

The Effect of Mode of Interview on Estimates from the 1993-94 Schools
and Staffing Survey (SASS) Public School Teacher Survey

Cornette L. Cole, Robert C. Abramson, Randall J. Parmer, Dennis J. Schwanz
Cornette L. Cole, U. S. Bureau of the Census, Washington, D.C. 20233

KEY WORDS: CATI, Significance tests, sign rank tests, mode of interview

I INTRODUCTION

To reduce the cost of data collection and to improve the efficiency, data was collected for the 1994 Schools and Staffing Survey's (SASS) Public School Teacher Survey by mail, telephone, and personal visits.

During the selection of sample teachers, a split-panel design was used where two-thirds of the teacher sample was randomly assigned for mail nonresponse follow-up interviewing from centralized computer-assisted telephone interviewing (CATI) facilities and the remaining one-third was assigned for telephone follow-up interviewing from decentralized facilities. The teachers were randomly assigned a teacher follow-up mode flag value of 1, 2, or 3 as indicated below. We used this teacher follow-up mode designation flag to separate teacher records and formed the CATI and nonCATI treatment groups we used for this analysis.

- 1 CATI
- 2 nonCATI
- 3 CATI (these records were initially held for possible sample reductions)

Data from the 1994 Public Teacher Survey was collected primarily through self-administered questionnaires, where sample teachers completed the questionnaires and returned them by mail. About 69% of the total interviews were mail returns.

Telephone calls were made from either CATI or decentralized facilities to teachers who did not return their questionnaires by mail. Personnel from the Census Bureau's Field Division were asked to determine the workload capacity of the centralized telephone interviewing (CATI) facilities. From the teachers who hadn't returned their questionnaires by mail, the indicated number of teacher records with follow-up mode designation flags which indicated that they had been designated for CATI follow-up interviewing, was sent to the CATI facilities to be interviewed.

The remaining CATI-designated cases, that is, those that CATI couldn't handle, were sent to be interviewed by decentralized telephone interviewing, along with mail nonrespondents previously designated for this follow-up mode. About 19% of the total interviews were completed in CATI interviews and 12% in decentralized (NON-CATI) telephone interviews.

A very small number of interviews, couldn't be interviewed through either of these telephone methods, and were completed during visits to schools by Census Bureau field representatives.

To be certain there was no bias in survey estimates because we used different modes in the telephone follow-up of mail nonrespondents, we initiated this study to compare the data we collected in CATI interviews with those collected in decentralized telephone interviews.

A. The Schools and Staffing Survey

The SASS is a periodic survey sponsored by the National Center for Education Statistics (NCES) and conducted by the U. S. Bureau of the Census. The SASS provides data on the policies and conditions of public and private elementary and secondary schools, principals, libraries, librarians, teachers and students in the United States.

The school, principal, library, librarian, teacher, and student samples were selected so that data from each of the components could be linked. For the 1993-94 school year, about 13,000 schools, 67,000 teachers, 7,600 libraries and librarians, and 6,900 students were selected¹ for SASS as follows:

- Private and public sample schools were selected first.
- All principals from SASS sample schools were in sample for the School Administrator Survey,
- A sample of teachers was selected within each of the

¹Kaufman et al, 1993-94 Schools and Staffing Survey: Sample Design and Estimation, U. S. Department of Education Office of Educational Research and Improvement, October 1996

SASS sample schools for the Teacher Survey.

- A subsample of SASS sample schools was selected for the Library and Librarian Surveys.
- And a subsample of SASS sample schools and teachers was selected to participate in the Student Record Survey.

B. Public School Teacher Survey Sampling Procedure

The sample of teachers for the Public School Teacher Survey was selected from SASS sample schools. Each sample school was asked to provide a list of its teachers with the information below for each teacher:

- whether the teacher was new (less than three years experience) or experienced,
- the teacher's race and ethnicity,
- whether he or she was considered a Bilingual or English as a Second Language (ESL) teacher
- his or her main field of teaching

Within each sample school, sample teachers were classified into one of the following five strata in the hierarchical order listed below. For example, if a teacher is both API and bilingual the teacher was assigned to the API stratum.

- (1) Asian or Pacific Islander (API)
- (2) American Indian, Aleut, or Eskimo (AIAE)
- (3) Bilingual
- (4) New
- (5) Experienced

Within each school and teacher stratum, teachers were selected with equal probability. From the lists of teachers provided by the schools, 56,736 public school teachers were selected.

C. Estimation

The weight used to produce estimates of public school teacher characteristics was a product of the following weight and factors:

- Basic Weight - the inverse of the probability of selection
- School Sampling Adjustment Factor - an adjustment to the school's probability of selection to account for school mergers, splits, and duplicates
- School Nonresponse Adjustment Factor - an adjustment to account for teachers whose schools did

not provide a list of its teachers

- Teacher Within School Noninterview Adjustment Factor - an adjustment that accounts for teacher nonrespondents
- Frame Ratio Adjustment Factor - a factor which adjusts teacher estimates to the total universe count of teachers from the public school sample frame
- Teacher Adjustment Factor - an adjustment which makes estimates of the weighted number of teachers from the SASS School and Teacher Survey consistent

II. METHODOLOGY

A. Estimates for the Analysis

A flag was assigned after the interviews were completed to indicate the actual mode of interview. The flag to indicate which telephone mode should be used to follow up mail nonrespondents was assigned prior to the initial questionnaire mailing. We used the follow-up mode designation flag to separate interviewed teacher records and form a CATI and NON-CATI treatment group for our analysis. Therefore, the treatment name (CATI or NON-CATI) is not necessarily an indicator of how the interview was actually completed.

The CATI treatment comprises all teachers who were designated for mail nonresponse follow-up interviewing from CATI facilities. Their interviews were actually either returned by mail or completed by telephone from CATI facilities.

The NON-CATI treatment comprises all teachers who were assigned for mail nonresponse follow-up interviewing from decentralized facilities. Their questionnaires may have actually been returned by mail or completed in interviews from decentralized facilities.

Recall that two-thirds of the teacher sample was assigned for CATI follow-up interviewing and the remaining one-third for telephone follow-up interviewing from decentralized facilities. To insure that the estimates we produce from records in our CATI and NON-CATI treatments would be approximately equal to the estimates we got for the entire sample, we increased the teacher basic weights on records in the CATI treatment by 1.5 and those in the NON-CATI treatment by 3.0. Then,

- We processed the CATI and NON-CATI data sets

(separately) through the same weighting procedure used to weight the regular Public School Teacher Survey data.

- We further separated the reweighted teacher records by actual mode of interview within each treatment.
- We finally produced CATI and NON-CATI treatment estimates for the Public School Teacher Survey questionnaire items.

We made CATI vs. NON-CATI comparisons of estimates for the following groups of teacher records:

Comparison Group 1: All interviews (all modes)

All teachers who were designated for telephone follow-up from centralized facilities (regardless of the mode the interview was completed in)

vs.

those who were designated for telephone follow-up from decentralized facilities (regardless of the mode the interview was completed in)

Comparison Group 2: Mail interviews only

Teachers who were designated for telephone follow-up from centralized facilities who returned their questionnaires by mail

vs.

those who returned their questionnaires by mail who had been designated for telephone follow-up from decentralized facilities

Comparison Group 3: Interviews completed during telephone follow-up only.

Teachers who were designated for telephone follow-up from centralized facilities and their interviews were completed in centralized telephone interviews

vs.

those who were designated for telephone follow-up from decentralized facilities and their interviews were completed in decentralized telephone interviews

Our primary interest was in Comparison Group 3. For these respondents, the telephone follow-up mode flag and the flag which indicates the actual interview modes have the same value.

B. Computing the Variances for the Analysis

We used the Balanced Repeated Replication (BRR)

method in *WESVAR*² to compute variances for each estimate. The *WESVAR* BRR procedure uses replication techniques to calculate variances for estimates using:

$$v(\theta) = \frac{1}{G'} \sum_{k=1}^{G'} (\theta_k - \theta)^2$$

where,

θ = the estimator for the teacher questionnaire item

$v(\theta)$ = the variance of the estimate

G' = the number of replicates

The replicate weights we used to compute the variance estimates in *WESVAR* were computed using the same replicate factors used to calculate variance estimates for the regular 1993-94 SASS publication estimates.

C. Comparing Treatment Estimates

We evaluated the magnitude of the differences between CATI and NON-CATI estimates to see if they were statistically significant. We formed the null hypothesis,

$$H_0: \Theta_{CATI} = \Theta_{NONCATI}$$

which says an estimate produced using the records of CATI teachers (θ_{CATI}) is the same as that produced using the records of NON-CATI teachers ($\theta_{NON-CATI}$).

To test the hypothesis, we used the 'z' statistic:

$$z = \frac{\theta_{cati} - \theta_{non-cati}}{\sqrt{var(\theta_{cati}) + var(\theta_{non-cati})}}$$

where,

- θ is the estimate of the teacher characteristic of interest, $var(\theta)$ is its variance,
- the numerator is the difference between the CATI and NON-CATI estimates and
- the denominator is an estimate of the standard error of the difference.

A negative value for the z meant the NON-CATI estimate was higher, while a positive z value meant the CATI estimate was higher. Results of the significance tests are presented in Section III.

²The *WESVAR* SAS Procedure, Version 1.2

C. Evaluating the Distribution of the Differences between Treatment Estimates

Our significance tests evaluated the magnitude of the difference between CATI and NON-CATI estimates individually for each teacher questionnaire item. We used the sign rank test to evaluate the distribution of the differences across the items.

We used the SAS *PROC UNIVARIATE*³ procedure to perform the Wilcoxon Signed-Rank Test. We assumed each difference was equally likely to be positive or negative and that the distribution of the differences is symmetrical. We tested the hypothesis that the median of the differences between the CATI and NON-CATI estimates is zero. The following steps are involved in the test:

- The absolute values of the differences are assigned ranks by magnitude, from smallest to largest, then the positive and negative signs are restored to the ranked values.
- The totals of the ranks with negative signs and those with positive signs are calculated.

The Wilcoxon signed rank statistic S is computed in SAS as follows:

$$S = \sum r_i^+ - \frac{n(n+1)}{4}$$

where,

- S is a sum of scaled binomial distributions
- r_i^+ is the rank of $|x_i|$ after discarding values of $x_i = 0$ and x_i is the difference between the CATI and NON-CATI estimates ($|\theta_{CATI} - \theta_{nonCATI}|$)
- n is the number of nonzero x_i values and
- the sum is computed over the values of x_i greater than zero.

The significance level of S is computed as:

$$\text{Significance level} = S \frac{\sqrt{n-1}}{\sqrt{nV-S^2}} \text{ where,}$$

$$V = \frac{n(n+1)(2n+1) - 0.5 \sum t_i(t_i+1)(t_i-1)}{24}$$

The sum is calculated over differences tied in absolute

value and t_i is the number of tied values with the i^{th} difference.

SAS outputs a probability or p-value that is a measure of the strength of the evidence against the null hypothesis. If the p-value is less than the significance level of the test, which in our case is 0.10, the null hypothesis should be rejected. The smaller the p-value, the stronger the evidence for rejecting the null hypothesis.

III. RESULTS

A. Tests of Significance

At the $\alpha = .10$ level of significance, we expect no more than 10 percent of the estimates within a group would be significantly different. Table 1 provides a summary of the results of our significance tests. The table shows that for all three groups, more than 10 percent of the comparisons yielded statistically significant results.

Table 1 also shows that the interviews completed during telephone follow-up (Comparison Group 3) had the higher proportion of significant differences. In this group, we are comparing the responses of teachers who were actually interviewed from CATI facilities with those interviewed from decentralized telephone facilities. We see in Table 1 that there were about twice the proportion of significant differences between these respondents than mail respondents.

B. Items with significant differences

Most of the significant differences were between the responses of teachers in the two treatments to the series of questions labeled "Perceptions and Attitudes Toward Teaching" (Section E of the 1993-94 Public School Teacher questionnaire).

In general, NON-CATI treatment estimates were higher for categories of items which have negative connotations, while CATI treatment estimates were higher for responses which suggest these teachers had a more positive outlook.

The NON-CATI treatment estimate was higher for items which say

- More of these teachers believed their principal did not enforce student rules, he did a poor job of getting resources, and he did not let them know what was expected of them.
- More of them reported they had little influence or no control over the curriculum, textbooks, homework, and over teacher evaluations and
- More of them said they would remain in the same school system, but would teach at another school the

³SAS Procedures Guide, Version 6, Third Edition

next year.

One the other hand, the CATI treatment estimates were higher for items which suggest the attitudes of most of these teachers were more positive.

- More of them reported their principals let them know what was expected from them and their schools' administrations treated them fairly and were supportive.
- More of them said their principals enforced school rules and backed them when they needed him to.
- More of them planned to continue teaching at the same school the next school year.

Responses to the question "*If you could go back to your college days and start over again, would you become a teacher or not?*" summarizes the contrast in attitude between teachers in the two treatments. The CATI estimate was higher for the category '*certainly would become a teacher*' and the NON-CATI estimate was higher for the category '*chances about even for or against*'.

B. Sign Rank Tests

In Table 2 below, probability values ($PR \geq |S|$) for the two-tailed tests are shown. Each p-value is greater than 10 percent, indicating that the hypothesis that the median of the differences between CATI and NON-CATI estimates is zero should not be rejected.

Table 2 also shows that the group consisting of interviews completed during telephone follow-up had the highest p-value. Thus, there is no evidence favoring the rejection of the hypothesis about the distribution of the differences for these respondents.

Our assumption for the sign-rank test was that each difference was equally likely to be positive or negative and that the distribution of the differences was symmetrical. The test results say there were about an equal number of differences with the CATI estimate higher as those with the NON-CATI estimate higher.

We stated earlier that most of the significant differences between the responses of teachers in the treatments were to attitude and perception items. Table 3 shows the two-tailed p-values from the sign-rank tests we performed using only the categories of attitude and perception items which have negative connotations.

Each p-value is less than 10 percent, suggesting we should reject the null hypothesis. The median difference for these items alone is different from zero and the distribution of the differences is skewed. This result agrees with our observation that NON-CATI treatment estimates were higher for these types of items.

IV. CONCLUSIONS

By randomly assigning teacher sample records between telephone modes, we gave each teacher record a chance of being assigned to CATI or decentralized telephone for follow-up. Our tests show that teachers within a treatment provided similar responses to attitude and perceptions questions and estimates of these responses were statistically different between treatments.

There are two possible explanations. One is that there was some periodicity in way the teacher records were ordered. This ordering resulted in the teachers assigned to the same treatment having similar characteristics. Another is that the assignment was truly random, but we were unlucky in the assignment, and the results of the assignment are due to the natural variability between teachers in the treatments.

There were also a higher proportion of significant differences between the responses of teachers in the third comparison group, the group with teachers interviewed during telephone follow-up. The majority of the significant differences for this group were to attitude and perception items, the same as we saw between CATI and NON-CATI respondents in the other two analysis groups.

Also, as seen in the other two groups, CATI treatment respondents reported more optimistic answers and the NON-CATI respondents reported more pessimistic answers.

Unlike the CATI and NON-CATI respondents in the other two groups, in the third group we isolated records by both designation mode and interview mode:

Actual Interview Mode

	<u>CATI</u> <u>Treatment</u>	<u>NON-CATI</u> <u>Treatment</u>
Group 1	all interviews	all interviews
Group 2	Mail	Mail
Group 3	CATI	Decentralized telephone and a very small number of CATI cases

The increase in the proportion of significant differences between respondents interviewed by telephone may be attributable to data in the CATI group being collected by CATI and data in the other primarily by decentralized telephone interviews. This result suggests that the attitude data we collected in CATI and decentralized telephone interviews were different.

LIMITATIONS

We cannot attribute the differences we observed between the responses of teachers in the analysis groups we formed for this study solely to the mode in which their interviews were completed or to the method we used to assign teacher records for mail nonresponse follow-up. There are other errors, such as those due to estimation, coverage, processing, nonresponse, etc., which may have influenced our results.

TABLE 1
Questionnaire Items with Significant Differences

Interview Mode	Proportion of the Differences between CATI and NON-CATI Treatment Estimates that were Statistically Significant
All Interviews	18%
Mail Returns	14%
Telephone Follow-up Interviews	29%

TABLE 2
Results of Sign Rank Tests --- All Items

Interview Mode	P-value ($PR \geq S $)
All Interviews	0.4092
Mail Returns	0.3782
Telephone Follow-up Interviews	0.7112

TABLE 3
Results of Sign Rank Tests --- Negative Responses
To Attitude and Perception Items Only

Interview Mode	P-value ($PR \geq S $)
All Interviews	0.0001
Mail Returns	0.0012
Telephone Follow-up Interviews	0.0019