observed periodically by the repeated survey data collections and that this process can be modeled as an ARIMA(0,1,1) time series process observed with sampling error. The formulation of this model is based on a general modeling procedure set forth in Smith (1980) and Smith and Barzily (1982) using Kalman filter concepts. The use of the first three models as decision aids in the choice of optimum intersurvey intervals is illustrated with data from the Schools and Staffing Survey (SASS).

We assume that data users will continue to use the data obtained from the most recent survey until a new survey is undertaken and the newly collected data are processed and released to data users. Thus, if the inter-survey period is long, "deterioration" of the data, if it is of considerable magnitude, could affect the quality of decisions made by users. On the other hand, if the survey is undertaken too frequently, the costs of conducting the survey and analyzing the data and the response burden may be judged to outweigh the benefits to be achieved in using fresh data. Typical analyses of cost-benefit tradeoffs tend to focus on the best use of a fixed resource amount over a time period that would include two or more survey data collections.

The usual cost model for a sample survey assumes a start-up cost, C_0 , and a per unit (ultimate sample unit) cost, C_1 . Thus, the total cost is represented as $C = C_0 + n\, C_1$. However, the start-up cost may be dependent on the periodicity. We represent it as C_0^k (where k is the periodicity) which may be regarded as increasing with increasing periodicity; i.e., the start-up cost is more if the periodicity is 3 years compared to the start-up cost if the periodicity is 2 years, and so on. On the other hand, the start-up cost may be considered to be constant; i.e., it may not depend on the periodicity of the survey.

In the family of statistical models that we develop below, we assume that the total resources are fixed. The different possible periodicities spend this total resources in different ways. This assumption then determines the possible sample sizes every time the survey is undertaken corresponding to different periodicities. Thus, if we are comparing two possible periodicities, say two years as against three years, we consider a six-year cycle (the least common multiple of the two periodicity numbers). In the six-year cycle, a survey with periodicity two years will be conducted three times while a survey with periodicity three years will be conducted only twice. If C_0^k and C_1 (where C_1 is assumed to be independent of the periodicity of the survey.) are known (whether the start-up cost is constant or increasing) we can calculate the possible sample sizes for these two alternatives where the total measure C is also known.

A Family of Error Models

We assume that the true value of a variable of interest remains constant for a year after the survey date. So the error "committed" in using the survey estimate is exactly equal to the difference between the survey estimate and the true value. So during one year from the survey date any user incurs an error which equals the difference between the true value and the survey estimate. The estimate of the standard error from the survey provides an indication of this difference. The survey estimate is normally distributed around the true value with a standard deviation which is the standard error of the estimate. The difference between the true value and the survey estimate is the deviation from the mean in the normal distribution of

the survey estimates considered as random variables. The average of these deviations is called the probable error. It is calculated as follows for any normal distribution:

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}|x-m|e^{-\frac{(x-m)^2}{2\sigma^2}}dx = \sqrt{\frac{2}{\pi}}\sigma = 0.8\sigma$$

Thus the average error incurred by any user during the first year after the survey is equal $0.8\sigma/\sqrt{n}$ where $\sigma/\sqrt{n}$ is the standard error of the estimate. At the end of one year, we assume that the true value undergoes a change denoted by D_1. So the expected value of the total error committed by all the users is the sum of the probable error and D_1. Proceeding in the same manner we denote the change in the second year as D_2 and so on.

In **Model 1** we ignore the direction of the change in the true value and just add the probable error to the sampling error for the change in the true value.

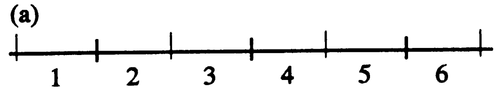
In **Model 2** we do not ignore the direction of the change. If the change occurs in the same direction as the survey estimate, we ignore the diminution in the shift due to the survey estimate already being in the same direction. If the shift occurs in the opposite direction the total error due to using the old survey estimate can be denoted as D_1 + probable error. Taking the average of the two possibilities we denote the expected error as D_1 + ½(probable error). Here the error terms D_1 and D_2 are treated as if they were random variables. Proceeding in the same manner we denote the change in the third year as D_3 and calculate the expected error as above.

In **Model 3** we add the square of the change to sampling error to denote the total error after the first year. We further assume that the change is normally distributed so the sum of the sampling error and the change is also normally distributed. This enables us to calculate the probable error of the normal distribution.

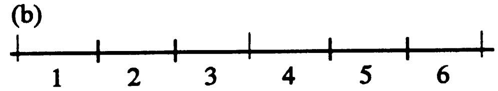
Determination of Periodicity of a Survey

We start with the assumption that the total resources are fixed and the problem is to determine the best periodicity of a survey. We illustrate the solution of this problem for the special case when the alternatives are: (a) every two years (biennial), or (b) every three years (triennial). We consider a cycle of six years with the survey taken at the starting point.

(a)



(b)



For a six year cycle, the biennial survey is conducted three times and the triennial survey is conducted twice. We do not take into account the survey after six years since a new cycle starts after the sixth year. We further assume that the true unobserved value remains unchanged for a year after the survey is completed. At the end of a year, the value changes by an amount D_1 and at the end of two years, the value changes again by an amount D_2. These D_1 and D_2 denote the shift in the true values. If the standard error of a variable in a survey (assuming SRS) is $\sigma/(n^{½})$ where σ is the standard deviation and n is the sample size, the average error or probable error of the estimate is $0.8{*}\sigma/(n^{½})$. That is, every time the estimated value is used (since the true value is unknown) an *error is committed*; the expected value of this error is $0.8\sigma/(n^{½})$. During the year after the survey, the survey value will be used for any decision, so the average error committed during the year is $0.8\sigma/(n^{½})$. When a year elapses the shift in the true value is added to the expected error to obtain the expected error committed during the second year and so on.

Let us examine the error committed for every year following the survey. These errors over the years are assumed to be additive. Let n_a and n_b be the sample sizes for the biennial and the triennial surveys respectively with simple random sampling. We further assume that the standard deviation in the population for the variable of interest remains unchanged during the whole cycle.

Model 1.

(a)

Year (Ordinal)	Average Error Committed
1	$0.8\sigma/(n_a^{1/2})$
2	$D_1 + 0.8\sigma/(n_a^{1/2})$
3	$0.8\sigma/(n_a^{1/2})$
4	$D_1 + 0.8\sigma/(n_a^{1/2})$
5	$0.8\sigma/(n_a^{1/2})$
6	$D_1 + 0.8\sigma/(n_a^{1/2})$
Average Total Error Committed (in six years)	$3D_1 + 4.8\sigma/(n_a^{1/2})$

(b)

Year (Ordinal)	Average Error Committed
1	$0.8\sigma/(n_b^{1/2})$
2	$D_1 + 0.8\sigma/(n_b^{1/2})$
3	$D_1 + D_2 + 0.8\sigma/(n_b^{1/2})$
4	$0.8\sigma/(n_b^{1/2})$
5	$D_1 + 0.8\sigma/(n_b^{1/2})$
6	$D_1 + D_2 + 0.8\sigma/(n_b^{1/2})$
Average Total Error Committed (in six years)	$4D_1 + 2D_2 + 4.8\sigma/(n_b^{1/2})$

Thus (a) is preferable if

$$3D_1 + 4.8\sigma/(n_a^{1/2}) < 4D_1 + 2D_2 + 4.8\sigma/(n_b^{1/2})$$

or

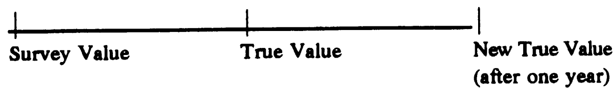
$$4.8\sigma[(n_a^{1/2}) - (n_b^{1/2})] < D_1 + 2D_2$$

and (b) is preferable if

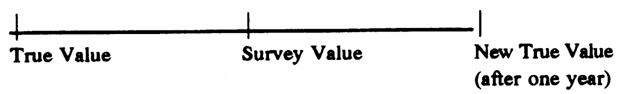
$$4.8\sigma[(n_a^{1/2}) - (n_b^{1/2})] > D_1 + 2D_2$$

Model 2

In Model 1, we assumed that the expected error and the shift in the value are additive for estimating the error in the second or the third year. Examine the following hypothetical case: In this case the addition of the errors seems reasonable.



Alternatively, examine the following case: In such a case, the average error in using the survey value after a year is definitely not $D_1 + 0.8\sigma/(n_b^{1/2})$, it is $D_1 - 0.8\sigma/(n_b^{1/2})$.



If we ignore this contribution of the survey error toward a diminution of the effect of the shift in the true value, the estimate of the average error committed after the first year is $D_1 + 0.4\sigma/(n_b^{1/2})$, and so on. So the errors look as follows:

Year (Ordinal)	(a)	(b)
1	$0.8\sigma/(n_a^{1/2})$	$0.8\sigma/(n_b^{1/2})$
2	$D_1 + 0.4\sigma/(n_a^{1/2})$	$D_1 + 0.4\sigma/(n_b^{1/2})$
3	$0.8\sigma/(n_a^{1/2})$	$D_1 + D_2 + 0.4\sigma/(n_b^{1/2})$
4	$D_1 + 0.4\sigma/(n_a^{1/2})$	$0.8\sigma/(n_b^{1/2})$
5	$0.8\sigma/(n_a^{1/2})$	$D_1 + 0.4\sigma/(n_b^{1/2})$
6	$D_1 + 0.4\sigma/(n_a^{1/2})$	$D_1 + D_2 + 0.4\sigma/(n_b^{1/2})$
Average Total Error Committed (in six years)	$3D_1 + 3.6\sigma/(n_a^{1/2})$	$4D_1 + 2D_2 + 3.2\sigma/(n_b^{1/2})$

Thus (a) is preferable if

$$3.6\sigma/(n_a^{1/2}) - 3.2\sigma/(n_b^{1/2}) < D_1 + 2D_2$$

and (b) is preferable if

$$3.6\sigma/(n_a^{1/2}) - 3.2\sigma/(n_b^{1/2}) > D_1 + 2D_2$$

Model 3

Let us assume that x_j is the value for the j^{th} year and

$$x_{j+1} - x_j = d_j$$

Let the variance of d_j's over the years be $D^2(1)$. For a Random Walk stochastic process, the d_j's are not normally distributed. Similarly, let $D^2(2)$ be the variance of differences over 2 years. For a Random Walk process, $D^2(2) = 2D^2(1)$. But, in general, this relation may not hold because of the autocorrelation of the changes between consecutive years. In general, $D^2(2)$ or $D^2(1)$ is not normally distributed. Never the less, we assume that the probable error from this process is 0.8D(1) or 0.8D(2), as in the case of normal distribution. Under the assumptions, the error looks as follows:

Year (Ordinal)	(a)	(b)
1	$0.8\sigma/(n_a^{1/2})$	$0.8\sigma/(n_b^{1/2})$
2	$0.8D(1) + 0.8\sigma/(n_a^{1/2})$	$0.8D(1) + 0.8\sigma/(n_b^{1/2})$
3	$0.8\sigma/(n_a^{1/2})$	$0.8D(2) + 0.8\sigma/(n_b^{1/2})$
4	$0.8D(1) + 0.8\sigma/(n_a^{1/2})$	$0.8\sigma/(n_b^{1/2})$
5	$0.8\sigma/(n_a^{1/2})$	$0.8D(1) + 0.8\sigma/(n_b^{1/2})$
6	$0.8D(1) + 0.8\sigma/(n_a^{1/2})$	$0.8D(2) + 0.8\sigma/(n_b^{1/2})$
Average Total Error Committed (in six years)	$2.4D(1) + 4.8\sigma/(n_a^{1/2})$	$1.6D(1) + 1.6D(2) + 4.8\sigma/(n_b^{1/2})$

Thus (a) is preferable if

$$4.8[\sigma/(n_a^{1/2}) - \sigma/(n_b^{1/2})] < 1.6D(2) - 0.8D(1)$$

and (b) is preferable if

$$4.8[\sigma/(n_a^{1/2}) - \sigma/(n_b^{1/2})] > 1.6D(2) - 0.8D(1).$$

Model 4

In Model 4 we introduce the concept of a loss parameter that converts the error whether sampling error alone is coupled with the shift over time. This converts the error into loss expressed as monetary units. The sum of average cost and average error over a period of years is minimized to determine the optimum periodicity. We present below the operation of each of these four models.

Let X_k be the true value of variable in the k^{th} year and

$\hat{X}_K$ *be the survey value*

$$\hat{X}_K = X_K + e_K,\ E(e_K) = 0$$

$$\begin{aligned} &E(\hat{X}_k - X_{k+T-1})^2 \\ &= E(\hat{X}_k - X_K + X_K - X_{K+1} + X_{K+1} - \dots - X_{K+T-1}) \\ &= E(e_K + (T-1)d)^2, \text{ under the Random Walk Model} \\ &= E(e_K^2) + E((T-1)d^2) \\ &= E(e_K^2) + (T-1)E(d^2) \\ &= V(e_K) + (T-1)E(d^2) \end{aligned}$$

If $\hat{X}_b$ *and* $\hat{X}_{b+p}$ *are two survey estimates p years apart, let*

$$M = \frac{(\hat{X}_b - \hat{X}_{b+p})^2}{b},\ M \text{ is an estimate of } E(d^2)$$

The total error in T years is the following:

Year (Ordinal)	Error
1	$S^2/n + 0 \cdot M$
2	$S^2/n + 1 \cdot M$
⋮	⋮
T	$S^2/n + (T-1)M$
Total Error (in T years)	$T(S^2/n) + \frac{1}{2}T(T-1)M$
Average Error Per Year (in a cycle of T years)	$S^2/n + \frac{1}{2}(T-1)M$

Let α be a weighting factor that converts error into cost or loss. Then

$$\text{Average Cost Per Year} = J = \frac{C_0 + nC_1}{T} + \alpha\left(\frac{S^2}{n} + \frac{T-1}{2}M\right)$$

$$\frac{\partial J}{\partial n} = \frac{C_1}{T} - \alpha\frac{S^2}{n^2} = 0, \text{ this gives } n = \sqrt{\frac{\alpha S^2 T}{C_1}}$$

$$\text{Average Cost} = J = C_0 + \frac{2\sqrt{C_1\alpha}S}{\sqrt{T}} + \alpha\frac{T-1}{2}M, \text{ for } T = 1, 2, 3, \ldots$$

The optimum T is the one for which the average cost is the minimum.

Model 5

In the above four models we have not assumed any underlying stochastic process for the variables that are measured in the surveys. In Model 5 we assume that the underlying process is consistent with an ARIMA (0,1,1) time series model. Consequently data users would be using a minimum mean square error forecast from the past data instead of the data of the last survey after the lapse of one or more intersurvey time intervals .

In this setup, let $e_k(j)$ be the j-step ahead forecast error based on data through time k. The mean square error is

$$E\left(e^2_{k-T}(T)\right) = M_{k-T}(0) + T{\cdot}E(d^2)$$

where $M_{k-t}(0)$ is the mean square error of the state estimate at the time k-T based on all data through time k-T.

If we assume that the survey system is in a steady state in the sense that

$$M_k(0) = M_{k+T}(0) = M$$

as a result of conducting surveys of constant sample size n every T periods. It can be shown from standard time series analysis techniques that

$$M = \left[\frac{T{\cdot}E(d^2)}{2}\right]\left[-1 + \frac{(1+4S^2)}{T{\cdot}n{\cdot}E(d^2)}\right]^{\frac{1}{2}}$$

We define the average cost per year as in Model 4

$$J = \frac{C_0+C_1 n}{T} + \frac{\alpha}{T}\sum_{j=0}^{T-1} M_k(j)$$

where $M_k(j)$ is the j-step ahead mean square error.

$$= \frac{C_0+C_1 n}{T} + \frac{\alpha}{T}\sum_{j=0}^{T-1}(M + j{\cdot}E(d^2))$$

$$= \frac{1}{T}[C_0 + C_1 n] + \alpha\left[-\frac{E(d^2)}{2} + E(d^2)\left[\frac{T{\cdot}S^2}{n{\cdot}E(d^2)} + \frac{T^2}{4}\right]^{\frac{1}{2}}\right]$$

Average cost J as a function of n and T can be minimized by solving formula for each T in a specified allowable set T={1, 2 ··· T_{max}} and adopting the n, T) for which J is minimized.

A Note on the Determination of α, the Weighting Factor

One procedure is to assign a value for α strictly based on judgment. If we want to develop a more sophisticated approach for determining a value for α we may argue as follows:

If $C_0 + C_1 n$ is the cost of implementing a survey and it results in sampling error of S^2/n for one variable, the total cost is

$$C_0 + C_1 n + \alpha\frac{S^2}{n}$$

Differentiating with respect to n and equating to zero, we get:

$$C_1 - \alpha\frac{S^2}{n^2} = 0$$

or

$$n = \sqrt{\alpha\frac{S^2}{C_1}} = \sqrt{\frac{\alpha}{C_1}}S, \text{ thus } \alpha = \frac{n^2 C_1}{S^2}$$

We note that the marginal gain from increasing the sample size from n to n+1 is

$S^2/n - \alpha S^2/(n+1)$. The sample size is optimum when the marginal cost equals marginal gain.

$$or\ C_1 = \alpha\frac{S^2}{n} - \alpha\frac{S^2}{n+1}$$

$$or\ C_1 = \alpha S^2\left(\frac{1}{n(n+1)}\right)$$

$$or\ n^2 + n - \alpha\frac{S^2}{C_1} = 0$$

$$or\ n = \frac{-1 + \sqrt{1 + \frac{4\alpha S^2}{C_1}}}{2},\ \textit{disregarding the other root}$$

$$or\ (2n+1)^2 = 1 + \frac{4\alpha S^2}{C_1}$$

$$or\ \frac{C_1(2n+1)^2 - 1}{4S^2} = \alpha$$

It can be seen that the two values of α are close to each other. If we look at the sample sizes employed in previous surveys and construct the cost function, we can get a value for α that has an objective basis.

Conclusion

These models have provided a direct approximate method for characterizing the decision problem of making a joint choice of inter-survey intervals and sample sizes under a fixed cost constraint.

References

Binder, D.A. and Dick, J.P. (1989), "Modeling and Estimation for Repeated Surveys," *Survey Methodology*, 15, 29-45.

Binder, D.A. and Hidiroglou, M.A. (1988), "Sampling in Time," in *Handbook of Statistics, Vol. 6*, ed. P.R. Krishnaiah and C.R. Rao, Amsterdam: Elsevier Science Publishers, 187-211.

Blight, B.J.N. and Scott, A.J. (1973), "A Stochastic Model for Repeated Surveys," *Journal of the Royal Statistical Society*, Ser. B, 35, 61-68.

Box, G.E.P. and Jenkins, G.M. (1970), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.

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

Harrison, P.J. (1967), "Exponential Smoothing and Short-Term Sales Forecasting," *Management Science*, 13, 821-842.

Pfefferman, D. and Burck, L. (1990), "Robust Small Area Estimation Combining Time Series and Cross-Sectional Data," *Survey Methodology*, 16, 217-237.

Rao, J.N.K., Srinath, K.P. and Quenneville, B. (1989), "Estimation of Level and Change Using Current Preliminary Data," in *Panel Surveys*, ed. D. Kasprzyk, G. Duncan, G. Kalton, and M.P. Singh, pp. 457-479. New York: Wiley.

Smith, W. (1980), "Sample Size and Timing Decisions for Repeated Socioeconomic Surveys," Unpublished D.Sc. dissertation, The George Washington University.

Smith, W. and Barzily, Z. (1982), "Kalman Filter Techniques for Control of Repeated Economic Surveys," *Journal of Economic Dynamics and Control*, 4, 261-279.

DISCUSSION

Gary Shapiro, Abt Associates
4800 Montgomery Lane, Bethesda, MD 20814

It is a pleasure to be a discussant for this session, as it gives me an opportunity to be involved again with SASS. I truly found these papers to very worthwhile. From the standpoint of what is interesting and useful to me, this is one of the very best sessions at the whole convention. I would particularly like to congratulate Steve Kaufman, who was remarkably a co-author of all four papers in the session.

Let me begin with a general comparison of the two estimation papers. Both papers deal with very difficult estimation problems, but take different philosophical approaches. The King paper takes the view that there is an operational problem for which the estimation method must be determined in time for the 93-94 SASS tabulations. In contrast, the Smith paper treats its estimation problem as a research issue - the problem is to be investigated and studied, with no rush to determine an immediate solution. Specific comments on these two papers, as well as the other two papers, follow.

I. Smith Paper on Intersurvey Inconsistency

This paper deals with a "simple" problem: controlling SASS figures to three sets of figures. The authors determined a generalized least squares (GLS) solution, which they could have just applied. However, they recognized that the "...real challenges...require statistical judgments". This is not an obvious conclusion that all investigators would have come to. I believe that many would have been satisfied with the initial GLS solution and would have applied it blindly without considering alternatives.

The authors began with a GLS method to minimize the sum of squares of the differences among the weights. I have observed instances where this was treated as the obvious and only possible quantity to minimize. I was very pleased to see that the authors of this paper did not do that and explored other minimizations as well. Personally, I find the motivation for this particular minimization weak.

I also commend the authors on working through the very simple example given in the paper. This was invaluable in assuring that the authors thoroughly understood what was going on, and also makes it very easy for a reader to understand.

I have one question. One of the alternatives considered was to reweight SASS to the Private School Survey by post-stratification, prior to applying the GLS procedure. I'm interested to know whether the post-stratification by itself gets SASS estimates close to Private School Survey estimates. If so, it might be feasible to only use post-stratification.

Finally, I wonder if there needs to be some movement towards the philosophy of the other estimation paper: If a decision is needed at some point, then the focus must be narrowed and a decision reached about which estimation methodology to use.

II. King Paper on Student Component Estimation

The student weighting in SASS is very difficult due to the complex survey methodology and the need to minimize the burden on schools. The weighting approximation that was derived appears to be a good choice to me, and I have no suggestions for improving it.

The original version of this paper stated that no further research was planned. I admired the honesty of this statement, as most papers talk about future research, even when there is little intent to conduct it. I was nonetheless pleased that the paper was revised to indicate that further research is planned. Since the need to estimate students is likely to be an issue for future years of SASS, it would be useful to evaluate how good the methodology here was. I suggest that an artificial data set be constructed, or/and that a full set of data be collected from a few schools. With such data sets, it will be possible to compare the "correct" estimates and the estimates using the methodology of the paper.

III. Kaufman Paper on Bootstrap Variance Estimator

Bootstrap variance estimation appears to a rather hot topic, in that there have been a number of papers at these meetings on the topic. In session #20, there were 3 papers on this topic:
Kovacevic, Yung and Pandher discuss the use of bootstrap variance estimation for quantile shares. Brodsky and Hughes provide a case study and a simulation. Robb also did a simulation study of bootstrap variance estimation.

Rao, in a different session, presented a review paper on re-sampling methods for variance estimation, including the bootstrap. Hinkins and Scheuren, in yet another session, included some rather disparaging remarks about bootstrap variance estimation in their wide-ranging paper.

This paper shows quite promising and encouraging results for bootstrap variance estimation, in that it does better than other methods. Robb, however, reported very much opposite results in his paper.

Perhaps Robb was not as clever as Kaufman in the application of the method.

Although I am not knowledgeable about bootstrap variance estimation, it appeared to me that determining j is rather cumbersome and difficult, and that this is an impediment to bootstrap variance estimation.

In general, this paper holds out the promise of making a substantial contribution towards the development of better variance estimates.

IV. Ghosh Paper on Optimal Periodicity

I found this an extremely interesting paper with a unique viewpoint. Agencies and policy makers may apply the objective approach presented in the paper to decide the periodicity of surveys, resulting in BIG efficiency gains. Of course, it is also possible that political considerations will preclude agencies from accomplishing any effective applications. I strongly encourage more research on the approach, with applications to additional surveys. I now make several specific comments and suggestions:

1. The paper assumes that survey estimates are unbiased. This is not realistic. I suggest that alternative assumptions are made, for example that there is a 5% relative bias. Such more realistic assumptions would lead towards relatively frequent periodicity as being optimal.
2. In Model 2, if the change is in the same direction as the periodicity bias, it is ignored. I do not see what the justification for this is, and suggest that the model be modified to not ignore the change in this case.
3. I recommend more study on SASS costs for the application of the methods. I realize that estimating cost components is quite difficult. Someone, perhaps Census Bureau staff, will need to spend a lot of time to produce good estimates of the cost components needed for the models.
4. Given the preliminary results of this work, I suggest that 1 year periodicity be evaluated as an alternative. Short periodicities of 1 or 2 years also have potential advantages of evening out survey costs among fiscal years.
5. I suggest the authors look at the work of Bob Fay on the Survey of Income and Education(SIE). Dr. Fay considered whether it was preferable to combine SIE and Current Population Survey for state estimates, or for SIE to stand alone. I believe his methods may also be useful for this work. I also suggest the authors look at the work currently being done by Chip Alexander and others at the Census Bureau on continuous measurement for the Census. Their methodology may have applications to this work.

SOME DATA ISSUES IN SCHOOL-BASED SURVEYS

Daniel Kasprzyk and Kerry Gruber, NCES, Sameena Salvucci, Mehrdad Saba, Fan Zhang, and Steven Fink, Synectics for Management Decisions
[Sameena Salvucci, Synectics Mgmt Decisions, 3030 Clarendon Blvd 305, Arlington VA 22201]

KEY WORDS: Frame coverage; Response error adjustments; Sampling unit definitions

A. Introduction

The Schools and Staffing Survey (SASS) and the Teacher Follow-up Survey are periodic mail surveys conducted by the U.S. Bureau of the Census for the National Center for Education Statistics (NCES), U.S. Department of Education (Gruber, Rohr, Fondelier, 1993; Whitener et al., 1994).

At the National Center for Education Statistics (NCES), SASS is regarded as a major data set for providing information on teachers, principals, and schools. Its periodicity, three years between the first three rounds and now scheduled for four years between the third and fourth round of SASS, allows NCES the opportunity to investigate and study the consequences of decisions made in earlier rounds of the survey in preparation for the next data collection cycle.

During the last three years, the SASS program has initiated a number of projects aimed either at improving understanding of the SASS data or at clarifying a long-standing issue. This paper summarizes the results of three recent studies whose purposes originated with those goals. The concern of the first study was to evaluate how and whether changing the school sampling frame (and the definition of a school) affected SASS estimates. Some understanding of this issue can help in the interpretation of change estimates from Round 1 to Round 2.

The second study aimed to quantify the magnitude of an edit necessary to bring survey information as collected by the SASS in correspondence with frame information for an individual school, as obtained through the Common Core of Data (CCD), an annual NCES database with comparable statistical information for all public schools and school districts in the U.S. (McMillen, Kasprzyk, and Planchon, 1994). While there can be legitimate reasons for SASS and CCD to differ, large discrepancies from CCD are often indicative of problematic survey questions, survey procedures, or response error. Large differences between SASS and CCD had been observed for State estimates in ten states during data review prior to public release. These differences were reduced somewhat through a post-processing edit (based on CCD data) of the individual school data for those ten states. This study extends the edit to the remaining 40 States and the District of Columbia and quantifies the changes in the estimates.

The third study identifies and compares estimates of the same or similar items across survey components. SASS has several built-in redundancies across its various components to allow researchers to use several components of SASS individually, thus eliminating processing steps. While such redundancies can be useful, they can also be confusing because estimates developed by researchers often differ, depending on the source of the data. The aim of the study was primarily to assist users and developers of SASS data to identify and understand differences in estimates of the same or similar items. The following sections describe the activities and results corresponding to the three studies.

B. Comparing SASS Estimates Using Different Sampling Unit Definitions

The public school sampling frame for the 1987-88 SASS was obtained from Quality Education Data, Inc. (QED). In this frame, a public school was defined as a physical unit or location. In the 1990-91 SASS, the public school sampling frame was based on the 1988-89 school year. The CCD-defined school is not a physical location, but an administrative unit. This difference in definition from the QED definition presented some concerns when the decision to change sampling frames was made. These concerns are well-founded, because some (CCD-defined) schools have two or more administrative units within one (QED-defined) physical location. This suggests that the estimates for the number of schools would be higher based on the CCD definition. The 1990-91 SASS sample design allows for the calculation of school, administrator, and teacher estimates using either the QED or the CCD definition of a school.

The purpose of this study was to measure the differences in estimates due to the difference in the CCD and QED definitions of a public school. Only 264 out of approximately 9,000 schools sampled in SASS were redefined. Knowing the extent of these differences and the characteristics of schools affected by

these definitional differences can guide the decision on how to make adjustments to the data for a trend analysis (Choy, Henke, Alt, Medrich, and Bobbitt, 1993) using the QED definition of school. Obtaining estimates based on the QED definition of school occurs by merging and identifying the multiple-CCD schools into the appropriate QED school, and summing the variables of interest across the CCD schools identified with the QED school. Weights for the QED schools are obtained by averaging all CCD schools' final weights within a QED-defined school.

Table 1 provides the QED- and CCD-defined estimates for the number of public schools and students for six states. These tables show the states most affected by the definitional change are North Dakota, South Dakota, Iowa, Nebraska, Minnesota, and Texas. This study showed only a small percentage of CCD-defined schools needed to be adjusted to meet the QED school definition. These schools, however, tended to be found in rural areas and states.

Table 2 provides the number of public schools and students by selected characteristics for rural/small towns and nationally under both definitions. The results showed more differences occur between the number of QED-defined schools and CCD-defined schools in small or rural towns versus urban fringe and large towns. The characteristics having the largest differences tend to occur as a result of the enrollment totals changing as two or more CCD schools are merged/defined as a QED school.

The most obvious ramification of this finding is that researchers analyzing rural trend data and some state trend data from the SASS need to be aware of the impact of these definitional differences on their analyses. For more details on this study see Holt and Scanlon (1994).

C. Effects of Post-Processing Edits on Survey Estimates

The initial review of the 1990-91 SASS data indicated the estimates of total teachers from the public school survey were at least 15 percent greater than the state Full-Time Equivalent (FTE) teacher counts reported on the 1990-91 CCD for nine states: Arkansas, Iowa, Missouri, Montana, Nebraska, North Dakota, Oklahoma, South Dakota, and Wisconsin; in addition, staff review of data from Arizona indicated data problems requiring further review (Gruber, Rohr, and Fondelier, 1993).

Two reasons were suggested for these overestimates. First, some schools did not appear to report data for their school but rather for their entire school district. At times this was due to vague or incorrect school names on the questionnaire label and at times the respondent misunderstood the instructions. The second factor contributing to the overestimates was that the survey respondents did not define schools in the same way that CCD did. For example, a school with grades K-8 at one address might be two CCD schools - an elementary school with grades K-6 and a middle school with grades 7 and 8; i.e., schools in SASS were reporting more grades than the same school had on the CCD (Gruber, Rohr, and Fondelier, 1993).

To make SASS state estimates of the number of teachers consistent with CCD, a post-processing edit was implemented to adjust the SASS data. The approach adopted was to edit SASS data to improve their consistency with CCD-reported data. The post-processing edit used the CCD school-level data for each school sampled in the 10 states to adjust the SASS data to CCD-appropriate grade ranges (Gruber,Rohr, and Fondelier, 1993) (table 3). The urgency to release the 1990-91 SASS data to the public precluded the NCES staff's ability to develop a comparable adjustment for the remaining 40 states and the District of Columbia. Thus, after the data were released a project was begun to develop a comparable adjustment and evaluate the impact of making adjustments to SASS estimates in the other 40 states. The principal concern with the released SASS data was the fact that the SASS data were processed differently in the two categories of states and that unknown biases existed in the data from the 40 states not included in the post-processing edit.

The study adjusted the 1990-91 SASS data to the appropriate CCD grade range following a set of decision rules intended to maintain the internal consistency of the reported data (Saba and Zhang, 1994), as was done with the ten states.

In comparing the CCD-adjusted and the original 1990-91 SASS estimates for FTE teachers (table 4) certain states stand out as being substantially affected by the CCD adjustment. The percent difference reflects the summed difference in SASS estimates and CCD-adjusted SASS estimates within each state.

D. Comparing Similar Estimates Across SASS Components

While the SASS survey is designed to be used across its school, district, administrator, and teacher components, researchers often conduct analyses using individual components. Reported results, therefore, would not usually uncover discrepancies from the same or similar survey items found in more than one component. Thus, the objectives of this study were to 1) identify and compare the same or similar survey items across the SASS and Teacher Follow-up Survey; and 2) compare national and state estimates for these items.

During the search for common variables across the surveys, attitudinal items were eliminated from the analysis. Results of this study are intended to assist researchers and users of the data to identify, help understand, and explain sources of variability on similar or the same survey items. They may also be of interest to persons responsible for various aspects of the design and operation of SASS.

After a review of the questionnaires, six variables were identified as being common on two or more surveys, including: school enrollment, teacher totals, teacher race/ethnicity, teacher certification, teacher training, and teacher attrition.

Public School K-12 Enrollment Comparisons. This section compares the enrollment figures reported in SASS by school district administrators and principals. In the School District Survey, school district staff were asked to report student enrollment (in head counts) in six categories (ungraded, prekindergarten, kindergarten, grades 1-6, grades 7-12, and postsecondary), plus the total of these categories. Principals responding to the Public School Questionnaire were asked to report their student enrollment (in head counts) for each of the grade levels (16 categories) plus a total. Question wording and percentage distribution are located in figure 1.

Total K-12 enrollment. The first comparison examines enrollment estimates provided by LEAs and by the schools. Nationally, school estimates of total elementary and secondary enrollment are lower than district estimates by about one million students (or 2.5 percent). Examining total enrollment by state (not shown but available in Fink, 1994) reveals that school estimates are higher than district estimates in 19 states by an average of 2.9 percent and lower in 32 states by an average of 5.0 percent. There is a statistically significant difference between the district and school enrollment estimates for 44 states. The District of Columbia shows the greatest difference with school totals almost 16 percent below district totals, followed by New Hampshire with district estimates greater than schools estimates by almost 11 percent.

Pre-Kindergarten enrollment. Nationally, pre-kindergarten enrollment estimates provided by schools are ten percent below district estimates (322,434 and 357,816, respectively). In 17 states, school estimates exceed district estimates by an average of 54 percent. In 32 states, school estimates are lower than district estimates by an average of 34 percent. In 11 states, the school estimates differ from the district estimates by more than 50 percent. Among the three states with the largest difference--Indiana, Montana, and Louisiana--school estimates are greater than twice the district estimates. All but seven states exceed the statistical significance level of .05. The detailed tables are available in Fink (1994).

Additional items were examined by Fink (1994). In general, estimates at the national level appear to differ by only a small percentage, though often being statistically significant. Comparing state estimates across SASS components often shows larger percentage differences. Individual categories, such as, ungraded, pre-kindergarten, and postsecondary also exhibit large differences across states.

Even though this study was initially aimed at assisting users of the SASS data, the most likely beneficiaries of the study are the data developers, who obviously must address serious conceptual and response issues for these items. Additional cognitive research, focus group research, pretesting, and user dialogue to determine the use of the various estimates in SASS is necessary.

Several reasons may account for the varying estimates from one survey to another. First, each component of SASS was completed by different respondents. The Teacher Demand and Shortage Survey was completed by school district personnel. Principals or headmasters/headmistresses completed the School Administrator Survey. The School Survey was completed by principals or individuals in the principal's office. Questions on The Teacher Survey were answered by currently employed school teachers. Finally, the Teacher Follow-up Survey questionnaires were sent a year later to a sample of participants in the SASS Teacher Survey. As a result, the quality of survey reports will differ.

Another reason why estimates on similar items may vary from one survey to another is the interview mode. SASS was designed to be primarily a mailout/mailback survey, but a substantial telephone follow-up was used for all sample units not returning the mail questionnaire (Jabine, 1994).

E. Endnote

The three studies summarized above provide an example of why data developers and data providers should try to maintain an inquisitive and questioning point of view. Each study aimed to provide a more thorough understanding of some aspect of the SASS data. Through these studies users can improve their understanding of the data they analyze, and data producers can take steps to improve the products they disseminate.

F. References

Choy, S., Henke, R., Alt, M., Medrich, E., and Bobbitt, S. (1993), **Schools and Staffing in the United States: A Statistical Profile, 1990-91.** NCES 93-146. Washington, DC: U.S. Government Printing Office. Stock no. 065-000-00581-6.

Fink, S. (1994), "1990-91 Schools and Staffing Surveys (SASS) Cross Questionnaire Analysis," NCES Working Paper, available from Special Surveys and Analysis Branch, NCES, Washington, D.C. 20208.

Gruber, K., Rohr, C., and Fondelier, S. (1993), **1990-91 Schools and Staffing Survey: Data File User's Manual, Volume 1: Survey Documentation**, NCES 93-144. Office of Educational Research and Improvement, Washington, DC: National Center for Education Statistics.

Holt, A. and Scanlon, B. (1994), "QED Estimates of the 1990-91 Schools and Staffing Survey: Deriving and Comparing QED School Estimates with CCD Estimates," NCES Working Paper, available from Special Surveys and Analysis Branch, NCES, Washington, D.C. 20208.

Jabine, T. (1994), **Quality Profile for SASS, Aspects of the Quality of Data in the Schools and Staffing Surveys (SASS)**, NCES 94-340. Office of Educational Research and Improvement, Washington, D.C.: National Center for Education Statistics.

McMillen, M., Kasprzyk, D. and Planchon, P.(1993), "Sampling Frames at the United States National Center for Education Statistics," **Proceedings of the International Conference on Establishment Surveys**, Alexandria, VA: American Statistical Association.

Saba, M., and Zhang, F., and Chang, M. (1994), "CCD Adjustment to the 1990-91 Schools and Staffing Survey (SASS)," NCES Working Paper, available from Special Surveys and Analysis Branch, NCES, Washington, D.C. 20208.

Whitener, S., Rohr, C., Bynum, L., Kaufman, S., and King, K. (1994), **1991-92 Teacher Follow-up Survey: Data File User's Manual.** NCES 94-331 (Public Use Version) and NCES 94-478 (Restricted Use Version). U.S. Department of Education, Office of Educational Research and Improvement, Washington, DC: National Center for Education Statistics.

Table 1.--CCD and QED-defined estimates in SASS for number of public schools and students for selected states

	Schools		Students	
	CCD	QED	CCD	QED
U.S. Total	**79,885**	**78,759**	**40,103,699**	**40,096,401**
North Dakota	647	516	118,778	118,799
South Dakota	732	579	148,790	147,591
Iowa	1,530	1,445	479,023	478,912
Nebraska	1,455	1,325	260,030	260,211
Minnesota	1,434	1,346	719,581	719,460
Texas	5,651	5,606	3,323,523	3,323,498

Source: U.S. Department of Education, Schools and Staffing Survey: 1990-91 (School Questionnaire)

Table 2.--QED & CCD defined estimates for number of public schools and students, 1990-1991

	QED		CCD		Percent Difference	
	Schools	Students	Schools	Students	Schools	Students
U.S. Total	**78,759**	**40,096,401**	**79,885**	**40,103,699**	**0.0**	**1.4**
Rural/small town	39,263	15,694,730	40,352	15,695,586	2.8	0.0
School Level						
Elementary	25,715	9,395,915	26,508	9,495,515	3.3	0.0
Secondary	10,967	5,359,209	11,170	5,257,121	1.9	-1.9
Combined	2,581	939,606	2,674	942,951	3.6	0.4
Minority Enrollment						
Less than 20%	29,021	10,938,818	29,974	10,938,435	3.3	0.0
20% or more	10,242	4,755,912	10,378	4,757,151	1.3	0.0
School Size						
Less Than 150	6,938	594,261	7,843	664,432	13.0	11.8
150 to 499	21,179	6,700,298	21,477	6,746,207	1.4	0.7
500 to 749	7,304	4,418,856	7,252	4,383,991	-0.7	-0.8
750 or More	3,842	3,981,315	3,780	3,900,956	-1.6	-2.0

Source: U.S. Department of Education, NCES, Schools and Staffing Survey: 1990-91 (School Questionnaire)

Table 3.– FTE Teachers for 1990-91 CCD and 1990-91 SASS After Adjustment (For Original 10 States)

State	CCD	SASS	SASS/CCD
U.S. Total	**2,282,398**	**2,381,944**	**104.36%**
Arizona	32,015	30,159	94.20%
Arkansas	25,787	27,091	105.06%
Iowa	31,795	33,402	105.05%
Missouri	51,115	52,632	102.97%
Montana	8,767	10,363	118.20%
Nebraska	18,771	18,107	96.46%
North Dakota	6,835	7,953	116.36%
Oklahoma	35,815	37,337	104.25%
South Dakota	8,389	9,863	117.57%
Wisconsin	50,724	55,207	108.84%

Source: Department of Education, NCES, 1990-91 CCD and 1990-91 SASS (School Questionnaire)
Note: All of the above states had a greater than 15 percent difference before adjustment.

Table 4.–FTE teachers for 1990-91 CCD, 1990-91 SASS Before and After CCD Adjustment

State	CCD	SASS Before Adjustment	SASS After Adjustment	Percentage Effect of Adjustment
U.S. Total	2,397,351	2,438,592	2,381,943	2.32%
Nevada	10,373	10,391	9,960	4.15%
Maine	15,513	16,069	15,289	4.85%
Louisiana	45,377	45,271	42,841	5.37%
Florida	108,088	105,167	99,479	5.41%
D.C.	5,950	5,543	5,956	7.45%
New Hampshire	10,637	10,852	9,924	8.55%
Minnesota	43,753	44,329	39,933	9.92%
Alaska	6,710	6,610	5,850	11.50%
Wyoming	6,784	7,349	6,151	16.30%

Source: U.S. Department of Education, NCES, 1990-91 CCD and 1990-91 SASS (School Questionnaire)

Figure 1.--Survey question wording, counts and percentage distributions

	School District Survey Questionnaire: Question 1		Public School Survey Questionnaire: Question 17	
Question Wording	What was the enrollment (in head counts) in this district on or about October 1 of THIS school year, and on or about October 1 of LAST school year?		How many students were enrolled in each grade on October 1 of this school year? (Report in head counts)	
Variables Used:	Counts	Distribution	Counts	Distribution
Ungraded	705,564	1.8%	321,721	0.8%
Kindergarten	3,237,854	7.9%	3,081,336	7.7%
Grades 1-6	19,419,747	47.5%	19,218,059	47.9%
Grades 7-12	17,482,583	42.8%	17,482,583	43.6%
Total	40,845,748	100.0%	40,103,699	100.0%

Source: NCES, Schools and Staffing Survey: 1990-1991 (School, District Questionnaire)

THE 1991-92 TEACHER FOLLOW-UP SURVEY REINTERVIEW AND EXTENSIVE RECONCILIATION

Cleo R. Jenkins and Angela-Jo Wetzel, Bureau of the Census
Cleo R. Jenkins, Bureau of the Census, Washington, D.C. 20233

Key Words: Measurement Error, Cognitive Research

I. INTRODUCTION

Traditionally, reinterviews have been designed for one (or more) of the following four purposes:

- to detect whether interviewers have deliberately falsified data,
- to evaluate interviewer performance,
- to estimate response variance, or
- to estimate response bias (Forsman and Schreiner, 1991).

Many reinterviews performed by the Census Bureau focus on estimating response variance. Although measuring response variance exposes inconsistencies in respondents' answers between interviews, it does little to explain why the inconsistencies occur.

Consequently, the 1991-92 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation was designed with a new objective in mind. Primarily, it focused on determining the reasons for respondent and instrument errors.

In this paper, we briefly describe the methods that were used to conduct this reinterview, followed by a discussion of both the methodology's benefits and limitations.

II. METHODOLOGY

A. Description of the 1991-92 TFS Reinterview Program

The Census Bureau conducted the 1991-92 TFS a year after collecting information from teachers in the 1990-91 Schools and Staffing Survey (SASS) for the National Center for Education Statistics (NCES). The TFS' purpose was to provide information about teacher attrition and to project teacher demand (Faupel et al., 1992). In general, the Census Bureau conducted the TFS Reinterview and Extensive Reconciliation two to three weeks after the TFS.

Both the TFS and the TFS Reinterview and Extensive Reconciliation contained two components: one for former teachers and another for current teachers. Each component had its own questionnaire (the TFS-2 for former teachers and the TFS-3 for current teachers), asking primarily different questions. The reinterview reasked a subset of questions from the TFS. The NCES chose the questions for reinterview. The Census Bureau offered suggestions, favoring factual over opinionated questions.

The TFS was a mixed-mode survey consisting of a first and second mail questionnaire, succeeded by a telephone follow-up of mail non-respondents. The TFS Reinterview and Extensive Reconciliation was conducted exclusively by phone.

B. Development of the Extensive Reconciliation Probes

The use of an extensive reconciliation distinguishes this reinterview from others. It contained a series of probes aimed at identifying the reason for response differences and a reconciliation question to determine the correct response.

Closed-ended probes offered respondents specific reasons for differences. They were not the same from question to question, but tailored to each reinterview question. We used closed-ended probes to capture the data efficiently.

Two methods were used to develop the closed-ended probes:

- An expert analysis was conducted in which potential problems with the reinterview questions or possible reasons for differences between the two interviews were identified (see Forsyth and Lessler, 1991, for a discussion of this method).
- The findings of previous cognitive research with the 1990 Field Test Teacher Questionnaire (see Bates and DeMaio, 1990) were used. This information was especially helpful in identifying questions that might be susceptible to misinterpretation.

If the respondent did not choose one of the closed-ended probes, they were asked the open-ended probe: "Or was there some other reason [for the difference]?". The open-ended reasons were professionally reviewed and clerically coded prior to data entry.

C. Reinterview and Extensive Reconciliation Procedure

Working from a paper questionnaire, supervisory field representatives (SFRs) administered the TFS Reinterview and Extensive Reconciliation by phone. The SFRs received their instructions in a home self-study manual. The manual instructed them to first administer all of the reinterview questions. Immediately after completing the reinterview, the SFRs compared the respondents' reinterview responses with their original responses. The original responses had been transcribed to the reinterview questionnaires. Because the original responses were visible during the reinterview, this made it a dependent reinterview.

When a difference between the two responses occurred, the SFRs continued with the extensive reconciliation by asking the series of probes and the reconciliation question.

D. Sample Selection

Our goal was to obtain completed reinterviews for approximately 500 former and 500 current teachers. To achieve this goal, Demographic Statistical Methods Division (DSMD) randomly selected approximately 800 former teachers and 700 current teachers from the TFS sample files. DSMD oversampled to compensate for any non-response from the original interview and the reinterview. The 1992 TFS Reinterview and Extensive Reconciliation achieved a 92 percent completion rate (number of completed reinterviews (1314) divided by the number of eligible reinterview cases (1425)). We obtained completed reinterviews from 685 former teachers and 629 current teachers.

E. Analysis

We used two measures to analyze our reinterview data for this paper.

1. Gross Difference Rate (GDR)

The GDR is the proportion of responses that differ between the original interview and the reinterview. We calculated the GDR before reconciliation for the overall question. The GDR provides a rough idea of how consistently respondents answer a question.

2. Net Difference Rate (NDR)

The NDR is the difference between the percent of original responses in a specific answer category and the percent of reinterview responses in that category. We calculated a NDR after reconciliation for each answer category for a question.

The NDR shows the direction of change in responses for an answer category. We tested each NDR to see if it was significantly different from zero at the 90 percent confidence level. If the NDR is significant and positive, the answer category was over-reported in the original interview. If the NDR is significant and negative, the answer category was under-reported in the original interview.

III. RESULTS AND DISCUSSION

A. Benefits of the Methodology

The reinterview and extensive reconciliation produced some meaningful information from which we were able to make recommendations for either improvements or further research for a number of the TFS questions. We identified 19 of the 49 reinterview questions as problematic. We considered a question problematic if 1) one or more of its answer categories had a significant NDR or 2) it had one or more notable reasons for response differences. Refer to Jenkins and Wetzel (1994a) for a complete analysis of each reinterview question.

In this paper we illustrate two types of problems that we were able to uncover: 1) comprehension and 2) information storage or retrieval.

1. Comprehension Problems

Respondents demonstrated difficulty understanding the meaning of some questions. We illustrate this using two questions: the grade level and the teaching assignment question. We present the original question followed by our recommendations for improving it. We offer the supporting data in a table that includes:

- the GDR before reconciliation,
- each answer category that has an after reconciliation NDR significantly different from zero at the 90% confidence level, and
- the complete list of respondents' answers to the series of probes.

a. The Grade Level Question:

In what grade levels are the students in your classes at THIS school?

The intent of this question is to learn what the grade levels are of all the students that the teacher teaches. Respondents were supposed to mark all grade levels that applied. For our analysis, we considered each of the 16 answer categories shown in Table 1 as a separate question with two possible answer categories: marked and unmarked.

Respondents demonstrated difficulties understanding the wording of this question. The NDRs in column 3 of this table suggest that respondents tended to overreport students in the 4th through 8th grades in the original interview. Respondents' reasons for inconsistent answers given in part 2 shed some light on this result:

- One-third (15) reported misunderstanding some aspect of the question. Specifically, four reported misunderstanding what was meant by "grade level" or "class." Another five were uncertain whether they should report the grade levels of students they sometimes teach or classes with only a few students. Six simply reported misunderstanding the question as a whole.
- Three respondents had difficulty because they taught special students. These respondents either had trouble reporting the equivalent grade levels for the students, or they were not certain whether they should report them as ungraded or in their equivalent graded level.

The reasons respondents gave for differences suggest that if the intent of this question is to learn what the grade levels are of all the students that the

teacher teaches, regardless of whether the student is in a formal "class" or not, then the question should be reworded: **In what grade levels are the students that you teach at THIS school?** This wording eliminates the confusing word "class," the definition of which gives respondents problems. Does a class need to meet regularly to be considered a class? Does it need to be a certain size before it qualifies as a class? Respondents are not certain of the answers to these questions.

b. The Teaching Assignment Question:

> **Which of the following categories best describes your teaching assignment?**
>
> [ ] **Regular full-time or part-time teacher**
> [ ] **Itinerant teacher (i.e., your assignment requires you to provide instruction at more than one school)**
> [ ] **Long-term substitute (i.e., your assignment requires that you fill the role of a regular teacher on a long-term basis, but you are still considered a substitute)**

In this question, respondents reported having difficulty with the question's wording and the answer categories. Part 3 of Table 2 shows that half (6) of the respondents who gave a reason for inconsistent answers said they misunderstood the question or thought the answer categories were confusing. The NDRs in part 2 of Table 2 suggest that the problem lies with the first two answer categories. Respondents tended to overstate being a regular full- or part-time teacher (1.6%) in the original interview, while understating being an itinerant teacher (-1.5%).

A possible explanation for this is that respondents chose the first answer category because they thought it fit their situation well enough. Perhaps they cued in on the words "full-time or part-time teacher," while overlooking, ignoring, or not understanding the word "regular." Without this word, itinerant and long-term substitute teachers might reasonably mistake themselves for full- or part-time teachers. This behavior of selecting the first response alternative that seems to constitute a reasonable answer is discussed by Krosnick (1991).

The word "itinerant" may be another problem. Cognitive research with the Public School Questionnaire revealed that many respondents did not know what an "itinerant" teacher was (Jenkins et al., 1992a, p. 26). They knew "itinerant" teachers by other names, including traveling, co-op, and satellite teachers.

Based on these results, we suggest the following changes to this question:

- Reorder the answer categories. The itinerant and long-term substitute teachers are more likely to consider themselves regular full- or part-time teachers than vice versa.
- Reword the "itinerant teacher" answer category. State the definition of "itinerant teacher" first, then the technical term in parentheses, instead of vice versa.
- Provide a more comprehensive list of familiar names for itinerant teachers, such as traveling, co-op, or satellite teachers.

Our suggested order and wording are:

[] **You provide instruction at more than one school (i.e., you are an itinerant, traveling, co-op, or satellite teacher).**
[] **You fill the role of a regular teacher on a long-term basis, but you are still considered a substitute (i.e., you are a long-term substitute teacher).**
[] **You are a regular full-time or part-time teacher.**

2. Information Storage or Retrieval Problems

Respondents demonstrated difficulty obtaining information to answer some questions. We illustrate this using two questions: the base year salary and the family income question. Again, we present the original question followed by our recommendations for improving it.

a. The Base-Year Salary Question:

> **The following questions refer to your before-tax earnings from teaching and other employment from the summer of 1991 through the end of the 1991-92 school year.**
>
> *Record earnings in whole dollars.*
>
> **DURING THE CURRENT SCHOOL YEAR--**
>
> **What is your academic base year salary for teaching in this school?**

This question requests a monetary value. The before reconciliation disagreement rate (14.8%) in part 1 of Table 3 shows that respondents had difficulty reporting this value. (According to reinterview instructions, the dollar values disagree if they exceed a $1,000.00 difference.) Part 2 of Table 3 shows that the predominant reason for monetary differences is that respondents were unsure of the exact amount of their earnings. This suggests that respondents do not have an easily accessible, precise figure stored in

memory to accurately answer this question. It also suggests an inability or unwillingness on the respondent's part to look up appropriate records which may exist.

We discuss these problems further after looking at the results from the next question.

b. The Family Income Question:

> **Which category represents the total combined income (include your own income) of ALL FAMILY MEMBERS age 14 and older in your household during 1991? Include money from jobs, net business or farm income, pensions, dividends, interest, rent, social security payments, and any other income received by family members in your household.**
>
> **[ ] less than $10,000**
> **[ ] .**
> **[ ] .**
> **[ ] .**
> **[ ] $100,000 or more**

This question requests categorical data. The GDR (16.2 percent) in part 1 of Table 4 is the largest of any of the closed-ended questions. Part 2 shows that nearly half (41) of the respondents who gave a reason for inconsistent answers said they were unsure of the exact amount. Again, this suggests that they do not have an easily accessible, precise figure stored in memory to accurately answer the question.

The fact that respondents had difficulties consistently answering an income question whether it requested a monetary value (base-year salary) or categorical data (family income) does not appear simple to solve. Initially we thought that asking respondents either 1) to obtain records to accurately answer the income questions or 2) to stop and think about them more carefully might be possible solutions to this problem. However, we now believe this to be a naive perspective. According to a recent experimental treatment, requiring the use of personal records may decrease response rates and increase follow-up costs without a large enough improvement in answer quality (Marquis, 1993).

We need to have a better understanding of respondents' use of records before we will be able to properly guide this process. Jenkins (1992b) concludes that 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. Perhaps asking respondents to gather appropriate records is more feasible with a self-administered questionnaire than other modes of administration. Certainly this is an area in need of further research.

Since asking respondents to use their records may have a detrimental effect on the data in other ways (i.e., increased nonresponse), the question becomes just how much measurement error in the data can the sponsor tolerate. Although responses to the family income question differ, they do so by a limited amount. A crosstabulation of inconsistent answers between the reinterview and original interview shows that almost 60 percent of them are due to respondents choosing answer categories that are next to each other in the two interviews. For instance, a respondent might choose the answer category $15,000-$19,000 in the original interview and $20,000-$24,000 in the reinterview, or vice versa.

B. Limitations of the Methodology

We believe the 1991-92 TFS Reinterview and Extensive Reconciliation had shortcomings involving the dependent-type reinterview and the closed-ended probes. Jenkins and Wetzel (in press) contains a complete report of the reinterview and extensive reconciliation's methodology and our recommendations for improving it.

1. The Dependent-Type Reinterview Produced Too Few Differences

In general, the 1991-1992 TFS Reinterview and Extensive Reconciliation produced too few differences. There are fourteen questions from the reinterview and extensive reconciliation that are the same as those from the 1989 TFS Reinterview, and all but two of them have before reconciliation GDRs significantly lower than their 1989 counterpart at the 90% confidence level. Evidence also exists from past research that dependent reinterviewing results in fewer differences (Schreiner, 1980; Koons, 1973).

Because of the low GDRs, our counts for specific reasons for differences are very small at times. This can be seen in the numbers we discuss in the previous section (Results and Discussion).

The 1989 and 1992 surveys had two major differences:

- The 1989 methodology used an independent reinterview, whereas the 1992 methodology used a dependent-type reinterview.
- The 1989 methodology used FRs in both the original and reinterview. In contrast, the 1992 procedures specified that SFRs conduct the reinterview.

We hoped that SFRs would be more likely to ignore the original response than FRs. The data suggest, however, that this was not the case and that the lower GDRs are due to the reinterview's dependency.

2. The Extensive Reconciliation Produced Too Many Open-ended Responses

Approximately 54% of the total number of reasons for differences were open-ended. This unexpectedly high percentage suggests that the series of closed-ended probes did a relatively poor job of providing respondents with adequate reasons for differences in their responses.

3. The Extensive Reconciliation Produced Too Many General Responses

An even larger deficiency with the extensive reconciliation was that respondents did not adequately verbalize the reasons for differences in their answers when the closed-ended questions did not apply. Approximately 43% of the open-ended responses were "don't know" or "misunderstood question." This is a much more serious error than obtaining open-ended responses that could be coded to specific reasons. The general responses led to the omission of useful data.

IV. CONCLUSION

The 1991-92 TFS Reinterview and Extensive Reconciliation represents the Bureau's first attempt to employ an extensive structured reconciliation. The ultimate goal was to identify problematic questions, to identify the sources of the problems, and to offer suggestions for improving the TFS questionnaires.

As demonstrated in this paper, we were able to identify some problem questions, particularly those exhibiting comprehension and information storage/retrieval difficulties. Moreover, we gained enough insight from the reinterview and extensive reconciliation to make recommendations for either improving the questions or for further research.

However, there were some methodological shortcomings. We showed that the reinterview and extensive reconciliation produced too few differences and, hence, too few reasons for differences between the original and reinterview responses. We believe this occurred because the reinterview was not independent from the original interview. In the future we strongly suggest employing: (1) an independent reinterview followed by a third visit small-scale unstructured extensive reconciliation, or (2) an independent reinterview followed by a large-scale extensive reconciliation using Computer Assisted Telephone Interview (CATI). We make these suggestions without having evaluated cost or respondent burden. However, given the correct methodology, the reinterview/extensive reconciliation may become an effective questionnaire evaluation technique.

NOTES

1. The SASS is a relatively new set of integrated surveys first launched in the 1987-88, 1990-91, 1993-94 school years, and scheduled every four years hence.

REFERENCES

Bates, N. and DeMaio, T.J. 1990. "Report on Cognitive Research on the Public and Private School Questionnaire for the Schools and Staffing Survey (SASS-3A and SASS-3B). " Internal Census Bureau memorandum. July 11, 1990.

Faupel, E., Bobbit, S. and Rohr, C. 1992. "1988-89 Teacher Followup Survey Data File User's Manual." National Center for Education Statistics. U.S. Department of Education. Office of Educational Research and Improvement. June 1992.

Forsman, G. and Schreiner, I. 1991. "The Design and Analysis of Reinterview: An Overview" in P.B. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman (eds), Measurement Errors in Surveys. New York: John Wiley and Sons, pp. 279-301.

Forsyth, B.H. and Lessler, J.T. 1991. "Cognitive Laboratory Methods: A Taxonomy" in P.B. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman (eds), Measurement Errors in Surveys. New York: John Wiley and Sons, pp. 393-418.

Jenkins, C. and Wetzel, A. 1994a. "The Results of the 1992 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation: Part One of a Two-Part Report. A report submitted to the National Center for Education Statistics, Census Bureau Memorandum, March 22, 1994.

Jenkins, C. and Wetzel, A. In press. "An Evaluation of the Methodology Utilized for the 1991-92 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation: Part Two of a Two-Part Report.

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

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

Koons, D.A. 1973. Quality Control and Measurement of Nonsampling Error in the Health Interview Survey. Washington, DC: National Center for Health Statistics. Series 2. Number 54 DHEW Publications No. HSM 73-1328. March, 1973.

Krosnick, J.A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys" Applied Cognitive Psychology. 5:213-236.

Marquis, K. 1993. Personal Communication. Preliminary Results from a Cognitive Approach to Redesigning Measurement in the Survey of Income and Program Participation. US Bureau of the Census. Washington, DC.

Schreiner, I. 1980. "Reinterview Results from the CPS Independent Reconciliation Experiment (Second Quarter 1978 through Third Quarter 1979). Internal Census Bureau memorandum. May 7, 1980.

Table 1. Grade Level Question - 629 Responses

Part 1. GDR, Significant NDR's and Confidence Limits (%)

Category	GDR	Limits	NDR	Limits
Ungraded	0.2	(-0.1, 0.4)		
Prekindergarten	0.6	(0.1, 1.2)		
Kindergarten	1.9	(1.0, 2.8)		
1st	2.5	(1.5, 3.6)		
2nd	3.0	(1.9, 4.1)		
3rd	2.5	(1.5, 3.6)		
4th	2.9	(1.8, 4.0)		
5th	3.2	(2.0, 4.3)	1.3	(0.1, 2.4)
6th	1.9	(1.0, 2.8)	1.6	(0.5, 2.7)
7th	2.7	(1.6, 3.8)	1.0	(0.1, 1.8)
8th	2.7	(1.6, 3.8)	1.4	(0.4, 2.5)
9th	2.5	(1.5, 3.6)	1.7	(0.7, 2.8)
10th	2.1	(1.1, 3.0)		
11th	1.7	(0.9, 2.6)		
12th	1.9	(1.0, 2.8)		
Postsecondary	0.5	(0.0, 0.9)		

Part 2. Reasons for Difference between Responses

Reason	Count	Percent
Total	49	100.0
Don't know	16	32.7
Misunderstood question	6	12.2
Unsure whether to report level of classes sometimes taught or with few students	5	10.2
Teaching different students since responding	4	8.2
Misunderstood what "grade level/class" meant	4	8.2
Forgot/remembered info	4	8.2
FR error	3	6.1
Teach special students - difficulty reporting/unsure whether to report equivalent grade levels	3	6.1
Other	2	4.1
Misunderstood reference period	2	4.1

Table 2. Teaching Assignment Question - 610 Responses

Part 1. Gross Difference Rates and Confidence Limits (%)

No. of Categories	GDR	Limits
3	2.0	(1.0, 2.9)

Part 2. Significant NDRs and Confidence Limits

Answer Category	NDR	Limits
Regular full/part-time teacher	1.6	(0.7, 2.6)
Itinerant teacher	-1.5	(-2.4, -0.6)

Part 3. Reasons for Difference between Responses

Reason	Count	Percent
Total	13	100.0
Misunderstood question	3	23.1
Category problems	3	23.1
Situation changed since responding	2	15.4
Don't know	2	15.4
FR/Manual/general error	2	15.4
Forgot/remembered info	1	7.7

Table 3. Base-Year Salary Question - 629 Responses

Part 1. Disagreement Rate and Confidence Limits (%)

No. of Categories	Rate	Limits
2	14.8	(12.5, 17.1)

Part 2. Reasons for Difference between Responses

Reason	Count	Percent
Total	109	100.0
Unsure of exact amount	71	65.1
Salary changed since responding	9	8.3
Don't know	9	8.3
Fr/manual/general error	5	4.6
Included other salary earnings	4	3.7
Misunderstood question	3	2.8
Included another source of income	2	1.8
Forgot/remembered info	2	1.8
Misunderstood reference period	2	1.8
Unsure how to report as an itinerant teacher	1	0.9
Gave after-tax earnings	1	0.9

Table 4. Family Income Question - 604 Responses

Part 1. Gross Difference Rate and Confidence Limits (%)

No. of Categories	GDR	Limits
13	16.2	(13.8, 18.7)

Part 2. Reasons for Difference between Responses

Reason	Count	Percent
Total	84	100.0
Unsure of exact amount	41	48.8
Don't know	11	13.1
Unsure what to include/exclude	8	9.5
Misunderstood reference period	7	8.3
FR/manual/general error	5	6.0
Wasn't sure whether to include adult children	4	4.8
Misunderstood question	2	2.4
Refused to answer in one interview	2	2.4
Other	1	1.2
Missed skip pattern/question	1	1.2
Forgot/remembered info	1	1.2
Misread question	1	1.2

IMPROVING COVERAGE IN A NATIONAL SURVEY OF TEACHERS

Daniel Royce and Irwin Schreiner, Bureau of the Census[1]
Daniel Royce, Bureau of the Census, Washington, D.C. 20233

Key Words: Teacher Lists, Accuracy

1. INTRODUCTION

The National Center for Education Statistics (NCES) sponsors the Schools and Staffing Survey (SASS) conducted by the U.S. Census Bureau. The Census Bureau first conducted the SASS during the 1987-88 school year and again during the 1990-91 and 1993-94 school years. The SASS is an integrated set of surveys, one of which is a survey of public and private school teachers.

At the beginning of the fall semester of the school year in which the SASS is conducted, the Census Bureau mails a Teacher Listing Record (TLR) to each sample public and private school. The instructions request that the schools list the teachers in their school on the TLR. The SASS then uses the TLRs to create the teacher frame for sampling teachers within the schools. Later during the school year, the Census Bureau mails a separate School Questionnaire to these same schools. This questionnaire asks for information about the school, including head counts of teachers within the school.

In the 1987-88 and 1990-91 SASSs, the schools, on average, reported a different number of teachers on the TLR than the School Questionnaire. This inconsistency in the reporting of teachers prompted the National Center for Education Statistics (NCES) to enlist the Census Bureau to conduct a special Teacher List Validity Study (TLVS).

The purpose of the TLVS was to evaluate the quality of the teacher lists on the TLR, and to provide insight into how teacher estimates could be improved. We designed the study to be primarily qualitative in nature. The Census Bureau conducted the TLVS during the 1992-93 school year. Specifically, the study tried to determine whether:

- the schools were filling out the TLR per our instructions (i.e. the instructions on the form)
- the schools were listing eligible in-scope teachers
- the school districts could provide more accurate listings of teachers
- the TLR or the School Questionnaire, if either, elicits a more accurate count of teachers
- certain types of teachers/non-teachers created problems for the schools when computing the teacher counts

We selected a small sample of schools primarily in those states that reported inconsistent teacher counts between the TLR and the School Questionnaire.

We employed reinterview as the primary technique in the study with reconciliation of differences between the original listing and the reinterview. In addition, we employed a "think aloud" technique during the reinterview. This technique, which is normally used in a cognitive interviewing setting, has respondents describe their thoughts while answering the questions.

We feel the study succeeded in providing insight into how to obtain more accurate coverage of teachers. For the 1993-94 SASS, we were able to field a much improved TLR. This study also demonstrates how reinterview can be used in a trouble-shooting capacity to help make a survey work better.

2. METHODOLOGY

The TLVS had two separate components involving different samples of schools. The first component consisted of a reinterview and reconciliation of the TLRs. The second component consisted of a reconciliation of differences between the number of teachers listed on the TLRs and the head counts of teachers on the School Questionnaires.

2.1 Sample Selection (Initial Stage)

We selected samples of both public and private schools. We selected a public school sample from the public school universe file that was planned for use in the school phase of the 1992-93 SASS (postponed until 1993-94). We selected a private school sample from the private school universe file that was current as of August 1992.

Before selecting the public and private school samples, we deleted schools in certain states because they had high field costs. We then selected the samples using the average teacher adjustment factor (TAF) from the 1990-91 SASS. This adjustment factor is based on a weighted average of the ratio between the number of teachers reported on the School

Questionnaire (numerator) and the number of teachers reported on the TLR (denominator).

For public schools, we defined each state's TAF as "good" if $0.9 \leq TAF \leq 1.1$. For private schools, we defined each affiliation's (i.e., Catholic, Episcopal, etc.) TAF as "good" if $0.8 \leq TAF \leq 1.0$. Anything outside these ranges, we defined as "bad." (The private school TAFs were all less than 1. After the sample was selected, errors were found on the teacher file which made those counts greater than they were supposed to be.)

Both the public and private school samples contained higher percentages of schools from the "bad" reporting states: 70 percent public, 75 percent private. We then alternated the assignment of the schools to the two components.

2.2 Component 1: Reinterview of the TLRs

In mid-November 1992, we mailed TLRs to the 300 private schools and 290 public schools in this component of our sample. We also mailed TLRs to the 254 school districts (Local Education Agencies, or LEAs) associated with the 290 public schools. We conducted telephone follow-up for mail nonreturns.

When we received about 85 percent of the TLRs, we selected the reinterview sample. We selected 100 public schools (with their corresponding LEA) and 100 private schools.

We selected the 100 public schools with the highest difference ratio as defined below:

L = teachers reported <u>only</u> on the **LEA TLR**
S = teachers reported <u>only</u> on the **school TLR**
B = teachers common on **both TLRs**

$$\text{difference ratio} = \frac{(L + S)}{(L + S + B)}$$

We obtained these counts by comparing name by name the LEA TLR to the school TLR. The ratios for the 100 public schools we selected for the reinterview ranged from .11 to .87.

We selected the 100 private schools with the highest difference ratio between what was reported on the TLR and what was reported as head counts (not names) in the 1991-92 Private School Survey (PSS).

S = teachers reported on the **school TLR**
P = teachers reported in **the 1991-92 PSS**

$$\text{difference ratio} = \left| \frac{(S - P)}{S} \right|$$

The difference ratios for the 100 private schools ranged from .18 to 23.5.

Reinterview began in mid-February 1993. We did not give the interviewers any formal training, but provided them with instructions to read before conducting the reinterviews. The interviewers we used were familiar with conducting reinterviews.

Of the 100 public schools selected, we assigned 50 for personal visit reinterview and 50 for telephone reinterview.

For the 50 personal visit cases, the reinterviewer asked the original respondent to fill out the TLR again, thinking aloud as he/she completed it. Our goal for these 50 cases was to determine how the respondent interpreted our instructions.

The reinterviewer then compared the reinterview TLR with the original TLR filled out in the previous Fall and reconciled any differences. We also instructed the reinterviewer to ask the school why the LEA reported certain teachers that they did not.

For the 50 telephone cases, the respondent did not complete another TLR. Instead, we instructed the reinterviewer to only reconcile differences between the TLR filled out by the school and the one filled out by the LEA.

Of the 100 private schools in our reinterview sample, we also assigned 50 for personal visit and 50 for telephone.

Here, the reinterviewers followed the same procedures as they did for the personal visit reinterviews for the public schools.

2.3 Component 2: Reconciliation of the TLRs and School Questionnaires

When we mailed the TLRs to the schools in the first component (in mid-November), we also mailed TLRs to a separate sample of 300 private schools and 290 public schools. (LEAs were not involved in this component.)

At the end of February we mailed School Questionnaires to each school and then followed-up by telephone any mail nonreturns.

When we received about 90 percent of the School Questionnaires, we selected the reinterview sample. We selected the public and private school reinterview samples the same way.

We selected the 100 public schools and 100 private schools with the highest difference ratio between what was reported on the TLR and what was reported on the School Questionnaire (as described below):

T = teachers reported on the **(TLR)**
X = teachers reported on **School Questionnaire**

$$\text{difference ratio} = \left| \frac{(T - X)}{T} \right|$$

The difference ratios ranged from .05 to .98 for the 100 public schools, and from .07 to 2.0 for the 100 private schools selected.

We sent out separate instructions to the interviewers in April. Reconciliation started at the beginning of May. The interviewers conducted all reconciliation by telephone.

We mailed back to the school a copy of the original TLR and School Questionnaire that they had completed. We also sent them a letter describing the study and letting them know that someone from the Census Bureau would be contacting them regarding the reconciliation.

2.4 Limitations

The major limitation of the study was that it was designed to be qualitative rather than quantitative. We selected a non-random sample of schools. Therefore, we cannot generalize our results to all schools. The discussions on significance tests apply ONLY to the schools in our sample. Even within the schools we did reinterview, we did not try to get specific numbers on how many teachers were erroneously missed or non-teachers that were erroneously included. Instead, we attempted to find out the types of teachers/non-teachers that the schools included or excluded in their counts.

We also tried to find out reasons why the schools excluded certain teachers and included persons who should not have been included. Unfortunately, the reinterview and reconciliation did not gather adequate reasons. Most of the respondents simply said they "forgot about that person" or "I thought this person should/shouldn't be included." Some didn't provide any reasons. Our Center for Survey Methods Research has implemented a program of cognitive research on the revised TLR which should provide this and other kinds of information.

3. Results

We present the types of teachers most often incorrectly excluded. and the types of non-teachers most often incorrectly included by the schools and LEAs on the TLRs and/or School Questionnaires. Non-teachers are those persons that were not supposed to be included in the counts. These results were instrumental in the development of the revised TLR for the 1993-94 SASS. We also compare results between the TLRs from the schools and LEAs in our reinterview component, and between the TLRs and School Questionnaires from the schools in our reconciliation component. While the statistical tests are limited to the sample only, the data suggest there are some differences in these comparisons.

Before we could analyze the data, we had to determine the actual count of teachers in each school. We used this count as the basis for our comparisons.

3.1 Types of Teachers/Non-teachers Erroneously Excluded/Included

We attempted to find out the types of teachers who were excluded in error from the teacher list or count, and the types of non-teachers who were included in error from the list or count. We gathered a wide variety of different types of teachers and non-teachers which we grouped into like categories.

The figures in the tables represent the number of schools and LEAs that mentioned that they excluded at least one teacher in the group, or included at least one non-teacher in the group. (i.e., If a school respondent said that he/she forgot to include 3 part-time teachers, then we would tally only once in the part-time teacher group, NOT three tallies. Or, if a respondent said that he/she included two pre-kindergarten teachers and three counselors by mistake, then we would tally once in the pre-kindergarten category and once in the guidance counselor category, NOT two and three, respectively.)

3.1.1 Public Schools vs. LEAs

When we compared the 99 public schools to their corresponding LEAs (there was one refusal during the reinterview), we found that 43 schools and 48 LEAs mentioned that they excluded at least one teacher from their list. **Table 1** shows that general full-time / general teachers, part-time teachers, and specialized subject matter teachers were among the types of teachers most often excluded.

The "general full-time / general teachers" category is a "catch all" category. Several schools and LEAs reported that they "forgot to include" or "missed" some teachers, but gave no explanation or description as to what type(s) of teachers. We wanted to account for these teachers, so we created this category. Unfortunately, it doesn't provide us with very much information, other than the fact that a large group of unknown teachers were missed.

Of the 99 schools and LEAs, 53 schools and 64 LEAs said they included at least one non-teacher on their list. **Table 2** shows "other" non-teachers (such as teachers on long-term leave and houseparents who teach their kids at home), librarians, speech therapists, and guidance counselors were among the types of non-teachers most often included in error.

There were several explanations of non-teachers that didn't fit into any of the non-teacher categories. Therefore, we created the "other non-teachers" category to capture those unique non-teachers.

Table 1. Types of Teachers Erroneously Excluded: Public Schools vs. LEAs

Teacher Groups	Number of Schools	Number of LEAs
general full-time / general teachers	22 (51.2%)	30 (62.5%)
part-time teachers	15 (34.9%)	21 (43.8%)
specialized subject matter teachers (i.e. voc. ed., art)	15 (34.9%)	17 (35.4%)
special education teachers	10 (23.3%)	10 (20.8%)
long-term substitutes	6 (14.0%)	10 (20.8%)
itinerant teachers	5 (11.6%)	9 (18.8%)
subject matter teachers (i.e. math, english)	3 (7.0%)	4 (8.3%)

Note: The percentages in the table add to over 100 due to schools and LEAs excluding more than one type of teacher. The bases used are the number of schools and LEAs excluding at least one teacher (43 schools and 48 LEAs).

Table 2. Types of Non-teachers Erroneously Included: Public Schools vs. LEAs

Non-teacher Groups	Number of Schools	Number of LEAs
"other" non-teachers	11 (20.8%)	18 (28.1%)
librarians	18 (34.0%)	10 (15.6%)
speech therapists	18 (34.0%)	10 (15.6%)
guidance counselors	9 (17.0%)	14 (21.9%)
principal / asst. principal	3 (5.7%)	6 (9.4%)
other school staff (i.e. secretary, social worker)	4 (7.5%)	5 (7.8%)
pre-kindergarten	2 (3.8%)	4 (6.3%)

Note: The percentages in the table add to over 100 percent due to schools and LEAs excluding more than one type of teacher. The bases used for the percentages are the number of schools and LEAs excluding at least one teacher (53 schools and 64 LEAs).

3.1.2 Teacher Listing Record (TLR) vs. School Questionnaire

We examined 198 schools (100 public and 98 private - we were unable to contact two private schools for the reconciliation) that completed both a TLR and a School Questionnaire. Of these, 72 TLRs and 59 School Questionnaires excluded at least one teacher from their teacher count. **Table 3** shows that respondents failed to report part-time teachers significantly more often than other types of teachers using both the TLR and the School Questionnaire.

Although the schools included several types of non-teachers in error using the TLR, **Table 4** shows the instances appear to be few and fairly spread out amongst several categories. While using the School Questionnaire, however, the respondents included librarians, "other" non-teachers, and pre-kindergarten teachers in error the most. Interestingly, of the 17 schools that erroneously included pre-kindergarten teachers using the School Questionnaire, the private schools did it significantly more often than the public schools (13 and 4, respectively).

3.2 Teacher Counts: Public Schools vs. LEAs

We compared the number of teachers in the school as reported by the school to the actual count of teachers in that school. We did the same with the LEA. We then looked at how many times each agreed with the actual count, and also how many times each agreed within ± 5 percent of the actual count.

Table 5 shows two-thirds (66 of 99) of the counts reported by the schools were within ± 5 percent of the actual count of teachers in the school. However, only about half (47 of 99) of the LEA reported counts were within ± 5 percent of the actual count of teachers in the school. The 66 schools is significantly greater than the 47 LEAs. This suggests that the public schools are more accurate listing teachers than their corresponding school district (LEA), at least for the schools in this study.

3.3 Teacher Counts: Teacher Listing Record (TLR) vs. School Questionnaire

We also wanted to find out whether the TLR or the School Questionnaire was a better instrument for obtaining the number of teachers in the school. In the 1990-91 SASS the teacher file weights (counts from the TLR) were adjusted so they equaled the teacher estimate (head count) from the school file (School Questionnaire count). This was done to make the

Table 3. Types of Teachers Erroneously Excluded: Teacher Listing Record (TLR) vs. School Questionnaire

Teacher Groups	Number of TLRs	Number of School Quest.
part-time teachers	27 (37.5%)	31 (52.5%)
general full-time / general teachers	15 (20.8%)	21 (35.6%)
special education teachers	11 (15.3%)	3 (5.1%)
specialized subject matter teachers (i.e. voc. ed, art)	10 (13.9%)	2 (3.4%)
subject matter teachers (i.e. math, english)	9 (12.5%)	1 (1.7%)
Chapter 1 teachers	6 (8.3%)	4 (6.8%)
itinerant teachers	3 (4.2%)	5 (8.5%)

Note: The percentages in the table add to over 100 percent due to schools excluding more than one type of teacher. The bases used for the percentages are the number of TLRs and School Questionnaires excluding at least one teacher (72 TLRs and 59 School Questionnaires).

Table 4. Types of Non-teachers Erroneously Included: Teacher Listing Record (TLR) vs. School Questionnaire

Non-teacher Groups	Number of TLRs	Number of School Quest.
librarians	8 (25.8%)	17 (22.4%)
"other" non-teachers	4 (12.9%)	18 (23.7%)
pre-kindergarten teachers	4 (12.9%)	17 (22.4%)
principal / asst. principal	4 (12.9%)	9 (11.8%)
guidance counselors	2 (6.5%)	8 (10.5%)
speech therapists	5 (16.1%)	4 (5.3%)
other school staff (i.e. secretary, social worker)	2 (6.5%)	7 (9.2%)

Note: The percentages in the table add to over 100 percent due to schools excluding more than one type of teacher. The bases used for the percentages are the number of TLRs and School Questionnaires excluding at least one teacher (31 TLRs and 76 School Questionnaires).

SASS estimated teacher counts from the School Questionnaire and TLR more consistent. Our hypothesis, however, was that the TLR would provide a more accurate count, since the respondent must list individual teacher names. The School Questionnaire simply asks for an overall "head count" of teachers in the school.

For each school, we compared the number of teachers in the school as reported using the TLR to the actual count of teachers in the school. We did the same for the School Questionnaire. We then looked at how many times each agreed with the actual count, and also how many times each agreed within ± 5 percent of the actual count.

Table 6 shows 70 percent (123 of 176) of the counts obtained using the TLR were within ± 5 percent of the actual count of teachers in the school. Only about 35 percent (61 of 176) of the counts obtained using the School Questionnaire were within ± 5 percent of the actual count of teachers in the school. The 70 percent using the TLR is significantly greater than the 35 percent using the School Questionnaire. This suggests that, for the schools in this study, the TLR is a better instrument than the School Questionnaire at getting a reliable count of teachers.

4. The Revised Teacher Listing Record

In the 1987-88 and 1990-91 SASSs, we obtained a list of teachers in each school from the school, not the LEA. Since the study suggests the schools are more accurate, we did the same for the 1993-94 SASS. Although the schools were not completely accurate, they were more accurate at listing teachers than their corresponding LEA. Because of this, we plan to continue to use the public schools, rather than the LEAs to obtain these lists.

The results of the TLVS gave us some insight on how to improve the TLR. We made substantial changes to the form for the 1993-94 SASS.

The instructions are more concise and easier to read. We feel that the changed wording made it easier for the respondent to decide who should and should not be included in the list of teachers. We felt that respondents were confused whether to include on the list a person who teaches sometimes, but mostly has non-teacher duties (i.e., a principal, a guidance counselor, a speech therapist, a librarian, etc.).

The TLR used during the TLVS stated to "... include full-time and part-time teachers whose MAIN assignment at this school is teaching." It also stated to "... exclude the principal or school administrator, regardless of whether he/she teaches ..." and "... exclude any staff member whose MAIN assignment at this school is an administrator, guidance counselor, ... or other position in which the major responsibilities are not teaching." We think the phrase "MAIN assignment" may have confused respondents. Also, we

Table 5. School and LEA Counts Compared to the Actual Counts

Difference from Actual Count	Number of Occurrences	
	school count	LEA count
Zero percent difference (complete agreement)	33 (33.3%)	17 (17.2%)
0 < difference ≤ 5 %	33 (33.3%)	30 (30.3%)
difference > 5 %	33 (33.3%)	52 (52.5%)
total	99	99

Table 6. TLR and School Questionnaire Counts Compared to the Actual Counts

Difference from Actual Count	Number of Occurrences	
	TLR count	School Quest. count
Zero percent difference (complete agreement)	106 (60.2%)	45 (25.6%)
0 < difference ≤ 5 %	17 (9.7%)	16 (9.1%)
difference > 5 %	53 (30.1%)	115 (65.3%)
total	176	176

Note: The total does not add up to 200 (100 public schools, 100 private schools) because we couldn't determine the actual count of teachers for 24 schools (12 public, 12 private).

think respondents may have been confused with who qualifies as a part-time teacher.

The instructions on the revised TLR used during the 1993-94 SASS were more specific in addressing these concepts. The instructions stated to "INCLUDE ON THE LIST: part-time teachers (including those who may teach only one class each week)," and "persons who teach a regularly scheduled class but whose main assignment is: principal or vice principal, guidance counselor, ..." It stated to "OMIT FROM THE LIST: persons who do not teach any regularly scheduled classes and whose main assignment is: principal or vice principal, guidance counselor, ..." These revised instructions help the respondent decide whether or not to list the person on the TLR.

The Census Bureau's Center for Survey Methods Research (CSMR) is conducting cognitive research on the revised TLR. The results will be available in the fall of 1994. We will use what we find from this to again revise and improve the TLR. We plan to test this TLR prior to the 1997-98 SASS.

References

Forsyth, B.H. and Lessler, J.T. 1991. "Cognitive Laboratory Methods: A Taxonomy" in P.B. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman (eds), Measurement Errors in Surveys. New York: John Wiley and Sons.

Kaufman, S. 1991. "1988 Schools and Staffing Survey Sample Design and Estimation." Technical Report. National Center for Education Statistics. U.S. Department of Education. Office of Educational Research and Improvement. May, 1991.

Kaufman, S. and Huang, H. 1993. "1990-91 Schools and Staffing Survey: Samplé Design and Estimation." Technical Report. National Center for Education Statistics. U.S. Department of Education. Office of Educational Research and Improvement. July, 1993.

Royce, D. 1993. "1993 Teacher List Validation Study (TLVS) for SASS: A Description." Internal Census Bureau memorandum. September 16, 1993.

Royce, D. 1994. "The Results of the 1993 Teacher List Validation Study (TLVS)." Internal Census Bureau Report. June 8, 1994.

Shapiro, G.M. 1992. "Sampling Specifications for Private Schools in the Schools and Staffing Surveys (SASS) Teacher List Validation Study and the Private Schools Survey Fiscal Component Feasibility Study." Internal Census Bureau memorandum. August 11, 1992.

Shapiro, G.M. 1992. "Sampling Specifications for Public Schools and LEAs in the Schools and Staffing Surveys (SASS) Teacher List Validation Study - Revised." Internal Census Bureau memorandum. September 14, 1992.

Acknowledgements

The authors would like to thank Charles Alexander, John Bushery, Deborah Fenstermaker, Karen King, Dennis Schwanz, and Gene Sebron for commenting on this paper.

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.

IMPROVING THE COVERAGE OF PRIVATE ELEMENTARY-SECONDARY SCHOOLS

Betty J. Jackson, Richard L. Frazier, Karen E. King, Dennis J. Schwanz, Bureau of the Census[1]
Betty J. Jackson, Bureau of the Census, Washington, D.C., 20233

Key Words: Data, Collection, Evaluation, Education

1. GENERAL

In September of 1986, members of the National Center for Education Statistics (NCES) along with Weststat and the Census Bureau met to discuss the formulation of a new survey to gather information, nationally, about public and private elementary and secondary schools in the United States. As a result the Schools and Staffing Survey was created. The Schools and Staffing Survey is a network of surveys that evolved from one survey. They include:

- Schools and Staffing Survey (SASS)
- Teacher Followup Survey (TFS)
- Private School Survey (PSS)

This paper attempts to address one component in updating the universe for the private school frame, the "List Frame".

Definition: Private schools in SASS are institutions which provide educational services for any of grades 1-12, have one or more teachers to give instruction, are not administered by a public agency and are not operated in a private home.

2. HISTORY

2.1 Private School Universe Creation

The Private School Universe was created in 1987 to select the private school sample for the Schools and Staffing Survey. The base for the private school universe is the Quality Education Data (QED) file. It is a commercial list of private schools compiled from handbooks, annual directories, and other materials which list private schools.

NCES purchased the file of private schools from the QED and provided it to the Census Bureau. In an attempt to improve coverage of private schools, the Census Bureau conducted two coverage improvement operations, (1) the "List Frame" consisting of contacting 17 national private school associations and obtaining from each a list of all schools affiliated with them; and (2) the "Area Search Frame" consisting of selecting 75 Primary Sampling Units (PSUs) (consisting of 94 counties).

2.2 Update of the Private School Universe

List Frame

Definition: Affiliation Lists are lists of private schools on the rolls of a specific private school association. These schools are affiliated with that association.

Between 1987 and 1992 the Census Bureau conducted three List Frame operations to update the private school universe. The first "List Frame" operation began in January 1987. Its purpose was to provide further coverage for the private school frame for SASS. NCES provided the Census Bureau with 22 private school associations to contact and obtain lists of schools from them. The Census Bureau then contacted these private school associations and asked for lists of their schools. The Census Bureau sent an explanation letter for the survey to the associations along with the request for their lists. We received 17 of the 22 lists requested.

Once the lists were received, we clerically matched them to the private school universe (QED). The match was done on school name, address and telephone number. The 1987 PSS operation resulted in 1,437 adds to the private school universe.

2.3 1989-90 Private School Survey

The Private School Survey (PSS) is a CENSUS of private elementary and secondary schools in the country. The purpose of the survey is to:

- build a universe frame of private schools that is of sufficient accuracy and completeness to serve as a sampling frame for other NCES private school surveys
- to generate bi-annual data on the total number of private schools, teachers and students.

The survey is conducted bi-annually. There were approximately 25,000 private schools contacted in the first PSS. Schools must be privately administered and contain at least a grade between 1 and 12 in the school to be classified as a private school in PSS or

SASS (see definition of private school on page 1). All schools are sent a questionnaire obtaining information about number of teachers, students, religious orientation, and association.

The first PSS was conducted in 1989-90. To prepare for the survey, we conducted a second coverage improvement operation on the private school universe. This consisted of a List Frame operation and an Area Search Frame operation, as was done for the 1988 SASS.

1989 List Frame Operation

The second List Frame operation for updating the private school universe began in March of 1989. Twenty-three affiliations were contacted to determine how many schools were associated with them. Due to budget constraints not all of the 23 affiliation lists were requested. We only requested affiliation lists from 12 of the associations. Eight of the 12 affiliations selected had sent lists in the first List Frame in 1987. Four affiliations sent lists for the first time. QED sent an updated list.

Our decision on which lists to request was based on the size of the lists. We chose association lists that were not too large because matching and unduplication are expensive. The largest list that we obtained contained about 2000 schools. Affiliations such as "Accelerated Christian Education" who reported 5000 schools were not requested to send a list.

The list frame was conducted similar to the one in 1987 with some minor changes. For the 8 affiliations that provided lists in 1987, we asked for updates (births and deaths) to those lists. If that was not possible, we took the complete list. We clerically matched the schools on the lists to the current private school universe. Non-matched schools to the universe were keyed to a separate file. After some editing was conducted, the file was merged with the universe.

2.4 1991-92 Private School Survey

The second PSS was conducted starting in 1991-1992. To prepare for it we updated the private school universe again. In the spring of 1991, we conducted a third List Frame operation.

1991 List Frame

The 1991 List Frame operation was more extensive than the first two. In 1991 we contacted 26 private school associations, the 50 states and the District of Columbia, QED and a private vender "Jostens" to obtain lists of private schools.

This time the budget was not a problem so we could do a matching and unduplicating operation on all 26 association lists and the lists from the 50 states and the District of Columbia as well as QED and Jostens.

Some state lists were on electronic files while others were in the form of books. Jostens sent a printout of their schools.

3. GOALS/OVERVIEW OF THE 1991 LIST FRAME UPDATING ANALYSIS

We will determine the characteristics of the list frame by religious orientation (Catholic, other Religious, Nonsectarian), school level (elementary, secondary, combined) and total student enrollment. We will be able to describe a typical list frame add.

Also, we will determine the characteristics of the list frame adds by cross-tabulating school characteristics (i.e., religious orientation by school level) and total student enrollment.

Finally, we will determine the effect of the list frame adds on private school characteristics as well as for cross-tabulations of school characteristics. The statistic of interest in this analysis is the percentage of the list frame universe estimate of each characteristic that is represented by the list frame adds (i.e., the numerator will be the list frame adds estimate of the characteristic and the denominator will be the list frame universe (original universe plus adds) estimate of the characteristic). We will show how the universe benefits from the list frame adds in general and by school characteristic.

4. ANALYSIS OF LIST SOURCES FOR ADDITIONS TO THE PRIVATE UNIVERSE

There are four main sources of lists that we contact when it is time to update the private school universe. These sources are the states (i.e. each of the fifty states plus the District of Columbia), the associations, Josten Education Data, and QED. We want to identify which sources of lists provided us with the most up-to-date and complete information about the types of school births we need. Our goal will be accomplished by answering the following questions.

- Which source provided the largest quantity of eligible or in-scope additions to the private universe?
- Which source provided the eligible or in-scope additions with the highest interview rate?

- Which source provided the largest quantity of ineligible or out-of-scope additions?
- Which source had the highest out-of-scope rates?

NOTE: If a school was found on more than one list then it was counted in the table for each list. In other words, if a school was found on a State list and on the Jostens list, that school was counted twice.

4.1 Highlights

- Evidence indicates that the lists from the states and the associations provide the highest quality and the largest quantity of additions to the universe for PSS than either the Quality Education Data or Josten Education Data lists.
- The fifty states and D.C. provided 8 out of 10 total additions to the private universe during the 1991 update. Among the individual state lists 7 out of 10 state additions came from California, Pennsylvania, New York, Florida, Illinois, New Jersey, Michigan, North Carolina, Indiana, Virginia, Georgia, and Wisconsin. These states were the heaviest providers of eligible schools.
- Twenty out of the forty-four association lists requested provided additions to the private universe. Their contribution to the private universe is on a smaller scale than the state lists. They have the highest out-of-scope rate but requesting the lists is good for public relations.
- The Quality Education Data and Josten Education Data lists make a minimal contribution to the private universe because most of their schools show up on either the state or association lists. Despite their small numbers, they have good in-scope school rates and good interview rates.

4.2 State Lists

Looking at the effect of state lists at the national level of in-scope, out-of-scope, and interview rates, roughly 84.2% of the 4,915 in-scope cases came from the State lists. The percentage of the 2,637 out-of-scope cases from this source is similar to the in-scope percentage given above. The top three out-of-scope reasons for State lists (excluding the "Other" category) is "School Closed" at 28% followed by "Duplicate" at 16.7% and "Private Home" at 10.7% The interview rates for the in-scope additions coming from the various state lists was 95.7%.

At the state level, the contributions made to the update differed by state. When we rank the states from largest to smallest contributors of additions, we find the following results. The top sixteen states listed are heavy contributors providing an above average number of schools (at least 121 schools) to the total state additions. After the lists were clerically matched to the current private universe, the top sixteen states account for 73% of the state additions. Approximately 2/3 or more of each of these 16 state's additions were eligible or in-scope with two exceptions: Arizona at 31% and Maryland at 52%. Of the schools in-scope, each state had at least a 90% interview rate. Thus, in general these heavy contributing states provided quality additions as well as a large quantity of additions.

For the remaining 35 states, their contribution was lighter to the overall total of state additions. Alaska, Maine, and North Dakota still had more than 50% of their lists remaining after unduplication with the universe, demonstrating the undercoverage we had in these states. Unfortunately, we found after interviewing that Alaska's and North Dakota's in-scope rates (15.2% and 19% respectively) were the lowest of all 50 states and District of Columbia. For the majority of light contributor states the in-scope rates and the interview rates were comparable to the heavier contributors mentioned above.

4.3 Association Lists

At the national level the percentage of the 4,915 in-scope cases coming from associations was 11.4%. The percentage breakdown of the 2,637 out-of-scope cases is roughly 15%. But 4 out of 10 schools contributed by the Association lists turned out to be out-of-scope after interviewing. Among the out-of-scope reasons for associations lists, "school closed" at 28.5% was number one (excluding other) but "Duplicate" has become a close second at 27.7% and "Private Home" at 4.8% as number three. The interview rates for the in-scope additions among the association lists was 95.7% (tied with state lists).

We ordered the 20 association lists that provided any additions from biggest to smallest provider.

The first eight association lists are the heavy contributors; providing an above average number of school (at least 48 schools) to the total association additions. These associations were:

- National Catholic Education
- National Association of Episcopal Schools
- General Conference of Seventh-Day Adventists
- National Independent Private School Association
- American Montessori Association
- National Center for Neighborhood Enterprise

- National Society for Hebrew Day Schools
- American Association for Christian Schools

They alone account for 76% of the association additions. The lists from these associations provided good quality additions as well as a large quantity. The impact of the list additions on the universe total for the majority of the associations was between 13-35 percent with one association at 92% (the National Center for Neighborhood Enterprises). The biggest contributor, National Catholic Education Association, has the smallest percentage of list additions on the universe at 2%.

The remaining twelve association lists were fairly light in the contribution to the total association additions as well as to their associations' total on the universe. New list additions as a percentage of the universe ranged from 4-16 percent with one exception at 100% the General Council Agudath Israel of America (probably the first time this list has been provided to us). This range is lower than the majority of heavier contributor's percentages (13-35). Yet all are larger than the impact percentage for the heaviest contributor; the National Catholic Education Association. For these smaller providers, the importance of these lists to these associations outweighs the fact that they provided only a small quantity of additions.

The in-scope rates (50%-100%) and interview rates (80%-100%) were similar for the heavy and light contributors with two exceptions. The National Association of Episcopal Schools (in-scope rate of 12.5%) and the National Center for Neighborhood Enterprise (in-scope rate of 28%), are among the top eight contributors with the smallest in-scope rates. However, at least 30% of the schools on the universe for these associations came from the list updating operation.

Requesting these lists may do more than just update the universe. List requests from associations may promote good public relations with the association heads and they in turn may encourage participation among their member schools.

4.4 Josten and Quality Education Data Lists

The Quality Education Data (QED) and the Josten lists are relatively small in term of the impact on the overall number of new list frame additions. The original QED list provided 49 school births. Only 20 were left after clerical unduplication with the existing universe. The Josten list provided 431 school births. Three hundred and six births were left after clerical unduplication with the existing universe.

The percentage breakdown of the 4,915 in-scope cases by these sources are QED at 0.3% and Josten's at 4%. The percentage breakdown of the 2,637 out-of-scope cases for these source is similar to the in-scope breakdown given above. The out-of-scope reasons most prevalent (excluding the "Other" category) are "school closed" and "duplicate". The interview rates for the in-scope additions among the two sources are QED list at 100% and Josten's list at 91.9%.

These lists come from professional list builders who supposedly use many of the resources we use. Since our resources are similar, overlap or duplication between them and the state/association lists becomes common. Refer to the next section for details.

4.5 List Overlap

Of the 20 schools obtained from QED, 14 were also on one of the state and/or association lists. Of the 6 schools found only on the QED list, 5 were out-of-scope leaving only one original QED school eligible for PSS.

Of the 306 schools obtained from Josten's, 72 were also on one of the state and/or association lists. Of the 234 schools found only on the Jostens list, 103 were out-of-scope.

The association list's overlap with the states' lists is about 30% of the total additions from the association lists. Why is it not higher? States have different criteria for licensing their private schools. Some states may exempt schools associated with churches to be licensed. Some states may list only a central administrative office, where the association lists would offer each site location associated with the administrative office. Both types of lists are needed to ensure coverage.

5. ANALYSIS OF THE CHARACTERISTICS OF ADDS AND THEIR IMPACT

5.1 Highlights

- Other Religious adds make up the largest percentage of adds for all variables (schools, students, teachers, graduates, and projected graduates) across all religious orientation categories.
- Combined school adds make up the largest percentage of adds for all variables (schools, students, teachers, graduates, and projected graduates) across all school levels.
- Updating had a big impact on Nonsectarian and Other Religious schools, but very little impact on

Catholic schools.

- Updating had the biggest impact on elementary schools although the impact on combined and secondary schools was significant as well.
- Updating had the biggest impact on the smallest schools. The impact decreased as the size of the school increased.

5.2 Goals

- Describe a typical list frame add.
- Show how the universe benefits from the list frame adds in general and by school characteristics.

5.3 Characteristics of Adds

Small schools contribute more significantly to the list frame adds than the larger ones. The overall percent contributions for schools for each of the size categories for the list frame adds schools are as follows: 0-75 students: 67%, 76-150 students: 18%, 151-225 students: 6%, 226 + students: 8%.

In general these percents hold true (in magnitude and direction) for each religious orientation and school level. The exception is the Catholic schools -- where the larger schools contribute more significantly (0-75 students: 20%, 76-150 students: 19%, 151-225 students: 19%, 226 + students: 40%).

The overall pattern for students, teachers, graduates, and projected graduates in the various size categories is similar to that of Catholic schools. It shows that the larger schools contribute a greater number of adds.

Graduates are defined as students who have already received a regular high school diploma. Projected graduates are defined as students who are expected to receive a regular high school diploma.

In general, the same size pattern as seen for Catholic schools holds for students, teachers, graduates, and projected graduates in the different size categories across religious orientation and school level. The exceptions are the following: students in Nonsectarian and elementary schools, and teachers in Other Religious, Nonsectarian, elementary, and secondary schools. Here the pattern is similar of the overall pattern for schools in the different size categories.

Other Religious adds contributed 2,688 schools (62%) of all school adds in the 1991 PSS list frame updating operation. This was followed by 1,430 Nonsectarian school adds (33%) and then 215 Catholic school adds (5%).

The pattern for schools across religious orientation is similar for the other four variables (students, teachers, graduates, and projected graduates).

Combined school adds contributed 2,926 schools (67%) of all school adds in the 1991 PSS list frame updating operation. This was followed by 1,107 elementary school adds (25%) and then 323 secondary school adds (7%).

These patterns are similar for the other four variables (students, teachers, graduates (when valid), and projected graduates (when valid).

In general, the patterns mentioned earlier for the different religious orientation and school level subgroups across all five variables (schools, students, teachers, graduates, and projected graduates) are the same when these variables are cross-tabbed. The exception is when the Catholic subgroup is cross-tabbed with school level. For this subgroup, Catholic secondary schools contribute more significantly than Catholic elementary schools.

Also, when religious orientation and school level are crosstabbed, the general trend by size of school (i.e., the smaller list frame schools contribute more significantly than the larger ones) is not as strong as before.

5.4 Impact of Adds on Private School Characteristics

The list frame adds represented 18% of schools, 8% of students, 11% of teachers, and 6% of both graduates and projected graduates. These percentages varied considerably for religious orientation and showed that this updating had a substantial impact on improving coverage of Nonsectarian and Other Religious schools and very little impact for Catholic schools. Nonsectarian led the way with 31% for schools, followed closely by Other Religious at 26%, and Catholic's considerably smaller 3%. These percentages were reduced somewhat for each religious orientation when you look at students, teachers, graduates and projected graduates. However, the general relationship seen for schools still held up in that the percentages for Nonsectarian and Other Religious were very close and significantly outdistanced the very small Catholic percentages. These percentages ranged from 11% to 18% for Other Religious, 10% to 17% for Nonsectarian and 2% for Catholic.

The previously-described relationship among religious orientation for schools, students, teachers, graduates and projected graduates generally held up within each school level as well with just a few exceptions. One exception was for combined students where the Nonsectarian percentage (37%) was

substantially larger than the 14% for Other Religious students. The other exceptions were for combined graduates and projected graduates where the 6% and 7% for Catholic was much closer to the corresponding percentages for the other religious orientation categories (13% for Other Religious and 9%-10% for Nonsectarian).

The school level percentages showed less variation and indicated that the list frame updating had a substantial impact on improving the coverage for all three school levels. Elementary schools lead the way with 26% for schools, followed by 17% for combined schools and 14% for secondary schools. As was seen for religious orientation, these percentages were reduced somewhat when looking at the other statistics (i.e., students, teachers, graduates and projected graduates) but this relationship seen for schools held up for all the other statistics. These percentages ranged from 17% to 19% for elementary, 8% to 11% for combined, and 3% to 6% for secondary.

The previously-described relationships among school levels for schools, students, teachers, graduates and projected graduates were generally seen within each religious orientation as well with just a few notable exceptions. One exception was for Nonsectarian students where the combined percentage (37%) was larger than the 28% for elementary and 11% for secondary. The other exceptions were for graduates and projected graduates for both Other Religious and Nonsectarian where the percentages for secondary and combined were much closer than those over all religion orientation categories.

The enrollment percentages showed considerable variation and reflected a very strong inverse relationship between the size of the school and the impact of this updating on improving the coverage. The smallest schools (0-75 students) led the way at 38% for schools indicating the updating had a very substantial effect on the coverage of these small schools. The second smallest schools (76-150 students) had the next largest percentage (16%), followed by 7% for 151-225 student schools and 5% for the largest schools (226 + students).

Unlike what had been seen for religious orientation and school level, the enrollment percentages for students, teachers, graduates, and projected graduates were similar to those for schools. This very high percentage for the smallest schools and the very strong inverse relationship between enrollment and the impact percentages also existed within each of the religious orientation and school level categories except the percentages for the smallest Catholic school were not very high. This enrollment relationship was also true within each of the school level categories for Nonsectarian and Other Religious schools. However, the inverse relationship was not always as strong and the percentages were not always as high for the Catholic school level categories.

VI. CONCLUSION

Evidence indicates that the state and association lists contributed more significantly to the quality and quantity of the universe for PSS than either the QED or Jostens list.

We should continue to collect lists of private schools from all the states in the future. We should give high priority to the lists from California, Pennsylvania, New York, Florida, Illinois, New Jersey, Michigan, North Carolina, Indiana, Virginia, Georgia, and Wisconsin who are heavy contributors of quality list adds.

We should also continue to collect lists of private schools from the associations in the future. The association lists do contribute to the universe on a smaller scale than the state lists. Requesting these lists may do more than just update the universe. List requests from associations may promote good public relations with the association heads and they in turn may encourage participation among their member schools.

ADDING VALUE TO THE VALUE-ADDED EDUCATIONAL PRODUCTION FUNCTION SPECIFICATION

Karen Akerhielm, Mathtech, Inc.
Suite 111, 202 Carnegie Center, Princeton, New Jersey 08540

KEY WORDS: NELS, Economics of Education

Introduction

The major criticism of using data from one year only in estimating an education production function is that there is no control for initial abilities or past learning experiences (Hanushek 1986). That is, since education is a cumulative process, school resources in a given year may not be affecting student achievement independently of the child's ability or of the school resources received by the child in previous years. One standard solution in the literature to the problem of using a "snapshot" of data from a single point in time has been to transform the basic cross-section regression model into a value-added specification. Instead of regressing an achievement measure (i.e., test score) from time t on a series of available inputs from time t, a test score from a previous time period is added to the model as an independent right-hand side covariate as a means to introduce "initial conditions" into the equation.

It is claimed that the initial test score must enter the value-added production function as an independent variable in order to control for omitted variables such as past learning experiences and initial ability. "Without such a measure our efforts are like attempting to measure the effectiveness of a beauty parlor without knowing what the clientele looked like to begin with," (Bowles 1970, p. 26). Thus, the purpose of the value-added specification is to estimate the effects of various inputs on student achievement, *given* past learning and any previously-determined abilities captured by the initial test score.

This paper argues that the conventional value-added model is misspecified since the initial test score is not exogenous. Not only will its own coefficient be unstable and uninterpretable, to the extent it is related to the other regressors, it will bias the other parameter estimates as well. Using the National Education Longitudinal Study (NELS), a new and extensive data set from the U.S. Department of Education, I propose and implement a new technique in this paper for a value-added educational production function specification that accounts for the endogeneity of initial ability. By comparing the two methods, I show that the conventional value-added model may mask the significant effect of school resources, such as teacher experience and class size, because of the possible misspecification caused by including the initial test score as an exogenous independent variable.

The Model

Using the NELS framework of eighth and tenth grade data, the "conventional" value-added specification can be written as follows:

$$(1)\ Y_{i10j} = \alpha + \beta_1 Y_{i8j} + \beta_2 X_{i10} + \beta_3 X_{i10j} + \epsilon_{i10j}$$

where Y_{i10j} is student i's test score in the tenth grade in subject j (where j equals math, reading, science and history), Y_{i8j} is student i's test score in the eighth grade in subject j, X_{i10} are those characteristics of student i in the tenth grade that are not subject-specific (such as family income, parental education, family composition, urbanicity of school, and sex and race of the student), X_{i10j} are those characteristics of student i in the tenth grade that are subject-specific (race, sex and years of experience of student i's teacher in subject j, and student i's class size in subject j), and ϵ_{i10j} is an unmeasured component that includes inputs such as innate ability and motivation that are not captured by the other variables. The error term can be thought of as "unobserved test-taking ability."

The purpose of estimating equation (1) is to determine what the effects are of school resources, such as class size and teacher experience, after controlling for family background characteristics (by including independent variables such as income and education) and after controlling for past learning experiences and initial ability (by including the eighth grade test score). However, this model is clearly misspecified because Y_{i8j} is not exogenous; its covariance with the error term is nonzero. Since Y_{i8j} embodies the effects of unobserved omitted inputs that are incidentally correlated with the included X terms, this may lead to biased estimates of the β's.

The method I propose in this paper is to instrument Y_{i8j} as a function of inputs in the eighth grade. Using a two-stage least squares (2SLS) framework, I then employ the predicted value of past achievement as the independent variable in the equation. In this way, the interim school, home and community inputs, if they have changed from eighth to tenth grade, would motivate the model's dynamics and permit the model to explain final achievement while avoiding statistically biased results. This method represents a departure from the current literature since it includes the predicted value of past achievement as a right-hand side variable instead of using the actual eighth grade test score itself. Equation

(1) can thus be rewritten as:

$$(2)\ Y_{i10j} = \alpha + \beta_1 \hat{Y}_{i8j} + \beta_2 X_{i10} + \beta_3 X_{i10j} + \epsilon_{i10j}$$

where the instruments for Y_{i8j} are family composition as of the eighth grade, urbanicity of student i's school in the eighth grade, the race, sex and experience of student i's teacher in eighth grade in subject j as well as the class size of student i's eighth grade class in subject j. The components of X_{i10} and X_{i10j} serve as their own instruments.

Estimation Results

This paper concentrates on public school students only (i.e., students who were in public school in both the eighth and tenth grade) since I have shown previously that the public and private school students in the NELS data set vary systematically from one another and thus the data on these students should not be pooled without a sample selection correction factor (Akerhielm 1993).

The dependent variable I use in the estimation of equation (2) is the tenth grade IRT (item response theory) test score. The IRT score is a transformation of the raw score (total number of right answers) such that scores in the two years are made comparable by placing them on a continuous scale. Specifically, the purpose of IRT is to calculate scores that could be compared regardless of which test form a student took. IRT compensates for the possibility of a low ability student guessing several hard items correctly, and it makes possible measurement of the gain in achievement from grade eight to grade ten even though the tests used were not identical at the two points in time (NCES January 1992). The IRT scores may be especially appropriate when estimating effects on math and English achievement since the tests for these subjects had more than one version in the follow-up year.

I break down the following analysis into all four curriculum areas to allow for the use of classroom and subject-specific data which were unavailable in older NCES data sets such as HS&B. The independent variables are those listed above under the discussion of equation (1). The two tenth grade school resource variables this paper focuses on are teacher experience (in years) and class size. The class size variable is constructed by taking the average class size in a given subject for all students in a school that responded to the NELS survey. This measure allows for the use of classroom-specific and subject-specific data (unlike the use of a pupil-teacher ratio), while avoiding problems of nonrandom allocation of students by ability into different class sizes (Akerhielm forthcoming). The instruments I use to predict the eighth grade IRT test score are the eighth grade inputs listed under equation (2).

Since education is a cumulative process, school resources in a given year may not be affecting student achievement independently of past school resources or of the student's initial ability. Thus, the focus of this estimation is on the effects of tenth grade teacher experience and class size on student achievement in the tenth grade, *given* the effects of past school resources and initial ability as embodied in the eighth grade test score, and controlling for family background.

Table 1 contains the value-added regression results. The first two columns of parameter estimates for each subject represent the "conventional" value-added model, as depicted in equation (1). The last two columns for each subject constitute the results from running the proposed model specification of equation (2), in which the eighth grade test score is a predicted value. Once again, the major difference in specification is that the conventional model uses the actual eighth grade test score as a right-hand side independent variable whereas the model proposed in this paper instruments the eighth grade score as a function of inputs in the eighth grade.

A number of points can be made from comparing the two methods. First, in the conventional value-added model, teacher experience only has statistically significant effects for history achievement. However, when using the new method, the effect of teacher experience increases in magnitude substantially and becomes positive and statistically significant (at the five percent level) for all four subjects. According to the results from the proposed model, teacher experience is important in raising the cognitive skills of tenth grade students, conditional on past learning conditions and ability. The same conclusion would not be made, however, when using the conventional method.

Second, the conventional value-added specification does not yield any statistically significant effects of class size on student achievement. Using the proposed value-added method, I find that the effect of class size becomes negative and significant for English/reading and science (at the ten percent level). When using the conventional method, however, one would conclude that there is no systematic relationship between class size and student achievement. As with teacher experience, the size of the effect increases in absolute magnitude for all four subjects (although the sign change for math and history is counter-intuitive) when using the proposed approach.

Third, initial conditions matter. In both models, the effect of the eighth grade test score is positively and significantly related to the tenth grade score; the magnitude of the effect of the initial test score decreases substantially when using the proposed model, as expected. For all four subjects in the proposed specification, the coefficient of the initial test score is significantly different than one.

TABLE 1: VALUE-ADDED REGRESSION RESULTS

	Math (n = 3966) Conventional Model		Math (n = 3966) Proposed Model		English (n = 3884) Conventional Model		English (n = 3884) Proposed Model	
Variable Name	Coeff.	T-stat.	Coeff.	T-stat	Coeff.	T-stat.	Coeff.	T-stat.
Constant	8.92	10.89	10.84	3.67	6.01	9.21	9.82	4.37
Family income (0,000)	.09	2.23	.67	9.72	.14	4.40	.45	10.09
Parent educ (1=>h.s.)	1.07	4.75	3.76	10.10	.67	3.94	2.54	10.64
Fam comp (1=married)	.99	4.20	1.80	3.79	.51	2.87	.66	2.30
Urban (1=yes)	.46	1.67	1.56	3.34	.48	2.34	.94	3.18
Student race (1=white)	.47	1.83	3.53	8.13	.33	1.76	1.96	7.24
Student sex (1=male)	-.09	-.42	.54	1.55	-.22	-1.36	-1.34	-5.95
Teacher race (1=white)	.18	.43	2.42	3.35	.23	.80	.98	2.36
Teacher sex (1=male)	-.65	-2.99	-.70	-1.93	-.32	-1.79	-.89	-3.50
Years of experience	-.01	-.29	.08	3.78	.01	.75	.05	3.59
Average class size	-.04	-1.38	.03	.61	-.04	-1.59	-.06	-1.73
Eighth grade test score	.85	84.40	.43	4.54	.77	62.48	.38	3.31

	Science (n = 3177) Conventional Model		Science (n = 3177) Proposed Model		History (n = 2329) Conventional Model		History (n = 2329) Proposed Model	
Variable Name	Coeff.	T-stat.	Coeff.	T-stat.	Coeff.	T-stat.	Coeff.	T-stat.
Constant	2.68	5.56	1.74	.58	6.11	11.49	8.01	3.20
Family income (0,000)	.14	6.14	.31	9.99	.06	2.08	.28	7.20
Parent educ (1=>h.s.)	.57	4.34	1.69	9.58	.85	5.13	2.18	9.19
Fam comp (1=married)	.32	2.33	.09	.36	.59	3.35	.77	2.36
Urban (1=yes)	.01	.03	.28	1.32	.22	1.06	-.01	-.03
Student race (1=white)	.80	5.36	1.73	8.62	.38	1.92	1.39	4.92
Student sex (1=male)	.72	5.85	1.49	8.98	.17	1.08	.57	2.57
Teacher race (1=white)	.81	3.13	1.47	4.18	-.72	-2.54	.10	.23
Teacher sex (1=male)	-.13	-1.00	-.16	-.91	-.15	-.87	-.30	-1.25
Years of experience	.01	1.55	.04	3.50	.03	2.86	.04	2.94
Average class size	-.02	-1.21	-.03	-1.69	-.01	-.91	.02	1.11
Eighth grade test score	.76	51.42	.60	2.42	.75	50.00	.38	2.52

Fourth, although school resources such as class size and teacher experience are important for some students, the magnitude of effects are small, especially as compared to the impacts of family background variables. For example, increasing teacher experience by one year will increase student achievement by .04 to .08 of a test point, depending on the subject. Likewise, reducing class size by one student will increase student achievement in English and science by .06 and .03 test points, respectively.

Finally, I estimated Hausman tests to examine the question of whether the eighth grade test score is exogenous. The Hausman specification test compares two estimators, the OLS and the 2SLS estimators. In this test, the null hypothesis states that both estimators are consistent but the 2SLS estimator is inefficient. By comparing the estimates from both estimators and noting that their difference is uncorrelated with the efficient estimator when the null hypothesis is true, a chi-square test statistic is derived based on the asymptotic distribution of the difference in the two estimators.

A large chi-square value indicates a large deviation from the null hypothesis. If the null hypothesis is rejected, this implies that the OLS model is misspecified (i.e., there may be a contemporaneous correlation between the eighth grade test score and the error term)

and that the two estimates are not equal. For three of the four subjects (all except science), the chi-square critical value of 3.84 (five percent significance level with one degree of freedom) was exceeded, suggesting that the OLS model is misspecified. That is, the null hypothesis of equality between the two estimators can be rejected for three of the four subjects and the two models can be distinguished on statistical grounds. (The Hausman tests were as follows: Mathematics – 19.9; English – 11.3; Science – 0.4; History – 5.9.)

When I estimated the tenth grade education production function as a cross-section model, without any control for initial ability or past learning experiences, the magnitude and significance of both family background and school resource effects are much higher than in either value-added model. Thus, it is essential to include an indicator of ability in the education production function to control for the links among ability, family and school inputs. Indeed, family and community effects, and to a lesser degree school resource impacts, may be upwardly biased in cross-section models. The question remains, however, as to what form the ability indicator should take.

Future Research

Further research is needed to determine whether other instruments may be more appropriate for instrumenting the initial test score. Research is also needed to determine the possible consequences of attrition bias. There are two potential sources of attrition in the follow-up tenth grade NELS sample. First, due to budgetary constraints that restricted the follow-up survey to 1,500 schools, not all students were followed up two years later. If attrition is not random, and the students who were not re-surveyed differ systematically from those who were, then the model estimates may be biased.

Second, the value-added analysis of this paper examines only those students who had their teacher surveyed in the same subject in both years. In the base year each student had two of their teachers (representing two of the four subject areas) surveyed. Although base year students were randomly assigned the combination of two subject areas, if a given base year student who was re-surveyed was not enrolled in the follow-up year in one or both of his or her preassigned subject areas, subjects were substituted. To the extent that certain subjects (such as science and history) are considered electives at the high school level and that students who take elective course are different from those who do not, the value-added analysis may provide biased estimates. Due to the possibility of attrition bias, the findings of this paper should be subjected to further testing and research.

Conclusions

While this paper upholds the need for a value-added model relative to a cross-section analysis, it questions the indicator commonly used to control for initial ability and past learning experiences. The value-added specification proposed and implemented in this paper estimates the effect of various school resources on tenth grade achievement conditional on past learning and any initial ability captured by a predicted eighth grade test score. The analysis finds that school resources, such as English and science class size and teacher experience in all subjects, affect achievement even after controlling for initial conditions. It was also shown that the value-added model proposed in the literature may obscure the significant effect of teacher and school inputs because of the misspecification of using the actual initial test score as an independent variable. Indeed, the results of this paper may help to explain why past economics research has failed to find any consistent effects of teacher experience and class size on student achievement.

Selected References

Akerhielm, Karen. (forthcoming) "Does Class Size Matter?" *Economics of Education Review.*

Akerhielm, Karen Blinder. (1993) "Comparing Public and Catholic Schools: Different Schools or Different Students?" Presented at the Annual Meeting of the Association of Public Policy Analysis and Management (Structural Change in Education Panel), Washington DC, October 1993.

Bowles, Samuel. (1970) "Towards an Educational Production Function," in *Education, Income, and Human Capital.* ed. W. Lee Hansen. New York: NBER, 11-70.

Hanushek, Eric A. (1986) "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature.* 24(3): 1141-1177.

Hanushek, Eric A. (1979) "Conceptual and Empirical Issues in the Estimation of Educational Production Functions," *The Journal of Human Resources.* 14(3): 351-388.

National Center for Education Statistics, U.S. Department of Education. (1992) *National Education Longitudinal Study (NELS) of 1988: First Follow-Up User's Manual -- Student Component.* NCES 92-007, January 1992.

Teacher Quality in Public and Private Schools

Dale Ballou and Michael Podgursky, University of Massachusetts
Podgursky, Department of Economics, University of Massachusetts, Amherst, MA 01003

Key Words: Education, Teachers, Private Schools

Many studies have compared the performance of public and private schools as measured by student achievement scores (e.g., Hoffer, Greeley and Coleman, 1986; Chubb and Moe, 1990). Far fewer have attempted to compare the quality of teacher inputs in the two sectors, largely due to a paucity of data on private schools. In this paper we analyze principals' assessments of the quality of the teaching staffs in public and private schools using data from the 1990-91 Schools and Staffing Survey (SASS).

Ratings of Teacher Quality

The 1990-91 Schools and Staffing Survey (SASS) was the second in a series which began in the 1987-88 school year, investigating staffing patterns in the nation's elementary and secondary schools. Survey responses were obtained from administrators of 8,969 public schools and 2,620 private schools. Additional information was obtained from a component of the survey sent to teachers in these schools. More than forty-six thousand teachers in public schools and six thousand in private schools responded. Survey items concerning working conditions and job satisfaction generally confirmed the patterns found in the earlier 1987-88 SASS. In particular, salaries differed sharply between the two sectors. The average base salary for public school teachers was $28,591; in the private sector, the corresponding figure was $18,741.

A unique feature of the 1990-91 SASS was an item requesting the school principal to rate the quality of the teaching staff on a five-point scale (poor = 1, excellent = 5). Figures 1 and 2 present principals' quality ratings for new teachers (those with three or fewer years experience) and experienced teachers (more than three years experience). As shown in Fig. 1, ratings of new teachers are similar across all four school types (public, Catholic, other religious, and non-sectarian). The modal response is four in each category. The mean rating in public schools (3.89) is slightly higher than that in any of the three private schools, as is the proportion of schools in which principals rate their new teachers "excellent." Given that teaching salaries are substantially lower in the private sector, it is perhaps surprising that the comparison is as favorable to private schools as it is. The comparison suggests that the workplace amenities as well as the greater freedom of private schools to recruit uncertified personnel largely offset the effects of the salary differential on recruitment.

When we turn to experienced teachers (Figure 2), the comparison becomes even more favorable to private schools. The public school mean rating (4.24) is now below all types of private schools. The proportion of private schools in which the experienced staff is rated excellent is dramatically higher than in the public sector -- almost twice as great among the non-sectarian schools. Again, the comparison suggests that private schools possess other advantages which enable them to recruit effectively, despite paying lower salaries. Moreover, while experienced teachers are rated higher than new teachers in all four types of school, the difference is considerably larger in the private sector. This may reflect more selective retention and/or better staff development, as poor teachers either improve or face dismissal.

These conclusions are, however, tentative, and depend on establishing the comparability of survey responses across sectors. There are two issues. First, to show that private schools benefit from operating in an environment relatively free of state regulation, bureaucratic encumbrances, etc., we will need to demonstrate that the comparatively high ratings received by their teachers are not due to other features of the private school environment. One notable feature is the practice of selective admissions, which enables private schools to recruit comparatively well-motivated and disciplined student populations. Hence some controls for the character of a school's students and the community from which they come are needed.

The second issue concerns the standards by which principals evaluate their staffs. Teacher ratings in SASS are shaped by educational goals and evaluative criteria which may vary widely across individuals and schools. Of course, the mere fact that ratings are subjective does not invalidate intersectoral comparisons, since a purely subjective component will average out in the data. However, evaluative criteria which vary systematically across sectors are of concern. Again, controls for the type of students and for the background and goals of the principal are required. One may wonder, however, if this is enough.

Some terminology will be useful here. We will say that standards are free of <u>sectoral bias</u> if public and private school heads who have similar characteristics would assign, on average, the same ratings to a given set of teachers working under given conditions. (An operational definition of "similar characteristics" appears below.) Conversely, if there are systematic differences in ratings under these conditions, sectoral bias is present.

A Model of Teacher Ratings

For the statistical analysis that follows, we assume a principal's rating of his staff is based on an underlying evaluation of teacher quality which varies continuously. Changes in the observed ratings are triggered when this continuous assessment crosses certain thresholds. Two latent assessments are defined for the i-th school, one pertaining to new (q_{in}) and one to experienced staff (q_{ie}). These latent measures of quality are in turn related to characteristics of the school (S_i), notably the mix of salary and working conditions offered to employees. Schools offering higher pay or a more attractive teaching environment, other things being equal, should succeed in attracting superior staff. Characteristics of the school may also influence the criteria by which teachers are assessed, as noted above.

Numerous items from SASS are included in the model for one or both of these reasons. Among them are school size, the ratio of teachers to students, the type of program provided by the school (general education, vocational, alternative schools, special education, and special emphasis in science, the arts, etc.), location (region as well as degree of urbanicity), the percentage of minority students, and the principal's assessment of the severity of student behavioral problems at the school (carrying of weapons, demonstrations of disrespect toward staff, physical and verbal abuse of teachers, and abuse of drugs and alcohol).

A school's success in attracting good teachers also depends on local labor market conditions (M_i). Variation in these background factors is picked up through indicators of region and of type and size of community, and through cost-of-living indices. In addition, the principal's own personal qualities may influence recruitment and the evaluative criteria applied to staff, so that we include a vector of principal characteristics as well (P_i). "Principal characteristics" is broadly construed to include statements of educational goals as well as demographic variables and measures of education and experience. Three goals are distinguished, depending on the which of several objectives the principal selects as the top priority for his school: academic achievement, moral or religious education, and all others.

We suppose that the latent quality assessment can be represented as a linear function of these variables plus a residual. Thus the quality of new teacersr satisfies

$$q_{in} = S_i \beta_{n1} + M_i \beta_{n2} + P_i \beta_{n3} + \delta_i \alpha_n + \epsilon_{in} ,$$

with an analogous expression for q_{ie}. The residual component of quality is represented as $\delta_i \alpha + \epsilon_{in}$, where α is a vector of sector-specific effects (public, Catholic, other private religious, and private non-sectarian), δ_i a vector of indicator variables picking out the sector to which school i belongs, and ϵ_{in} is an error term. We assume α is free of sectoral bias: that is, differences in the elements of α represent variation in teacher quality which administrators in all sectors would recognize. Of course, if teacher quality does not vary across sectors given S, M, and P, the elements of α will be equal to a common population mean. Our hypothesis is that key features of the environment in which private schools function -- freedom from bureaucratic control, reduced state regulation, non-unionized work forces -- will cause significant differences in the elements of α even after one has taken account of school, market, and principal characteristics.

An observed rating is triggered when the latent continuous assessment exceeds a particular threshold. Let t_j (j = 1,4) denote the four thresholds against which q_{ie} and q_{in} are measured. For example, new teachers are rated "1" when $q_{in}<t_1$, rated "2" when $t_1<q_{in}<t_2$, etc. On the assumption that the error ϵ_{in} is an i.i.d. logistic disturbance with mean zero and unit variance, the parameters β_1, β_2, and β_3 can be estimated by maximum likelihood methods.

Maximum likelihood estimates of the sector coefficients (α_n) are reported in Tables 1 and 2 below (a full set of coefficient estimates is repored in Ballou and Podgursky, 1994). Three variants of the model are shown i each table. Model 1 contains indicators of sector only. In Model 2 we add elements of S, M, and P except for measures of salary. Two measures of teacher salary and the cost of living index are added in Model 3. The salary variables are the pay offered inexperienced teachers with a BA and the pay offered teachers with a master's degree and twenty years experience.

Table 1
Ordered Logit Coefficients: New Teachers

model	(1)	(2)	(3)
Public	—	—	—
Catholic	-100 (.078)	-.349*** (.097)	.015 (.116)
Oth. Religious	-.091 (.061)	-.321*** (.083)	-.031 (.108)
Non-Religious	-.005 (.083)	-.139 (.096)	-.026 (.129)
Other Covariates	none	34	37
sample size	10,878	10,406	9,237

*,**,*** significant at 10%, 5%, and 1% respectively

The sector coefficients in the first column of Table 1 show that ratings of new teachers are slightly lower in each type of private school than they are in the public sector. However, the differences are small and none are significant

statistically. We argued above that the ability of private schools to recruit on more or less equal terms with the public schools reflects the balancing of superior working conditions against lower pay. This conjecture receives strong support from the estimates in this table. When school, community, and principal characteristics are added to the model, the coefficients on the sector indicators fall (column two). Thus, in public and private schools which offer similar working conditions and levels of job satisfaction, publicschools have a significant advantage in recruiting teachers. The source of this advantage is revealed in turn when pay is added to the model, as the differences between sectors are once again small and insignificant (column three).

Estimates for experienced teachers are presented in Table 2. The sector coefficients are large and positive: as noted above, experienced teachers in private schools receive significantly higher ratings than their counterparts in the public sector. The differences remain large even when working conditions and pay are added to the model.

Table 2
Ordered Logit coefficients:
Experienced Teachers

model	(1)	(2)	(3)
Public	—	—	—
Catholic	.714*** (.083)	-.507*** (.104)	.451*** (.124)
Other Religious	.815*** (.065)	.607*** (.089)	.548*** (.115)
Non-Religious	1.148*** (.093)	.844*** (.107)	.026*** (.143)
Other Covariates	none	34	37
sample size	10,878	10,406	9,237

Controlling for Sectoral Bias

Interpretation of the results in Tables 1 and 2 rested on the assumption that there was no sectoral bias in the standards by which teachers are judged. If this assumption is violated, inferences about the comparative quality of teachers across sectors are problematic. We allow sectoral bias to enter the model by respecifying the contents of the vector α as $\alpha + \mu$. The elements of μ are sectoral biases, components of q_i which reflect the sector of origin of the evaluator. (As always, both α and μ are residual components of the quality assessment conditional on S_i, M_i, and P_i.) Since α and μ are combined in a single term, it is no longer possible to determine whether differences in the elements of this vector are due to differences in teacher performance which would be recognized in all sectors or to variation in the standards prevailing in different sectors.

This conclusion is unduly pessimistic, however. If α differs between new and experienced teachers, while μ does not, it is possible to estimate at least the difference $\alpha_e-\alpha_n$ by exploiting the fact that each administrator is observed twice. Precisely this specification is suggested by the pattern displayed in Figs. 1 and 2. While the ratings of experienced teachers exceed those of new teachers in all types of schools, the gap varies across sectors, being widest among the private non-sectarian schools, smallest in the public sector. It is reasonable to suppose that this gap represents a genuine difference in quality (again, as perceived by the principal), since it is unlikely that an administrator would apply inconsistent criteria in evaluating two groups of teachers within the same school.

This would be little value if $\alpha_e-\alpha_n$ held no policy interest. However, the opposite is true. Given that teachers learn on the job, it is to be expected that experienced teachers will outperform new teachers. When the reverse occurs, it is a sign that the school is failing to retain many of the best new teachers and/or to improve the performance of the others. Similarly, the more often experienced teachers are rated above new teachers, the more likely it is that some deliberate policy, either selective retention or staff development, is a contributing factor.

To keep the analysis tractable, we collapse the ratings given teachers to a 2-point scale: less than excellent and excellent. Let y_{in} (y_{ie}) = 1 if new (experienced) teachers in school i are rated excellent, 0 otherwise. The possible outcomes for the ordered pair (y_{in}, y_{ie}) are the set {(0,0), (0,1), (1,0), and (1,1)}. The model of the latent quality assessment is amended to

$$q_{ie} = S_i \beta_{e1} + M_i \beta_{e2} + P_i \beta_{e3} + \delta_i \alpha_e + \delta_i \mu + \epsilon_i + \epsilon_{ie}$$

$$q_{in} = S_i \beta_{n1} + M_i \beta_{n2} + P_i \beta_{n3} + \delta_i \alpha_n + \delta_i \mu + \epsilon_i + \epsilon_{in}$$

in which ϵ_i represents a subjective component of teacher evaluations common to both evaluations (say, the principal is a hard rather than an easy grader). The sum $\delta_i\mu + \epsilon_i$ can be regarded as an unobserved fixed effect at the school-level. Fortunately, by conditioning on the sum $y_{in} + y_{ie}$, it is possible to remove these nuisance parameters from expressions for Prob(y_{in}, y_{ie}). On the assumption that ϵ_{in} and ϵ_{ie} are independent, logistic disturbances, the probability of the event (y_{in},y_{ie}) = (1,0), conditional on y_{in} + y_{ie} = 1, is

$$\exp(Z_i\pi_n - Z_i\pi_e) / (1+\exp(Z_i\pi_n - Z_i\pi_e)), \quad (1)$$

where $Z_i\pi_e = (S_i \beta_{e1}+M_i \beta_{e2}+P_i \beta_{e3} +\delta_i \alpha_e)$ and

$Z_i\pi_n = (S_i\,\beta_{n1}+M_i\,\beta_{n2}+P_i\,\beta_{n3}+\delta_i\,\alpha_n)$. Note that (1) does not contain μ or ϵ_i, as the fixed effect common to both q_e and q_n is eliminated in deriving the conditional probability. It follows from (1) that

$$\mathrm{Prob}(y_{in}=0, y_{ie}=1 \mid y_{in}+y_{ie}=1) = 1/(1+\exp(Z_i\pi_n - Z_i\pi_e)).$$

Maximization of the likelihood function formed of conditional probabilities was proposed for panel data with fixed effects by Chamberlain (1980), who also showed that the inverse information matrix provides a consistent estimator of the asymptotic covariance matrix. Since the probabilities of the outcomes (0,0) and (1,1) conditional on $y_{in} + y_{ie}$ are both one, the value of the likelihood function is not affected by observations in which new and experienced teachers are rated alike. Note also that only those elements of S, M, and P which have a differential impact on quality of new and experienced teachers ($\pi_e \neq \pi_n$) will affect the outcome.

Estimates of sector coefficients are presented in Table 3. The dependent variable is defined so that a positive coefficient increases the probability that experienced teachers will be rated above new teachers. In all three formulations of the model, this outcome is more likely in the private sector. While adding controls for working conditions and salary reduce the magnitude of the effect, it remains strong and statistically significant (though only at 10% for parochial schools in Model 3). Sample sizes are, of course, considerably smaller than in Tables 1 and 2, since observations in which experienced and new teachers receive the same rating are not used for estimation.

Table 3
Ordered Logit Coefficients:
Experienced/New Ratings

model	(1)	(2)	(3)
Public	—	—	—
Catholic	1.150*** (.212)	1.053*** (.250)	.538** (.292)
Other Religious	1.298*** (.171)	1.197*** (.218)	.628*** (.264)
Non-Religious	1.830*** (.289)	1.545*** (.312)	1.010*** (.362)
Other Covariates	none	34	37
sample size	3,688	3,525	3,121

Conclusion

Analysis of principals' evaluations of their new and experienced teaching staffs from the 1990-91 Schools and Staffing Survey reveals significant differences between public and private schools. In spite of their much lower rates of pay in private schools principals rate the quality of their inexperienced teachers similarly in the public and private sectors. The experienced teaching staff, however, is rated significantly higher in private schools, a difference which does not seem to be accounted for by student or principal characteristics. A review of additional evidence points to possible reasons for the superior performance of private schools in this regard: greater flexibility in structuring pay, more supervision and mentoring of new teachers, and freedom to dismiss teachers for poor performance (Ballou and Podgursky, 1994).

References

Ballou, Dale and Michael Podgursky. 1994. "Teacher Quality and Personnel Policy in Public and Private Schools" Department of Economics. University of Massachusetts at Amherst. mimeo.

Chubb, John E. and Terry M. Moe. 1990. Politics, Markets, and America's Schools. Washington, D.C. Brookings Institution.

Hoffer, Thomas, Andrew Greeley, and James Coleman. 1985. "Achievement Growth in Public and Private Schools" Sociology of Education 58 (April), 74-97.

Chamberlain, Gary. 1980. "Analysis of Covariance with Qualitative Data." Review of Economic Studies. 47, 225-238.

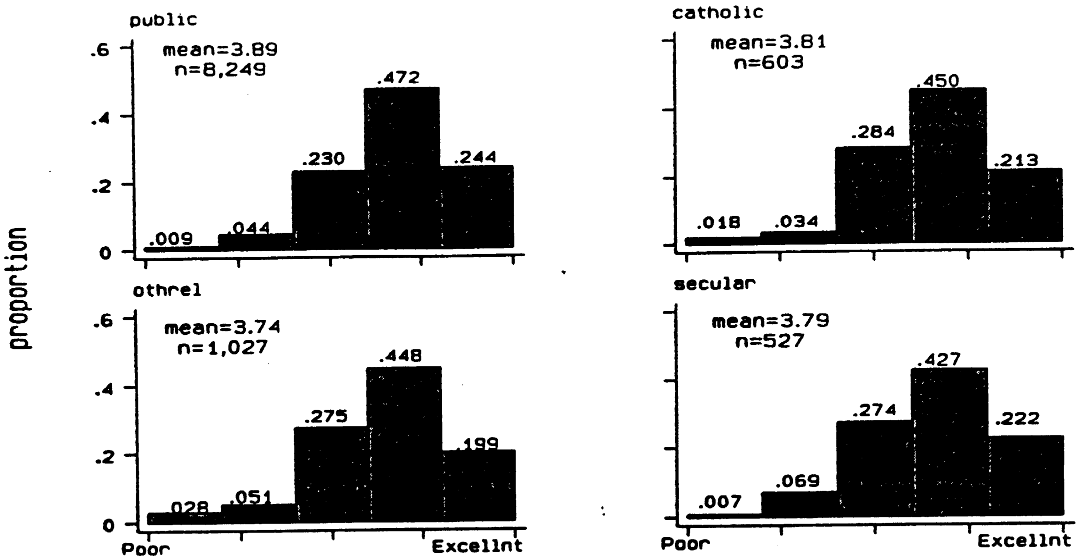


Figure 1: New Teacher Ratings by School Type

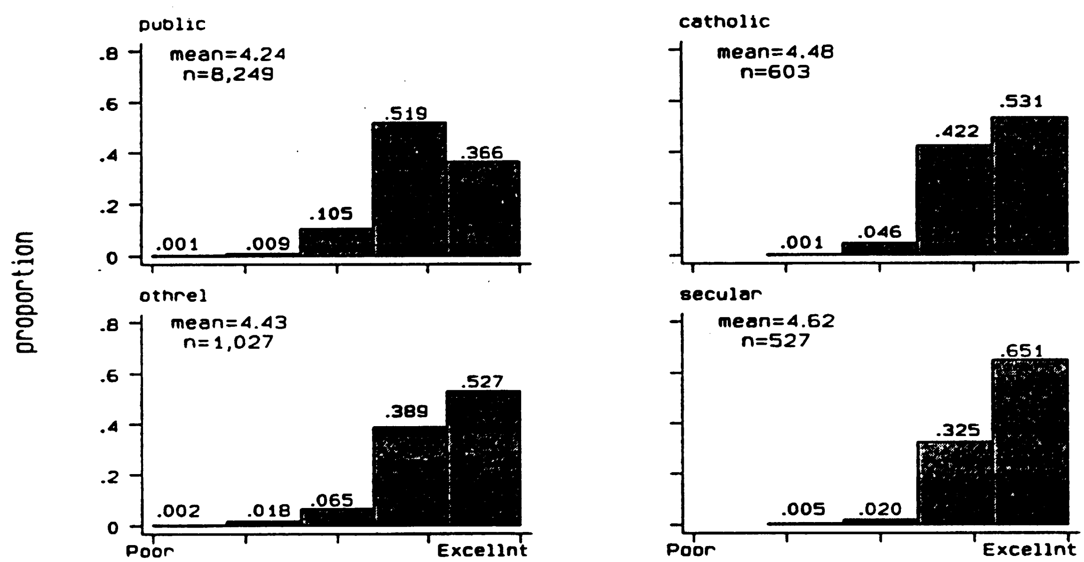


Figure 2: Experienced Ranking by School Type

TEACHER SHORTAGES AND TEACHER QUALITY

Richard M. Ingersoll, American Institutes for Research
AIR, 3333 K St. NW, Washington, DC 20007

Key Words: teacher shortage, teacher quality, SASS

Introduction

Beginning in the early 1980s, a series of highly publicized reports focussed national attention on the imminent possibility of widespread shortages of elementary and secondary school teachers in the U.S. (e.g. Darling-Hammond 1984; Good and Hinkel 1983; National Commission on Excellence in Education 1983). These predictions came as a complete surprise to many. Throughout much of the 1970s, there had appeared to be a surplus of school teachers. Indeed, reductions in the teaching force through layoffs had been common to many schools and districts in the U.S. But the new research on teacher supply and demand made a compelling case that through the 1980s teacher supply would drastically decrease, while demand for new teachers would steadily increase, resulting in shortages.

The shortage argument was that fewer and less qualified college graduates were choosing to teach, while more children of the "baby boom" generation were entering the school system, driving enrollments and, hence, hiring up. Moreover, a growing imbalance between supply and demand would be exacerbated, according to this view, because of problems of teacher retention. A high level of teacher attrition, these analysts argued, was a large source of demand for new teachers and a key factor behind the predicted shortages (e.g. Grissmer and Kirby 1987; Murnane et al. 1992; National Academy of Sciences 1987).

These reports arrived in a context of widespread concern and criticism surrounding the adequacy of the elementary and secondary school system as a whole. Critics linked declining U.S. economic performance, especially in the international arena, to declining school performance (National Commission on Excellence in Education 1983). The apparent inability of schools to attract and retain qualified teachers appeared to be one more in a host of symptoms of the "crisis" besetting schools. As a result, the imminent possibility of teacher shortages gained widespread coverage in the national media.

The education research community was, however, not unanimous in its assessment of the threat of teacher shortages. Several analysts argued that teacher supply was and would continue to be adequate and that attrition was not particularly high (e.g. Feistritzer 1986). A study of Indiana conducted in the late 1980s seemed to provide empirical support for these arguments. It suggested that teacher supply was up, due to increased re-entry of former teachers and that attrition was actually at its lowest point in years, due to a stable work force and a decline in turnover among new teachers and women (Grissmer and Kirby 1992).

As a result of these contradictory claims, since the late 1980s there has been widespread confusion about whether teacher shortages have been or will be a reality and education policymakers have not known what to believe. One source of the confusion and irresolution, almost all involved have agreed, has been a lack of data, especially at the national level, on the disputed phenomena: the demand for teachers, the supply of teachers and the gap between the two (e.g. Darling-Hammond and Hudson 1990; Haggstrom et al. 1988; Boe and Gilford 1992).

In order to address these shortcomings, the National Center for Education Statistics (NCES), the statistical agency of the U.S. Department of Education, fielded a major new survey of schools and teachers in the late 1980s - the Schools and Staffing Survey (SASS). This paper presents data from SASS that directly address the debate as to whether there are shortages of teachers in the U.S. The story they tell is both provocative and unsettling. In brief, our analysis suggests that there has not been shortages in the quantity of available elementary and secondary school teachers in this country. But, our analysis suggests there have been, in fact, distinct inadequacies in how well schools are staffed. Schools have filled teaching positions, but only at the expense of minimal standards of teacher qualification. The result: teacher quality has been sacrificed for teacher quantity.[1]

Data

The Schools and Staffing Survey is the largest and most comprehensive data source available on the staffing, occupational and organizational aspects of schools in the U.S. It includes a wide range of information on the characteristics, work, and attitudes of school faculty, and on the characteristics of a nationally representative sample of schools and districts. SASS was designed to be administered triennially; at this point two waves are available - for the 1987-88 and 1990-91 school years.[2]

SASS includes four sets of integrated questionnaires: a school survey; a central district office survey for public schools; a principal survey, and a teacher survey.

Response rates have been high, ranging from about 84 percent for private school teachers to 95 percent for public school administrators. The samples utilized in this analysis contain about 4,800 public school districts, 9,000 public schools, 2,600 private schools, 46,700 public school teachers, and 6,600 private school teachers. All of the data reported here are weighted to be representative of the national population of teachers and schools in the year of the survey.

The 1987-88 and 1990-91 waves of SASS obtained a rich array of information on issues at the heart of the shortage debate: the numbers of and fields of teaching position vacancies in schools; the degree to which schools experienced difficulties in filling vacancies; the numbers of unfilled positions; the methods that schools used to respond to difficulties in filling vacancies; the sources of new teachers; the background, characteristics, qualifications and assignments of newly hired and already employed teachers. In order to provide context, I also utilize selected data from several other NCES surveys and reports.

Results

Shortages of teachers, most simply put, occur where demand, or the number of teaching positions funded, outstrips supply, or the number of teachers available. Analyses of shortages then must begin by assessing demand and supply.

Demand for teachers appears to be on the rise. After a decade and a half of decline, since the mid 1980s school enrollments have steadily increased and are projected to continue to do so (NCES 1992). Total public school enrollment, for example, rose about 5 percent from 1984 to 1990. As a result, schools are hiring teachers. At the beginning of both the 1987-88 and 1990-91 school years, an overwhelming majority of schools had job openings for teachers. These openings have not simply been replacements of teachers who left. The number of employed elementary and secondary teachers has steadily increased since the mid 1980s (NCES 1993). For example, from 1987-88 to 1990-91, the total population of elementary and secondary teachers jumped from 2,630,000 to 2,915,000.

Changes in teacher supply are more difficult to assess. This is because the quantity of potential teachers - the reserve pool - is large, diverse and probably, unknowable. Newly qualified teachers who have recently graduated from state-approved teacher training programs at colleges and universities are perhaps the most obvious and quantifiable source. But these only comprised about 20 percent of those hired in 1987-88 and 1990-91. There are numerous other sources of teachers for teaching jobs. For instance, over half of those teachers newly hired in both 1987-88 and 1990-91 were re-entrants — former teachers who were returning, or delayed entrants — trained teachers who did not seek a position immediately after their schooling. Indeed, data from NCES's Recent College Graduates Survey indicate that as many as 40 percent of newly trained and qualified teachers do not seek teaching positions immediately after their schooling (Gray et al. 1993; Frankel and Stowe 1990). Some delay their entrance into teaching and some never teach. All of these newly qualified teachers are potential members of the reserve pool.

The real supply issue is, of course, not the number of potential teachers but how many candidates are ready and willing to apply to teaching openings. In order to assess the supply of those ready and willing to teach, principals were asked if their schools had difficulty hiring suitable candidates to fill openings.

Of those schools reporting openings in 1987-88, principals in 44 percent of the public and 56 percent of the private schools reported they experienced difficulties in filling their vacancies. The situation was comparable in 1990-91. In fact, in 1990-91, 15 percent of principals reported that they had vacancies that were simply impossible to fill with a qualified teacher in the grade level to be taught. Despite these widespread difficulties in finding suitable candidates, however, there were very few teaching positions left unfilled or withdrawn because suitable candidates could not be found in the 1987-88 or 1990-91 school years in the U.S. Why?

In reality, schools often simply cannot and do not leave teaching positions unfilled, regardless of supply. There are two general strategies by which school officials can reduce shortfalls between the supply of and demand for particular kinds of teachers. One involves altering demand and the other involves altering supply (Haggstrom et al. 1988).

The first strategy is to decrease the demand for certain kinds of teachers by either eliminating positions, or shifting students to existing staff. This would result in increases in teachers' courseloads, school class sizes or pupil-teacher ratios. Data from SASS indicate this mechanism has not been used with frequency in recent years.

A second possible strategy is to increase or alter the supply of particular kinds of teachers. One version of this strategy increases supply by increasing salaries. The evidence for this is mixed. Average starting salaries for public school teachers have increased (in real dollars) over the past decade. But this only came after steady decreases (in real dollars) through the 1970s. In fact, the average starting salary for public school teachers in 1991

was about equal to that in 1972 (NCES 1992) (see Table 1). Moreover, the salaries of new college graduates who have become teachers in recent years have been considerably below that of new college graduates who chose most other occupations (Cahalan and Gray 1993) (see Table 2).

Table 1.--Average Starting Salary for Public School Teachers (in constant 1991 dollars): Selected Years 1972-1991

School Year Ending	
1972	$22,761
1974	$22,311
1976	$21,794
1978	$21,065
1980	$19,342
1982	$19,151
1984	$20,340
1986	$22,003
1988	$22,582
1989	$22,715
1990	$22,708
1991	$22,830

Table 2.--Average Annual Salaries of New Bachelor Degree Recipients in Teaching and Other Selected Occupations, 1990-91

Occupation	Salary	Difference
Teaching	$19,913[1]	—
Computer Science	30,419	$10,504
Math, Physical Sciences	26,040	6,125
Business/Management	25,961	6,046
Writers/Artists	22,353	2,438
Biologists		21,325
1,420		
Communications	19,584	- 329
Public Affairs/Social Studies	19,227	- 686
All occupations	$23,632	$3,717

[1] Scheduled salary based on average contract length of 9.7 months.

Another version of the second strategy alters supply by filling a position with an underqualified candidate. This could be accomplished by shifting existing staff to areas of greater need; that is, assigning teachers trained in one field to teach in another. For example, social studies teachers could be assigned to teach mathematics courses. Alternatively, school officials could hire available teacher candidates, regardless of qualifications.

Data from SASS indicate that this supply strategy has been commonly used. For both public and private schools, among the most common methods of coping with difficulties in filling openings in 1987-88 and 1990-91 were to hire less qualified teachers, to assign other teachers and to use substitute teachers. For instance, in 1990-91, 50 percent of public school principals, who indicated they had difficulty filling openings, reported using substitute teachers as a remedy.

The widespread use of this latter supply strategy necessitates a shift in focus for teacher supply assessments. Rather than focus on whether or not there are or will be sufficient numbers of potential teachers, supply assessments need to examine the actual fit between the needs of schools and the qualifications of the teachers currently employed. That is, the focus shifts from assessing the adequacy of the quantity of potential teachers to assessing the adequacy of the quality of employed teachers. (also see Kennedy 1992; Darling-Hammond and Hudson 1990).

Assessing levels of teacher qualifications and quality, like assessing quantity, is a difficult and ambiguous task. How to define and measure a qualified teacher and quality teaching are subjects of great controversy (Haney et al. 1987; Ingersoll 1994; Kennedy 1992). There is, however, almost universal agreement that one of the most important characteristics of a qualified teacher is training and preparation in the subject or field in which they are teaching. Research has shown moderate but consistent support for the reasonable proposition that subject knowledge is an important predictor of both teaching quality and student learning (for reviews of this research, see Shavelson et al. 1989; Darling-Hammond and Hudson 1990; Murnane and Raizen 1988). Knowledge of subject matter does not guarantee qualified teachers and quality teaching, but is a necessary prerequisite.

SASS data indicate that inadequacies in teacher quality were not due to a lack of basic training in subject matter. In 1990-91, for example, 99 percent of high school teachers employed in the United States held a bachelor's degree and 46 percent had obtained a graduate degree. The issue in question is the phenomenon of out-of-field teaching - teachers assigned to teach in fields for which they do not have adequate or appropriate training.

Of course, some degree of out-of-field teaching may be unavoidable and may not be an indicator of a shortage of qualified teaching candidates. School administrators charged with the task of offering programs in a range of required and elective subjects may often be forced to make spot decisions concerning the assignment of available faculty to an array of changing course offerings. But even low levels of out-of-field teaching are meaningful to teacher quality assessments. This is especially true for the case of high

schools and for the core academic fields. In high schools, teachers are divided by fields into departments; faculties are thus more specialized than in elementary schools, and therefore the differences between fields are more distinct and, perhaps, greater. Moreover, the level of mastery in different subjects is higher in high schools, and hence a clear case has been made by policy analysts and researchers that teachers ought to have adequate background in the subjects they teach (e.g., Shavelson et al. 1989; Murnane and Raizen 1988; Darling-Hammond and Hudson 1990). In the following section I focus on the levels of and variations in out-of-field teaching in high schools.

SASS data show, in fact, that substantial numbers of high school teachers were assigned to teach out of field or out of department in both 1987-88 and 1990-91. The data indicate that, while most high school teachers had a undergraduate or graduate major in their main teaching assignment field, large numbers of teachers were assigned to teach courses in additional fields for which they did not have a major or even a minor. In 1990-91, public high school teachers taught, on average, about 15 percent of their class schedules in fields for which they did not have a minor. This amounted to about one course in six. Private high school teachers taught far more of their classes without minimal qualifications. On average, for about one-quarter of their scheduled classes, they did not have at least a minor in the field. These percentages all substantially increase (sometimes double) if the standard is raised from a minor to a major in the field taught. As a result, substantial numbers of high school students were taught core academic classes by teachers without even minimal training in the field. These levels of out-of-field teaching, however, varied substantially by field.[3]

In 1990-91, fifteen percent of all high school English students — almost 2.25 million high school students in this country — were taught by teachers who did not have at least a college minor in English, language arts, journalism or communication. Twenty-one percent of all high school mathematics students, or over 2.5 million, were taught mathematics by teachers without at least a minor in mathematics or mathematics education. Eleven percent of high school students were taught science by teachers without at least a minor in <u>any</u> of the biological, physical or natural sciences or science education. Eleven percent of high school students were taught social studies by teachers without at least a minor in history, any of the social sciences or social studies education.

Out-of-field levels also varied considerably across different types of schools. Notably, public schools with a high proportion of poverty-level students (those with over 50 percent eligible for the federal free lunch program) had a higher proportion of students taught by out-of-field faculty in mathematics, science, and English than schools with less than 20 percent poverty-level students (Table 3).

Small schools (less than 300 students) in both the public and private sector tended to have relatively higher levels of out-of-field teaching. On one extreme were small private schools with 41 percent of mathematics students and 38 percent of English students out of field. On the other extreme were large public schools (600 or more students). Even these schools, however, had substantial levels of out-of-field teaching (Table 4).

Table 3.-- Percentage of public high school students enrolled in classes taught by teachers without at least a minor in the field, by poverty level of students*: 1990-91

	Math	Science	Social Studies	English
Total Public	20.5	10.2	9.7	13.8
% Poverty Level				
Less than 20%	18.8	7.7	9.3	12.1
20-49%	23.4	12.6	11.1	16.5
50% or more	24.2	14.1	8.3	18.0

* Percent students eligible for federal free lunch program.

Table 4.-- Percentage of public high school students enrolled in classes taught by teachers without at least a minor in the field, by school sector and size: 1990-91

	Math	Science	Social Studies	English
Total Overall	21.1	11.2	11.0	14.7
Total Public	20.5	10.2	9.7	13.8
Size				
Less than 300	26.6	16.7	14.2	16.2
300-599	20.8	11.1	11.4	17.7
600 or more	20.1	8.8	8.9	13.1
Total Private	25.9	19.5	22.2	22.7
Size				
Less than 300	41.4	28.7	34.3	37.7
300-599	23.2	8.0	19.1	15.2
600 or more	18.5	7.6	10.0	19.7

Conclusion

This paper addresses the ongoing debate as to whether there are shortages of teachers in the U.S. The analysis suggests that, in body counts alone, there are not shortages in the quantity of available school teachers in this country because the reserve pool of teachers is large and the supply of teachers is highly manipulable.

But, our analysis suggests there are, in fact, distinct inadequacies in how well schools are staffed. Schools have been able to fill available teaching positions, but only at the expense of minimal teacher qualifications. If one accepts the premise that adequate staffing requires high school teachers, for example, to hold at least a college minor in the fields which they teach, then this analysis suggests that many of the nation's high schools have not been adequately staffed. These inadequacies, however, were not an issue of teacher training. Most school teachers in the United States had completed a basic level of education and training. The inadequacies lay in the fit between teacher's fields of training and their teaching assignments. Many teachers were assigned to teach classes which did not match their education or training. As a result, there were substantial numbers of high school students taught by teachers who did not have even a college minor in the field taught. The result: teacher quality has been sacrificed for teacher quantity.

But these data do not establish, for example, to what extent out-of-field teaching is a short-term condition resulting from teacher shortages or to what extent it is a normal and ongoing practice in particular schools. It is quite likely that out-of-field assignments are both a chronic practice and also one that is increasingly utilized in shortage situations. Moreover, if out-of-field teaching is a remedy for difficulties in hiring, the problem is most likely not due to insufficient numbers of adequately trained teachers, but to the unwillingness of existing trained teacher candidates to seek positions. These issues warrant further investigation.

The extent to which schools employ underqualified teachers has, of course, important implications not only for the shortage debate, but for contemporary education reform efforts seeking to improve teacher and teaching quality. Such efforts have sought to raise the standards, increase the training and upgrade the work of teachers. From this viewpoint, widespread assignment of teachers to teach subjects for which they are not trained is an example of an inappropriate utilization of costly resources. Moreover, the cross school variations in the utilization or under-utilization of these human resources, illustrated in Tables 3 and 4, have implications for several streams of current education research and reform.

Equity is one of the central concerns of contemporary educational researchers and policymakers (e.g., National Commission on Excellence in Education 1983). Concern centers around disparities in the resources and quality of schooling provided to different student subgroups. This analysis draws attention to differences in the distribution of one such resource—qualified teachers. These data suggest that poorer student populations more often receive less qualified teachers. This raises questions about the impact of out-of-field teaching levels on the achievement of students from such schools.

Private/public school differences is another central theme in much current education research. In particular, analysts have focused on the widespread differences in the ways public and private schools are organized and operated (e.g. Coleman and Hoffer 1987). This analysis draws attention to distinct differences in an important but overlooked aspect of school organization—the management and utilization of teachers as professionals. These data suggest many private schools are characterized by high levels of underqualified teaching. This raises questions about differences in the degree of teacher professionalism between public and private schools.

Finally, the state of mathematics and science educational quality and achievement in the United States is another important topic in contemporary education research. There is a growing constituency who have looked to mathematics and science education as a key example of what is wrong with the American education system, and hence, a target for education reform (Darling-Hammond and Hudson 1990; Murnane and Raizen 1988). This analysis draws attention to the especially high levels of out-of-field teaching in mathematics. This raises questions concerning the distinct variations in levels of out-of-field teaching among fields and the impact of teacher background on student achievement.

References

Boe, E. & Gilford, D. (1992). Teacher supply, demand and quality. Washington, DC: National Academy Press.

Cahalan, M. & Gray, L. (1993). Occupational and Educational Outcomes of Recent College Graduates 1 Year After Graduation: 1991. Washington, DC: US Department of Education, National Center for Education Statistics.

Choy, S., Henke, R., Alt, M., Medrich, E. & Bobbitt, S. (1993). Schools and Staffing in the U.S: A Statistical Profile, 1990-91. Washington, DC: US Department of Education, National Center for Education Statistics.

Coleman, J. & Hoffer, T. (1987). Public and Private Schools: The Impact of Communities. New York: Basic.

Darling-Hammond, L. (1984). Beyond the commission reports: The coming crisis in teaching. Santa Monica, CA: Rand Corporation.

Darling-Hammond, L., Hudson, L. (1990). "Pre-college Science and Mathematics Teachers: Supply, Demand and Quality. Review of Research in Education. Washington, DC: American Educational Research Association.

Feistritzer, E. (1986). Teacher Crisis: Myth or reality? Washington, D.C.: National Center for Education Information

Frankel, M. & Stowe, P. (1990). New teachers in the job market, 1987 update. Washington, DC: US Department of Education, National Center for Education Statistics.

Good, T. & Hinkel, G., (1983). Teacher shortage in science and mathematics: myths, realities and research. Washington, DC: National Institute of Education.

Gray, L, Cahalan, M., Hein, S., Litman, C., Severynse, J., Warren, S., Wisan, G., Stowe, P. (1993). New teachers in the job market, 1991 update. Washington, DC: US Department of Education, National Center for Education Statistics.

Grissmer, D. & Kirby, S. (1987). Teacher attrition: the uphill climb to staff the nation's schools. Santa Monica, CA: Rand Corporation.

Grissmer, D. & Kirby, S. (1992). Patterns of attrition among Indiana teachers, 1965- 1987. Santa Monica, CA: Rand Corporation.

Haggstrom, G. W., Darling-Hammond, L., & Grissmer, D. (1988) Assessing teacher supply and demand. Santa Monica CA: Rand Corporation.

Haney, W., Madus, G. & Kreitzer, A. (1987). Charms talismanic: Testing teachers for the improvement of American education. Review of Research in Education. (pp.169-238). Washington, DC: American Educational Research Association.

Ingersoll, R. (1994). National assessments of teacher quality. Washington, DC: US Department of Education, National Center for Education Statistics.

Ingersoll, R. & Chambers, J. (1994).Teacher Supply, Demand and Quality: 1991. Washington, DC: US Department of Education, National Center for Education Statistics.

Kaufman, Steven, Huang, H., (1993). 1990-91 Schools and Staffing Survey: Sample Design and Estimation. Washington, DC: US Department of Education, National Center for Education Statistics.

Kennedy, M. 1992. "The problem of improving teacher quality while balancing supply and demand." pp. 63-126 in Boe, E. & Gilford, D., eds. Teacher supply, demand and quality. Washington, DC: National Academy Press.

Murnane, R., Singer, J., Willett. J., Kemple, J., Olsen, R. (1992). Who will teach?: Policies that matter. Cambridge, Ma.: Harvard University Press.

Murnane, R. & Raizen, S. (1988). eds. Improving indicators of the quality of science and mathematics education in grades k-12. Washington, DC: National Academy Press.

National Academy of Sciences. (1987). Toward understanding teacher supply and demand. Washington, DC: National Academy Press.

National Center for Education Statistics. 1992 The Condition of Education . Washington, DC. US Department of Education.

National Center for Education Statistics. (1993). The Digest of Education Statistics . Washington, DC. US Department of Education.

National Commission on Excellence in Education. (1983). A Nation at Risk: The Imperative for Educational Reform. Washington, DC: Govt. Printing Office.

Shavelson, R., McDonnell, L. & Oakes, J. (1989). Indicators for monitoring mathematics and science education. Santa Monica, CA: Rand Corporation.

Endnotes

[1] This paper is drawn from a larger report on teacher supply, demand and quality sponsored by NCES (contract number RN93140001). This paper does not constitute an official NCES publication. The views expressed here are solely those of the author. A more detailed and comprehensive analysis is contained in the official report, see Ingersoll and Chambers 1994.

[2] SASS data tapes, survey questionnaires and user's manuals are available from NCES, U.S. Department of Education, 555 New Jersey Ave., Washington, D.C. 20208-5641. For information concerning the survey design and sample estimation of SASS see Kaufman and Huang (1993). For an extensive report, summarizing the items used in this investigation and providing an overview of the entire survey see Choy et al. (1993).

[3] Out-of-field teaching can be empirically measured in a number of ways. Here, I focus on (1.) a minimal level of (2.) substantive training in (3.) broadly defined fields. Thus: (1.) At least a minor in the field is defined as adequate. (2.) The focus is on substantive training; I do not focus on formal training in teaching methods and pedagogy i.e. certification. (3.) Fields are defined parallel to conventional departmental divisions in high schools. That is, fields include all within-department disciplines. Hence, for example, a minor in any of the natural, physical or biological sciences is considered adequate training to teach any science course. See Ingersoll and Chambers (1994) for a detailed discussion of a range of out-of-field teaching measures.

WORK EXPERIENCE, LOCAL LABOR MARKETS, AND DROPPING OUT OF HIGH SCHOOL

Paul Swaim, Economic Research Service
Room 340, 1301 New York Ave., N.W., Washington, DC 20005

Key Words: Education, Dropouts, Employment

I. Introduction. This paper extends the existing literature on high school completion in several ways. First, the relationship between working while in school and dropping out is analyzed for the early 1990's. Second, the potential importance of local variations in employment opportunities for high school completion is analyzed. High dropout rates are often found in localities offering few good jobs, but it is not known whether spatial differences in the availability of jobs or wage rates affect youths' educational outcomes. Finally, I make use of an important new longitudinal data set, namely, the National Education Longitudinal Study of 1988 (NELS) and its first two followups, followup 1 (F1) in 1990 and followup 2 (F2) in 1992.

The first premise of this research is that America has a serious dropout problem. On the one hand, a significant number of youths continue to drop out of high school. On the other, individuals lacking a secondary education are at increasing risk of economic impoverishment and other hardships, and represent an economic and social burden for society. The second premise of this research is that employment experiences and opportunities are potentially important determinants of academic progress while in high school.

An extensive empirical literature has verified the large and increasing importance of schooling for individual earnings and, hence, the living standards of workers and their families (e.g., Levy and Murnane, 1992). Indeed, the inflation-adjusted wages of high school dropouts have fallen precipitously since the early 1970's, particularly for new cohorts of males. Although more speculative, it is possible that the low educational levels of a significant minority of American workers may be an important drag on aggregate productivity and international competitivity (MIT Commission on Industrial Productivity, 1989).

The failure to complete high school also affects a variety of nonmarket outcomes (Haveman and Wolfe, 1984; Astone and McLanahan, 1992). Research indicates that the lack of a high school education can lead to a lower investment in one's own children, an increased risk of divorce, less efficient contraceptive use and higher mortality rates. Dropping out of high school is not simply a problem for the individual but one for their family and for society.

Recent decades have witnessed a slow decline in the percentage of young people failing to graduate from high school. In 1970, 16.6 percent of all persons aged 20-21 reported less than a high school education (Table 1). In 1991, the dropout rate had fallen to 14.8%. Despite this downward trend, substantial numbers of young people continue to leave high school without graduating. Large numbers of minorities, in particular, still drop out of high school. Over one third of Hispanics aged 20-21 in 1991 reported having less than a high school education. For Blacks, the percentage of dropouts is much improved from 29% in 1970, but approximately one fifth of blacks aged 20-21 currently report having no high school diploma. Dropout rates also differ greatly across localities. Tabulations for all U.S. counties from the 1990 Census show that the share of 16 to 19 year olds out of school but not high school graduates ranged from 2 to 38 percent.

Table 1
Percentage of High School Dropouts Among Persons 20-21 Years Old, 1970-1991

Group	1970	1980	1991
All	16.6	15.6	14.8
Male	16.1	17.8	15.5
Female	16.9	14.3	14.2
White[1]	14.6	12.1	10.6
Black[1]	29.6	24.6	19.1
Hispanic	NA	41.6	37.5

[1]Excludes Hispanics in 1980 and 1991.
Source: Current Population Survey.

Why have dropout rates remained so high, especially for particular subgroups of the population and communities, when the individual consequences of dropping out are so negative? Educational researchers have examined this issue in detail (e.g., Natriello, 1987). Regression studies using rich, longitudinal data sets such as High School and Beyond have identified a large number of correlates of dropping out. It has proven more difficult, however, to sort out the key causal mechanisms at work. Given the high economic stakes of schooling outcomes, it is particularly unfortunate that the importance of work experience while in high school and the labor market returns to educational investments have typically received little attention in this research.

II. Labor Markets and High School Completion. Proponents of educational reform recently have emphasized the need to better manage the school-to-work transition. Under the current "nonsystem" it appears that many noncollege-bound high school students see little relationship between their class work and their future career prospects and, hence, do exert little effort in school. Comparisons with secondary schools in Germany, Japan, and elsewhere have motivated proposals to strengthen the ties between high schools and surrounding employers. The federal Educational Reform Act of 1994 moves in this direction, by supporting youth apprenticeship programs that integrate work experience and vocational preparation with high school study. Local initiatives by employers and schools, such as the Boston Compact, have also pioneered closer links between high schools and the world of work.

Of course, many high school students have long mixed study and work. Unfortunately, little is known about how the experience of holding one or more jobs while attending high school typically affects academic progress. Both good and bad effects have been conjectured. On the positive side, working might reinforce behavioral traits, such as punctuality and diligence that contribute to school success. Direct exposure to the labor market may also result in a better appreciation of the importance of schooling for occupational advancement and, hence, result in greater effort while in school. On the negative side, jobs can absorb time and energy that would be better directed toward study and other school activities, especially if weekly hours worked are high, and increase the risk of dropping out.

Independent of work experience while in high school, labor markets may influence schooling decisions through the incentive effects implicit in the structure of wages. Human capital theory, as developed by Becker (1975) and others, has provided economists with a rigorous framework for studying how labor market incentives affect educational attainment. Education is viewed as a purposive investment activity which is pursued up to the point at which the marginal return to more schooling equals the return to the best alternative investment. The primary economic incentive to become more educated is that more educated workers qualify for better paying jobs. I will refer to this labor market incentive to stay in school as the educational wage premium effect. However, staying in school causes an immediate loss of income to the extent that less time is available for paid-employment. I will refer to the labor market disincentive to stay in school associated with foregone earnings as the opportunity cost effect.

During the 1980's, a rising wage premium effect should have encouraged greater investment in education overall, but may not have had much effect on dropout rates. A large number of studies have shown that the association between educational attainment and wages strengthened. However, a closer look at the evolution of relative wages by education levels during the 1980's suggests that the strengthening of the wage premium effect was much stronger for decisions about continuing on to college after completing high school than for decisions about dropping out versus completing high school. The hourly wages of high school graduates with no college were approximately 25 percent higher than dropouts' wages in 1973, 1979, and 1988 (Table 2). The college wage premium did increase during the 1980's, but that premium typically may not be relevant for students on the margin between dropping out and finishing high school.

Table 2

Hourly Wages in 1988 Dollars and Relative Wages by Education, 1973-1988: Workers with 0-9 Years of Potential Labor Market Experience

Group	1973	1979	1988
Men:			
Dropout	$7.52 (1.0)	$7.20 (1.0)	$5.54 (1.0)
H.S. Grad.	$9.69 (1.3)	$8.96 (1.2)	$7.31 (1.3)
College	$12.96 (1.7)	$11.38 (1.6)	$12.16 (2.2)
Women:			
Dropout	$5.80 (1.0)	$5.48 (1.0)	$4.82 (1.0)
H.S. Grad.	$7.15 (1.2)	$6.87 (1.2)	$6.18 (1.3)
College	$10.42 (1.8)	$9.29 (1.7)	$10.00 (2.1)

Source: Bound and Johnson's (1992) tabulations from the Current Population Survey outrotation files.

The large fall in the real wages of high school dropouts during the 1980's reduced the opportunity cost of staying in high school rather than dropping out in order to work more hours (Table 2). Employers have also shown an increased preference for employing part-time workers. The increased availability of part-time jobs may have further reduced the opportunity cost of high school attendance because it probably has become easier to mix schooling with work. The opportunity cost effect should thus have lowered dropout rates. This observation suggests that recent dropout data may be difficult to reconcile with human capital theory unless the opportunity cost effect is quantitatively small.

The human capital approach is subject to two, potentially important, limitations. First, high school dropouts--particularly those dropping out well before graduation age--may not be well enough informed about the labor market consequences of their schooling decisions to pursue their long-run economic interests in a systematic fashion. A second qualification is that

high school dropouts may not be disposed to pursue their long-run economic interest, even if they can identify it. The possibility that dropping out may reflect self destructive behavior means that the applicability of a rational choice model, such as human capital theory, should be assessed and not simply assumed. As argued above, work experience while attending high school might be a valuable source of information about employment opportunities and encourage a more responsible attitude toward vocational preparation. Thus, work experience could lead to better informed and more forward looking educational choices, more consonant with human capital theory. As already noted, however, working while in school might instead reduce the time and energy available for study and increase the risk of dropping out.

III. Data. The data set used is the National Education Longitudinal Study of 1988 (NELS). In the base year (BY) of 1988, approximately 25,000 eighth graders were surveyed with followups conducted in 1990 (F1) and 1992 (F2). At the time this analysis was conducted, NCES had not released the final version of the second followup data. Accordingly, the interim version of the F2 data is used.

NELS is a particularly well-suited data set for this study for several reasons. First, NELS begins following students in eighth grade; earlier than most other data sets which follow high school students. Second, NELS continues to interview dropouts after they have left school, which is unprecedented in a data set of this size and richness. Third, NELS provides considerable information on work experience. Fourth, NELS contains not only a student questionnaire but data from parents, teachers and school administrators allowing for many levels of analysis.

The dependent variable in the analysis below is dropout status as measured at the times of the two followups. The F1 and F2 followup interviews took place in the Spring of the sophomore and senior years, respectively, of those sample members progressing at a typical rate. At the time of each interview, an individual is classified as a dropout if that individual has been out of school for 20 or more consecutive days, has not completed high school or an equivalent credential (e.g., a GED), and is not enrolled in an alternative program preparing for an equivalent credential. For short, I will refer to these groups as sophomore and senior dropouts. Many of these individuals probably will eventually complete high school or earn at equivalency certificate. A "dropout" in this context is best interpreted as an individual who is not progressing steadily toward completing high school and is at risk of never completing high school.

Table 3 reports sophomore and senior dropout rates according to this definition. Six percent of the F1 individuals were dropouts, as were 10.2 percent of the F2 individuals. In both years, sex differences in dropout rates were small, but Blacks, Hispanics, American Indians, and nonnative English speakers had significantly higher risks of dropping out, while Asians were less likely to dropout. Dropout rates were highest in urban areas, intermediate in rural areas, and lowest in suburbs.

Table 3
Sophomore and Senior Dropout Rates in the 1988 National Longitudinal Education Survey (NELS)

Group	Sophomore Dropout	Senior Dropout
All	6.0	10.2
Female	5.7	10.9
Male	6.2	9.5
White	5.4	9.8
Black	10.2	12.8
Asian/Pacific	2.9	6.0
American Indian	11.1	23.3
Hispanic	9.1	17.4
Non-native English	8.8	15.2
Urban	8.8	11.9
Suburban	4.8	8.7
Rural	6.1	10.9

The analyses of the affects of local job markets on educational attainment in the next section is restricted by the availability of geographic codes for the schools surveyed. To date, I have obtained geographic locations from NCES only for the a large share of the public schools in NELS. When variables representing conditions in the local labor market are added to the empirical models, the estimation sample is confined to respondents attending public schools in the base year (1988) whose geographic locations I have been able to obtain. The final sample size for the sophomore (senior) panel is 12,414 (11,752) individuals with 609 (954) dropouts. Models not incorporating area controls can be estimated on larger samples of 17,316 (16,396) individuals, of whom 756 (1,159) are dropouts.

IV. Results. Table 4 provides descriptive tabulations on work experience at the time of the F1 and F2 followup interviews. These data confirm that many high school students work, particularly in their senior year. Perhaps the biggest surprise from the perspective of human capital theory is that a higher proportion of continuing students are employed than of dropouts, 32.7 versus 29.9 percent in F1 and 79.6 versus 65.5 percent in F2. This pattern is particularly strong for females and for F1 Blacks. Conditional on employment, hours worked per week are approximately

twice as high for dropouts as for continuing students, but the hourly wage received by dropouts is a only a little higher than the wage received by students. It appears that the major opportunity cost in foregone earnings associated with staying in high school is that individuals who would work in any case can work more hours if they drop out of school.

Table 4
Work Experience of Students and Dropouts in NELS

	Sophomores		Seniors	
	Students	Dropouts	Students	Dropouts
All:				
Employed (%)	32.7	29.9	79.6	65.5
Hours/week[1]	16.7	32.8	18.9	36.3
Hourly wage[1]	$4.44	$5.18	$5.66	$5.94
Females:				
Employed (%)	32.2	20.0	81.2	55.1
Hours/week[1]	15.3	28.7	17.6	32.3
Hourly wage[1]	$4.10	$4.52	$5.46	$5.42
Males:				
Employed (%)	33.2	39.1	77.9	77.0
Hours/week[1]	18.1	34.8	20.3	39.4
Hourly wage[1]	$4.77	$5.49	$5.87	$6.35
Blacks:				
Employed (%)	24.8	6.6	68.6	56.2
Hours/week[1]	19.2	32.9	20.6	35.3
Hourly wage[1]	$4.40	$6.44	$5.74	$5.61
Hispanics:				
Employed (%)	22.5	23.2	77.6	56.8
Hours/week[1]	17.7	28.0	21.2	37.1
Hourly wage[1]	$4.41	$5.17	$5.62	$5.79

[1]For individual with jobs.

Table 5 reports select maximum likelihood coefficients for a series of logit models of the probability of being a sophomore or senior dropout. The coefficients reported in the table correspond to measures of either individual work experience or county-level measures of wage incentives, with positive coefficients correspond to higher predicted probabilities of dropping out. All of the logit models also contain 28 additional controls for personal, family, and school characteristics that previous research suggests are related to dropout behavior. To conserve space, these coefficients are not reported here, but it bears noting that some of the demographic differences in dropout rates reported in table 3 differ in sign from the corresponding coefficients in the logit model. For example, controlling for family resources and achievement test scores, Blacks and Hispanics are less --not more-- likely to dropout than are Whites and nonhispanics.

Table 5
Select Maximum Likelihood Coefficients for Logit Models of the Probability of Dropping Out[1]

Independent Variable	Model 1	Model 2	Model 3
A. Sophomore Dropouts:			
Weekly hours of work:			
Total	.005*		
Dummy for total weekly hours:			
1-19		-.088	-.075
20+		.617**	.701*
County Wage Structure:			
Opportunity cost			.021
High school premium			.041
College premium			-.058
B. Senior Dropouts:			
Weekly hours of work:			
Total	-.033**		
School day	.084***		
Dummy for total weekly hours:			
1-19		-.315*	-.301*
20+		.934***	.801**
County Wage Structure:			
Opportunity cost			.006
High school premium			.091
College premium			.001

***, **, * denote significance at 1, 5, 10 percent.
[1]All models also contain 28 individual, family, and school level controls. Models 1-2 estimated on the full NELS panels and Model 3 estimated on restricted sample of public school students for which county identifiers were obtained from NCES.

Because Table 4 clearly indicates that hours currently worked are higher for dropouts, it would be inappropriate to use the association between current work hours and dropout status to assess whether working more hours while in school increases the risk of dropping out. Accordingly, I estimate the effect of working, or working more hours, while a student at the time of the immediately prior survey interview on the probability of dropping out by the time of a given followup. That is, I relate the probability of becoming a sophomore (senior) dropout to hours worked in the 8th (10th) grade.

The significance levels reported in Table 5 embody a conservative adjustment for the deviation of the complex NELS survey design from a simple random sample design. First, the logit regressions were estimated using relative weights (which sum to 1 and are proportional to the final survey weights supplied by NCES and, hence, account for the oversampling of certain populations). Second, I then multiplied the resulting standard error estimates by the square root of

the average design effects calculated by NCES for sample means of variables in the BY-F1 and BY-F2 panel samples (1.9 and 2.0, respectively).

The results are fairly straight-forward for the work experience variables. Working more than 20 hours per week while in school leads to an increased risk of dropping out in the next two years, and this effect is stronger for senior dropouts than for sophomore dropouts. It also appears that seniors who work more hours on school days, holding their total weekly hours fixed, are at an increased risk of dropping out. Finally, students working 1-19 hours a week are less likely to dropout than students not working at all, although this effect is statistically significant only for senior dropouts. Consistent with D'Amico's (1984) results for a decade earlier, paid employment in moderation appears to be good for school progress, but too much time at work increases the chance of school failure.

The results for county-level measures of wage incentives are much less clear cut. Model 3 incorporates an estimated wage rate for workers who have not completed high school, including high school students and dropouts. This variable is intended to capture the local opportunity cost effect. Thus, a positive coefficient is predicted, because an increase in the opportunity cost of schooling should encourage more dropping out. This model also includes two measures of the return to education, the ratio of high school graduate wages to dropout wages and the ratio of college graduate wages to high school graduate wages. Increases in either ratio should produce an educational wage premium effect that leads to greater educational investments and, hence, less dropping out and a negative coefficient. Neither the opportunity cost coefficient nor the educational wage premium coefficients attain statistical significance. Variants of the specifications reported in Table 5 were estimated that used alternative measures of the county labor market variables or added additional measures of local labor market conditions, but these variables usually were not statistically significant, so long as an extensive set of individual, family, and school controls were also included in the model.

The failure of the opportunity cost and educational wage premium coefficients to attain significance cast some doubt on the human capital model of dropout behavior. However, neither concept is measured very precisely and measurement error provides an alternative explanation for this finding. It could also be argued that migration rates are high enough that local measures of educational wage premiums, which should be assessed in light of the full span of a working life, are irrelevant to education choices because so many individuals will spend much of their careers some where other than where they grew up. Nonetheless, local variations in wages available to high school students should affect the opportunity costs of high school because these cost are immediate. One indication that local labor market conditions do matter is that when industry mix variables are added to the model many of the associated coefficients are quite large and attain statistical significance, although a coherent explanation of the indicated pattern of industry effects is not obvious. In sum, the results concerning the human capital model and the impact of local employment conditions are inconclusive at best.

V. Conclusions. This paper analyzes the influence of work experience while in high school and spatial differences in labor market conditions on dropout behavior. Data from the National Education Longitudinal Study of 1988 (NELS) are used to estimate separate logit models of the probability of dropout status in the Spring of the sophomore and senior years. Working more than twenty hours a week during the school year increases the probability of subsequently dropping out, but employment for fewer than twenty hours per week appears to encourage timely progression toward high school graduation. Dropout probabilities are significantly affected by the industrial composition of employment in the home county, suggesting that local labor markets matter for school attainment, even after controlling for a long list of individual, family, and school characteristics. However, no evidence is found that the local labor market effects operate through the opportunity cost and educational wage premium effects emphasized by the human capital theory of educational attainment.

REFERENCES.

Astone, N.M. and S.S. McLanahan, 1992, "Family Structure, Parental Practices and High School Completion," **American Sociological Review** 56:309-320.

Bound, J. and G. Johnson, 1992, "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations," **American Economic Review** 82(3):371-392.

Becker, G.S., 1984, **Human Capital**, 2nd edition, New York: Columbia University Press.

D'Amico, R., 1984, "Does Employment During High School Impair Academic Progress?" **Sociology of Education** 57:152-164.

Haveman, R.H., and B. Wolf, 1984, "Schooling and Economic Well-Being: The role of Nonmarket Effects," **Journal of Human Resources** 19(3):377-407.

Levy, F., and R.J. Murnane, 1992, U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations," **Journal of Economic Literature** 30(3):1333-1381.

The MIT Commission on Industrial Productivity, 1989, **Made in America: Regaining the Productive Edge,** Cambridge, MA: The MIT Press.

Natriello, G. (ed.), 1987, **School Dropouts: Patterns and Policies,** New York: Teachers College Press.

DISCUSSION: EDUCATION RESEARCH USING THE SCHOOLS AND STAFFING SURVEY AND THE NATIONAL EDUCATION LONGITUDINAL SURVEY

Thomas M. Smith, National Center for Education Statistics, U.S. Department of Education
555 New Jersey Ave. NW, Washington, DC 20208-5650

Key Words: Education, Teacher quality, Working while in school, Educational production function

These four papers are an excellent example of how data collected by the National Center for Education Statistics (NCES) can be used in policy relevant analyses. Teacher quality, the consequences of working while in school, and the effects of school organization on student achievement are all timely topics in the current policy debate.

Teacher shortages and teacher quality (Ingersoll and Han)

This paper examines an issue underlying many of the discussions involving the issue of "teacher shortage," that is whether or not qualified teachers are standing in front of America's classrooms. In their paper, Ingersoll and Han, make use of two very nice perspectives in addressing the issue of "adequate" teacher education and certification—first, what percentage of students are taught a core subject course by a teacher not formally trained in that subject and second, what percentage of teachers are assigned to teach courses outside their area of expertise.

The student perspective provides a framework in which to examine an important opportunity to learn (OTL) issue. . . Do students have equal access to quality teaching? The matching of classrooms of students to teachers trained in the appropriate subject matter can be considered a minimum standard for OTL. If the percentage of students taught by teachers who did not have a major or minor in this subject matter varies by type of school (high vs. low poverty, urban vs. suburban), questions about the equity of inputs provided to different schools arise. Ingersoll and Han find that students enrolled in public high schools with a greater proportion of either poverty-level or of minority students are more likely to have teachers not formally trained to teach math, science, English, and social studies. Here we see how the classroom level inputs approach is probably more useful to equity debates than other more aggregate measures, such as district level revenues per pupil comparisons.

Second, if we believe that teacher working conditions have an effect on student learning, then the assignment of teachers to fields in which they are untrained can have an adverse effect on their morale (possibly increasing the likelihood of attrition) and could change the allocation of preparation time across all of their courses (decreasing the amount of time they spend preparing for their other courses in order to prepare for the one(s) they have no background in). In this way, incidents of "skills mismatch" can potentially effect the learning environment for all students in the school, not just for those students unlucky enough to be taught by an untrained teacher.

I would recommend focusing this paper on issues central to the OTL debate by emphasizing comparisons across different types of schools—even to the extent of looking at poverty differences within locale—thereby contributing to the debate over whether poor or minority children are less likely to receive a quality education than their affluent or white counterparts.

Teacher Quality and Personnel Policy in Public and Private Schools (Ballou and Podgursky)

This paper addresses an important education policy issue, teacher quality, and employs a clever approach to analyze the effects of personnel policy on quality across public and private schools. I do, nonetheless, think that there are several forces, which have not been taken into account, that may affect the results. These criticisms are relatively minor, however, and may not affect the general conclusions, those being that public schools would benefit from greater flexibility in structuring pay, more supervision and mentoring of new teachers, and the freedom to dismiss teachers for poor performance. Two other factors may be at work here though, 1) principals' conflicting goals of student achievement and conflict minimization, and 2) the real productivity effects of increased "teacher power."

One of the theories to come out of the sociology of organizations is that a primary goal of managers, especially in non-market environments, is to minimize

conflict. It is not too far fetched to think of principals trying to minimize conflict between the school board and teachers, parents and teachers, teachers and students, and themselves and all of these groups. Various forms of conflict minimization may be at odds with the goal of maximizing student performance (for example., think about the effect on teachers of a comprehensive parental involvement program). If new teachers, in general, provide minimal problems for their principal, then a principal's perceptions of the quality of his or her new teachers many not be influenced by their goal of conflict minimization. However, if experienced faculty are more likely to cause a principal grief, then the degree to which the principal can control the school environment may influence his or her perception of the quality of their experienced faculty. In fact, principals in private schools are more likely than their public school counterparts to report that they have a great deal of influence over establishing curriculum (63 vs. 26 percent) and setting disciplinary policy (81 vs. 58 percent—See Indicator 47 in *The Condition of Education 1993*). Another factor relating to "principal power", the principal's role in hiring, is already a piece of your model and is strongly predictive. Other measures of "principal power" could easily be added to your model.

Another angle from which to examine the learning environment within schools is from the teachers' perspective. Increased teacher control over classroom policies may improve the quality of their work environment, influencing either teachers' effectiveness in their classrooms or at least decreasing the likelihood that they spend a lot of time complaining to their prin cipals. In fact, teachers in private schools are more likely than their public school counterparts to report that they have control over classroom decisions such as selecting textbooks, selecting course content and topics, selecting teaching techniques, and disciplining students (See Indicator 47 in *The Condition of Education 1993*). It would be interesting to see how much of the variance in the principal's perception of experienced teachers' effectiveness was soaked up by "teacher power" variables (though this could increase the complexity of the modeling exponentially.

Also, in addition to looking at whether or not the state requires private school teachers to be certified, are there any other variables that might measure the degree of state regulation or control of bureaucracy in private schools *relative* to public schools within a state?

I would also recommend plotting some of the expected probabilities (for principals perception of teacher power) for varied levels of the most interesting predic tors (such as teacher salary) since the ordered logit coefficients themselves have no obvious interpretations (since the magnitude of a change in the independent variables is determined by both the beta's and the logistic probability density function).

Work Experience, Labor Market Conditions, and the Decision to Drop Out of High School: Evidence from the NELS:88 (Swaim)

This is a very nice paper, contributing further evidence to prior research finding that working a little while in high school may provide just enough information on the world of un-skilled work to keep kids in school, while working too much (20 or more hours per week) may be too much for kids to handle. This is a framework from which research on the value of vocational education would benefit. Even though a vocational curriculum may not have a positive benefit on gain scores (Rasinski 1994), high school programs which offer "in-field work experience" or "cooperative education" may limit student dropout rates (and would be a nice follow-up analysis).

Suggestions: I would try alternative formulations of "dropping out" or conversely "school engagement." Since Cameron and Heckman (1993) show reduced earnings for GED graduates relative to terminal dropouts (and in a working paper with Nabeel Alsalam (1993) I have shown reduced benefit to late completion) it would be useful to see if working while in school is positively related to *continuous* enrollment.

It would also be interesting to examine the association between student employment and dropping out for students of different ability levels. Is working long hours a bigger problem for low achieving students than high achieving students? (the Akerhielm paper provides a nice way to instrument this to try and avoid endogeneity problems)

The 8th—10th and the 10th—12th grade gain scores could also be used a measure of how much students are benefiting from the time they spend in high school. One might expect positive local labor market conditions to pull only those with small achievement gains out of school. By interacting gains with some of your labor market variables you may be able to get at this issue.

Still another angle would be to study the effect of working while in school on achievement. Does a student employment negatively affect gain scores? Although working less that 20 hours per week may help kids stay in school, it may hurt their longer term possibilities of higher educational attainment (e.g. getting into a good college).

Adding value to the value-added educational production function specification (Akerhielm)

This paper tackles the problem of endogeneity in modeling factors affecting achievement growth and provides a workable solution in using instrumental variables. I have two major comments that I hope will be helpful. First, I would encourage you try out gain scores as the dependent variable in your model. In an experimental framework, we would really be interested in the effect (achievement gain) resulting from a treatment (smaller class sizes, more teacher experience). I think that your statistical model (education production function) should try to do the same thing, student achievement gains based on variability in the level and quality of inputs. The IRT technology that places 10th grades scores on the same scale as 8th grade scores allows us to avoid entering the "pretest" as a right hand side variable (where it has the problem of measurement error in addition to endogeneity). If you are concerned that achievement growth rates differ for those starting at different levels, your instrumental variable for "pre-test" could work here.

Second, the sample drawn for the base year cohort of NELS:88 is both stratified and clustered, not a simple random sample. Although standard regression techniques will produce unbiased coefficients, the fact that students are clustered within schools will produce an error term in your model that is not independent (in most cases resulting in an underestimate of the true standard error). There are several ways to "fix" the resulting standard errors. First you could apply a design effect adjustment available in the NELS documentation. Second, you could use a Taylor Series Estimation Procedure (such as is employed in SUUDAN) to estimate efficient standard errors. You could also employ a random effects model to account for unobserved school characteristics which affect achievement or you could use a hierarchical linear modeling (HLM) technique (which would allow the added benefit of allowing you to partial out individual from school effects).

I also have a few suggestions for further analyses. Do minority or low SES students benefit more from lower class size than their white or high SES counterparts? There is some experimental evidence of this from project STAR in Tennessee. Also, I think that a control for course-taking over the past 2 years (available from the transcripts, which should be available now) may be important if students are not randomly assigned to classes and teachers.

We should continue to try to understand the situations and contexts in which resources can be effectively targeted, so that we do not just "throw money at schools" (Hanushek 1989 and 1994). This paper is a good first step.

References

Cameron, S. and J. Heckman. (1993). "The Non-equivalence of High School Equivalents." *Journal of Labor Economics*, Volume 11, Number 1, Part 1, January.

Hanushek, E. A. (1989). "The impact of differential expenditures on school performance." *Educational Researcher, 18*(4), 45-51.

Hanushek, E. A. (1994). "Money might matter some where: A response to Hedges, Laine, and Greenwald." *Educational Researcher, 23*(4), 5-8.

Rasinski, K (1994). "The effect of high school vocational education on academic achievement gain and high school persistence: Evidence form NELS:88." Draft 2 of Report to U.S. Department of Education, Office of Research, National Assessment of Vocational Education.

Smith, T. and A. Alsalam. (1993) "Does the GED help those who get it?" presented at the Eastern Economics Association meeting in Washington, DC, March.

U.S. Department of Education, National Center for Education Statistics, *The Condition of Education* 1993. Washington, DC: 1993.

Listing of NCES Working Papers to Date

Number	Title	Contact
94-01	Schools and Staffing Survey (SASS) Papers Presented at Meetings of the American Statistical Association	Dan Kasprzyk
94-02	Generalized Variance Estimate for Schools and Staffing Survey (SASS)	Dan Kasprzyk
94-03	1991 Schools and Staffing Survey (SASS) Reinterview Response Variance Report	Dan Kasprzyk
94-04	The Accuracy of Teachers' Self-reports on their Postsecondary Education: Teacher Transcript Study, Schools and Staffing Survey	Dan Kasprzyk
94-05	Cost-of-Education Differentials Across the States	William Fowler
94-06	Six Papers on Teachers from the 1990-91 SASS and Other Related Surveys	Dan Kasprzyk
94-07	Data Comparability and Public Policy: New Interest in Public Library Data Papers Presented at Meetings of the American Statistical Association	Carrol Kindel
95-01	Schools and Staffing Survey: 1994 papers presented at the 1994 Meeting of the American Statistical Association	Dan Kasprzyk