SOME STATISTICAL FINDINGS FROM NATIONWIDE TEACHER POLLING

Glen Robinson, National Education Association
Chester H. McCall, Jr., Booz-Allen Applied Research, Inc.

Introduction

Data collection and processing are becoming increasingly important in the field of education. To be of maximum value, such data must be obtained rapidly, accurately, and on a nationwide basis.

During the past two and one-half years, the Research Division of the National Education Association has been engaged in a project to develop and test data collection and analysis procedures to meet the rising demands in the field of education. The purpose of this paper is to describe some of the results of this project.

Background of Project

During the early 1950's, there were few timely and accurate nationwide statistics that pertained to teacher and school principal populations. Those that were reported represented either compilations of statewide figures or compilations of statewide estimates. There were practically no studies based on probability samples drawn from all the school districts in the nation. One reason for this was the lack of appropriate listings of all the school districts.

A new era was opened in nationwide educational sampling when the 43,500 school districts in the nation enumerated in the 1957 Census of Governments were placed on magnetic tape by the Governments Division of the U. S. Bureau of the Census. For the first time probability sampling of the nation's school districts could be achieved at a reasonable cost.

Meanwhile, new research techniques in probability sampling had been developing in areas outside education, and the NEA Research Division staff believed that the application of these techniques to the Division's work could result in greater efficiency, reliability, and speed. Consequently, the staff undertook an intensive project to apply and test the best of these new sampling and polling techniques in the field of educational research. The study was considered to be of prime importance to the future work of the Research Division.

Since the extensive work involved in a project of this scope had to be carried on concurrently with the Division's regular operation, special assistance was needed for the project. The firm of Booz-Allen Applied Research, Inc. was retained as technical consultants.

The project had three primary purposes:

1. To study and evaluate the sampling procedures then being used by the NEA Research Division
2. To examine, develop, and test recent sampling and polling techniques which could be properly applied to the types of studies made by the Division
3. To develop within the Division a continuous training program in the use of new techniques.

To date the following nationwide surveys have been made as a part of this project:

1. Three polls of teachers' opinions
2. Two studies of teacher mobility and turnover
3. One study of the status of the American teacher
4. One poll of elementary-school principals' opinions

All seven studies were mailed surveys. These studies were carefully designed and co-ordinated so as to yield a maximum of data relating to sampling problems such as the effect of nonresponse bias, the effect of sampling error, the time required for response, and the effectiveness of certain techniques in improving response rates. These data have been placed on punch cards for detailed analysis. Such analyses are presently under way.

General Procedures

The design used in the studies was a two-stage cluster sample. The first stage consisted of a stratified random sample of school systems (clusters) drawn from the 43,500 school districts enumerated by the Bureau of the Census in its 1957 Census of Governments. The first-stage sample was selected for the NEA Research Division according to the Division's specifications by the Bureau of the Census.

Letters were sent to the superintendents in each of the sample systems requesting a copy of their directory listing the names and school addresses of teachers and principals in their school system.

The second stage consisted of a random sample of teachers or principals drawn from these school districts. Samples were selected so that all teachers in the nation had the same probability of being selected. Hence, the samples were self-weighting.

The survey instruments were mailed directly to the sample of teachers or school principals. These questionnaires were accompanied by a personally addressed and typed letter, signed by the Director

of the Research Division, explaining the nature of the survey and the importance of the teacher's response. The general pattern of follow-up procedures used consisted of a personal letter sent by air mail to nonrespondents at the end of two weeks. Telegrams were sent to nonrespondents at the end of four and five weeks. The mean net response rate (useable questionnaires returned) for the seven studies was 95.7 percent. Some persons did not respond because they were critically ill, had died, had moved without leaving a forwarding address, or had written saying they would not take part in the poll. When this group was added to the net response group, a gross response rate of 97.7 percent was obtained for the seven studies.

Survey instruments were pre-tested on samples of teachers drawn from school systems not selected in the first-stage sample.

Some of the statistical findings resulting from these procedures will be described later in this paper.

General Results and Implications of Project

Although much work remains to be done in analyzing the data received and in making improvements in the sampling procedures, a number of benefits are already resulting from the sampling project. These are:

1. Improved accuracy in data collection and analysis by the NEA Research Division. Since these samples were drawn in accordance with probability theory, it is possible to report the results with a stated degree of accuracy and confidence. The high response rates have greatly reduced the possible nonresponse bias. Analysis of data is yielding more knowledge of nonresponse groups.

2. Development and use of better sampling frames. Although the first stage of the samples for the seven studies was drawn for the NEA Research Division by the Bureau of the Census from 1957 lists, the Division has since purchased from the Census Bureau a complete listing of all school systems as of the school year 1959-60. This more recent information is on IBM cards and is presently being used by the Division for sampling purposes.

3. More timely reporting of data. As a result of the project, it is now possible to report survey data within the same school year in which the data were collected. These surveys have demonstrated that with proper sampling and follow-up procedures, response rates in excess of 90 percent can be achieved within two months of initial mailing date.

4. Reduced unit cost. A substantial reduction in the unit cost of surveys has been achieved by the project. An appraisal of the full extent of this accomplishment is forthcoming. A work-sample survey was conducted by the Research Division staff for a six-month period to estimate the unit cost of conducting surveys by the new and old procedures. These data are scheduled for analysis in the near future.

5. Implications for future operation. The sampling project has initiated a new period in improved data collection and analysis within the NEA Research Division. Improved sampling techniques will facilitate the use of survey designs not previously feasible. Improved sampling techniques have also indicated that further improvements can be achieved through the use of automated data processing. Action is being taken to secure computer equipment (of the IBM 1620 class) in the immediate future.

Some Statistical Considerations

The previous sections of this paper have presented a brief description and background of the present sample survey project within the NEA Research Division. This portion of the paper will emphasize the statistical and related aspects of the program.

The test vehicle chosen to examine the usefulness of a two-stage sampling procedure was the Teacher Opinion Poll. These polls consisted of approximately 10 questions designed to stimulate interest and hence elicit a high response. Eleven status questions were selected to test the consistency of the survey findings with external estimates and the reproducibility of survey findings from one year to another. The first Teacher Opinion Poll was mailed February 1960. The following section describes the method by which the sample was selected.

Teacher Opinion Poll Design

The first problem encountered in selecting a design for sampling teachers was the lack of a complete listing of all public-school teachers. It was necessary, therefore, to examine other possible procedures. Since the Bureau of the Census maintained a listing of all school systems in the United States on magnetic tape, it was believed that a suitable design would be the selection of school systems by some random device and then, having selected these systems, to obtain a sample of individual teachers. This design was selected out of necessity rather than desired optimum sampling. The two-stage plan was executed in the following manner:

1. School systems were stratified into eight strata according to the number of teachers in the system as of October 1956.

2. A sample of 31 systems was selected from each of the eight strata. From these systems a random selection of teachers was then obtained. The teachers selected constituted the sample for the teacher opinion polls and other studies of teacher populations.

With only limited information available on the important characteristics of the population

stratified by this method, it was felt that a first estimate of an appropriate sample size would be to assume the required sampling for a purely random selection from among all teachers in the United States. This method was utilized to ascertain the required sample size. It is recognized that minimum accuracy requirements may not have been met in some cases because of the effect of the inter-cluster contribution to the variance of the estimates. As the information from these studies is analyzed, it will be possible to revise the sampling procedure and bring it more in line with optimum allocation techniques.

In 1960, two teacher samples were selected from the same 250 systems. These are referred to as Teacher Opinion Poll No. 1 and Teacher Opinion Poll No. 2. In 1961, a new set of systems was selected from the Bureau of Census records. Teacher Opinion Poll No. 3, The Status of the American Teacher, and The Elementary-School Principal Opinion Poll were all selected from the 1961 sample of systems. In 1961, the respondents to Teacher Opinion Polls 1 and 2 were mailed a questionnaire designed to measure teacher mobility and turnover. These latter two studies were made in order to assess the amount of change occurring in teacher sampling frames within a one-year period.

The design used in the Teacher Opinion Polls has been criticized by some persons as not being the most efficient design. We are aware of this possible weakness and would point out that the major reason for selecting the sample in this way was the restriction imposed on the project by the lack of an appropriate listing of all teachers. As the data from the studies of the past two years are more thoroughly analyzed, more appropriate statistical designs will be introduced where feasible.

Response Rates

One of the criticisms which is frequently directed at those who use mailed survey techniques is the generally poor response. The project has overcome this deficiency through careful planning, efficient control of mailing lists, appropriate processing of incoming questionnaires, and use of effective follow-up procedures. Remarkable response rates were achieved in seven surveys conducted during 1960 and 1961. Table 1 shows the response rates for each of the seven surveys, the mailing date, the cut-off date for accepting replies, and the sample size. It is interesting to note that the lowest net response rate was approximately 93 percent. While we have run no statistical tests on this, we find little reason to believe that within reasonable limits the length of a questionnaire necessarily influences the tendency to respond. It is possible that the time of mailing within the school year influences response as much as the length of the questionnaire.

TABLE 1.--MAIL AND RESPONSE INFORMATION FOR SEVEN SAMPLE SURVEY STUDIES, 1960 AND 1961

Sample survey	Length of questionnaire	Date mailed	Date of last reply	Sample size	Response rate	
					Net[a/]	Gross[b/]
1	2	3	4	5	6	7
Teacher Opinion Poll 1	2 pages	23 February 1960	20 April 1960	1149	.968	.981
Teacher Opinion Poll 2	2 pages	29 March 1960	24 May 1960	1147	.949	.982
Teacher Opinion Poll 3	2 pages	3 February 1961	7 April 1961	1633	.966	.975
Teacher Mobility and Turnover 1	1 page	27 February 1961	13 June 1961	1112	.950	.989
Teacher Mobility and Turnover 2	1 page	29 March 1961	14 June 1961	1089	.965	.983
Status of American Teacher	4 pages	27 March 1961	27 June 1961	2104	.926	.949
Elementary-School Principal Opinion	3 pages	17 February 1961	18 May 1961	721	.976	.978

a/ Net response rate includes only those persons who returned useable questionnaires. Average net response rate for seven studies was .957.

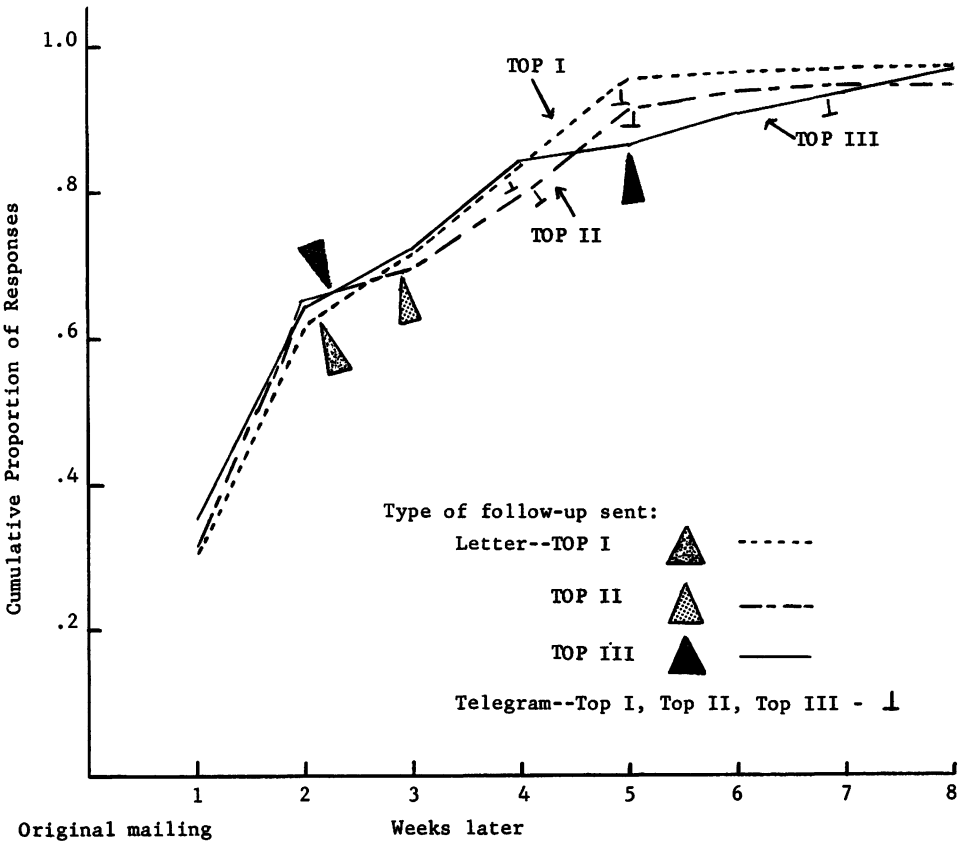
b/ Gross response rate includes persons who returned useable questionnaires plus those persons whose reason for nonresponse could be determined--deceased, critically ill, moved without leaving forwarding address, or wrote saying they refused to answer questionnaire. Average gross response rate for seven studies was .977.

A system of questionnaire numbering was used to maintain records and to identify nonrespondents.

In conducting the studies, special effort was made to glean as much information as possible about the populations studied and their response patterns. For example, Figure I shows the cumulative response curves for the three Teacher Opinion Polls along with the mailing dates of the follow-ups. The striking agreement among these three curves suggests that this pattern might be expected when similar follow-up procedures are used with the same type of population and questionnaires. On the other hand, the effect of deviation from the general follow-up procedure, described previously, is clearly evident in the chart.

There was also need for research on the tendency to respond by members of various subgroups within the teacher population. For example, little was known about the response patterns of men teachers versus women teachers or elementary-school teachers versus secondary-school teachers. In

FIGURE I--RESPONSE PATTERNS FOR THREE NATIONWIDE TEACHER OPINION POLLS, 1960 AND 1961



Cumulative response rates[a]			
Week	TOP I	TOP II	TOP III
1	.301	.315	.351
2	.614	.648	.640
3	.711	.694	.722
4	.841	.797	.842
5	.953	.912	.868
6	.963	.940	.909
7	.968	.949	.933
8	.968	.949	.966

a/ Response rates plotted are cumulative net response rates which include only those persons returning useable questionnaires.

each of the three Teacher Opinion Polls the date on which the questionnaire was mailed, the date the questionnaire was received, and the postmark date on the return envelope were noted on the returned questionnaires. This information yielded approximate estimates of the length of time the questionnaire was actually in the respondent's hands; these data were punched into the IBM card for each questionnaire. Some interesting observations were made on the basis of a preliminary analysis of the time-in-respondent's-hands data. This analysis indicated that there was a highly significant difference (probability less than .001) between the response curves for men and women teachers in both elementary and secondary schools. Men were found to respond more rapidly.

Although not a part of the seven studies, another interesting experiment, conducted in 1961, pertained to factors affecting response rates. This was done in a study of small high schools in the United States. Instruments were sent to both the principals and the teachers in a sample of small high schools. Half the questionnaires to each group were sent in stamped envelopes. The other half were sent in metered envelopes. The follow-up procedures were identical for both groups. Utilization of the Kolmogoroff-Smirnov test demonstrated that in both groups the response distribution was significantly different at the .001 level, with the persons receiving stamped envelopes answering in greater proportion. The inference here is that the use of postage stamps created either a more personal impression or an impression of importance that stimulated greater response. Further study is needed to substantiate this.

Selected Estimates

Several choices in the selection of statistical estimates were available for analyzing the two-stage sample design used in the Teacher Opinion Polls. An excellent reference for such estimates is a book by Sukhatme.[1/] Of the various estimates which were available, it was believed than an unbiased estimator would be the most appropriate. For this reason the design was set up so that the probability of being selected in the sample was identical for all teachers in the United States. This procedure, in effect, yielded a self-weighting sample.

An estimate for the mean (in the general case) of the i^{th} stratum when k_i clusters were subsampled is:

$$\bar{X}_i = \frac{1}{N_i} \times \frac{K_i}{k_i} \sum_{j=1}^{k_i} N_{ij}\bar{X}_{ij} \qquad (1)$$

where $$\bar{X}_{ij} = \sum_{t=1}^{n_{ij}} x_{ijt}/n_{ij}$$

The variance for these sample estimates, however, is much more complicated. An estimate of the variance of the sample mean in the i^{th} stratum, for example, is given by the following expression:

$$\hat{\sigma}^2_{\bar{X}} = \left\{\frac{1}{k_i} - \frac{1}{K_i}\right\} s^2_{bi} + \frac{1}{k_i K_i} \sum_{j=1}^{k_i} \left\{\left(\frac{N_{ij}}{N_i}\right)^2 \left(\frac{1}{n_{ij}} - \frac{1}{N_{ij}}\right) s^2_{ij}\right\} \qquad (2)$$

where $$s^2_{ij} = \sum_{t=1}^{n_{ij}} \frac{(X_{ijt} - \bar{X}_{ij})^2}{n_{ij} - 1} \qquad (3)$$

$$s^2_{bi} = \frac{1}{k_i - 1} \sum_{j=1}^{k_i} \left\{\frac{N_{ij}}{\bar{N}_i} \bar{X}_{ij} - \bar{X}_i\right\}^2 \qquad (4)$$

and $$\bar{N}_i = \sum_{j=1}^{K_i} \frac{N_{ij}}{K_i} \qquad (5)$$

The terms in the equation are as follows:

- K number of strata in population.
- K_i number of clusters in i^{th} stratum.
- k_i number of clusters selected from i^{th} stratum. (note that $k_i \leq K_i$).
- N_{ij} number of observations in j^{th} cluster in i^{th} stratum.
- N_i number of observations in i^{th} stratum.
- N total population size; $N = \sum_{i=1}^{K} N_i$.
- N_i' total number of observations in clusters which are being sampled in i^{th} stratum.
- n_{ij} number of observations in sample from j^{th} cluster in i^{th} stratum (note that n_{ij} may equal zero).
- n_i number of observations in total sample from i^{th} stratum; $n_i = \sum_{j=1}^{k_i} n_{ij}$.

1/ Sukhatme, Pandurang. Sampling Theory of Surveys with Applications. Ames: Iowa State College Press, 1958.

n total sample size; $n = \sum_{i=1}^{K} n_i$.

X_{ijt} the t^{th} sampled observation in the j^{th} sampled cluster in the i^{th} stratum.

Since $\bar{N}_i$, the average number of observations per cluster in the i^{th} stratum, is not available in this particular equation, it is estimated by the following equation:

$$\hat{\bar{N}}_i = \bar{n}_i = \sum_{j=1}^{k_i} \frac{N_{ij}}{k_i} \qquad (6)$$

We recognize that our use of this term may be subject to some error.

It should be noted that S_{ij}^2 is the variance within the j^{th} sampled cluster in the i^{th} stratum. S_{bi}^2 is the weighted variance among the sampled cluster means in the i^{th} stratum. In addition, the value of K_i is not known for the two years under study. This figure has been estimated by the projection and interpolation of data for the school years 1956-57 and 1959-60. We are of the opinion that the errors in these estimates do not contribute significantly to the amount of error in the estimates of the variance of sample means.

An estimate for the mean in the total population is given by the expression

$$\bar{\bar{X}} = \sum_{i=1}^{K} \frac{N_i}{N} \bar{X}_i$$

$$= \sum_{i=1}^{K} \frac{K_i}{k_i} \sum_{j=1}^{k_i} \frac{N_{ij}}{N} \bar{X}_{ij}. \qquad (7)$$

While an estimate for the variance of $\bar{\bar{X}}$ is then seen to be

$$\hat{\sigma}^2_{\bar{\bar{X}}} = \sum_{i=1}^{K} \left\{\frac{N_i}{N}\right\}^2 \hat{\sigma}^2_{\bar{X}_i} \qquad (8)$$

The statistics which have been tabulated for presentation in this paper are the following mean characteristics for the classroom-teacher population: years of teaching experience and age of teachers. Table 2 summarizes the estimates of these parameters for 1960 and 1961 as found in Teacher Opinion Polls 1 and 3. In addition, included in Table 2 are the estimates of the standard errors of these characteristics from equation (8) and also under the assumption of a purely unrestricted random sample, as computed from equation (9), namely,

$$\hat{\sigma}^2_{\bar{\bar{X}}} = \frac{\sum\sum\sum \left(X_{ijt} - \bar{\bar{X}}\right)^2}{n(n-1)} \qquad (9)$$

Although the estimates of the standard error, as derived from equation (8), indicate variances greater than we feel are desirable, they provide a good base from which to make adjustments in the sample design. From a detailed analysis of the strata and cluster data used in computing these estimates, it should be possible to improve the sample design and adjust the sample size so as to reduce these variances. As a consequence of the observed standard errors, it appears worthwhile to consider the use of estimates which will have smaller variation; for example, an estimator which will provide for a reduction in the magnitude of S_{bi}^2. One possible estimator would yield

$$s_{bi}^2 = \frac{1}{k_i - 1} \sum_{j=1}^{k_i} \left(\frac{N_{ij}}{\bar{N}_i}\right)^2 \left(\bar{X}_{ij} - \bar{X}_i\right)^2. \qquad (10)$$

Unfortunately, for this presentation time did not permit an evaluation of other possible estimates and standard errors.

We would welcome any suggestions or comments that might be helpful in improving either the sample design or the method of assessing sampling variability.

TABLE 2.--NATIONWIDE ESTIMATES OF SELECTED CLASSROOM-TEACHER POPULATION CHARACTERISTICS AND THE STANDARD ERRORS OF THESE CHARACTERISTICS

Characteristics	Survey[a]	Estimate of mean	Standard error	
			Computed by Equation (8)[b]	Computed by Equation (9)[c]
1	2	3	4	5
Teaching experience	TOP 1	14.16 years	.58	.35
Teaching experience	TOP 3	13.74 years	.66	.28
Age	TOP 1	40.76 years	1.31	.37
Age	TOP 3	40.35 years	1.85	.32

a/ Teacher Opinion Poll 1 was mailed February 1960; Teacher Opinion Poll 3 was mailed February 1961.

b/ Equation (8) provides an estimate of the variance of the sample mean which takes into account the amount of variance derived from the use of a cluster sample. These numbers are based upon a preliminary machine run and are subject to revision.

c/ Equation (9) provides an estimate of the variance of the sample mean which assumes an unrestricted random sample.